WORKING PAPER SERIES

Robust Opinion Aggregation
and its Dynamics

*Simone Cerreia-Vioglio, Roberto Corrao,
Giacomo Lanzani*

Working Paper n. 662

This Version: March, 2020

Robust Opinion Aggregation and its Dynamics[*]

Simone Cerreia-Vioglio[a], Roberto Corrao[b], Giacomo Lanzani[b]

[a]Università Bocconi and Igier, [b]MIT

March 2020

Abstract

We study agents in a social network who receive initial noisy signals about a fundamental parameter and then, in each period, solve a robust non-parametric estimation problem given their previous information and the most recent estimates of their neighbors. The resulting robust opinion aggregators are characterized by simple functional properties: normalization, monotonicity, and translation invariance. These aggregators admit the linear DeGroot's model as a particular parametric specification. However, robust opinion aggregators allow for additional features such as overweighting/underweighting of extreme opinions, confirmatory bias, as well as discarding information obtained from sources perceived as redundant. We show that under this general model, it is still possible to link the long-run behavior of the opinions to the structure of the underlying network. In particular, we provide sufficient conditions for convergence and consensus and we offer some bounds on the rate of convergence. In some parametric cases, we derive the influence of the agents on the limit opinions and we stress how it depends on their centrality as well as on their initial signals. Finally, we study sufficient conditions under which a large society learns the true parameter while also highlighting why this property may fail.

1 Introduction

In many social network settings of economic interest, individuals fail to adjust their opinions in a Bayesian fashion. In this case, non-Bayesian social learning models offer a better description of the opinions' dynamics.[1] A reasonable explanation is that fully rational inference is not an easy task to implement under the complex information structures that arise in social networks. Moreover, agents are sometimes just trying to either adapt to each other or to adjust to a belief shared by the entire society.[2] Finally, when modeling the evolution of Bayesian updates in a network, tractability is easily lost, especially outside the standard quadratic-Gaussian setting.[3]

In this paper, we propose and analyze a general and highly flexible model of non-Bayesian social learning. We highlight the key role of the network structure for the limit behavior of opinions in terms of convergence, consensus (i.e., all the agents end up sharing the same opinion), and information aggregation. In doing so, we also derive a set of effects and predictions that are novel to this literature.

Robust estimation in networks We consider a network of agents in which each of them observes the signals' realizations of her neighbors. These signals are equal to a common fundamental location parameter plus some agent specific noise. Motivated by the fact that it might be hard for real-life individuals to assess the informational content encoded in others' actions and opinions, we model agents that are not Bayesian. Specifically, in the face of their uncertainty about the data-generating process and the global network structure, the agents pool their neighbors' opinions as to solve a *robust estimation* problem in each period. In other words, given their neighbors signals' realizations, in the first period, agents try to estimate a la Huber [48] the true parameter and in the following periods they repeatedly update their opinions using their neighbors' estimates coming from the previous round of updating. This estimation procedure does not require them to postulate a probabilistic model for the data observed. Instead, each agent minimizes a loss function which penalizes larger *residuals* between what she observes and her estimate. This captures the idea that individuals are aware that the data observed contain some valuable information on the parameter, plus some noise.

---

[1]See the empirical evidence in Breza, Chandrasekhar, and Tahbaz-Salehi [15], and Chandrasekhar, Larreguy, and Xandri [23], and the references therein.

[2]In these cases, adhering to Bayesian updating might not be reasonable even from a normative point of view. See Section 7 for more details on how our model also addresses this issue.

[3]For a discussion of the difficulties of Bayesian modelling of social networks see Breza, Chandrasekhar, Golub, and Parvathaneni [14]. A notable exception is the Gaussian case studied by Mossel, Olsman, and Tamuz [70].

Robust opinion aggregators The estimation procedure described above, depending on the loss functions used by the agents, generates a map from profiles of last period opinions to the updates. Our first main result, Theorem 1, is a characterization of these maps which we call *robust opinion aggregators*.[4] These aggregators feature properties that make the analysis still tractable:

1. **Normalization**: Every time the agents have reached a consensus none of them further updates her opinion;

2. **Monotonicity**: If two profiles of opinions are such that the first one dominates (according to the coordinatewise order) the second, then this relation is preserved after aggregation;

3. **Translation invariance**: If the opinion of *each* agent is shifted by the *same constant*, then the updates are shifted as well.

The first two properties have a straightforward interpretation, whereas translation invariance is a natural consequence of the fact that agents try to estimate a location parameter. In general, these simple properties are appealing because they arise from the estimation procedure we consider, and, at the same time, they nest several classes of opinion aggregators such as the celebrated *DeGroot's model*.[5]

Our opinion aggregators are general enough to capture economic phenomena such as dislike for (or attraction to) extreme opinions, confirmatory bias, disregard for redundant information as well as assortativeness. Indeed, a key feature of our model is that the reciprocal influence among agents also depends on their original estimates\opinion. For example, if the robust opinion aggregator considered features attraction to extreme opinions, then the influence of each agent depends *both* on her centrality and on how extreme was her original opinion relative to the entire population. This novel effect might have immediate and relevant implications for designing intervention policies in networks. The intrinsic level of extremism among the agents (captured, for example, by asymmetric loss functions) might challenge the establishment of a moderate consensus even when the latter is proposed by central agents.[6] To the best of our knowledge, the current work is the first to propose a unifying approach that deals systematically with each of the aforementioned effects.

---

[4]In a related project, [19], we show how the same dynamics arise from a stochastic coordination game between the agents, provided that they keep some minimal inertia in favor of their own action played in the previous period (in the current setting, the opinion they stated).

[5]Our model reduces to the standard DeGroot's model when all the agents use a quadratic loss function.

[6]These policy interventions can assume different forms such as incentive distortions (Galeotti, Golub, and Goyal [37]) or information design (Galperti and Perego [39]).

On the empirical side, the recent field studies that compare Bayesian to non-Bayesian social learning models have obtained evidence consistent with our properties. For instance, Chandrasekhar, Larreguy, and Xandri [23] find that most of the time, if the sampled subjects come to a consensus, then they remain stuck on their beliefs even when such a behavior is objectively suboptimal: this is consistent with normalization. Similarly, they also find that the overwhelming majority of subjects responds monotonically to changes in their neighbors' opinions.

The dynamics of robust opinion aggregation The opinions' dynamics induced by this more general form of social learning are different from the ones described by the standard linear updating rule of DeGroot. Thus, few fundamental questions naturally arise. Are these new dynamics completely undisciplined? Is it still possible to obtain convergence of opinions? Also, if the answer is yes, can we say anything on the rate of convergence and the formation of consensus? Does the crowd become wise in the limit a la Golub and Jackson [42]? The second goal of the present work is to answer these and other questions regarding our more general opinions' dynamics. In answering these questions, we systematically highlight the differences from the DeGroot's model: the benchmark for non-Bayesian opinion aggregation.

In our second main result, Theorem 2, we show that the opinions' *time averages*, induced by robust opinion aggregators, *uniformly* converge. Intuitively, the agents' updates either converge or eventually oscillate, whereas the bound on the rate of convergence is independent of the particular signals' realization.

Despite being an essential result of what an external observer can learn by observing the evolution of opinions, the convergence of time averages is usually not satisfactory for the analysis of learning and aggregation of agents' opinions. Therefore, we next look for conditions that ensure proper convergence of the iterates of our maps. In particular, we provide conditions that have transparent economic interpretations, and that can be derived from our foundations. Hence, we analyze the properties of the network structure that guarantee convergence of the limit opinions, for example:

1. **Self-influence**: For every profile of opinions, the update of each agent is influenced by her own past opinion;

2. **Uniform common influencer**: There is at least one source of information that is trusted by the entire society;

3. **Strong connectedness**: For every pair of agents in the network, there is a sequence of agents connecting them.

In our third main result, Theorem 3, we provide minimal sufficient conditions for standard convergence: be it to consensus or not. These conditions are expressed in

terms of the network's connections and are implied by each one of the previous three properties where strong connectedness has to be paired to aperiodicity. If either the uniform common influencer property or strong connectedness is satisfied, then a consensus is reached at the limit. Under these two assumptions, we are also able to provide bounds for the corresponding rates of convergence.

As mentioned, all these conditions are related to a network structure among agents that we obtain from the opinion aggregator at hand. We show that this derived network, under mild assumptions, coincides with the primitive one disciplining which signals the agents observe. In our derived network, an agent is linked to another if the former is *always* responsive to changes in the opinion of the latter. At the same time, we highlight that it is also possible to capture different network layers by considering links that are active only under certain stances stated by the agents (e.g., some of the agents might be listened to only when they state an extreme opinion). As an implication, our analysis suggests caution in concluding that convergence to consensus depends on the *local* network properties as in DeGroot's model.

Next, we focus our attention on a subclass of robust opinion aggregators that satisfy *comonotonic additivity*. In this case, the aggregation is linear whenever restricted to comonotonic vectors of opinions. We call the elements of this class *Choquet aggregators* because they have a representation in the form of a Choquet integral. The median, together with all the quantile functions as well as the order statistics, are examples of such aggregators. Aside from the representation, these aggregators have also some useful properties and interpretations. For example, each iteration of a Choquet aggregator corresponds to a linear aggregation using a matrix that is selected from a finite set of possible alternatives. The network structure and the current ranking of opinions interact in determining which matrix is used and make these aggregators particularly suitable to model agents that overweight either the extreme or the intermediate opinions in the network.

Vox populi, vox Dei Finally,[7] we explore whether or not robust opinion aggregation allows the agents to learn the true parameter. In other words, we explore under which conditions our non-Bayesian social learning model yields the *wisdom of the crowd* a la Golub and Jackson [42]. Our findings are mixed. We can provide sufficient conditions for wisdom and we can also link them to our foundation. Also in our model, wisdom occurs when the influence of each agent vanishes in the limit. But, we find new factors, which are irrelevant in the linear model, that might prevent the wisdom of the crowd.

---

[7]Vox populi, vox Dei is a Latin sentence meaning: The voice of the people is the voice of God. It is often shortened to just "Vox populi" as in the original paper of Galton [36] on the wisdom of the crowd. Incidentally, in that paper, Galton "aggregated" opinions using the empirical median rather than the average.

For example, symmetry of the distribution of agents' signals plays a key role under general robust opinion aggregation. Similarly, even if a small fraction of agents uses a distorted aggregation procedure (e.g., they put excessive weight on extreme realizations), the long-run opinion emerging in large networks might be biased with respect to the true parameter. Finally, we note how the nonlinearity of our opinion aggregators makes higher order moments relevant for the volatility of the limit consensus.

Related literature   The taxonomy of social learning outlined in Breza, Chandrasekhar, Golub, and Parvathaneni [14] would categorize our model as a model of *aggregation* of information represented by *continuous* levels (e.g., the intensity of beliefs) in a network of agents with an *intermediate* level of sophistication.[8]

Within this class, the DeGroot's model [26] is the benchmark.[9] It is a discrete-time dynamic model where a group of agents starts with initial estimates or opinions and then periodically updates them by taking weighted averages of the estimates of their neighbors. In this simple model, there is a clear link between the properties of the underlying network structure and the long-run evolution of opinions. These features are exploited in Golub and Jackson [42] to fully characterize convergence and convergence to consensus in terms of the network structure. However, the DeGroot's model makes the somewhat unrealistic assumption that the weights assigned to agents are fixed and do not depend on the opinion stated. These and other issues are addressed in our more flexible and richer model of information aggregation in networks. For example, we show that the sufficient conditions for convergence of [42, Theorem 2] are still valid and we point out how they might fail to be necessary.

DeMarzo, Vayanos, and Zwiebel [27] provide a microfoundation of the DeGroot's model as a repeated naive *maximum likelihood estimation* procedure of an underlying parameter that captures a form of persuasion bias.[10] In their model, the linearity of aggregation crucially relies on the assumption that the error terms are normal and independent. Our approach improves on the work of DeMarzo, Vayanos, and Zwiebel in two dimensions. First, we do not impose any parametric specification, but we still encompass the Gaussian case as a particular specification.[11] Second, the iteration of

---

[8]A subclass of our opinion aggregators (see Section 5.4) are flexible enough to deal also with *discrete* states.

[9]For a comprehensive treatment of this literature see Acemoglu and Ozdaglar [1], Golub and Sadler [45], Mobius and Rosenblat [68], and the references therein.

[10]They also allow agents to vary over time the weight they give to their own past beliefs relative to the others. For the generalization of their procedure in our model see Section D.2 in the Online Appendix. Banerjee, Breza, Chandrasekhar, and Mobius [8] consider a different departure from the DeGroot's model by allowing for heterogeneity of the timing at which the agents receive their first piece of information.

[11]Note that maximum likelihood estimators fall within the class of robust estimators, that is, the

robust estimation requires a much lower degree of bounded rationality of the agents, as argued above and in Section 3.

Among the most recent papers, the one closest to ours is Molavi, Tahbaz-Salehi, and Jadbabaie [69]. The first difference concerns the stochastic component of the model. They follow Jadbabaie, Molavi, Sandroni, and Tahbaz-Salehi [52] in considering social learning when agents both repeatedly receive signals about an underlying state of the world and naively combine the beliefs of their neighbors. Instead, we follow the wisdom of the crowd approach of [42], and we study the long-run opinions as the size of the society grows to infinity. The second difference regards the direction of the relaxation of the linearity in the naive-updating rules of the agents. Both papers take an axiomatic approach, postulating some properties of the opinion aggregators, the main differences being between the assumptions of translation invariance and label neutrality. In the Online Appendix, we show that for the questions we explore, i.e., the convergence of limit opinions and the wisdom of the crowd, log-linear aggregators a la [69] can be studied in an equivalent linear system, thus making use of the results developed for DeGroot's model and its time-varying versions. Since our class of robust opinion aggregators encompasses the linear model, our results cover their aggregators too. However, notice that the equivalence with a linear system may be lost for a problem of learning with repeated signals like the one they analyze in their paper.

Our results and their proofs relate to three different strains of literature in Mathematics: namely, discrete dynamical systems, fixed points approximation, and nonexpansive selfmaps (i.e., selfmaps which are Lipschitz continuous of order 1). We provide here a brief overview. The literature on discrete dynamical systems/repeated averaging shares a common theme. Agents aggregate opinions at each point in time following DeGroot's rationale, but with a potential time-varying averaging procedure which could be linear or not. These works are typically concerned in providing the more general conditions possible on the sequence of averaging procedures which guarantee convergence to consensus. One of the first papers dealing with such a problem is Chatterjee and Seneta [24] where agents aggregate opinions using time-varying aggregation matrices. Krause [56] provides an excellent textbook exposition of the topic and a full characterization of convergence to consensus.[12] In a nutshell, [56, Theorem 8.3.4] shows that convergence to consensus is achieved if and only if a form of strict internality is satisfied, that is, the range of opinions of the agents eventually shrinks no matter what is the initial vector of opinions. Our results differ from the ones above in two dimensions. First, in our Theorems 2 and 3, we tackle the issue of convergence in general and we do not restrict

_____

general class considered in this paper. In the Online Appendix, we also rationalize robust opinion aggregation as a repeated maximum likelihood estimation.

[12]Both this monograph and the survey [1] provide detailed references about the generalizations of DeGroot's model in the engineering and computer science literature.

ourselves just to convergence to consensus. This significantly complicates the analysis and we need to resort to completely different techniques coming from functional analysis and which we discuss below. The overlap with Krause's result is therefore restricted to Corollaries 2–4 and limited to the part of convergence. As we comment in Remark 1, the convergence part of these results can be easily obtained from our more general Theorem 3, rather than proving any form of strict internality a la Krause. At the same time, this brings us to the second difference. Since our opinion aggregators are microfounded, under mild conditions, they inherit the primitive network structure of the foundation. In turn, this imposes a strong discipline on the averaging process and the corresponding sequence of replicating matrices. This fact was never fully exploited before. Mathematically, this turns out to significantly simplify our proofs, which for these results rely on a combination of operatorial arguments and simple Markov chains techniques. Moreover, it allows us to provide bounds on the rate of convergence which are function of the underlying network. Finally, Krause's results are related to an underlying network structure by Muller-Frank [71]. The results obtained in [71] rely on a form of internality to the opinions in the neighborhood of each agent. Our opinion aggregators, in general, do not have any strict internality property. However, note that part of our Corollary 4 can also be obtained from Theorem 4 in [71].

The other two literatures relevant to our work are the ones about nonexpansive maps and fixed points approximation. The goal there is to find solutions to functional equations induced by operators or more in general fixed points of operators which are not contractions. We exploit some of the techniques originating from these literatures and in Appendices B and D.4, in several remarks (see Remarks 5–7 and 9), we comment more in detail about the differences and the common aspects with our work. In particular, we comment on the relation with the work of Baillon, Bruck, and Reich [7] on the Cesaro convergence of the iterates of nonexpansive selfmaps. More in general, even though we deal with issues present also in the two aforementioned literatures, our opinion aggregators have properties which do not seem to have been studied in these fields and which in turn allow us to obtain either sharper or novel results.

Our results also make use of techniques and concepts coming from decision theory. The three papers which are mostly related to our work are Ghirardato, Maccheroni, and Marinacci [41], Maccheroni, Marinacci, and Rustichini [63], and Schmeidler [79]. The first two papers were the first to study functionals which satisfy the properties of normalization, monotonicity, and translation invariance and to fruitfully use nonstandard differential techniques. In our case, these techniques turn out to be extremely useful when we discuss the wisdom of the crowd. The third paper was the first one to study comonotonic additive functionals.

Outline   The paper is structured as follows. Section 2 presents the definitions of the mathematical objects used in our analysis. Section 3 introduces our estimation model and characterizes the class of robust opinion aggregators. In Section 4, we provide some illustrative examples. Section 5 describes the long-run evolutions of opinions and the properties of the limit. Section 6 explores the conditions for obtaining the wisdom of the crowd a la Golub and Jackson [42]. Finally, Section 7 discusses some of our main assumptions and how to relax them. Most of the proofs are in the Appendix, except for some instrumental results, whose proofs are relegated to the Online Appendix.

2   Preliminaries

Consider a finite set of agents $N = \{1, ..., n\}$. We denote by I a closed interval of $\mathbb{R}$ with nonempty interior. For example, if $I = [0, 1]$, then we interpret a number in this interval as either a measurement of agreement on a particular instance or a subjective probability about a specific event. In what follows, we study selfmaps $T : B \to B$ where $B = I^n$. We call these selfmaps *opinion aggregators*. Given an opinion aggregator T and $i \in N$, we denote the i-th component of T by T_i. In other words, given $x \in B$, $T_i(x)$ is the i-th component of the vector $T(x)$. With a small abuse of notation, we denote by the letter I two objects: a closed interval with nonempty interior and the identity map $I : B \to B$. The context will always clarify unambiguously to which object we are referring to.

Given two vectors $x, y \in \mathbb{R}^n$,[13] recall that they are comonotonic if and only if $[x_i - x_j][y_i - y_j] \geq 0$ for all $i, j \in N$. By $e \in \mathbb{R}^n$, we denote the vector whose components are all 1s. We denote by Δ the collection of probability vectors in $\mathbb{R}^n$, that is, $p \in \Delta$ if and only if $p_i \geq 0$ for all $i \in N$ and $\sum_{i=1}^n p_i = 1$. We endow B with the topology induced by the supnorm $\| \ \|_\infty$. Given an opinion aggregator $T : B \to B$ and $x \in B$, the sequence $\{T^t(x)\}_{t \in \mathbb{N}}$ will be called the *sequence of updates of x*.

Among other things, we are concerned about the limit and the rate of convergence of these sequences. We will be dealing with two kinds of limit: the standard one induced by the supnorm as well as the one of Cesaro, that is,

$$C - \lim_t T^t(x) \stackrel{\text{def}}{=} \lim_\tau \frac{1}{\tau} \sum_{t=1}^\tau T^t(x)$$

where the limit on the right-hand side of the definition is the standard limit.

We denote by $\mathcal{W}$ the collection of stochastic matrices, that is, all $n \times n$ square matrices whose entries are positive and rows sum up to 1.

We say that an opinion aggregator T is:

---

[13]In this paper, vectors will always be column vectors, unless otherwise specified.

1. *normalized* if and only if $T(ke) = ke$ for all $k \in I$;

2. *monotone* if and only if for each $x, y \in B$

$$x \geq y \implies T(x) \geq T(y);$$

3. *translation invariant* if and only if

$$T(x + ke) = T(x) + ke \qquad \forall x \in B, \forall k \in \mathbb{R} \text{ s.t. } x + ke \in B;$$

4. *constant affine* if and only if

$$T(\lambda x + (1 - \lambda) ke) = \lambda T(x) + (1 - \lambda) ke \qquad \forall x \in B, \forall k \in I, \forall \lambda \in [0, 1];$$

5. *comonotonic additive* if and only if

$$x \text{ and } y \text{ comonotonic} \implies T(x + y) = T(x) + T(y);$$

6. *linear* if and only if there exists a matrix $W \in \mathcal{W}$ such that

$$T(x) = Wx \qquad \forall x \in B;$$

7. *odd* if and only if

$$T(-x) = -T(x) \qquad \forall x \in B \text{ s.t. } -x \in B;$$

Our foundations yield opinion aggregators that have the following properties: normalization, monotonicity, and translation invariance.

Definition 1 *Let T be an opinion aggregator. We say that T is* robust *if and only if T is normalized, monotone, and translation invariant.*

We call these aggregators robust for two reasons: 1) our foundation builds on the theory of robust statistics (see Section 3 and Theorem 1), 2) more in general, our foundation naturally generalizes the one of the linear model, without taking a parametric approach, that is without committing to any specific functional form.

A final important mathematical object is the notion of directed graph/network. We opt for a matrix representation. So a network will be the set of agents N paired with an adjacency matrix A, that is, a matrix such that $a_{ij} = 1$ if there is a directed arc from i to j, and $a_{ij} = 0$ otherwise.

3   The model

We assume that the agents in our population N try to estimate a fundamental parameter $\mu \in \mathbb{R}$. Each agent $i \in N$ initially observes a signal

$$X_i^0(\omega) = \mu + \varepsilon_i(\omega) \tag{1}$$

where $\varepsilon_i : \Omega \to \mathbb{R}$ is a random variable defined over a common probability space $(\Omega, \mathcal{F}, P)$. We assume that $X_i^0(\omega) \in I$ for all $i \in N$ and for all $\omega \in \Omega$. The period-0 estimate of each agent i coincides with the realization $X_i^0(\omega)$ of her signal that, for simplicity, we denote by x_i^0.

In period 1, the agents communicate with each other to acquire new information on the parameter μ. We model the communication through a directed network (N, A). In particular, each agent i collects the sample of realizations of signals in her neighborhood $N_i = \{j \in N : a_{ij} = 1\}$ and then solves an estimation problem about μ, based on these data. We consider a generalization of the class of M-estimators for location parameters considered in Huber [48].[14] Formally, we endow each agent i with a lower semicontinuous loss function $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ and we assume she solves

$$\min_{c \in \mathbb{R}} \phi_i\left(x^0 - ce\right) \tag{2}$$

where $x^0 = \left(x_j^0\right)_{j=1}^n$.[15] Given the profile of loss functions $\phi = (\phi_i)_{i=1}^n$, the updates x^1 at period 1 belong to the set

$$\mathbf{T}^\phi\left(x^0\right) = \prod_{i=1}^n \mathrm{argmin}_{c \in \mathbb{R}} \, \phi_i\left(x^0 - ce\right). \tag{3}$$

In all of our results below, the assumptions on ϕ will guarantee the "internality" of $\mathbf{T}^\phi$ and its nonemptiness, that is, $\emptyset \neq \mathbf{T}^\phi(x) \subseteq B$ for all $x \in B$. Therefore, (3) defines an updating correspondence $\mathbf{T}^\phi : B \rightrightarrows B$ that satisfies a property of translation invariance (see, e.g., [48])

$$\mathbf{T}^\phi(x + ke) = \mathbf{T}^\phi(x) + ke \qquad \forall x \in B, \forall k \in \mathbb{R} \text{ s.t. } x + ke \in B$$

---

[14] Our generalization falls within the class of extremum estimators.

[15] The network structure (N, A) can be reflected in the profile of loss functions $\phi = (\phi_i)_{i=1}^n$ by assuming that for each $i \in N$ and for each $z, z' \in \mathbb{R}^n$

$$z_j = z_j' \qquad \forall j \in N_i \implies \phi_i(z) = \phi_i(z').$$

It is a natural assumption since $\phi_i(z)$ of each agent i is solely influenced by the components z_j where j is in the neighborhood of i (see, e.g., Babichenko and Tamuz [6]). Nevertheless, it is not important from a mathematical point of view and it can be dispensed with. Finally, throughout the rest of the paper, we maintain the semicontinuity assumption of each ϕ_i.

where $\mathbf{T}^{\phi}(x) + ke$ is the set of vectors in $\mathbf{T}^{\phi}(x)$ shifted by ke. We define by $T^{\phi} : B \to B$ an arbitrary selection of $\mathbf{T}^{\phi}$, that is, a function such that $T^{\phi}(x) \in \mathbf{T}^{\phi}(x)$ for all $x \in B$. With a standard abuse of notation, we identify single-valued correspondences with functions and, in this case, write $\mathbf{T}^{\phi} = T^{\phi}$.

The minimization problem in (2) has the following interpretation: agent i optimally selects the estimate for μ as to minimize a loss function of the induced residuals' vector $\epsilon = x^0 - ce$. In particular, the function ϕ_i represents a belief-free form of the ex-ante information of agent i about both the network structure and the objective distribution of errors (both potentially unknown to each agent apart from their neighborhoods N_i). For example, if i believes that the signal of j is highly informative, then her loss function ϕ_i will penalize relatively more the residual $\epsilon_j = x_j^0 - c$. Here, we implicitly assume that the complexity of the environment does not allow the agents to attach probabilistic beliefs to the data generating process, including the network structure. This assumption is backed by the empirical evidence reported in Breza, Chandrasekhar, and Tahbaz-Salehi [15].

In the subsequent periods, the agents do not receive any additional external information on μ, but rather keep iterating the same estimation procedure for a new set of data points given by the last-period estimates of their neighbors. Formally, we have that $x^t \in \mathbf{T}^{\phi}(x^{t-1})$ for all periods $t \in \mathbb{N}$. In particular, whenever $\mathbf{T}^{\phi}$ is single-valued (denoted by T^{ϕ}), the deterministic dynamics of the estimates in the population given the initial realizations' vector x^0, are described by the iteration of the operator $T^{\phi} : B \to B$ at x^0, that is,

$$\left\{ x^t \right\}_{t \in \mathbb{N}} = \left\{ \left(T^{\phi} \right)^t \left(x^0 \right) \right\}_{t \in \mathbb{N}}.$$

A justification of this iteration procedure with a quadratic ϕ has been proposed by DeMarzo, Vayanos, and Zwiebel [27] and is related to a form of persuasion bias. Under this interpretation, the agents ignore the information redundancies in their neighbors' estimates and consider what they observe as brand new information. Despite the convincing arguments presented in [27] in favor of this kind of behavior, this interpretation requires a certain degree of bounded rationality. Indeed, their estimation approach is optimal when the agents know the specific parametric form of the error terms. However, this form is lost after the first round of estimation. In contrast, robust estimation does not depend on any parametric specification of the errors. To see this, assume for simplicity that the profile of loss functions ϕ is such that $\mathbf{T}^{\phi} = T^{\phi}$ is single-valued and T is translation invariant (cf. Theorem 1). If we define $\varepsilon^1 = T^{\phi}(\varepsilon)$, then the random vector describing the profile of period 1 estimates is given by

$$X^1 = T^{\phi}\left(X^0 \right) = T^{\phi}\left(\mu e + \varepsilon \right) = \mu e + \varepsilon^1.$$

Given that, at period 1, each agent i observes the realizations of the estimates in her neighborhood $\left(X_j^1\right)_{j \in N_i}$, it follows that the agents face a new location experiment with error terms equal to $\varepsilon^1 = (\varepsilon_i^1)_{i=1}^n$. The uncertainty about the distribution of ε^1 is qualitatively similar to the one of the previous period. Due to this similarity, it seems then natural that each i would repeat the same estimation procedure of period 1 in all the subsequent periods. The random vector describing the t-period agents' estimates is $X^t = \left(T^\phi\right)^t (X^0) = \mu e + \varepsilon^t$ and is the result of a location experiment with errors $\varepsilon^t = \left(T^\phi\right)^t (\varepsilon)$. Therefore, for each initial realization x^0 of signals X^0, the dynamics followed by the agents' estimates are described by $\{x^t\}_{t \in \mathbb{N}}$.

Finally, we note that the updating procedure proposed by DeMarzo, Vayanos, and Zwiebel is easily nested in our framework by considering quadratic loss functions. Formally, each agent i minimizes

$$\phi_i (x - ce) = \sum_{j=1}^n w_{ij} (x_j - c)^2 = \sum_{j \in N_i} w_{ij} (x_j - c)^2 \tag{4}$$

for some vector of weights $w_i \in \Delta$ such that $w_{ij} = 0$ for all $j \notin N_i$. In their context, w_{ij} represents the subjective belief of i about the precision of the j-th signal.

3.1 Robust opinion aggregators: a foundation

Here, we study the general properties of ϕ that *characterize* robust opinion aggregators. The following definition captures the most elementary form of trust in the signals observed.[16]

Definition 2 *The profile of loss functions ϕ is* sensitive *if and only if $\phi_i (he) > \phi_i (0)$ for all $i \in N$ and for all $h \in \mathbb{R} \backslash \{0\}$.*

In words, if agent i observes a unanimous opinion (including herself) her loss is minimized by declaring this same opinion. Indeed, if residuals are zero, then the agent is perfectly matching all the observations, thus minimizing the loss. The next definition is a form of complementarity in disagreeing with two or more agents from the same side.

Definition 3 *The profile of loss functions ϕ has* increasing shifts *if and only if for each $i \in N$, $z, v \in \mathbb{R}^n$, and $h \in \mathbb{R}_{++}$*

$$z \geq v \implies \phi_i (z + he) - \phi_i (z) \geq \phi_i (v + he) - \phi_i (v).$$

---

[16]With a small abuse of terminology, sometimes we will either say that a profile of loss functions is sensitive (resp., lower semicontinuous, robust, etc) or that it is a profile of sensitive (resp., lower semicontinuous, robust, etc) loss functions.

It has strictly increasing shifts *if and only if the above inequality is strict whenever* $z \gg v$.

We consider the property of increasing shifts because it is very permissive and naturally emerges from the characterization of Theorem 1. Intuitively, it requires that a raise in the agent i own estimate by h induces a higher loss the lower are the opinions of the other agents. It is implied by stronger properties usually required on games played on networks, such as supermodularity and degree complementarity (see, e.g., Galeotti, Goyal, Jackson, Vega-Redondo, and Yariv [38]). Moreover, we will see momentarily how this corresponds to the generalization of the convexity assumption imposed in robust statistics.

We call *robust* a profile of lower semicontinuous loss functions $\phi = (\phi_i)_{i=1}^n$ which is sensitive and has increasing shifts. The collection of all these profiles is denoted by Φ_R. Next, we formalize the relationship between robust opinion aggregators and profiles of robust loss functions.

Theorem 1 *Let T be an opinion aggregator. The following statements are equivalent:*

(i) *There exists $\phi \in \Phi_R$ which has strictly increasing shifts and is such that $T = T^\phi$, that is, for each $i \in N$*

$$T_i(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce) \qquad \forall x \in B; \tag{5}$$

(ii) *T is a robust opinion aggregator.*

Our first theorem provides a foundation for robust opinion aggregation as the result of an estimation process. The foundation is tight since, at the same time, each estimation process as in (5) yields a robust opinion aggregator. In Theorem 1, the property of strictly increasing shifts guarantees that $\operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce)$ is a singleton. This is a desirable property, nonetheless it is violated for few interesting specifications of ϕ (see, e.g., (7) where each agent minimizes the absolute deviations). In Appendix A (see Proposition 12), we treat the more general case, showing that the correspondence $\mathbf{T}^\phi$, defined as in (3), always admits a selection which is a robust opinion aggregator.

Even though sensitivity and (strictly) increasing shifts are the properties characterizing robust opinion aggregation, it might not be immediate to verify that a given profile of loss functions satisfies them. The following result is a useful tool which allows to recognize loss functions that induce robust opinion aggregators. Let Φ_R^* denote the set of profiles of continuous, convex, and supermodular loss functions which are sensitive. Lemma 1 proves that $\Phi_R^* \subseteq \Phi_R$. Moreover, in Proposition 2, we discuss an important class of natural loss functions which turn out to belong to the class Φ_R^*.

Lemma 1 *Let $\phi = (\phi_i)_{i=1}^n$ be a profile of loss functions. If $\phi \in \Phi_R^*$, then $\phi \in \Phi_R$. Moreover, if ϕ_i is strictly convex for all $i \in N$, then ϕ has strictly increasing shifts.*

Typically, loss functions can satisfy extra properties that might be natural in light of the estimation procedure at hand. In the next few results, we study how these properties translate into additional features of T, where the latter play a key role in establishing convergence of the sequence of updates (cf. Proposition 4 and Section 6). We begin by studying two properties of homogeneity.

Definition 4 *The profile of loss functions ϕ is* positively homogeneous *if and only if for each $i \in N$ there exists a positive function $\eta_i : \mathbb{R}_+ \to \mathbb{R}_+$ such that*

$$\phi_i(\lambda z) = \eta_i(\lambda) \phi_i(z) \qquad \forall z \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}_+.$$

For instance, positive homogeneity is satisfied whenever loss functions are seminorms (or suitable monotone transformations). For example, this is the case of either the absolute deviation or the quadratic loss function.

Definition 5 *The profile of loss functions ϕ is* symmetric *if and only if for each $i \in N$*

$$\phi_i(z) = \phi_i(-z) \qquad \forall z \in \mathbb{R}^n.$$

Symmetry is always satisfied when the loss functions only depend on the absolute value of the residuals and not on their sign.

Proposition 1 *Let T be an opinion aggregator. The following facts are true:*

1. *T is a constant affine robust opinion aggregator if and only if there exists a positively homogeneous $\phi \in \Phi_R$ with strictly increasing shifts such that $T = T^\phi$, that is, for each $i \in N$*

$$T_i(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce) \qquad \forall x \in B.$$

2. *T is an odd robust opinion aggregator if there exists a symmetric $\phi \in \Phi_R$ with strictly increasing shifts such that $T = T^\phi$, that is, for each $i \in N$*

$$T_i(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce) \qquad \forall x \in B.$$

As noted for Theorem 1, also here in proving the "if" parts, we can dispense with the assumption of strictly increasing shifts (see Lemma 7). In this case, $\operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce)$ fails to be a singleton, but $\mathbf{T}^\phi$ admits a selection which is constant affine (resp., odd). The opinion aggregators induced by seminorms as well as those in which each component is a quantile, like the median, are constant affine. Another interesting case are Choquet aggregators (see Section 5.4).

L-estimators An interesting class of estimators which yield robust opinion aggregators are L-estimators. When using an L-estimator, agent i takes a convex linear combination of the order statistics in the observed sample. These estimators seem particularly appealing for two reasons. First, they are robust to misspecification of the data generating process of the observed opinions. Second, they are straightforward to compute since they consist of a weighted average procedure. In addition to these desirable features, they allow for descriptively relevant biases in information aggregation in a network framework. Indeed, they can be naturally used to capture the overweighting (as well as the neglecting) of extreme realizations (see Section 5.4 and Example 6). Finally, we also note that L-estimators are comonotonic additive. Moreover, it is particularly easy to link their behavior to the network structure.[17] These mathematical properties simplify the analysis of their long-run limit.

Lemma 2 *If each agent $i \in N$ uses an L-estimator, the induced opinion aggregator is robust and comonotonic additive.*

In the Online Appendix, we derive sufficient conditions under which also maximum likelihood and Bayesian estimators are robust opinion aggregators.

3.2 Independent signals and additive separable loss functions

In this section, following Huber [48], we assume that the agents commonly know that the errors $\varepsilon = (\varepsilon_i)_{i=1}^n$ are independently and symmetrically distributed according to some objective distribution, where the latter is still unknown from their perspective. In this case, at period 1, every agent i solves an estimation problem with the sample $(X_j)_{j \in N_i}$ that she perceives as independent. Each agent i uses an estimator T_i^ϕ that, for each signals' realization x^0, solves

$$\min_{c \in \mathbb{R}} \phi_i \left(x^0 - ce \right) = \min_{c \in \mathbb{R}} \sum_{j \in N_i} \rho_i \left(x_j^0 - c \right) \tag{6}$$

where $\rho_i : \mathbb{R} \to \mathbb{R}_+$ is continuous, convex, strictly decreasing on $\mathbb{R}_-$ and strictly increasing on $\mathbb{R}_+$. This shows that the above problem (2) is a proper generalization of the estimation method proposed by Huber. Indeed, most of the loss functions used in robust statistics satisfy the properties we studied above: e.g., the quadratic loss $\rho_i(s) = s^2$, the absolute loss $\rho_i(s) = |s|$, the p-loss $\rho_i(s) = |s|^p$ with $p \geq 1$, and the Huber loss defined as

$$\rho_i(s) = \begin{cases} s^2 & \text{if } |s| \leq k \\ 2k|s| - k^2 & \text{if } |s| > k \end{cases}$$

---

[17]See Proposition 8 and Example 6.

where $k > 0$.

Even if in a network setting the agents should take into account that the information they receive is correlated, their imperfect knowledge of the correlation structure may lead them to use an additively separable loss function. For this class, it is often easier to recognize when our properties hold.

Definition 6 *The profile of loss functions ϕ is* additively separable *if and only if there exist a stochastic matrix $W \in \mathcal{W}$ and a profile of lower semicontinuous functions $\rho = (\rho_i : \mathbb{R} \to \mathbb{R}_+)_{i=1}^n$ such that for each $i \in N$*

$$\phi_i(z) = \sum_{j=1}^n w_{ij} \rho_i(z_j) \qquad \forall z \in \mathbb{R}^n.$$

Problem (6) coincides with the case in which each agent gives uniform weight to each signal in her neighborhood: a natural assumption when signals are also perceived as identically distributed.[18] At the same time, it is reasonable to expect that agents might perceive the signal of some of their neighbors as either more or less precise. This can be captured by the weights w_{ij} not being uniform over N_i. As we discuss in the Online Appendix, this also is the result of a maximum likelihood estimation procedure.

Given a profile of additively separable loss functions ϕ, we often identify it with the corresponding pair: $\phi = (W, \rho)$. We denote the set of profiles of *robust additively separable* loss functions with Φ_A. The following proposition characterizes the elements of Φ_A.

Proposition 2 *Let $W \in \mathcal{W}$ and $\rho = (\rho_i : \mathbb{R} \to \mathbb{R}_+)_{i=1}^n$. The following statements are equivalent:*

(i) $(W, \rho) \in \Phi_A$;

(ii) $(W, \rho) \in \Phi_A \cap \Phi_R^*$;

(iii) ρ_i is convex, strictly decreasing on $\mathbb{R}_-$, and strictly increasing on $\mathbb{R}_+$ for all $i \in N$.

Thus, if each agent uses an additively separable loss function where ρ_i is like in (iii), the resulting opinion aggregator is robust (see Theorem 1 and Proposition 12; the former covers the case in which ρ_i is strictly convex while the latter covers the general case).

---

[18]The profile of loss functions is additively separable where $w_{ij} = a_{ij}/\sum_{l=1}^n a_{il}$ for all $i, j \in N$. Clearly, for W to be a well defined stochastic matrix, one needs to make the economically natural assumption that each agent is connected to at least another agent, that is, for each $i \in N$ there exists $l \in N$ such that $a_{il} > 0$.

4 Examples

Median and quantiles In the framework of robust estimation, assume that agents use the following profile of additively separable loss functions

$$\phi_i(z) = \sum_{j=1}^{n} w_{ij} |z_j| \qquad \forall z \in \mathbb{R}^n \tag{7}$$

where the values w_{ij} are the entries of a stochastic matrix $\mathcal{W}$. By Proposition 2, observe that $\phi = (\phi_i)_{i=1}^n$ is a profile of robust loss functions. In this case, agents minimize the weighted absolute deviations. It is well known that the solution correspondence $\mathbf{T}^\phi$ in (2) admits as a selection T^ϕ, defined by

$$T_i^\phi(x) = \min \left\{ c \in \mathbb{R} : \sum_{j:x_j \leq c} w_{ij} \geq 0.5 \right\} \qquad \forall x \in \mathbb{R}^n, \forall i \in N. \tag{8}$$

Clearly, $T_i^\phi(x)$ is the (weighted) median of x. Our results yield that T^ϕ is a robust opinion aggregator. This is a very minor departure from the quadratic losses approach, yet it induces dramatically different dynamics as we next discuss (see also Section 5.4).[19]

Example 1 A group of agents $N = \{1, 2, 3, 4\}$ share their opinions $x^0 \in B = [0, 1]^4$. The weights assigned to the other agents are represented by the matrix

$$W = \begin{pmatrix} 0.4 & 0.3 & 0.3 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \\ 0.1 & 0.1 & 0.2 & 0.6 \\ 0 & 0 & 0.6 & 0.4 \end{pmatrix}.$$

where the entry in row i, column j, is the value w_{ij} above. It is immediate to see that aggregation through weighted *averages* would achieve consensus in the limit (see, e.g., Golub and Jackson [42, Proposition 1]).[20] However, the dynamics induced by using the median are qualitatively different.

- If $x^0 = (0, 1, 1, 1)$, then the block of agents agreeing on the higher opinion is sufficiently large to attract agent 1 to the same opinion, and the limit (consensus) opinion of $(1, 1, 1, 1)$ is reached in one round of updating.

---

[19]For a similar but asynchronous updating rule see Mei, Bullo, Chen, and Dorfler [65].
[20]That is, $\lim_t T^t(x) = \lim_t W^t x$ exists for all $x \in B$. Moreover, the limit is a constant vector (i.e., consensus).

- However, the prediction of consensus is lost if the initial opinion of player 2 is slightly lowered. Let $x^0 = (0, 1/2, 1, 1)$, then their first round of updating gives $x^1 = (1/2, 1/2, 1, 1)$, and this polarization will be the limit outcome: a strongly connected society fails to reach consensus without a sufficiently large block of an initial agreement.

- Finally, consider $x^0 = (0, 1/2, 0, 1)$. Then the agents' first update is $x^1 = (0, 0, 1, 0)$ and agents 1 and 2 will never change opinion again, whereas agents 3 and 4 will keep reciprocally switching their opinions. This shows that also convergence may not be guaranteed. At the same time, this latter feature allows us to say that the *time averages* of the updates converge, that is,

$$\lim_\tau \frac{1}{\tau} \sum_{t=1}^\tau T^t \left(x^0 \right) = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

In Theorem 2, we show that each robust opinion aggregator is convergent in this weaker sense, no matter what is the initial condition. ▲

Quantiles Another restriction imposed by a quadratic loss as in (4) is that upward and downward discrepancies from the observed opinions are felt as equally harming by every agent. However, it might well be the case that (some) agents dislike more one or the other. One easy example of this kind of behavior is the asymmetric version of the absolute deviations in (7). Formally, for each $i \in N$ we consider

$$\phi_i \left(z \right) = \alpha_i \sum_{j : z_j \geq 0} w_{ij} z_j + (1 - \alpha_i) \sum_{j : z_j < 0} w_{ij} \left(-z_j \right) \qquad \forall z \in \mathbb{R}^n, \tag{9}$$

where $\alpha_i \in (0, 1)$. If $\alpha_i = 1/2$, then the loss function in (9) reduces, up to a multiplicative factor, to the one in (7). By Proposition 2, observe that $\phi = (\phi_i)_{i=1}^n$ is a profile of robust loss functions.[21] It is well known that the solution correspondence $\mathbf{T}^\phi$ in (2), when each agent uses a loss function as in (9), admits as a selection the (weighted) quantile function for the distribution of observed opinions.[22] Quantile functions capture

---

[21]Observe that $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ can be written as

$$\phi_i \left(z \right) = \sum_{j=1}^n \alpha_i w_{ij} \max \left\{ z_j, 0 \right\} + \sum_{j=1}^n (1 - \alpha_i) w_{ij} \max \left\{ -z_j, 0 \right\}$$

$$= \sum_{j=1}^n w_{ij} \left(\alpha_i \max \left\{ z_j, 0 \right\} - (1 - \alpha_i) \min \left\{ z_j, 0 \right\} \right) = \sum_{j=1}^n w_{ij} \rho_i \left(z_j \right) \qquad \forall z \in \mathbb{R}^n$$

where $\rho_i \left(s \right) = \alpha_i \max \left\{ s, 0 \right\} - (1 - \alpha_i) \min \left\{ s, 0 \right\}$ for all $s \in \mathbb{R}$. It is easy to see that ρ_i satisfies the properties of point (iii) of Proposition 2 for all $i \in N$.

[22]More formally, the selection T^ϕ is defined as in (8) where 0.5 is replaced by α_i for all $x \in B$ and for all $i \in N$.

19

the behavior of agents who have a bias in favor of relatively extreme stances (α_i close to 0 or 1) or relatively moderate ones (α_i close to $1/2$). Our results yield that T^ϕ is a robust opinion aggregator (see also Section 5.4).

A quasi-arithmetic mean The quantile functionals are nondifferentiable and, even though we do not rely on differentiability properties for our main results, smooth aggregators might be more easily analyzed in applications. A smooth and tractable robust opinion aggregator is obtained by considering the following loss function

$$\phi_i^\lambda (z) = \sum_{j=1}^{n} w_{ij} \left[\exp \left(\lambda z_j \right) - \lambda z_j \right] \qquad \forall z \in \mathbb{R}^n \tag{10}$$

where $\lambda \neq 0$ and the values w_{ij} are the entries of a stochastic matrix $\mathcal{W}$. In particular, whenever $\lambda > 0$, upward deviations from i's current opinion c are more penalized than downward deviations and vice versa whenever $\lambda < 0$. Figure 1 compares the quadratic loss with the asymmetric one with $\lambda = 1$. Some interesting comparative statics hold
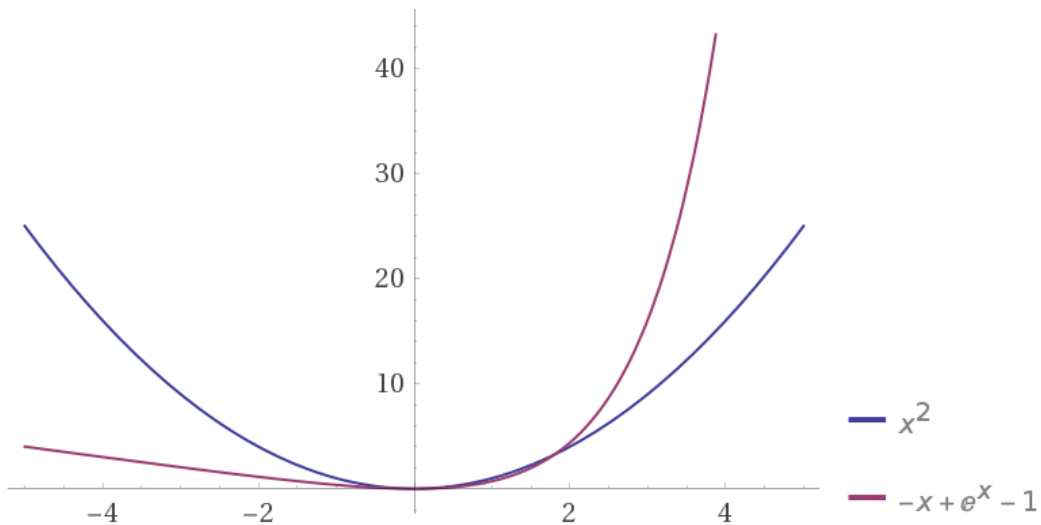

Figure 1: Smooth asymmetric loss

for the loss function in (10). We first present the formal result and then discuss its interpretation. The next proposition provides a glimpse into our more general findings about long run opinions and how they differ from the linear case. First, by Proposition 2, observe that $\phi = \left(\phi_i^\lambda \right)_{i=1}^{n}$ is a profile of robust loss functions. Since $\phi = \left(\phi_i^\lambda \right)_{i=1}^{n}$ has strictly increasing shifts too, for ease of notation, we denote the unique selection of $\mathbf{T}^\phi$ by T^λ.

Proposition 3 *Let ϕ be the profile of loss functions $\left(\phi_i^\lambda \right)_{i=1}^{n}$ as in (10) with $W \in \mathcal{W}$ and $\lambda \in \mathbb{R} \backslash \{0\}$. The following statements are true:*

20

1. *For each $i \in N$ we have that*

$$T_i^\lambda(x) = \frac{1}{\lambda} \ln \left(\sum_{j=1}^n w_{ij} \exp(\lambda x_j) \right) \qquad \forall x \in B. \tag{11}$$

2. *For each $i \in N$ we have that*

$$\lim_{\lambda \to \hat{\lambda}} T_i^\lambda(x) = \begin{cases} \max_{j:w_{ij}>0} x_j & \text{if } \hat{\lambda} = \infty \\ \sum_{j=1}^n w_{ij} x_j & \text{if } \hat{\lambda} = 0 \\ \min_{j:w_{ij}>0} x_j & \text{if } \hat{\lambda} = -\infty \end{cases} \qquad \forall x \in B.$$

3. *If there exists a vector $s \in \Delta$ such that*

$$\lim_{t \to \infty} W^t x = \left(\sum_{i=1}^n s_i x_i \right) e \qquad \forall x \in \mathbb{R}^n, \tag{12}$$

then we have that $\bar{T}^\lambda : B \to B$, defined by

$$\bar{T}^\lambda(x) = \frac{1}{\lambda} \ln \left(\sum_{i=1}^n s_i \exp(\lambda x_i) \right) e \qquad \forall x \in B,$$

is such that

(a) *$\bar{T}^\lambda(x) = \lim_t \left(T^\lambda \right)^t(x)$ for all $x \in B$.*

(b) *$\frac{\partial \bar{T}_i^\lambda}{\partial x_j}(x) = \frac{s_j \exp(\lambda x_j)}{\sum_{l=1}^n s_l \exp(\lambda x_l)}$ for all $x \in B$ and for all $i, j \in N$.*

Point 1 gives an explicit functional form for the opinion aggregator. Point 2 shows that this functional form encompasses the linear case as a limit, but also allows for behavior which is nonneutral toward the direction of disagreement. In point 3, we see another prediction of the linear model getting reversed. It is not just the network structure W that determines the limit influence of each agent, but the initial opinion also plays a key role. Indeed, when $\lambda > 0$, the higher the initial signal realization of an individual is, the higher is her marginal contribution to the limit. This fact has extremely relevant consequences. For example, consider one of the classical applications of non-Bayesian learning, technology adoption in a village of a developing country, with an opinion vector representing how much the agents have invested in the new technology (e.g., the share of land cultivated with the new technology). There $\lambda > 0$ captures the idea that the most innovative members of the society have a disproportionate influence on the others, maybe because their performance attracts relatively more attention. In that case, if resources are limited, i.e., if the external actor can only increase adoption for an agent directly, relying on the network diffusion for the rest, the policy prescription is qualitatively different. Indeed, she should choose the agent j for which the index

in point 3.b is maximized, combining the standard eigenvector centrality s_j with a distortion increasing in the initial opinion of agent j. Finally, one can easily show that, given two agents i and j sharing the same influence under the linear model, that is such that $s_i = s_j$, if their initial opinions are more dispersed (resp., more concentrated), then the limit consensus is higher (lower) when $\lambda > 0$.[23]

5   Convergence

In the rest of the paper, we analyze the dynamics induced by iterated robust estimation. From the point of view of an external analyst, these dynamics are stochastic as they depend on the realizations of the initial signals. To address this issue, we follow a two-step analysis. First, in this section, we analyze the opinions' convergence for a given population size. Our strategy is to study the global properties (i.e., properties that hold for every realization of the signals) of the corresponding deterministic dynamical system. It follows that the same properties are inherited by the original stochastic system of opinion dynamics. In doing so, we also highlight some of the local properties of robust opinion aggregation that are qualitatively different from the linear case. For example, the extent of polarization, whether or not consensus is attained in the limit, and the influence of each individual might depend on the initial distribution of opinions within the society (see Example 1 and Proposition 3). Second, in Section 6, in the spirit of the wisdom of the crowd a la Golub and Jackson [42], we let the size n of the network grow in order to analyze the asymptotic properties of the consensus opinion emerging from aggregation. Also in this case, we stress the differences between robust and linear opinion aggregation.

We begin by showing that robust opinion aggregators always induce a weaker form of opinions' convergence: namely, their time averages convergence. Building on this result, we give general conditions under which standard convergence is obtained, and we study the stability and consensus properties of the limit.

5.1   The convergence of time averages

Section 3 shows that a robust statistical foundation of opinion aggregation leads to aggregators that typically are not linear. In light of this, given an initial opinions' vector x^0 the study of its evolution via the sequence of updates $\{T^t(x^0)\}_{t \in \mathbb{N}}$ cannot rely on the results developed for the classical DeGroot's model (e.g., DeGroot [26], Jackson [51, Chapter 8], and Golub and Jackson [42]). For example, a priori one

---

[23]Formally, if x majorizes y, that is $\sum_{i=1}^{n} \varphi(x_i) s_i \geq \sum_{i=1}^{n} \varphi(y_i) s_i$ for all real-valued, continuous, and convex functions, then $\bar{T}^\lambda(x) \geq \bar{T}^\lambda(y)$.

cannot rule out that the behavior of the sequence of updates might depend heavily on the initial condition x^0 (e.g., convergence, rate of convergence, consensus). Indeed, despite the name robust, we might well wonder whether or not our opinion aggregators generate chaotic dynamics. This is not the case since robust opinion aggregators and their iterates are nonexpansive (see Lemma 8 in Appendix B).[24]

We are interested in the convergence of the sequence $\{T^t(x^0)\}_{t \in \mathbb{N}}$ given a specific x^0 or, more in general, in the convergence of $\{T^t(x)\}_{t \in \mathbb{N}}$, irrespective of the x chosen.

Definition 7 *Let T be an opinion aggregator. We say that T is* convergent *at x^0 if and only if $\lim_t T^t(x^0)$ exists. Moreover, we say that T is* convergent *if and only if T is convergent at each x in B.*

We cannot expect to obtain that robust opinion aggregators are convergent in general as Example 1 already clarified. That example, using a simple opinion aggregator, illustrates this as well as the gist of our first convergence result: the sequence of updates might not converge, yet their time averages do.

In dealing with the issue of convergence, we thus first focus on Cesaro convergence of the updates. We do so for two reasons:[25]

1. If given an initial condition x^0 the updates $T^t(x^0)$ converge, then $\{T^t(x^0)\}_{t \in \mathbb{N}}$ converges a la Cesaro and

$$C - \lim_t T^t(x^0) = \lim_t T^t(x^0).$$

   Therefore, conditions which yield that $C - \lim_t T^t(x^0)$ exists are conceptually the weaker counterpart of assumptions which deliver the convergence at x^0. This all the more is true if we are interested in global convergence.

2. Example 1 illustrates how the opposite might not be true. Therefore, we first study which properties of T yield the existence of $C - \lim_t T^t(x)$ for all $x \in B$

---

[24] A common requirement for chaotic behavior is the following property, termed *sensitive dependence on initial conditions*,

$$\exists r > 0, \forall x \in B, \forall \varepsilon > 0, \exists y \in B, \exists t \in \mathbb{N} \text{ s.t. } \|x - y\|_\infty < \varepsilon \text{ and } \|T^t(x) - T^t(y)\|_\infty \geq r.$$

In other words, a small change ε in the initial condition, say from x to y, might generate very different dynamics. Nonexpansive opinion aggregators violate this property. For a textbook treatment, see Devaney [28, p. 49] and Robinson [76, Section 3.5].

[25] See, e.g., [18, Lemma 15.5] for the simple relation between convergence and Cesaro convergence. The notion of Cesaro limit has been used already in the networks literature by Golub and Morris [44]. In particular, they explore the convergence of Abel averages. Under their assumptions, this convergence is equivalent to Cesaro convergence. Other works that emphasize the importance of Cesaro convergence of the actions played in a game include Fudenberg and Levine [35].

and, at a later stage, we study additional conditions on T, which turn Cesaro convergence into standard convergence. These extra requirements can be mapped into properties of the loss functions of our foundation, as discussed below.

The following result shows that, for each $x \in B$, the time averages of the sequence of updates generated by robust opinion aggregators always converge. Moreover, the last part of the result shows that, whenever the opinions of the agents are bounded, the initial realizations of their signal do not affect the rate of convergence of time averages. Therefore, the time needed for the information to stabilize on average does not depend on the objective data generating process, but only on the estimation procedures (i.e., the profile of loss functions ϕ) used by the agents.

Theorem 2 *If T is a robust opinion aggregator, then*

$$C - \lim_t T^t(x) \ \ exists \qquad \forall x \in B. \tag{13}$$

Moreover, if $\bar{T} : B \to B$ is defined by

$$\bar{T}(x) = C - \lim_t T^t(x) \qquad \forall x \in B, \tag{14}$$

then $\bar{T}$ is a robust opinion aggregator such that $\bar{T} \circ T = \bar{T}$, and if $\hat{B}$ is a bounded subset of B, then

$$\lim_\tau \left(\sup_{x \in \hat{B}} \left\| \frac{1}{\tau} \sum_{t=1}^\tau T^t(x) - \bar{T}(x) \right\|_\infty \right) = 0. \tag{15}$$

The conceptual message of Theorem 2 is linked to our results on the wisdom of the crowd. In Section 6, we give conditions under which the *Cesaro limit* of the updates converges in probability to the true underlying parameter μ, provided the size of society goes to infinity. If the robust opinion aggregator T happens to be convergent, then this implies the *wisdom of the crowd*: agents are going to learn the true parameter. Instead, if T is not convergent, still there is *wisdom from the crowd*: an external observer that can compute the time averages in a part of the society, can extract enough information to learn the truth.

From a mathematical point of view, in order to address the problem of standard convergence, all we need now is a condition that paired with the (uniform) convergence of time averages yields the usual convergence in norm. In light of the classic paper of Lorentz [61], we know that such a condition exists (cf. Definition 8).

5.2 Standard convergence

We first discuss the technical condition of asymptotic regularity which characterizes standard convergence. Asymptotic regularity is hard to interpret, yet useful. In Section

24

5.3, we study sufficient conditions on the network structure induced by T (see Definition 10) which are easier to interpret and imply asymptotic regularity. Moreover, we also provide natural conditions on the profile of loss functions ϕ and on the network (N, A) which imply them.

Definition 8 *Let T be an opinion aggregator. We say that T is asymptotically regular if and only if for each $x \in B$*

$$\lim_t \left\| T^{t+1}(x) - T^t(x) \right\|_\infty = 0. \tag{16}$$

On the one hand, (16) is weaker then the Cauchy property.[26] On the other hand, it is still enough to grant convergence. More formally, as we have seen, if T is a robust opinion aggregator, then the time averages of $\{T^t(x)\}_{t \in \mathbb{N}}$ Cesaro converge uniformly. If $\{T^t(x)\}_{t \in \mathbb{N}}$ further satisfies (16), then this is enough to show that $C - \lim_t T^t(x) = \lim_t T^t(x)$.[27] The following result contains the above observation and elaborates on the rate of convergence.

Proposition 4 *Let T be a robust opinion aggregator. The following statements are equivalent:*

(i) T is asymptotically regular;

(ii) T is convergent.

Moreover, if T is constant affine, then they are also equivalent to the following:

(iii) There exists $\{c_t\}_{t \in \mathbb{N}} \subseteq [0, \infty)$ such that $c_t \to 0$ and

$$\left\| \bar{T}(x) - T^t(x) \right\|_\infty \le c_t \left\| x \right\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B. \tag{17}$$

Proposition 4 shows that asymptotic regularity characterizes convergence for robust opinion aggregators over B. Moreover, constant affinity, which is satisfied in several relevant cases (cf. Proposition 1 and Section 5.4), yields that if T is convergent, then the rate of convergence is *independent* of the initial condition x. In this case, the rate of convergence is given by the sequence $\{c_t\}_{t \in \mathbb{N}}$, which we can bound in some prominent cases (cf. Corollaries 2–4).

---

[26] An example is available upon request.

[27] Conditions, such as asymptotic regularity, that turn Cesaro convergence (or other weaker forms of convergence) into standard convergence are also called Tauberian (see, e.g., Korevaar [54]).

5.3 Network structure

In the standard DeGroot's linear model, convergence is implied by the properties of an underlying network structure. In this case, the underlying network structure can either be implicit in T and given by the indicator matrix $A(W)$ of W (e.g., Golub and Jackson [42]) or be explicit and given by a primitive network assumed in the foundation of the opinion aggregator (e.g., DeMarzo, Vayanos, and Zwiebel [27]).[28] When T is linear, both approaches are tightly linked.

In what follows, we explore both paths and show they are linked also when T is robust. In terms of the first approach, we begin by extending the notion of indicator matrix to the nonlinear case. To start, we need to define when an agent is connected to another one. A piece of notation: $e^j \in \mathbb{R}^n$ denotes the j-th vector of the canonical basis.

Definition 9 *Let T be an opinion aggregator. We say that j strongly influences i if and only if there exists $\varepsilon_{ij} \in (0,1)$ such that for each $x \in B$ and for each $h > 0$ such that $x + he^j \in B$*

$$T_i\left(x + he^j\right) - T_i(x) \geq \varepsilon_{ij} h. \tag{18}$$

The interpretation of (18) is simple: the update of i increases at least linearly in the opinion of j. This is a strict monotonicity type of property which from a marginal point of view, under differentiability, is tantamount to assume that the j-th component of the gradient of T_i is bounded away from 0. This allows us to define the underlying network structure of T. The adjacency matrix $A(T)$ below represents the edges in the directed network induced by a generic opinion aggregator T.

Definition 10 *Let T be an opinion aggregator. We say that $A(T)$ is the* adjacency *matrix induced by T if and only if for each $i, j \in N$ the ij-th entry is such that*

$$a_{ij} = \begin{cases} 1 & \text{if } j \text{ strongly influences } i \\ 0 & \text{otherwise} \end{cases}.$$

The directed network given by the adjacency matrix $A(T)$ is intuitively the *minimal* network underlying the opinion aggregator T. This is because the condition for a_{ij} to be equal to 1 is strong (cf. Example 5 and Section 5.4).[29] Indeed, it requires that the updates of i are responsive to changes in the opinion of j, starting from every vector

---

[28]Recall that the indicator matrix of W, $A(W)$, is such that the ij-th entry is 1 if $w_{ij} > 0$ and 0 otherwise.

[29]In Remark 7 in Appendix B, we discuss alternative notions of underlying network which were explored in the mathematical literature.

of opinions x. When T is linear, that is, $T(x) = Wx$ for all $x \in B$ with $W \in \mathcal{W}$, we have that

$$a_{ij} = 1 \iff w_{ij} > 0$$

and, in particular, $A(T)$ coincides with the indicator matrix $A(W)$ of W. As observed, $A(T)$ is a *minimal* network. This observation naturally leads us to wonder whether other layers of weaker networks can be derived from T. We briefly discuss this idea in Remark 2.

Natural conditions on the profile of loss functions yield that the induced network $(N, A(T))$ coincides with the primitive network (N, A) where the latter represents the links connecting the agents (see Section 3). We formalize this fact in the next result focusing on additively separable loss functions.[30]

Proposition 5 *Let $\phi = (W, \rho) \in \Phi_A$. If I is compact and ρ_i is twice continuously differentiable and strongly convex for all $i \in N$, then $\mathbf{T}^\phi = T^\phi$, defined as in (3), is single-valued and*

$$w_{ij} > 0 \iff j \text{ strongly influences } i.$$

In particular, $A(W) = A(T^\phi)$.

The result above confirms that, under mild conditions, an agent j opinion affects the update of i if and only if i observes j. In particular, recall that our assumption is that the indicator matrix of W coincides with the underlying network given by A. Thus, the endogenous network we defined for T coincides with the primitive one and justifies Definition 10 in terms of our foundation. Finally, as already argued, if T is linear, Definition 10 coincides with the classical one, providing a further consistency check.

5.3.1 Convergence

Given a robust opinion aggregator T, the network $(N, A(T))$ is instrumental in providing sufficient conditions for convergence. In light of our previous discussions, and in contrast to asymptotic regularity, those conditions have an immediate interpretation in terms of connections among agents. We here recall some terminology from the network literature. Consider the network $(N, A(T))$ and $M \subseteq N$. A path in M is a finite sequence of agents $i_1, i_2, ..., i_K \in M$ with $K \geq 2$, not necessarily distinct, such that

---

[30] The proof of Proposition 5 is at the end of Appendix A. In general, we can prove a similar result for profiles of loss functions which are not additively separable. In this case, the assumptions of differentiability and strong convexity can be also weakened and replaced with a coercivity condition and a Lipschitz property of the difference quotients. These two alternative properties mirror the assumptions of "bounded derivative" and "state monotonicity" of Frankel, Morris, and Pauzner [31].

$a_{i_k i_{k+1}} > 0$ for all $k \in \{1, ..., K - 1\}$. In this case, the length of the path is $K - 1$. A cycle in M is a path such that $i_1 = i_K$. A cycle is simple if and only if the only agent appearing twice in the sequence is the starting (and ending) one. We say that M is closed if and only if for each $i \in M$, $a_{ij} > 0$ implies $j \in M$. Finally, M is aperiodic if and only if the greatest common divisor of the lengths of its simple cycles is 1.

A group of agents M is closed if an agent of the group only observes agents in M. Aperiodicity instead has a simple interpretation: the cycles present do not generate periodic behaviors. These two notions play a key role for convergence.

Definition 11 *Let T be an opinion aggregator with adjacency matrix $A(T)$. We say that T is* strongly aperiodic *if and only if each closed group M is aperiodic.*

This definition coincides with the definition of strongly aperiodic proposed by Golub and Jackson [42, Definition 7] for the linear case. One key difference with the linear case though is that it might be emptily satisfied: for example, if $A(T)$ coincides with the null matrix. In fact, when T is not linear it might happen that $A(T)$ might have some null row. This is always ruled out in the linear case since W is always assumed to have a nonzero element in each row. In what follows, we want to avoid the presence of agents which are not influenced by any other agent. We formalize this in the next definition.

Definition 12 *Let T be an opinion aggregator with adjacency matrix $A(T)$. We say that T has a* nontrivial network *if and only if no row of $A(T)$ is null, that is, for each $i \in N$ there exists $j \in N$ such that $a_{ij} > 0$.*

Intuitively, the next result says that, given a robust opinion aggregator with nontrivial network, if each closed group does not have cycles which are "problematic", then T is convergent, thus linking the properties of the underlying network with the stability of the long-run opinions.

Theorem 3 *Let T be a robust opinion aggregator. If T is strongly aperiodic and has a nontrivial network, then T is asymptotically regular and, in particular, convergent.*

Our convergence theorem significantly generalizes in one direction Golub and Jackson [42, Theorem 2]. Nevertheless, differently from [42], it is not true that these conditions are also necessary for convergence (see Example 5). Next, we provide a parametric class of convergent robust opinion aggregators.

Example 2 (Confirmatory bias) It is often argued that in some societies, individuals tend to trust more those sources of information whose opinion confirms their original

prior.[31] This phenomenon can be captured by a modification of DeGroot's linear model which is a generalization of the one proposed by Jackson [51]. Let $B = [0, 1]^n$ and assume that the network structure is represented by an adjacency matrix A with nonnull rows. As in the linear model, society is represented by a stochastic matrix $W(x)$ where $w_{ij}(x)$ is the weight assigned by individual i to agent j. Differently from the linear model, the weight is allowed to depend on the vector x. Moreover, it is assumed that each individual downweights the agents who disagree the most with her and only if these agents belong to her neighborhood:

$$T(x) = W(x) x \qquad \forall x \in B$$

where

$$w_{ij}(x) = \frac{a_{ij} e^{-\gamma_{ij}|x_i - x_j|}}{\sum_{l=1}^{n} a_{il} e^{-\gamma_{il}|x_i - x_l|}} \qquad \forall x \in B$$

with $\gamma_{ij} \in (0, 1]$ for all $i, j \in N$ and $\sum_{j=1}^{n} \gamma_{ij} \leq 1$ for all $i \in N$. Here, $1/\gamma_{ij}$ captures the relative strength of the weight assigned by individual i to agent j net of the difference in their opinions. It is easy to see that the aggregator T is robust and $A(T) = A$. Thus, in particular, by Theorem 3, if each closed group M is aperiodic under A, then T is convergent. ▲

If each agent strongly influences herself, that is $a_{ii} > 0$ for all $i \in N$, then the presence of problematic cycles is ruled out and the presence of a nontrivial network is also guaranteed.

Corollary 1 *Let T be a robust opinion aggregator with adjacency matrix $A(T)$. If T is* self-influential, *that is $a_{ii} > 0$ for all $i \in N$, then T is convergent.*

Self-influentiality characterizes a situation where the opinion of each agent has a form of own-history dependence. Indeed, if we focus on a generic agent i, given two instances x and y, if the only difference is the agent's opinion, say $x_i > y_i$, then her revised opinion is strictly higher in the first instance than in the second one. In a repeated setting, information gathered in the past is not entirely dismissed in light of new evidence.

The previous corollary shows how a very intuitive and weak condition is sufficient to obtain opinions' convergence in the long-run. Moreover, Chandrasekhar, Larreguy, and Xandri [23] find in their field study that the behavior of most of the subjects is consistent with self-influentiality, even when this is objectively suboptimal.[32] So one

---

[31] See, e.g., the evidence presented in Benjamin [10], Lord, Ross, and Lepper [60], Yariv [84].

[32] In particular they write: *"[...] over 82.9% of the time that subjects in the Indian sample who have an information set that is dominated by a network neighbor fail to simply copy their neighbor, which is what Bayesians would do in an all-Bayesian environment. In contrast, this failure occurs 54.5% of the time in the Mexican data."*

might be left to wonder what type of aggregators might fail to be self-influential. The next example discusses two critical cases.

Example 3 Assume T is linear with matrix W. Clearly, T is *not* self-influential if and only if there exists an agent i such that $w_{ii} = 0$, that is, an agent whose own opinion never enters in her updating rule. Other important robust opinion aggregators which may fail to be self-influential are those such that each T_i corresponds to a quantile functional (see Section 5.4 for a characterization of the dynamics induced by these opinion aggregators). The intuition, in this case, is simple. Quantiles tend to disregard outliers, be those the opinions of the agent or not. In this case, an aggregator of this kind is self-influential if and only if it is the identity (cf. Proposition 7). $\blacktriangle$

5.3.2 Equilibria

The previous section provides sufficient conditions on the network structure $(N, A(T))$ which guarantee convergence. At the same time, two natural questions pertain to the limit itself. In other words, if $\lim_t T^t(x^0)$ exists, what is it, and what are its properties?

In order to answer these questions, fixed points of the opinion aggregator T play a fundamental role and so does the network $(N, A(T))$. Moreover, fixed points have a standard interpretation: they characterize situations where a profile of opinions does not change once reached.

Definition 13 *Let T be an opinion aggregator. The point $\bar{x} \in B$ is an* equilibrium *of T if and only if $T(\bar{x}) = \bar{x}$. The set of equilibria is denoted by $E(T)$.*

The notions of equilibrium and convergence are tied to each other. If a sequence of updates converges, then it necessarily converges to an equilibrium.

Lemma 3 *Let T be a robust opinion aggregator. If T is convergent and $\bar{T}$ is defined as in (14), then $\bar{T}(x) = \lim_t T^t(x) \in E(T)$ for all $x \in B$.*

This simple lemma clarifies the role played by the operator $\bar{T} : B \to B$, as defined in (14). Indeed, given $x^0 \in B$, if T is robust, then $\bar{T}(x^0)$ is the opinion to which the time averages of the updates converge. If T further happens to be convergent, then $\bar{T}(x^0)$ is the equilibrium opinion to which the sequence of updates converges to. It is a simple mathematical result, yet it stresses how strong is the property of convergence. Indeed, in contrast as next example shows, if $C - \lim_t T^t(x^0)$ exists, but $\lim_t T^t(x^0)$ does not, then it might not be true that $C - \lim_t T^t(x^0)$ is an equilibrium.[33]

---

[33] An important case where $\bar{T}(x) = C - \lim_t T^t(x)$ is an equilibrium of T for all $x \in B$ is given by linear opinion aggregators (see, e.g., Aliprantis and Border [3, Theorem 20.14]).

Example 4 Let $B = [0,1]^3$ and assume that $T : B \to B$ is defined by

$$T(x) = (\max\{x_2, x_3\}, x_3, x_2) \qquad \forall x \in B.$$

It can be checked immediately that the opinion aggregator T is robust. Set $x^0 = (1, 0, 1)$ and $\hat{x} = \left(1, \frac{1}{2}, \frac{1}{2}\right)$. It follows that $C - \lim_t T^t(x^0) = \left(1, \frac{1}{2}, \frac{1}{2}\right) = \hat{x}$ and $T(\hat{x}) = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right) \neq \hat{x}$. ▲

An essential subset of equilibria that are always present in our framework is the *consensus* subset. Formally, these vectors are the constant ones, and they represent a situation in which all agents share the same opinion. We denote them by D, that is, $x \in D \subseteq B$ if and only if $x_i = x_j$ for all $i, j \in N$.

Definition 14 *Let T be an opinion aggregator. We say that T is a* consensus operator *if and only if $E(T) \subseteq D$.*

In words, a robust opinion aggregator is a consensus operator if and only if its only equilibria are the constant vectors in B. The properties of convergence and being a consensus operator are separate and independent. To wit, there are linear opinion aggregators which are consensus operators, but are not convergent,[34] while the identity operator is convergent, but not a consensus operator. Some sufficient properties that yield a consensus operator are easily expressed in terms of the network $(N, A(T))$. We report them in the next definition and discuss them right below.

Definition 15 *Let T be an opinion aggregator with adjacency matrix $A(T)$. We say that:*

1. *T has the* uniform common influencer property *if and only if there exists $k \in N$ such that $a_{ik} > 0$ for all $i \in N$.*

2. *T has the* pairwise common influencer property *if and only if for each $i, j \in N$ there exists $k \in N$ such that $a_{ik}, a_{jk} > 0$.*

3. *T is* strongly connected *if and only if for each $i, j \in N$ there exists a path in N such that $i_1 = i$ and $i_K = j$.*

The uniform common influencer property states that there exists an agent k (an influencer) which is observed by all agents. Thus, such an agent has an extreme centrality in the network. Indeed, many standard centrality measures used in the social network

---

[34]For example, set $W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

literature would be maximal for k. If T were to be linear, this would be tantamount to have the k-th column of the matrix W strictly positive.

The pairwise common influencer property is a weakening of the former property. We do not require the existence of an extremely central agent k. However, each pair of agents must share an individual whose opinion is observed by both of them. Intuitively, this is a minimal requirement about the presence of a direct source of information relied on by both agents. A typical situation where we expect the pairwise common influencer property to hold is one of asymmetric networks with a bunch of media followed by the agents. If there is a minimal overlapping in the media trusted by the agents, the property holds. If T were to be linear, the pairwise common influencer property would be equivalent to assume that the representing matrix W is scrambling.

Strong connectedness takes a simple interpretation: after enough periods each agent i is influenced by each agent j. What might happen is that it might be that there are cycles and so, loosely speaking, agent i might lose sight of j. If T were to be linear, strong connectedness of T would be equivalent to the irreducibility of W.

The next result shows that all the three properties above are sufficient to yield a consensus operator. Next section shows that they also play a key role in obtaining convergence.

Proposition 6 *Let T be a robust opinion aggregator. T is a consensus operator provided one of the following holds:*

   a. T has the pairwise common influencer property;

   b. T has the uniform common influencer property;

   c. T is strongly connected.

5.3.3   Rate of convergence and consensus

Theorem 3 and Corollary 1 provide sufficient conditions for convergence, but are mute in terms of rate of convergence and properties of the limit. Proposition 6, paired with Lemma 3, provides conditions that guarantee that if T is convergent, then the limit outcome is consensus. The next three results show how the properties yielding consensus also guarantee stronger forms of convergence.

Corollary 2 *Let T be a robust opinion aggregator. If T has the uniform common influencer property, then $\bar{T}(x) = \lim_t T^t(x) \in D$ for all $x \in B$. Moreover, there exists $\varepsilon \in (0, 1)$ such that*

$$\left\| \bar{T}(x) - T^t(x) \right\|_\infty \leq 2(1 - \varepsilon)^t \|x\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B.$$

In words, if there is a uniform common influencer in the society, then the sequence of updates converges, to consensus, and exponentially fast. As we have seen above, the uniform common influencer property has a weaker version which we term pairwise common influencer property. Next theorem proves that also this latter condition yields convergence. But in this case, its rate is almost exponential.[35]

Corollary 3 *Let T be a robust opinion aggregator. If T has the pairwise common influencer property, then $\bar{T}(x) = \lim_t T^t(x) \in D$ for all $x \in B$. Moreover, there exist $\varepsilon \in (0,1)$ and $\hat{t} \in \mathbb{N}$ such that $\hat{t} \leq n^2 - 3n + 3$ and*

$$\left\| \bar{T}(x) - T^t(x) \right\|_\infty \leq 2(1-\varepsilon)^{\left\lfloor \frac{t}{\hat{t}} \right\rfloor} \left\| x \right\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B. \tag{19}$$

In the previous results about convergence to consensus, only local properties of the underlying network $A(T)$ were considered, in the sense that agents have to share (pairwise or uniformly) a first-hand source of information. Instead, the next result leverages the connection structure of the social network.

Corollary 4 *Let T be a robust opinion aggregator. If T is strongly connected and strongly aperiodic, then $\bar{T}(x) = \lim_t T^t(x) \in D$ for all $x \in B$. Moreover, there exists $\varepsilon \in (0,1)$ and $\hat{t} \in \mathbb{N}$ such that $\hat{t} \leq (n-1)^2 + 1$ and*

$$\left\| \bar{T}(x) - T^t(x) \right\|_\infty \leq 2(1-\varepsilon)^{\left\lfloor \frac{t}{\hat{t}} \right\rfloor} \left\| x \right\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B. \tag{20}$$

Moreover, if T is also self-influential, then $\hat{t} \leq n - 1$.

Under strong connectedness of T, observe that T is strongly aperiodic if and only if N is aperiodic as in Golub and Jackson [42].

Remark 1 The previous three corollaries can be partially obtained via our main convergence result. If T has either the uniform common influencer property (resp., pairwise common influencer property) or is strongly connected and strongly aperiodic, then T has a nontrivial network and is strongly aperiodic. By Theorem 3, Lemma 3, and Proposition 6, we have that $\bar{T}(x) = \lim_t T^t(x) \in E(T) = D$ for all $x \in B$. This proof strategy has the drawback of not elaborating on the rate of convergence. For this reason, we obtain the previous three corollaries as a by-product of another general result: Theorem 7 in Appendix B. The aforementioned result is a nonlinear version of a well known fact: in DeGroot's model, convergence to consensus happens if and only if there exists $\hat{t} \in \mathbb{N}$ such that some column of $W^{\hat{t}}$ is strictly positive (see, e.g., Jackson [51, Corollary 8.2]). In terms of underlying network, this is equivalent to say

---

[35]Recall that, given $s \in (0,\infty)$, $\lfloor s \rfloor$ is the integer part of s, that is, the greatest integer $l \in \mathbb{N}_0$ such that $s \geq l$.

that there exists an agent k which *eventually* strongly influences every other agent in the population. Formally, for each $t \geq \hat{t}$ there exists a path from any agent i to such an agent k.[36] In Theorem 7, we generalize one implication of this result by a) allowing for nonlinearities in T and b) offering some estimate about the rate of convergence. This should also clarify the origin and the nature of the different bounds on $\hat{t}$. They are all obtained via the adjacency matrix induced by T and the implied network structure specified by $A(T)$. In particular, in Corollary 4, $\hat{t}$ can be chosen to be the index/exponent of primitivity of $A(T)$, that is, the smallest integer such that each entry of $A(T)^{\hat{t}}$ is strictly positive (in this case, each agent eventually strongly influences any other agent). In other words, $\hat{t}$ is the smallest integer such that for each $i, j \in N$ there exists a path of length $\hat{t}$ from i to j. This allows also to provide other bounds, for example, it is known that $\hat{t} \leq d^2 + 1$ where d is the diameter of the network $(N, A(T))$ (see, e.g., Neufeld [72]) or $\hat{t} \leq n + s(n-2)$, provided the shortest (simple) cycle has length $s \geq 1$ (see, e.g., Horn and Johnson [47, Theorem 8.5.7]). ▲

All the convergence results we have discussed until now, which involved conditions on the network $A(T)$, provided sufficient conditions for convergence. At the same time, for the linear case, some of these conditions turn out to be necessary and sufficient. For example, Golub and Jackson [42, Theorem 2 and Proposition 1] prove that strong aperiodicity is necessary and sufficient for convergence while (strong) aperiodicity is necessary and sufficient for convergence to consensus, under strong connectedness. In general, we do not expect the necessity parts to continue to hold in the nonlinear case. Intuitively, this is due to our strong definition of underlying network $(N, A(T))$ (cf. Remark 7 in Appendix B). We illustrate this in the next example.

Example 5 Let $B = \mathbb{R}^2$ and let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be defined to be such that

$$T_1(x) = \begin{cases} x_2 & \text{if } x_1 > x_2 \\ \frac{1}{2}x_1 + \frac{1}{2}x_2 & \text{if } x_2 \geq x_1 \end{cases} \quad \text{and } T_2(x) = x_1 \quad \forall x \in B.$$

On the one hand, agent 2 always discards her own opinion. On the other hand, agent 1 discards her own signal if and only if it has the highest realization. Easy computations yield that T is a robust opinion aggregator such that

$$A(T) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Thus, T is strongly connected, but it clearly fails to be strongly aperiodic. At the same time, it is easy to show that $\lim_t T^t(x) = \min\{x_1, x_2\} e$ for all $x \in B$, showing

---

[36] In terms of adjacency matrices, we have that the ik-th entry of $A(T)^{\hat{t}}$ is strictly positive for all $i \in N$ and so is, as a consequence, the ik-th entry of $A(T)^t$ for all $t \geq \hat{t}$.

that, under strong connectedness, strong aperiodicity is sufficient for convergence to consensus, but not necessary. ▲

Example 5 suggests the presence of multiple layers of networks in our model. Therefore, our model is flexible enough to capture more complex social structures. As an example, the network $A(T)$ may be used to model the "strong ties" among agents a la Granovetter [46], that is those clustered links that are always active regardless of the particular stances of the agents. However, there might be additional links (i.e., "weak ties") not in $A(T)$ that are active only under exceptional circumstances (e.g., a_{ij} might be 0, but j might influence i when the former takes an extreme position).[37] More generally, our strong notion of network $A(T)$ is the base for sufficient conditions for convergence, in particular, to consensus. At the same time, weaker notions may be used to obtain necessary conditions for convergence. A complete study of these conditions is beyond the scope of the paper, but we offer a glimpse on how to undertake this task in the next remark.

Remark 2 Consider a robust opinion aggregator T. We say that j *weakly influences* i if and only if there exist $x \in B$ and $h > 0$ such that $x + he^j \in B$ and

$$T_i\left(x + he^j\right) - T_i(x) > 0.$$

With this, we can define the adjacency matrix of weak ties $\bar{A}(T)$, that is, for each $i, j \in N$ the ij-th entry is such that

$$\bar{a}_{ij} = \begin{cases} 1 & \text{if } j \text{ } weakly \text{ influences } i \\ 0 & \text{otherwise} \end{cases}.$$

If T is linear, it is plain that $A(W) = A(T) = \bar{A}(T)$. Moreover, in the linear case, global convergence to consensus is characterized by the existence of a unique strongly connected, closed, and aperiodic group of agents (see, e.g., Jackson [51, Corollary 8.1]). For robust opinion aggregators, the necessity of a unique closed group for convergence to consensus is preserved, with the caveat that the network to be used is $\left(N, \bar{A}(T)\right)$.

5.4 Choquet aggregators

In this section, we consider a particular example of robust opinion aggregators. This class encompasses linear opinion aggregators as well as those aggregators whose components are either any quantile functional (e.g., the median) or any order statistics.

---

[37]These layers can be mapped into the data collected on the field. For example, in their analysis of the network structure of the Indonesian villages Alatas, Banerjee, Chandrasekhar, Hanna, and Olken [2] identify both the strong familiar ties and the links due to the extreme relative wealth of some agents.

Definition 16 *Let T be an opinion aggregator. We say that T is a* Choquet aggregator *if and only if T is normalized, monotone, and comonotonic additive.*

It is routine to show that Choquet aggregators are constant affine and robust. Right below, we offer a characterization which turns out to be useful in exploring the behavior of aggregators such as the one in Example 1. Moreover, it justifies our terminology. Indeed, if T is a Choquet aggregator, it follows that each $T_i : B \to \mathbb{R}$ is normalized, monotone, and comonotonic additive.[38] Given Schmeidler [78, p. 256], it is well known that a map T_i has these properties if and only if there exists a unique capacity $\nu_i : 2^N \to [0, 1]$ such that

$$T_i(x) = \int_N x d\nu_i \qquad \forall x \in B \tag{21}$$

where the latter is a Choquet expectation/average. On the one hand, a capacity is a set function with the following properties:

1. $\nu_i(\emptyset) = 0$ and $\nu_i(N) = 1$;

2. $A \supseteq B$ implies $\nu_i(A) \geq \nu_i(B)$.

On the other hand, the Choquet average for a function defined over a finite set of points (i.e., a vector) is a rather simple object. First, in words, one should order the components of x from the highest to the lowest. Formally, this is done with the help of a permutation π over N such that $x_{\pi(1)} \geq ... \geq x_{\pi(n)}$. Then, one computes the weight given to $x_{\pi(l)}$ in terms of ν_i which is

$$p_{\pi(l)} \overset{\text{def}}{=} \nu_i\left(\cup_{j=1}^l \{\pi(j)\}\right) - \nu_i\left(\cup_{j=1}^{l-1} \{\pi(j)\}\right) \qquad \forall l \in N \tag{22}$$

with the assumption that $\nu_i\left(\cup_{j=1}^0 \{\pi(j)\}\right) = 0$. If the values taken by x were pairwise distinct and we interpret ν_i as a measure of likelihood, (22) is exactly the likelihood of observing a value greater than or equal to $x_{\pi(l)}$ minus the likelihood of observing strictly higher values. The Choquet expectation is then the average of the ordered values of x using the probability vector p computed in (22):

$$\int_N x d\nu_i = \sum_{l=1}^n x_{\pi(l)} \left[\nu_i\left(\cup_{j=1}^l \{\pi(j)\}\right) - \nu_i\left(\cup_{j=1}^{l-1} \{\pi(j)\}\right)\right] = \sum_{l=1}^n x_{\pi(l)} p_{\pi(l)} \,\forall x \in \mathbb{R}^n. \tag{23}$$

---

[38] With a small abuse of notation, we use the same name for similar properties that pertain to, respectively, functionals and operators. At the same time, no confusion should arise. A functional $T_i : B \to \mathbb{R}$ is normalized if and only if $T_i(ke) = k$ for all $k \in I$ and comonotonic additive if and only if $T_i(x + y) = T_i(x) + T_i(y)$, provided x and y are comonotonic. Monotonicity is the usual notion.

Note that if ν_i is a standard additive probability, then the Choquet average coincides with the standard notion of weighted average. If T is a Choquet aggregator, then we denote by ν_i the capacity that represents T_i as in (21).

In some situations, it is reasonable to start with the capacity as the primitive objective. This is the case if an agent can assign an informational value to the subsets of the society, but such a value is not necessarily additive. Additivity may fail if two different sources are perceived as strongly correlated, since in that case, given the observation from the first source, the additional information obtained by observing the second source is (perceived as) much lower than if observed alone (see, e.g., Liang and Mu [59]).

The next proposition shows that for comonotonic aggregators with $\{0,1\}$-valued capacities, if convergence happens, then it will happen in a finite number of periods. These aggregators are important for two reasons. First, they encompass aggregators in which each agent aggregates opinions, for example, according to one of the following criteria: the median, any more general quantile, max, min, or any more general order statistic. Second, even if this is outside the scope of this paper, they can be used to study the evolution of opinions that lie in a discrete set of possible opinions O. Indeed, the comonotonic aggregators with $\{0,1\}$-valued capacities have the property that $\{T_i(x) : i \in N\} \subseteq \{x_i : i \in N\}$. For example, the triggering model by Kempe, Kleinberg, and Tardos [53] is a particular case of Choquet aggregator with $\{0,1\}$-valued capacities over the discrete set $O = \{0,1\}$.

Theorem 4 *Let T be a Choquet aggregator such that ν_i is a $\{0,1\}$-valued capacity for all $i \in N$. If $x \in B$, then either $\{T^t(x)\}_{t \in \mathbb{N}}$ converges or it is eventually periodic, that is, there exists $\bar{t}, p \leq n^n$ such that*

$$T^{t+p}(x) = T^t(x) \qquad \forall t \geq \bar{t}.$$

Moreover, $\{T^t(x)\}_{t \in \mathbb{N}}$ converges if and only if it becomes constant after at most n^n periods.

Remark 3 The previous result provides an easy criterion to discern the behavior of the sequence of updates $\{T^t(x)\}_{t \in \mathbb{N}}$. Set $t = n^n$ where n is the number of agents in the population, and so the maximum value of distinct components x can have. If $T^t(x) = T^{t+1}(x)$, then $\{T^t(x)\}_{t \in \mathbb{N}}$ converges. If $T^t(x) \neq T^{t+1}(x)$, then $\{T^t(x)\}_{t \in \mathbb{N}}$ is eventually periodic with period smaller than or equal to n^n. ▲

One might wonder what additional restrictions the network conditions we studied in Section 5.3 impose on the Choquet aggregators considered in Theorem 4. The result below is a negative one.

Proposition 7 *Let T be a Choquet aggregator such that ν_i is a $\{0,1\}$-valued capacity for all $i \in N$. The following statements are equivalent:*

(i) j strongly influences i;

(ii) $T_i(x) = x_j$ for all $x \in B$.

Moreover, T is self-influential if and only if $T(x) = x$ for all $x \in B$ and T has the pairwise common influencer property if and only if there exists $k \in N$ such that $T_i(x) = x_k$ for all $x \in B$ and for all $i \in N$.

We have noticed already how convergence is implied by the seemingly natural assumption of self-influence (cf. Corollary 1). However, for Choquet aggregators which are represented by $\{0,1\}$-valued capacities, self-influence is too strong of an assumption, since the only Choquet aggregator of this type is the one that coincides with the identity. More in general, for this type of Choquet aggregators, we have that each agent i can be influenced by at most only one other individual j. Self-influence yields that j coincides with i, while the pairwise common influencer property returns the existence of a unique uniform common influencer $k = j$.

It is important not to overstate the reach of this negative result. Indeed, a parallel with decision theory suggests a natural specification for Choquet aggregators, consistent with the properties of Section 5.3. In fact, consider a stochastic matrix $W \in \mathcal{W}$, W is assumed to capture the relative weights agents assign to each other as well as the network structure among agents, as in DeGroot's model. But differently from the latter, we might allow for agents who do not compute linearly the opinions' average, but rather use a distorted collection of weights, as in cumulative prospect theory. In this case, the capacity used by each agent i is $\nu_i : 2^N \to [0,1]$ defined by

$$\nu_i(A) = f_i\left(\sum_{l \in A} w_{il}\right) \qquad \forall A \subseteq N, \tag{24}$$

where $f_i : [0,1] \to [0,1]$ is strictly increasing, continuous, and such that $f_i(0) = 0$ as well as $f_i(1) = 1$. The interesting feature of these Choquet aggregators is that the underlying network of T coincides with the primitive one, that is $A(T) = A(W)$. In light of Theorem 3, T is therefore convergent whenever W is.

Proposition 8 *Let T be a Choquet aggregator such that ν_i is defined as in (24) for all $i \in N$ where $W \in \mathcal{W}$. The following statements are equivalent:*

(i) j strongly influences i;

(ii) $w_{ij} > 0$.

In particular, we have that $A(T) = A(W)$.

Example 6 Choquet aggregators like the ones of the previous result encompass L-estimators. In fact, consider a directed network (N, A) and construct the stochastic matrix of weights W defined by $w_{ij} = a_{ij}/\sum_{l=1}^{n} a_{il}$ for all $i, j \in N$.[39] For each agent i define $\nu_i : 2^N \to [0, 1]$ as in (24). Let $x \in B$ and consider p as in equation (22). We have that $p_{\pi(l)} = 0$ if and only if $a_{i\pi(l)} = 0$, that is, $\pi(l)$ does not belong to the neighborhood N_i of i. Otherwise, $p_{\pi(l)} = f(k/N_i) - f((k-1)/N_i)$ where k belongs to $\{1, ..., N_i\}$ and also corresponds to the k-th highest value of x once the latter is restricted to N_i. It follows that $T_i(x)$ is the convex linear combination of the order statistics of x restricted to N_i with weights $\{f(k/N_i) - f((k-1)/N_i)\}_{k=1}^{N_i}$, that is, $T_i(x)$ is an L-estimator. ▲

The specification in (24) allows for exploring economically relevant phenomena that are precluded by linear aggregators, possibly using tools already developed in decision theory. As an example, if f_i is set to be equal to the prominent Prelec's probability weighting function [74], (i.e., $f_i(s) = \exp(-(-\ln(s))^{\alpha_i})$ with $\alpha_i \in (0, 1)$) we obtain a one-parameter function with a clear psychological foundation.[40] Indeed, such a functional specification characterizes an agent who is particularly sensitive to the *range* of opinions in the distribution, and that assigns a disproportionately high weight to extreme stances, with the size of the distortion decreasing in $\alpha_i \in (0, 1)$ (cf. Figure 2). More generally, using an f_i different from the identity map is a way to introduce a *perception bias* a la Banerjee and Fudenberg [9] in a model of naive and nonequilibrium learning.

We conclude the section on Choquet aggregators by discussing another subclass for which the issue of convergence is rather easy to settle. Before doing so, observe that for any permutation $\pi : \{1, ..., n\} \to \{1, ..., n\}$ we can construct the convex set

$$B_\pi = \left\{ x \in B : x_{\pi(1)} \geq x_{\pi(2)} \geq ... \geq x_{\pi(n)} \right\}.$$

In other words, $x \in B_\pi$ if and only if the i-th highest opinion belongs to agent $\pi(i)$ for all $i \in N$. If we denote by Π the collection of all permutations, B is the union of these subsets, that is, $B = \cup_{\pi \in \Pi} B_\pi$. Given (23), it follows that for each $\pi \in \Pi$ there exists a stochastic matrix $W_\pi \in \mathcal{W}$ such that

$$T(x) = W_\pi x \qquad \forall x \in B_\pi.$$

---

[39] Clearly, for W to be a well defined stochastic matrix, one needs to make the economically natural assumption that each agent is connected to at least another agent, that is, for each $i \in N$ there exists $l \in N$ such that $a_{il} > 0$.

[40] Clearly, f_i is defined only on $(0, 1]$, but it also admits a unique continuous extension to $[0, 1]$. Moreover, the extension takes value 0 in 0.
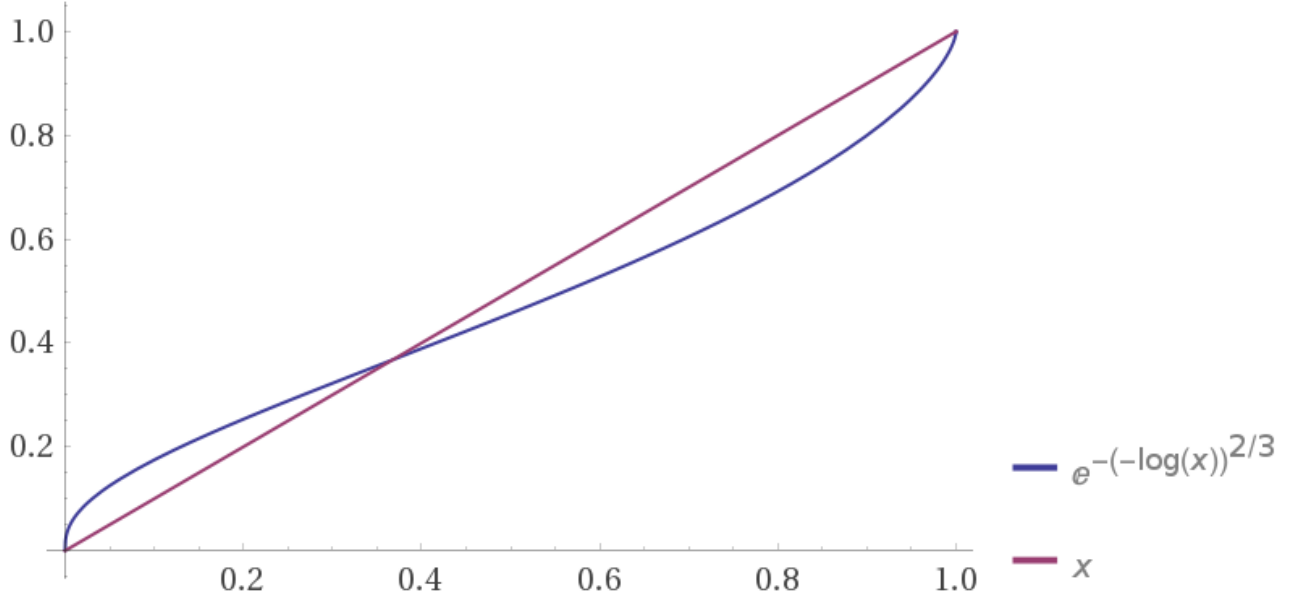
Figure 2: Probability Distortion Function

Thus, for each $x \in B$ there exists $\pi \in \Pi$ such that $x \in B_\pi$ and $T(x) = W_\pi x$, and the update $T(x)$ belongs to a set $B_{\pi'}$ for some $\pi' \in \Pi$. A priori, we might have $B_{\pi'} \neq B_\pi$, yielding that $T^2(x) = W_{\pi'} W_\pi x$. More in general, at each round, the updating is done via a stochastic matrix that might change, but comes from the finite set $\{W_\pi\}_{\pi \in \Pi}$. The next condition guarantees that only the first matrix W_π depends on $x \in B_\pi$ and then $T^t(x) = W_\pi^t x$ for all $t \in \mathbb{N}$.

Definition 17 *Let T be a Choquet opinion aggregator. We say that T is* assortative *if and only if for each $\pi \in \Pi$ and for each $i, i' \in N$*

$$i' \geq i \implies \nu_{\pi(i)} \left(\cup_{j=1}^{l} \{\pi(j)\} \right) \geq \nu_{\pi(i')} \left(\cup_{j=1}^{l} \{\pi(j)\} \right) \qquad \forall l \in N. \tag{25}$$

In words, this means that individuals with higher opinions assign a higher weight to individuals with high opinions. To see this consider $x \in B$ and assume that $\pi \in \Pi$ is such that $x_{\pi(1)} \geq ... \geq x_{\pi(n)}$. Consider the agents $\pi(i)$ and $\pi(i')$ where $i' \geq i$. By construction, the opinion of $\pi(i)$ is higher than the one of $\pi(i')$. By (23), we also have that

$$T_{\pi(i)}(x) = \int_N x d\nu_{\pi(i)} = \sum_{l=1}^{n} x_{\pi(l)} p_{\pi(l)}^{\pi(i)} \text{ and } T_{\pi(i')}(x) = \int_N x d\nu_{\pi(i')} = \sum_{l=1}^{n} x_{\pi(l)} p_{\pi(l)}^{\pi(i')}$$

where $p^{\pi(i)} \in \Delta$ (resp., $p^{\pi(i')}$) is computed via (22) by replacing ν_i with $\nu_{\pi(i)}$ (resp., $\nu_{\pi(i')}$). Condition (25) amounts to require that $p^{\pi(i)}$ dominates in first order stochastic dominance $p^{\pi(i')}$ once their components are rearranged according to π. Thus, our

40

definition captures the idea that, in an assortative society, a change in the stance of the individuals with a higher opinion affects more the individuals with a higher opinion, because they assign to them more weight.

Assortative societies naturally arise when agents are allowed to endogenously select their network and try to trade-off the benefit of interactions with dislike to disagreement (see, e.g., Bolletta and Pin [13] and Frick, Iijima, and Ishi [32]). It turns out that when a Choquet aggregator T is assortative, given the starting point x, the evolution of the system is described by iteration of a stochastic matrix W as in the linear case.

Proposition 9 *Let T be a Choquet aggregator. If T is assortative, then for each $x \in B$ there exists $W \in \{W_\pi\}_{\pi \in \Pi}$ such that*

$$T^t(x) = W^t x \qquad \forall t \in \mathbb{N}.$$

As a consequence, in assortative societies, it is much easier to compute the long-run dynamics of the opinions' updates. Indeed, for every initial opinion $x \in B_\pi$, the results of the linear case (see Golub and Jackson [42]), applied to $W = W_\pi$, characterize the limit weight of each agent in the society. However, these weights are dependent on the initial distribution: if one starts from $y \in B_{\pi'}$ with $\pi \neq \pi'$, then there is no guarantee that the limit influence of each agent is the same.

Example 7 (Assortativeness) Consider an arbitrary capacity $\bar{\nu} : 2^N \to [0,1]$ and let $\delta_i : 2^N \to [0,1]$ be the Dirac at i for all $i \in N$.[41] Let $\gamma \in [0,1]$ and define the capacity $\nu_i : 2^N \to [0,1]$ by

$$\nu_i(A) = \gamma \delta_i(A) + (1-\gamma)\bar{\nu}(A) \qquad \forall A \subseteq N, \forall i \in N. \tag{26}$$

Easy computations show that the Choquet aggregator T (where ν_i is defined as in (26) for all $i \in N$) is assortative. Thus, we have that for each $\pi \in \Pi$ and $x \in B_\pi$

$$T^t(x) = W_\pi^t x$$

where the i-th row of W_π is $\gamma e^i + (1-\gamma)\bar{p}^\pi$ for all $i \in N$ and $\bar{p}^\pi \in \Delta$ is derived from $\bar{\nu}$ via the formula (22).[42] Here, $\bar{\nu}$ captures the commonly perceived informativeness of the signals of agents' groups. However, every agent overweights her own opinion. Therefore, the resulting opinion dynamics are described by the powers of the matrix W_π whose entries depend on the initial ordering of signals' realizations. Finally, by Corollary 1, if $\gamma \in (0,1)$, then T is convergent and, in particular,[43]

$$\bar{T}_i(x) = \bar{p}^\pi \cdot x \qquad \forall \pi \in \Pi, \forall x \in B_\pi, \forall i \in N.$$

---

[41] That is, $\delta_i(A) = 1$ if $i \in A$ and 0 otherwise for all $A \subseteq N$.

[42] To signal the dependence on π, we add the superscript π to $\bar{p}$.

[43] Observe that $T = \gamma I + (1-\gamma)S$ where $S_i(x) = \int_N x d\bar{\nu}$ for all $i \in N$ and for all $x \in B$. By Lemma 3 and since T is convergent and robust, set $\bar{x} \stackrel{\text{def}}{=} \bar{T}(x) = \lim_t T^t(x) \in E(T)$. Since $\gamma \in (0,1)$, this implies that $\bar{x} = T(\bar{x}) = \gamma\bar{x} + (1-\gamma)S(\bar{x})$, yielding that $S(\bar{x}) = \bar{x}$.

6 Vox populi, vox Dei

Motivated by our robust statistics foundation, we next study if the updating procedure $\{T^t(x^0)\}_{t \in \mathbb{N}}$ leads to estimates which allow either the agents in the network or an external observer to learn the truth as the size of the population becomes larger and larger. More formally, we consider the same setup of Section 3, with the caveat that here everything is parametrized by the size n of the population. Recall that there exists a true parameter $\mu \in I$ and each agent $i \in N = \{1, ..., n\}$ observes a signal $X_i(n)$ defined over a probability space $(\Omega, \mathcal{F}, P)$. In this section, we make three assumptions:

Assumption 1 $I = \mathbb{R}$.

Assumption 2 For each $n \in \mathbb{N}$ we assume that $X_i(n) = \mu + \varepsilon_i(n)$ for all $i \in N$ where $\{\varepsilon_i(n)\}_{i \in N, n \in \mathbb{N}}$ is an array of uniformly bounded, symmetric, and independent random variables.[44]

Assumption 3 $\bar{T}_i(n) = \bar{T}_j(n)$ for all $i, j \in N$ and for all $n \in \mathbb{N}$.

By Lemma 3, Assumption 3 is satisfied whenever $T(n)$ is a convergent consensus robust opinion aggregator (cf. Section 5.3.1 for conditions which guarantee convergence and Proposition 6 for conditions which grant consensus).[45]

Given $n \in \mathbb{N}$, if agents update their estimates via a convergent robust opinion aggregator $T(n)$, then each agent i will reach a final estimate $\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega))$ for all $\omega \in \Omega$. It is then natural to ask whether or not this estimate gets arbitrarily close to the true parameter μ as the population size increases. Following Golub and Jackson [42], we are interested in whether or not the society becomes wise in the limit, where the latter property is formalized as follows:

Definition 18 *Let* $\{T(n)\}_{n \in \mathbb{N}}$ *be a sequence of robust opinion aggregators. The sequence* $\{T(n)\}_{n \in \mathbb{N}}$ *is* wise *if and only if*

$$\bar{T}_i(n)(X_1(n), ..., X_n(n)) \xrightarrow{P} \mu \qquad \forall i \in N. \tag{27}$$

---

[44]Formally, the property of symmetry means that for each $i \in N$ and for each $n \in \mathbb{N}$

$$P(\{\omega \in \Omega : \varepsilon_i(n)(\omega) \in B\}) = P(\{\omega \in \Omega : -\varepsilon_i(n)(\omega) \in B\})$$

for all Borel sets $B \subseteq \mathbb{R}$.

[45]Let $T(n) : \mathbb{R}^n \to \mathbb{R}^n$ be a convergent consensus robust opinion aggregator. This implies that $\lim_t (T(n))^t(x) = \bar{T}(n)(x) \in E(T(n)) = D$ for all $x \in B$. Since $\bar{T}(n)(x)$ is a constant vector for all $x \in B$, this implies Assumption 3.

Even though Assumption 3 is not strictly necessary, having it facilitates our reasoning, and it is also conceptually relevant. In fact, in order for the entire set of agents to learn the true parameter μ, it seems natural to consider a situation in which the result of the updating procedure is common across agents.

We first review the wisdom of the crowd for the linear case. But, before doing so, we introduce some notation.

Notation With $\hat{I}$, we denote a bounded open interval such that $X_i(n)(\omega) \in \hat{I}$ for all $\omega \in \Omega$, $i \in N$, and $n \in \mathbb{N}$. We denote by $\ell \overset{\text{def}}{=} \sup \hat{I} - \inf \hat{I}$ the length of $\hat{I}$. Moreover, we denote the collection of probability vectors in $\mathbb{R}^n$ by Δ_n, rather than just Δ.

If $T(n)$ is linear, so is $\bar{T}(n)$ and the latter is represented by a matrix $\bar{W}(n)$. In this case, Assumption 3 yields that all the rows of $\bar{W}(n)$ coincide with the left Perron-Frobenius eigenvector $s(n) \in \Delta_n$ associated to the eigenvalue 1. DeMarzo, Vayanos, and Zwiebel [27] as well as Golub and Jackson [42] call $s(n)$ the *vector of influence weights* and the latter show that if $\lim_{n \to \infty} \max_{k \in N} s_k(n) \to 0$, then $\{T(n)\}_{n \in \mathbb{N}}$ is wise. In generalizing this result to the nonlinear case, we face two main difficulties. From a mathematical point of view, it is not obvious how to generalize the notion of left eigenvector. From an interpretive point of view, the conceptual relevance of the vector $s(n)$ comes from the immediate computability in terms of the primitive matrix $W(n)$. But, other than specific cases, it is not immediately evident in terms of primitives what is its meaning. To wit, it is not immediate by a mere inspection of the components of $W(n)$, if one can obtain useful bounds on $\max_{k \in N} s_k(n)$.

For such a reason, we make the following trivial observation: in the linear case, the vector $s(n)$ coincides with the gradient $\nabla \bar{T}_i(n)$ of $\bar{T}_i(n)$ at any point $x \in \mathbb{R}^n$. This simple observation is useful in two dimensions. For starters, it allows us to overcome the difficulty of not having a viable notion of left eigenvector for nonlinear operators, and it clarifies how the j-th component of the influence vector $s(n)$ captures the idea of "marginal contribution of agent j" to the final opinion $\bar{T}_i(n)(X_1(n)(\omega),...,X_n(n)(\omega))$. Finally, via the chain rule, if $\bar{T}(n)$ is the pointwise limit of $T^t(n)$ as t runs to infinity, it allows us to bound these marginal contributions via the marginal contributions of each agent at each round, that is, via the gradient of $T(n)$. Unfortunately, in proceeding this way, we face some technical complications. Our opinion aggregators might be non-differentiable. Nevertheless, being Lipschitz continuous, by Rademacher's Theorem, they are almost everywhere differentiable. Let $\mathcal{D}(\bar{T}(n)) \subseteq \hat{I}^n$ (resp., $\mathcal{D}(T(n))$) be the subset of $\hat{I}^n$ where $\bar{T}(n)$ (resp., $T(n)$) is differentiable.

Definition 19 *Let $T(n) : \mathbb{R}^n \to \mathbb{R}^n$ be a robust opinion aggregator. We say that $s(T(n)) \in \mathbb{R}^n$ is* the influence vector of $T(n)$ *if and only if*

$$s_i(T(n)) = \sup_{x \in \mathcal{D}(\bar{T}(n))} \frac{\partial \bar{T}_1(n)}{\partial x_i}(x) \qquad \forall i \in N.$$

As we mentioned, the above definition of influence vector coincides with the one of Golub and Jackson whenever $T(n)$ is linear since $s(T(n)) = s(n)$. For, in this case, one has that

$$s_i(T(n)) = \frac{\partial \bar{T}_1(n)}{\partial x_i}(x) = s_i(n) \qquad \forall i \in N, \forall x \in \mathcal{D}(\bar{T}(n)). \qquad (28)$$

In the robust case, $s_i(T(n))$ captures the maximal marginal contribution of a change of the opinion i on the final consensus estimate. Intuitively, the next result shows that under the assumptions of the current section, if the influence weight of each agent goes to zero, then the estimates of the network become more and more accurate.

Proposition 10 *Let $\{T(n)\}_{n \in \mathbb{N}}$ be a sequence of odd robust opinion aggregators. If there exist two sequences $\{c(n)\}_{n \in \mathbb{N}}$ and $\{w(n)\}_{n \in \mathbb{N}}$ such that for each $n \in \mathbb{N}$, $c(n) \in \mathbb{R}$, $w(n) \in \Delta_n$, and*

$$s(T(n)) \leq c(n) w(n) \ \text{ as well as } \ c(n)^2 \max_{k \in N} w_k(n) \to 0 \ \text{ as } n \to \infty, \qquad (29)$$

then $\{T(n)\}_{n \in \mathbb{N}}$ is wise.

Compared to the linear case, we must observe that Proposition 10 differs only in one central aspect: our result relies on the signals being symmetric around μ. In fact, first, linear opinion aggregators are trivially odd. Second, our condition in (29) is equivalent to the one of Golub and Jackson.[46] On the one hand, the assumption of symmetric errors, paired with $T(n)$ being odd, guarantees that $\bar{T}_i(n)$ is an unbiased estimator, that is,

$$\mathbb{E}\left(\bar{T}_i(n)(X_1(n), ..., X_n(n))\right) = \mu \qquad \forall i \in N, \forall n \in \mathbb{N}. \qquad (30)$$

On the other hand, via McDiarmid's inequality, (29) guarantees that $\bar{T}_i(n)(X_1(n), ..., X_n(n))$ converges in probability to its expectation as n becomes larger, yielding that $\bar{T}_i(n)$ is a consistent estimator, that is, (27) holds. Finally, observe that odd aggregators naturally arise when the profiles of loss functions used by the agents are symmetric (cf. Proposition 1). The next corollary provides a simple condition which yields (29) and, in turn, that $\{T(n)\}_{n \in \mathbb{N}}$ is wise.

---

[46]In the linear case, given (28), note that

$$1 = \sum_{i=1}^{n} s_i(n) = \sum_{i=1}^{n} s_i(T(n)) \leq c(n) \sum_{i=1}^{n} w_i(n) = c(n) \quad \forall n \in \mathbb{N}.$$

This implies that for each $n \in \mathbb{N}$

$$c(n)^2 \max_{k \in N} w_k(n) \geq c(n) \max_{k \in N} w_k(n) \geq \max_{k \in N} s_k(n) \geq 0,$$

yielding that $\lim_{n \to \infty} \max_{k \in N} s_k(n) \to 0$. As for the opposite implication, note that, in the linear case, we can always set $c(n) = 1$ and $w(n) = s(n)$ for all $n \in \mathbb{N}$.

Corollary 5 *Let* $\{T(n)\}_{n \in \mathbb{N}}$ *be a sequence of odd robust opinion aggregators. If we have that*

$$\max_{k \in N} s_k(T(n)) = o\left(\frac{1}{\sqrt{n}}\right),$$

then $\{T(n)\}_{n \in \mathbb{N}}$ *is wise.*

Next remark elaborates on how fast the crowd becomes wise.

Remark 4 We can also provide bounds on both the variance of $\bar{T}_i(n)(X_1(n), ..., X_n(n))$ and the probability of

$$\left| \bar{T}_i(n)(X_1(n), ..., X_n(n)) - \mu \right| \geq \delta.$$

This helps in elucidating the convergence result contained in Proposition 10. If $T(n)$ is an odd robust opinion aggregator such that $s(T(n)) \leq cw(n)$ for some $c \in (0, \infty)$ and $w(n) \in \Delta_n$, we have that (30) holds and for each $\delta \in (0, \ell]$

$$P\left(\left\{\omega \in \Omega : \left| \bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu \right| \geq \delta \right\}\right) \leq 2e^{-\frac{2\delta^2}{\ell^2 c^2 \max_{k \in N} w_k(n)}}. \quad (31)$$

Since our random variables take values in a bounded interval $\hat{I}$, the difference

$$\left| \bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu \right|$$

can be at most the length of the interval, that is, ℓ. Thus, the previous inequality might provide a useful bound in terms of controlling for deviations from the actual parameter. Mathematically, (31) is McDiarmid's inequality. In turn, this allows us to control the variance of $\bar{T}_i(n)(X_1(n), ..., X_n(n))$. Indeed, we have that for each $i \in N$

$$\mathrm{Var}\big(\bar{T}_i(n)(X_1(n), ..., X_n(n))\big) \leq \ell^2 c^2 \max_{k \in N} w_k(n)\left[1 - e^{-\frac{2}{c^2 \max_{k \in N} w_k(n)}}\right]. \quad (32)$$

Thus, if $\max_{k \in N} w_k(n)$ gets smaller, then the bound on the variance gets smaller. ▲

Before moving on, we note that the above results and remark encompass also the case when $T(n)$ is not a convergent operator (an assumption we indeed never made in this section). In such a case, recall that $\bar{T}(n)$ is the limit of the time averages of $\{T^t(n)\}_{t \in \mathbb{N}}$. Conceptually, this extra layer of generality is interesting if we think about the following question: can an external observer learn the true parameter μ by observing only part of the updating dynamics of a subset of the agents? More formally, assume that the external observer from a specific point in time, say m, gets to see the updating process $\{T_i^{t+m}(n)(X_1(n)(\omega), ..., X_n(n)(\omega))\}_{t=1}^{\tau}$ of agent i. By Theorem 2, we know that as $\tau \to \infty$

$$\frac{1}{\tau} \sum_{t=1}^{\tau} T_i^{t+m}(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) \to \bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) \; \forall \omega \in \Omega.$$

Our results up so far show that the external observer can use $\bar{T}(n)$ to extract information about the underlying parameter, even if the opinions of the individual agents in the network are not stabilizing.

The next obvious question we tackle relates to the possibility of verifying condition (29) in terms of the original sequence of robust opinion aggregators $\{T(n)\}_{n \in \mathbb{N}}$.

Proposition 11 *Let $\{T(n)\}_{n \in \mathbb{N}}$ be a sequence of odd robust opinion aggregators. If there exist two sequences $\{c(n)\}_{n \in \mathbb{N}}$ and $\{w(n)\}_{n \in \mathbb{N}}$ such that for each $n \in \mathbb{N}$, $c(n) \in \mathbb{R}$, $w(n) \in \Delta_n$, and for each $i, h \in N$*

$$\sup_{x \in \mathcal{D}(T(n))} \frac{\partial T_h(n)}{\partial x_i}(x) \le c(n) w_i(n) \ \text{ and } \ c(n)^2 \max_{k \in N} w_k(n) \to 0 \ \text{ as } n \to \infty, \qquad (33)$$

then (29) holds and $\{T(n)\}_{n \in \mathbb{N}}$ is wise.

We next illustrate how our previous result can become handy in checking condition (29), particularly in the context of our statistical foundation. As a by-product, we will obtain that, under the assumptions of this section, the wisdom of the crowd can be achieved as long as the minimum degree of connections gets larger as the population size increases.

Example 8 Assume that each agent processes signals following Huber's estimation procedure (see Sections 3.1 and 3.2), that is,

$$T_i(n)(x) \in \operatorname{argmin}_{c \in \mathbb{R}} \sum_{j \in N_i(n)} \rho_i(n)(x_j - c) \qquad \forall i \in N, \forall x \in \mathbb{R}^n, \forall n \in \mathbb{N}.$$

Assume that the profile of loss functions $\phi(n) = (W(n), \rho(n))$ used by the agents is as in Proposition 5 where the agents' weights $w_{ij}(n)$ are uniform over their neighborhoods $N_i(n)$. Moreover, assume that there exists $\bar{c} \in \mathbb{R}$ such that

$$\frac{\rho_i''(n)(s)}{\rho_i''(n)(s')} \le \bar{c} \quad \forall i \in N, \forall n \in \mathbb{N}, \forall s, s' \in [-2\ell, 2\ell].$$

By the Implicit Function Theorem,[47] we have that $T(n)$ is Frechet differentiable and

$$\frac{\partial T_h(n)}{\partial x_i}(x) \le \bar{c} \frac{1}{\min_{k \in N} |N_k(n)|} \qquad \forall i, h \in N, \forall x \in \hat{B}, \forall n \in \mathbb{N}.$$

---

[47] Let $s, s' \in [-2\ell, 2\ell]$ be such that

$$\rho_h''(n)(s) = \max_{\tilde{s} \in [-2\ell, 2\ell]} \rho_h''(n)(\tilde{s}) \ \text{ and } \ \rho_h''(n)(s') = \min_{\tilde{s} \in [-2\ell, 2\ell]} \rho_h''(n)(\tilde{s}).$$

Easy computations yield that

$$\frac{\partial T_h(n)}{\partial x_i}(x) = \frac{a_{hi}(n) \rho_h''(n)(x_i - T_h(x))}{\sum_{l=1}^{n} a_{hl}(n) \rho_h''(n)(x_l - T_h(x))} \le \frac{\rho_h''(n)(s)}{|N_h(n)| \rho_h''(n)(s')} \le \frac{\bar{c}}{\min_{k \in N} |N_k(n)|}.$$

By Proposition 11, wisdom is reached at the limit (i.e., (27) holds) if the minimal degree in the society is growing sufficiently fast,[48] that is,

$$\frac{1}{\min_{k \in N} |N_k(n)|} = o\left(\frac{1}{\sqrt{n}}\right). \tag{34}$$

Note that condition (34) allows each agent to be connected to a vanishing fraction of the society.   ▲

A word of caution must be mentioned here about the possibility of obtaining the wisdom of the crowd in the realm of robust opinion aggregation. It can be shown that given any sequence of linear opinion aggregators $\{T(n)\}_{n \in \mathbb{N}}$ even if this sequence is wise, we can slightly perturb it and obtain a new sequence which is not wise.[49] In particular, in this latter case, the limit estimate will be bounded away from the true parameter μ. The intuition of this result is simple. We obtain the wisdom of the crowd relying on the errors' symmetry as well as on the fact that our robust opinion aggregators recognize this symmetry, that is, they are odd. But as soon as this condition is not satisfied, it is not difficult to imagine that the limit estimate might not be consistent: for example, assume that each agent observes every other agent and aggregates opinions according to the median computed with uniform weights. If errors are not symmetric, the wisdom of the crowd cannot be achieved. What it is more difficult to show is that small perturbations, even of wise sequences, are enough to generate such type of behavior, thus showing that the wisdom of the crowd relies on special assumptions which might or might not be satisfied in applications.

Linear opinion aggregators are a knife-edge case in terms of symmetry. The wisdom of the crowd result of Golub and Jackson [42] does not depend on the informativeness of the signals received by the individuals (i.e., it only assumes that signals have a finite second moment). For instance, suppose that the array of signals $(X_i(n))_{i \in N, n \in \mathbb{N}}$ analyzed above is replaced with an array $(Y_i(n))_{i \in N, n \in \mathbb{N}}$ where $Y_i(n)$ is a mean preserving spread of $X_i(n)$. Such a change is irrelevant for a linear opinion aggregator, even if the signals are less informative, the wisdom of the crowd is obtained if and only if it was obtained under the original signal structure. Instead, robust opinion aggregators do not implicitly assume symmetry in the way in which extreme realizations are weighted. For example, they leave open the possibility that extremely negative realizations are

---

[48]Similarly to the proof of Corollary 5, set

$$w_i(n) = \frac{1}{n} \text{ and } c(n) = n\bar{c}\frac{1}{\min_{k \in N} |N_k(n)|} \quad \forall i \in N, \forall n \in \mathbb{N}.$$

[49]Moreover, the sequence of loss functions justifying the sequence of perturbed robust opinion aggregators can also be chosen to be "close" to the profiles of the quadratic loss functions justifying $\{T(n)\}_{n \in \mathbb{N}}$ (cf. Theorem 1).

overweighted by the agents. Therefore, by making the tails of the distribution fatter, a mean preserving spread of the individuals' signals may shift the limit consensus downward. In general, the effect of less precise signals is ambiguous, but a precise result can be obtained in the case of concave (resp., convex) opinion aggregators.[50] In this case, we will have that for each $i \in N$ and for each $n \in \mathbb{N}$

$$\mathbb{E}\left(\bar{T}_i\left(n\right)\left(X_1\left(n\right), ..., X_n\left(n\right)\right)\right) \geq \mathbb{E}\left(\bar{T}_i\left(n\right)\left(Y_1\left(n\right), ..., Y_n\left(n\right)\right)\right) \ \left(\text{resp.}, \ \leq\right).$$

Concave and convex opinion aggregators encompass some of the aggregators described in Sections 4 and 5.4. In the first case, the aggregator obtained in equation (11) is concave (resp., convex) whenever $\lambda < 0$ (resp., $\lambda > 0$). In the second case, the Choquet aggregator in which each agent uses a capacity as in (24) is concave (resp., convex) provided each f_i is convex (resp., concave).

7   Discussion

In this section, we discuss two main aspects of our model and how they can be relaxed.

Strategic interaction   In this paper, we have provided a justification for robust opinion aggregation in terms of repeated robust estimation of a location parameter, in line with the approach of DeMarzo, Vayanos, and Zwiebel [27]. The other classical motivation for DeGroot's model is a process of *myopic* best response dynamics of agents that are learning to play a coordination game with quadratic payoffs. Note that the period by period minimization of a robust loss function described in Section 3.1, can be alternatively interpreted as a best response dynamic process under a much broader class of payoff functions. It is then clear that all our results leading to convergence (resp., consensus) can be recasted as procedures to select a Nash equilibrium of the corresponding coordination game (resp., full coordination equilibrium). Nevertheless, best response dynamics assume some degree of bounded rationality in that agents best respond to *past* coplayers' actions. In our related work [19], we show that the action dynamics induced by robust aggregators arise also when the agents care about the *current* and *future* action profiles and act strategically. In particular, this Markovian behavior is derived under the additional assumption of a minimal amount of *inertia*

---

[50]An opinion aggregator $T\left(n\right) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is concave (resp., convex) if and only if for each $\lambda \in \left(0, 1\right)$ and for each $x, y \in \mathbb{R}^n$

$$T\left(n\right)\left(\lambda x + \left(1 - \lambda\right) y\right) \geq \lambda T\left(n\right)\left(x\right) + \left(1 - \lambda\right) T\left(n\right)\left(y\right) \quad \left(\text{resp.}, \ \leq\right).$$

of the agents with respect their own past actions. This result provides a fully rational foundation of the dynamics studied in this paper in terms of (forward looking) equilibrium behavior.

Translation invariance In our microfoundation, we have considered agents that initially observe the realizations of different location experiments. In this setting, it is natural to consider loss functions defined over the residuals (i.e., the differences between observations and estimate). It immediately follows that the induced opinion aggregator satisfies translation invariance, which might be seen as too restrictive. Indeed, under the strategic interaction interpretation considered above, one might desire to allow for more general properties of the best response function obtained, for instance, from a payoff function that does not only depend on actions' differences. In our related work [19], we relax the assumptions on the payoff functions of the agents, naturally obtaining equilibrium maps that satisfy translation *subinvariance*, that is, agents react less than proportionally to uniform shifts in actions. It turns out that most of the results of this paper (e.g., Theorems 1 and 2) would continue to hold under minor adaptations.

A   Appendix: A robust foundation and examples

In this appendix, we prove all the results of Sections 3 and 4 as well as Proposition 5 in Section 5. Recall that a loss function is a functional from $\mathbb{R}^n$ to $\mathbb{R}_+$. We next prove two ancillary lemmas which will highlight some useful properties satisfied by a profile of robust loss functions $\phi = (\phi_i)_{i=1}^n$. Recall that ϕ is a profile of robust loss functions, denoted by $\phi \in \Phi_R$, if and only if it is sensitive and has increasing shifts (Definitions 2 and 3) and each ϕ_i is lower semicontinuous.

Lemma 4 *Let $\phi = (\phi_i)_{i=1}^n$ be a profile of loss functions. If $\phi \in \Phi_R$, then for each $i \in N$ and $\tilde{z} \in \mathbb{R}^n$*

$$\tilde{z} \gg 0 \implies \phi_i(\tilde{z}) > \phi_i\left(\tilde{z} - \min_{j \in N} \tilde{z}_j e\right),$$

and

$$0 \gg \tilde{z} \implies \phi_i(\tilde{z}) > \phi_i\left(\tilde{z} - \max_{j \in N} \tilde{z}_j e\right).$$

Proof. Fix $i \in N$. Consider $\tilde{z} \in \mathbb{R}^n$ such that $\tilde{z} \gg 0$. Define $z = \tilde{z} - \min_{j \in N} \tilde{z}_j e$, $v = 0$, and $h = \min_{j \in N} \tilde{z}_j$. Note that $z \geq v$ as well as $h \in \mathbb{R}_{++}$. Since ϕ has increasing shifts and is sensitive, we obtain that

$$\phi_i(\tilde{z}) - \phi_i\left(\tilde{z} - \min_{j \in N} \tilde{z}_j e\right) = \phi_i(z + he) - \phi_i(z)$$

$$\geq \phi_i(v + he) - \phi_i(v) = \phi_i\left(\min_{j \in N} \tilde{z}_j e\right) - \phi_i(0) > 0,$$

proving the first inequality. Consider now $\tilde{z} \in \mathbb{R}^n$ such that $0 \gg \tilde{z}$. Define $z = \max_{j \in N} \tilde{z}_j e$, $v = \tilde{z}$, and $h = -\max_{j \in N} \tilde{z}_j$. Note that $z \geq v$ as well as $h \in \mathbb{R}_{++}$. Since ϕ is sensitive and has increasing shifts, we obtain that

$$0 > \phi_i(0) - \phi_i\left(\max_{j \in N} \tilde{z}_j e\right) = \phi_i(z + he) - \phi_i(z)$$

$$\geq \phi_i(v + he) - \phi_i(v) = \phi_i\left(\tilde{z} - \max_{j \in N} \tilde{z}_j e\right) - \phi_i(\tilde{z}),$$

proving the second equality. ∎

Next lemma shows that the property of increasing shifts is a property of convexity of each loss function ϕ_i.

Lemma 5 *Let $\phi = (\phi_i)_{i=1}^n$ be a profile of loss functions. If $\phi \in \Phi_R$, then for each $i \in N$ and for each $x \in \mathbb{R}^n$ the function $f_{i,x} : \mathbb{R} \to \mathbb{R}_+$, defined by $f_{i,x}(c) = \phi_i(x - ce)$ for all $c \in \mathbb{R}$, is convex. Moreover, if ϕ has strictly increasing shifts, then $f_{i,x}$ is strictly convex for all $i \in N$ and for all $x \in \mathbb{R}^n$.*

Proof. Fix $i \in N$ and $x \in \mathbb{R}^n$. Define $g_{i,x} : \mathbb{R} \to \mathbb{R}_+$ by $g_{i,x}(c) = \phi_i(x + ce)$ for all $c \in \mathbb{R}$. Consider $c_1, c_2 \in \mathbb{R}$ such that $c_1 > c_2$ and $h > 0$. Since $\phi \in \Phi_R$ and $x + c_1 e \geq x + c_2 e$, it follows that

$$g_{i,x}(c_1 + h) - g_{i,x}(c_1) = \phi_i((x + c_1 e) + he) - \phi_i(x + c_1 e)$$

$$\geq \phi_i((x + c_2 e) + he) - \phi_i(x + c_2 e)$$

$$= g_{i,x}(c_2 + h) - g_{i,x}(c_2).$$

By [75, Problem N, pp. 223–224], it follows that $g_{i,x}$ is midconvex. Since ϕ_i is lower semicontinuous, we have that $g_{i,x}$ is measurable. By [75, Theorem C, p. 221], it follows that $g_{i,x}$ is continuous. By [75, Theorem A, p. 212], this implies that $g_{i,x}$ is convex. Finally, observe that $f_{i,x} = g_{i,x} \circ h$ where $h(c) = -c$ for all $c \in \mathbb{R}$, yielding that $f_{i,x}$ is convex being the composition of a convex function with an affine function. Next, assume that ϕ has strictly increasing shifts and, in particular, has increasing shifts. By the previous part of the proof, $g_{i,x}$ is convex. By contradiction, assume that $g_{i,x}$ is not strictly convex. This implies that there exists an interval $[d_2, d_1]$, with $d_2 < d_1$, where $g_{i,x}$ is affine. Define $c_1 = \frac{1}{2}d_1 + \frac{1}{2}d_2$, $c_2 = d_2$, and $h = (d_1 - d_2)/2$. Note that $c_1 > c_2$ and $h > 0$. Since ϕ has strictly increasing shifts, by the same computations of the previous part of the proof, we have that

$$g_{i,x}(d_1) - g_{i,x}\left(\frac{1}{2}d_1 + \frac{1}{2}d_2\right) = g_{i,x}(c_1 + h) - g_{i,x}(c_1) > g_{i,x}(c_2 + h) - g_{i,x}(c_2)$$

$$= g_{i,x}\left(\frac{1}{2}d_1 + \frac{1}{2}d_2\right) - g_{i,x}(d_2),$$

yielding that $g_{i,x}\left(\frac{1}{2}d_1 + \frac{1}{2}d_2\right) < \frac{1}{2}g_{i,x}(d_1) + \frac{1}{2}g_{i,x}(d_2)$, a contradiction with affinity. Since $g_{i,x}$ is strictly convex, so is $f_{i,x} = g_{i,x} \circ h$. ∎

We next show that an estimation procedure like the one in (2) and (3) yields a selection of $\mathbf{T}^\phi$ which is a robust opinion aggregator. This is a generalization of (i) implies (ii) of Theorem 1.

Proposition 12 *Let $\phi = (\phi_i)_{i=1}^n$ be a profile of loss functions. If $\phi \in \Phi_R$, then the correspondence $\mathbf{T}^\phi$ admits a selection T^ϕ which is a robust opinion aggregator. Moreover, if ϕ has strictly increasing shifts, then $\mathbf{T}^\phi = T^\phi$ is single-valued and, in particular, is a robust opinion aggregator.*

Proof. Fix $i \in N$. We begin by considering the correspondence $\mathbf{T}_i^\phi : B \rightrightarrows I$ defined by

$$\mathbf{T}_i^\phi(x) = \operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce) \qquad \forall x \in B.$$

We next show that $\mathbf{T}_i^\phi$ is well defined, nonempty-, convex-, and compact-valued, and such that for each $x, y \in B$

$$x \geq y \implies \mathbf{T}_i^\phi(x) \geq_{\text{SSO}} \mathbf{T}_i^\phi(y) \tag{35}$$

where $\geq_{\text{SSO}}$ is the strong set order. Fix $x \in B$. We next show that

$$\forall d \notin \left[\min_{j \in N} x_j, \max_{j \in N} x_j\right], \exists c \in \left[\min_{j \in N} x_j, \max_{j \in N} x_j\right] \text{ s.t. } \phi_i(x - ce) < \phi_i(x - de). \tag{36}$$

Consider d as above. We have two cases either $d < \min_{j \in N} x_j$ or $d > \max_{j \in N} x_j$. In the first case, we have that $x - de \gg 0$, in the second case, we have that $0 \gg x - de$. By Lemma 4 and since $\phi \in \Phi_R$, if we set $\tilde{c} = \min_{j \in N} x_j - d$ (resp., $\max_{j \in N} x_j - d$), we obtain that

$$\phi_i(x - de) > \phi_i(x - de - \tilde{c}e) = \phi_i(x - ce)$$

where $c = \min_{j \in N} x_j \in [\min_{j \in N} x_j, \max_{j \in N} x_j]$ (resp., $c = \max_{j \in N} x_j \in [\min_{j \in N} x_j, \max_{j \in N} x_j]$), proving (36). By (36), we can conclude that

$$\min_{c \in \mathbb{R}} \phi_i(x - ce) = \min_{c \in I} \phi_i(x - ce) = \min_{c \in \left[\min_{j \in N} x_j, \max_{j \in N} x_j\right]} \phi_i(x - ce) \tag{37}$$

as well as

$$\operatorname{argmin}_{c \in \mathbb{R}} \phi_i(x - ce) = \operatorname{argmin}_{c \in I} \phi_i(x - ce) = \operatorname{argmin}_{c \in \left[\min_{j \in N} x_j, \max_{j \in N} x_j\right]} \phi_i(x - ce).$$

By a standard generalization of Weierstrass' Theorem (see, e.g., [3, Theorem 2.43]) and since ϕ_i is lower semicontinuous, this implies that the above minimization problems admit solution and each argmin is a compact set. By Lemma 5, $\operatorname{argmin}_{c \in \left[\min_{j \in N} x_j, \max_{j \in N} x_j\right]} \phi_i(x - ce)$

is also convex. Since x was arbitrarily chosen, this implies that $\mathbf{T}_i^\phi$ is well defined, nonempty-, convex-, and compact-valued and, in particular,

$$\emptyset \neq \mathbf{T}_i^\phi(x) \subseteq \left[\min_{j \in N} x_j, \max_{j \in N} x_j\right] \subseteq I \quad \forall x \in B. \tag{38}$$

We next prove (35). In order to do so, we rewrite explicitly (37) as a problem of parametric optimization/monotone comparative statics. Define $f : I \times B \to \mathbb{R}$ by

$$f(c, x) = -\phi_i(x - ce) \qquad \forall(c, x) \in I \times B.$$

It is immediate to see that

$$\mathbf{T}_i(x) = \operatorname{argmax}_{c \in I} f(c, x) \qquad \forall x \in B.$$

We next show that f has increasing differences in (c, x). Consider $x, y \in B$ as well as $c, d \in I$ such that $c \geq d$ and $x \geq y$. Define $z = x - ce$, $v = y - ce$, and $h = c - d$. Note that $z \geq v$ and $h \in \mathbb{R}_+$. Since $\phi \in \Phi_R$, it follows that

$$\begin{aligned} f(c, x) - f(d, x) = \phi_i(x - de) - \phi_i(x - ce) &= \phi_i(z + he) - \phi_i(z) \\ &\geq \phi_i(v + he) - \phi_i(v) = \phi_i(y - de) - \phi_i(y - ce) \\ &= f(c, y) - f(d, y). \end{aligned}$$

This shows that f satisfies the property of increasing differences in (c, x). By [67, Theorem 5], $\mathbf{T}_i$ satisfies (35). We finally show that $\mathbf{T}_i^\phi$ is such that for each $x \in B$ and for each $k \in \mathbb{R}$ such that $x + ke \in B$

$$c^\star \in \mathbf{T}_i^\phi(x) \iff c^\star + k \in \mathbf{T}_i^\phi(x + ke). \tag{39}$$

Fix $x \in B$. Consider $k \in \mathbb{R}$ such that $x + ke \in B$. Consider $c^\star \in \mathbf{T}_i^\phi(x)$. By definition, it follows that $\phi_i(x - c^\star e) \leq \phi_i(x - ce)$ for all $c \in \mathbb{R}$. This implies that

$$\phi_i(x + ke - (c^\star + k)e) = \phi_i(x - c^\star e) \leq \phi_i(x - (d - k)e) = \phi_i(x + ke - de) \quad \forall d \in \mathbb{R}.$$

By definition of $\mathbf{T}_i^\phi$, this implies that $c^\star + k \in \mathbf{T}_i^\phi(x + ke)$. Vice versa, if $c^\star + k \in \mathbf{T}_i^\phi(x + ke)$, then

$$\phi_i(x + ke - (c^\star + k)e) \leq \phi_i(x + ke - de) \qquad \forall d \in \mathbb{R},$$

yielding that

$$\phi_i(x - c^\star e) = \phi_i(x + ke - (c^\star + k)e) \leq \phi_i(x - ce) \qquad \forall c \in \mathbb{R},$$

proving that $c^\star \in \mathbf{T}_i^\phi(x)$.

To sum up, since $i \in N$ was arbitrarily chosen, we proved that, for each $i \in N$, $\mathbf{T}_i^\phi$ is well defined, nonempty-, convex-, and compact-valued, and satisfies (35) as well as (39). Observe also that $\mathbf{T}^\phi : B \rightrightarrows B$ is the product correspondence $\mathbf{T}^\phi = \prod_{i=1}^n \mathbf{T}_i^\phi$. We are ready to show that $\mathbf{T}^\phi$ admits a selection T^ϕ which is a robust opinion aggregator. Define $T^\phi : B \to B$ to be such that

$$T_i^\phi(x) = \min \mathbf{T}_i^\phi(x) \qquad \forall x \in B, \forall i \in N.$$

Since $\mathbf{T}_i^\phi(x)$ is compact for all $x \in B$, it follows that $T_i^\phi(x)$ is well defined and, in particular, $T_i^\phi(x) \in \mathbf{T}_i^\phi(x)$ for all $x \in B$ and for all $i \in N$, proving that T^ϕ is a selection of $\mathbf{T}^\phi$. By (38), it follows that $\mathbf{T}_i^\phi(ke) = \{k\}$ for all $k \in I$ and for all $i \in N$, proving that $T_i^\phi(ke) = k$ for all $k \in I$ and for all $i \in N$, that is, that T^ϕ is normalized. Next, consider $x, y \in B$ such that $x \geq y$. By (35), we have that $T_i^\phi(x) \geq T_i^\phi(y)$ for all $i \in N$, proving monotonicity of T_i^ϕ for all $i \in N$ and so of T^ϕ. Finally, consider $x \in B$ and $k \in \mathbb{R}$ such that $x + ke \in B$. By (39) and definition of $T_i^\phi(x)$ as well as $T_i^\phi(x + ke)$, we have that $T_i^\phi(x) \in \mathbf{T}_i^\phi(x)$ for all $i \in N$, yielding that $T_i^\phi(x) + k \in \mathbf{T}_i^\phi(x + ke)$ for all $i \in N$ and, in particular, $T_i^\phi(x) + k \geq T_i^\phi(x + ke)$ for all $i \in N$. This implies that $T_i^\phi(x + ke) = T_i^\phi(x) + k$ for all $i \in N$, proving translation invariance.[51]

Finally, by Lemma 5, if ϕ has strictly increasing shifts, then the map $c \mapsto \phi_i(x - ce)$ is strictly convex, yielding that each $\mathbf{T}_i^\phi$ is single-valued and so is $\mathbf{T}^\phi$. ∎

We are now ready to prove the main theorem of Section 3. Its proof (as well as others) would be extremely facilitated if I were to be equal to $\mathbb{R}$. In our case though, I is only assumed to be a closed interval with nonempty interior. Nevertheless, we can always extend T from B to the entire space $\mathbb{R}^n$. Next lemma, proved in the Online Appendix, guarantees this. Moreover, even though T might have many extensions, all the extensions generate the same sequence of updates and therefore, the same limiting behavior.

Lemma 6 *Let T be an opinion aggregator. The following statements are true:*

1. *If T is robust, then it admits an extension $S : \mathbb{R}^n \to \mathbb{R}^n$ which is also robust.*

2. *If T is robust and constant affine, then it admits a unique extension $S : \mathbb{R}^n \to \mathbb{R}^n$ which is robust and constant affine.*

---

[51]Fix $i \in N$. By the previous part of the proof, for each $x \in B$ and for each $k \in \mathbb{R}$ such that $x + ke \in B$, we have that $T_i^\phi(x + ke) \leq T_i^\phi(x) + k$. Next, note that if $x \in B$ and $x + ke \in B$, then $(x + ke) - ke = x \in B$. It follows that

$$T_i^\phi(x) = T_i^\phi((x + ke) - ke) \leq T_i^\phi(x + ke) - k,$$

proving the opposite inequality.

3. If T is normalized and monotone, then $\|T^t(x)\|_\infty \leq \|x\|_\infty$ for all $x \in B$ and for all $t \in \mathbb{N}$.

4. If $x \in B$, then there exists $\tilde{I} \subseteq I$ which is a compact subinterval of I with nonempty interior and $x \in \tilde{I}^n \overset{\text{def}}{=} \tilde{B}$. Moreover, if T is robust, the restriction $\tilde{T} = T_{|\tilde{B}}$ is a robust opinion aggregator and $\tilde{T}^t(x) = T^t(x)$ as well as $\bar{\tilde{T}}(x) = \bar{T}(x)$ for all $t \in \mathbb{N}$ and for all $x \in \tilde{B}$.

Proof of Theorem 1. (i) implies (ii). By Proposition 12 and since $\phi \in \Phi_R$ and has strictly increasing shifts, the implication follows.

(ii) implies (i). Let $T : B \to B$ be a robust opinion aggregator. By Lemma 6, there exists an extension from $\mathbb{R}^n$ to $\mathbb{R}^n$. With a small abuse of notation, we denote it by the same symbol T. By Lemma 8, we have that T is nonexpansive and, in particular, continuous and so are the maps T_i. Fix $i \in N$. By [20, Corollary 3], we have that there exists a closed and convex set of probability vectors, $C_i \subseteq \Delta$, and a function $\alpha_i : \mathbb{R}^n \to [0,1]$ such that

$$T_i(x) = \alpha_i(x) \min_{w \in C_i} w \cdot x + [1 - \alpha_i(x)] \max_{w \in C_i} w \cdot x \qquad \forall x \in \mathbb{R}^n. \tag{40}$$

Fix $x \in \mathbb{R}^n$. Since C_i is compact, define $w^{i,1}(x), w^{i,2}(x) \in C_i$ to be such that $w^{i,1}(x) \cdot x = \min_{w \in C_i} w \cdot x$ and $w^{i,2}(x) \cdot x = \max_{w \in C_i} w \cdot x$. By (40) and since C_i is convex, we have that $w^i(x) = \alpha_i(x) w^{i,1}(x) + [1 - \alpha_i(x)] w^{i,2}(x) \in C_i$ and $T_i(x) = w^i(x) \cdot x$. Since x and i were arbitrarily chosen, it follows that for each $i \in N$ and for each $x \in B$ there exists $w \in C_i$ such that $T_i(x) = w \cdot x$. For each $i \in N$ and for each $x \in \mathbb{R}^n$ define the correspondence $\Gamma_i : \mathbb{R}^n \rightrightarrows \Delta$ by

$$\Gamma_i(x) = \{w \in C_i : T_i(x) = w \cdot x\} \qquad \forall x \in \mathbb{R}^n.$$

We first prove the following ancillary claim.

Claim: For each $i \in N$ the correspondence Γ_i is nonempty-, convex-, and compact-valued, upper hemicontinuous, and such that

$$\Gamma_i(x) = \Gamma_i(x + he) \qquad \forall x \in \mathbb{R}^n, h \in \mathbb{R}. \tag{41}$$

Proof of Claim. Fix $i \in N$. Consider $x \in \mathbb{R}^n$. By the previous part of the proof, we have that $w^i(x) \in \Gamma_i(x)$, proving that $\Gamma_i(x)$ is nonempty. Since C_i is convex and compact, it is immediate to check that $\Gamma_i(x)$ is convex and compact. Since x was arbitrarily chosen, it follows that Γ_i is nonempty-, convex- and compact-valued. We next show that Γ_i is upper hemicontinuous. Consider a sequence $\{(x^n, w^n)\}_{n \in \mathbb{N}}$ such that $x^n \to x$ and $w^n \in \Gamma_i(x^n)$ for all $n \in \mathbb{N}$. By definition, we have that

$w^n \in C_i$ as well as $T_i(x^n) = w^n \cdot x^n$ for all $n \in \mathbb{N}$. Since C_i is compact, there exists a subsequence $\{w^{n_k}\}_{k \in \mathbb{N}}$ such that $w^{n_k} \to w \in C_i$. Since T_i is continuous, we have that $T_i(x) = \lim_k T_i(x^{n_k}) = \lim_k w^{n_k} \cdot x^{n_k} = w \cdot x$, proving that $\{w^n\}_{n \in \mathbb{N}}$ has a limit point in $\Gamma_i(x)$. By [3, Theorem 17.20] and since Γ_i is compact-valued, we can conclude that Γ_i is upper hemicontinuous. Finally, consider $x \in \mathbb{R}^n$ and $h \in \mathbb{R}$. Since T is translation invariant and $C_i \subseteq \Delta$, note that

$$
\begin{aligned}
w \in \Gamma_i(x) &\iff w \in C_i \text{ and } T_i(x) = w \cdot x \\
&\iff w \in C_i \text{ and } T_i(x) + h = w \cdot x + w \cdot he \\
&\iff w \in C_i \text{ and } T_i(x + he) = w \cdot (x + he) \iff w \in \Gamma_i(x + he),
\end{aligned}
$$

proving (41). $\qquad\square$

Fix $i \in N$. Define $f : \mathbb{R}^n \times \Delta \to \mathbb{R}_+$ by $f(x, w) = \sum_{j=1}^n w_j x_j^2$ for all $(x, w) \in \mathbb{R}^n \times \Delta$. It is immediate to see that f is continuous in the product topology. Define $\phi_i^T : \mathbb{R}^n \to \mathbb{R}_+$ by

$$
\phi_i^T(x) = \min_{w \in \Gamma_i(x)} f(x, w) = - \max_{w \in \Gamma_i(x)} -f(x, w) \qquad \forall x \in \mathbb{R}^n.
$$

By [3, Lemma 17.30] and since f is continuous and Γ_i is nonempty- and compact-valued as well as upper hemicontinuous, it follows that ϕ_i^T is lower semicontinuous. Next, consider $h \in \mathbb{R} \setminus \{0\}$. It follows that $f(he, w) = h^2$ for all $w \in \Delta$. We can conclude that

$$
\phi_i^T(he) = h^2 > 0 = \phi_i^T(0).
$$

Since i and h were arbitrarily chosen, this implies that $\phi = (\phi_i^T)_{i=1}^n$ is sensitive. Next, we move to the property of strictly increasing shifts. By (41), we have that

$$
\phi_i^T(x + he) = \min_{w \in \Gamma_i(x+he)} \sum_{j=1}^n w_j (x_j + h)^2 = \min_{w \in \Gamma_i(x)} \left[\sum_{j=1}^n w_j x_j^2 + 2h \sum_{j=1}^n w_j x_j + h^2 \right]
\tag{42}
$$

$$
= \min_{w \in \Gamma_i(x)} \sum_{j=1}^n w_j x_j^2 + 2h T_i(x) + h^2 \qquad \forall x \in \mathbb{R}^n, \forall h \in \mathbb{R}.
\tag{43}
$$

Consider $z, v \in \mathbb{R}^n$ and $h \in \mathbb{R}_{++}$. By (42) and (43) and since T is monotone, we can conclude that

$$
z \geq v \implies \phi_i^T(z + he) - \phi_i^T(z) = 2h T_i(z) + h^2 \geq 2h T_i(v) + h^2 = \phi_i^T(v + he) - \phi_i^T(v).
$$

Since i was arbitrarily chosen, it follows that $\phi = (\phi_i^T)_{i=1}^n$ has increasing shifts and, in particular, $\phi \in \Phi_R$. Next, consider $z, v \in \mathbb{R}^n$ such that $z \gg v$. Set $k = \min_{j \in N}(z_j - v_j)$. It follows that $k > 0$ and $z \geq v + ke$. Since T is monotone and translation invariant

and $k > 0$, we can conclude that $T(z) \geq T(v + ke) = T(v) + ke$. Since $z, v \in \mathbb{R}^n$ were arbitrarily chosen, it follows that

$$z \gg v \implies T(z) \gg T(v).$$

By (42) and (43), this implies that if $z, v \in \mathbb{R}^n$ and $h \in \mathbb{R}_{++}$, then

$$z \gg v \implies \phi_i^T(z + he) - \phi_i^T(z) = 2hT_i(z) + h^2 > 2hT_i(v) + h^2 = \phi_i^T(v + he) - \phi_i^T(v).$$

Since i was arbitrarily chosen, it follows that $\phi = (\phi_i^T)_{i=1}^n$ has strictly increasing shifts. We are left to prove (5), that is, for each $i \in N$

$$T_i(x) = \mathrm{argmin}_{c \in \mathbb{R}} \phi_i^T(x - ce) \qquad \forall x \in B.$$

By Proposition 12 and since $\phi = (\phi_i^T)_{i=1}^n \in \Phi_R$ has strictly increasing shifts, we have that $\mathbf{T}_i^\phi(x) = \mathrm{argmin}_{c \in \mathbb{R}} \phi_i^T(x - ce)$ is well defined and single-valued for all $x \in B$ and for all $i \in N$. We are left to prove that it coincides with $T_i(x)$ for all $x \in B$ and for all $i \in N$. Fix $i \in N$ and $x \in B$. By (42) and (43), we have that

$$\phi_i^T(x - ce) = \min_{w \in \Gamma_i(x)} \sum_{j=1}^n w_j x_j^2 - 2cT_i(x) + c^2 \quad \forall c \in \mathbb{R}$$

which, as a function of c, is quadratic and minimized at $c = T_i(x)$, proving the statement. ■

Proof of Lemma 1. Since $\phi \in \Phi_R^*$, we have that ϕ is sensitive and each ϕ_i is continuous. To prove the first part of the statement, we only need to show that ϕ has increasing shifts. By [64, Corollary 4.1] and since ϕ_i is convex and supermodular for all $i \in N$, we have that ϕ_i is ultramodular, that is, for all $z, v \in \mathbb{R}^n$ and $h \in \mathbb{R}_+^n$

$$z \geq v \implies \phi_i(z + h) - \phi_i(z) \geq \phi_i(v + h) - \phi_i(v),$$

yielding that, in particular, ϕ has increasing shifts. As for the second part of the statement, assume that ϕ_i is strictly convex for all $i \in N$. From the previous part of the proof, we have that ϕ has increasing shifts. Fix $i \in N$ and $z, v \in \mathbb{R}^n$. Consider the map $g_{i,v} : \mathbb{R} \to \mathbb{R}_+$ defined by $g_{i,v}(c) = \phi_i(v + ce)$ for all $c \in \mathbb{R}$. Clearly, $g_{i,v}$ is strictly convex. Consider $c_1, c_2 \in \mathbb{R}$ and $h > 0$ such that $c_1 > c_2$. Define $d = c_1 + h$. Since $d = c_1 + h > c_2 + h > c_2$, note that there exists $\alpha \in (0, 1)$ such that $c_2 + h = \alpha c_2 + (1 - \alpha) d$. Straightforward computations yield also that $c_1 = (1 - \alpha) c_2 + \alpha d$. Since $g_{i,v}$ is strictly convex, $\alpha \in (0, 1)$, and $c_2 \neq d$, it follows that

$$g_{i,v}(c_2 + h) < \alpha g_{i,v}(c_2) + (1 - \alpha) g_{i,v}(d)$$

as well as

$$g_{i,v}(c_1) < (1 - \alpha) g_{i,v}(c_2) + \alpha g_{i,v}(d)$$

Adding the two inequalities, we obtain that $g_{i,v}(c_2 + h) + g_{i,v}(c_1) < g_{i,v}(c_2) + g_{i,v}(d)$, that is,

$$g_{i,v}(c_2 + h) - g_{i,v}(c_2) < g_{i,v}(c_1 + h) - g_{i,v}(c_1). \tag{44}$$

Finally, consider $h > 0$ and assume that $z \gg v$. It follows that there exists $\delta > 0$ such that $z \geq v + \delta e$. By (44) and since ϕ has increasing shifts, we can conclude that

$$\phi_i(z + he) - \phi_i(z) \geq \phi_i((v + \delta e) + he) - \phi_i(v + \delta e) = g_{i,v}(\delta + h) - g_{i,v}(\delta)$$
$$> g_{i,v}(0 + h) - g_{i,v}(0) = \phi_i(v + he) - \phi_i(v).$$

Since i was arbitrarily chosen, it follows that ϕ has strictly increasing shifts. ∎

In order to prove Proposition 1, we state and prove an ancillary result which allows us to obtain a similar result for loss functions which admit more than a minimizer.

Lemma 7 *Let $\phi \in \Phi_R$. The following facts are true:*

1. *If ϕ is positively homogeneous, then the correspondence $\mathbf{T}^\phi$ admits a selection T^ϕ which is a constant affine robust opinion aggregator.*

2. *If ϕ is symmetric, then the correspondence $\mathbf{T}^\phi$ admits a selection T^ϕ which is an odd robust opinion aggregator.*

Proof. Before starting, we make four observations adopting the same notation of the proof of Proposition 12. Since $\mathbf{T}_i^\phi : B \rightrightarrows I$ is nonempty-, convex-, and compact-valued for all $i \in N$, we have that

$$\mathbf{T}_i^\phi(x) = \left[\min \mathbf{T}_i^\phi(x), \max \mathbf{T}_i^\phi(x)\right] \quad \forall x \in B, \forall i \in N. \tag{45}$$

By Proposition 12 and its proof, we know that $T^\phi : B \to B$, defined by $T_i^\phi(x) = \min \mathbf{T}_i^\phi(x)$ for all $x \in B$ and for all $i \in N$, is a robust opinion aggregator which is a selection of $\mathbf{T}^\phi$. A quick inspection of the proof yields that $\tilde{T}^\phi : B \to B$, defined by $\tilde{T}_i^\phi(x) = \max \mathbf{T}_i^\phi(x)$ for all $x \in B$ and for all $i \in N$, would satisfy the same properties. Finally, since $\mathbf{T}_i^\phi$ is convex-valued for all $i \in N$, $\hat{T}^\phi : B \to B$, defined by $\hat{T}_i^\phi = \frac{1}{2}T_i^\phi + \frac{1}{2}\tilde{T}_i^\phi$ for all $x \in B$ and for all $i \in N$, would satisfy the same properties.

1. Since $\phi \in \Phi_R$, observe that ϕ is sensitive. Since ϕ is positively homogeneous, this implies that $0 < \phi_i(\lambda e) = \eta_i(\lambda)\phi_i(e)$ for all $\lambda \in \mathbb{R}_{++}$ and for all $i \in N$, that is $\eta_i(\lambda) > 0$ for all $\lambda \in \mathbb{R}_{++}$ and for all $i \in N$. Fix $i \in N$. Consider $x \in B$, $k \in I$, and $\lambda \in (0, 1)$. Let $c^\star \in \mathbb{R}$. Since ϕ is positively homogeneous, note that

$$\phi_i(x - c^\star e) \leq \phi_i(x - ce) \quad \forall c \in \mathbb{R} \iff \eta_i(\lambda)\phi_i(x - c^\star e) \leq \eta_i(\lambda)\phi_i(x - ce) \quad \forall c \in \mathbb{R}$$
$$\iff \phi_i(\lambda x - \lambda c^\star e) \leq \phi_i(\lambda x - \lambda ce) \quad \forall c \in \mathbb{R}$$
$$\iff \phi_i(\lambda x + (1 - \lambda)ke - (\lambda c^\star + (1 - \lambda)k)e)$$
$$\leq \phi_i(\lambda x + (1 - \lambda)ke - (\lambda c + (1 - \lambda)k)e) \quad \forall c \in \mathbb{R},$$

proving that $c^\star \in \mathbf{T}_i^\phi(x)$ if and only if $\lambda c^\star + (1 - \lambda) k \in \mathbf{T}_i^\phi(\lambda x + (1 - \lambda) ke)$. By definition of T_i^ϕ and since $T_i^\phi(x) \in \mathbf{T}_i^\phi(x)$, this implies that

$$T_i^\phi(\lambda x + (1 - \lambda) ke) \leq \lambda T_i^\phi(x) + (1 - \lambda) k.$$

By definition of T_i^ϕ and since $T_i^\phi(\lambda x + (1 - \lambda) ke) \in \mathbf{T}_i^\phi(\lambda x + (1 - \lambda) ke)$, we also have that

$$\frac{T_i^\phi(\lambda x + (1 - \lambda) ke) - (1 - \lambda) k}{\lambda} \in \mathbf{T}_i^\phi(x) \text{ and } T_i^\phi(x) \leq \frac{T_i^\phi(\lambda x + (1 - \lambda) ke) - (1 - \lambda) k}{\lambda},$$

proving that $\lambda T_i^\phi(x) + (1 - \lambda) k \leq T_i^\phi(\lambda x + (1 - \lambda) ke)$. Since i, x, k, and λ were arbitrarily chosen, we can conclude that

$$T^\phi(\lambda x + (1 - \lambda) ke) = \lambda T^\phi(x) + (1 - \lambda) ke \quad \forall x \in B, \forall k \in I, \forall \lambda \in (0, 1),$$

proving constant affinity.

2. Fix $i \in N$. Consider $x \in B$ such that $-x \in B$. We next show that $\hat{T}_i^\phi(-x) = -\hat{T}_i^\phi(x)$. Let $c^\star \in \mathbb{R}$. Since ϕ is symmetric, note that

$$\phi_i(x - c^\star e) \leq \phi_i(x - ce) \quad \forall c \in \mathbb{R} \iff \phi_i(-x + c^\star e) \leq \phi_i(-x + ce) \quad \forall c \in \mathbb{R}$$
$$\iff \phi_i(-x - (-c^\star) e) \leq \phi_i(-x - de) \quad \forall d \in \mathbb{R},$$

proving that $c^\star \in \mathbf{T}_i^\phi(x)$ if and only if $-c^\star \in \mathbf{T}_i^\phi(-x)$. By (45), this implies that

$$-\min \mathbf{T}_i^\phi(x) = \max \mathbf{T}_i^\phi(-x) \text{ and } -\max \mathbf{T}_i^\phi(x) = \min \mathbf{T}_i^\phi(-x).$$

By definition of $\hat{T}_i^\phi$, it follows that

$$\hat{T}_i^\phi(x) = \frac{1}{2}\left(-\max \mathbf{T}_i^\phi(-x) + -\min \mathbf{T}_i^\phi(-x)\right)$$
$$= -\frac{1}{2}\left(\max \mathbf{T}_i^\phi(-x) + \min \mathbf{T}_i^\phi(-x)\right) = -\hat{T}_i^\phi(-x).$$

Since i and x were arbitrarily chosen, we can conclude that $\hat{T}^\phi$ is odd. ∎

Proof of Proposition 1. 1. "If". By Proposition 12 and since $\phi \in \Phi_R$ has strictly increasing shifts, then $\mathbf{T}^\phi = T^\phi$ is single-valued. By point 1 of Lemma 7 and since ϕ is also positively homogeneous, we can conclude that $T = T^\phi$ is a constant affine robust opinion aggregator. "Only if". Consider the profile of loss functions $\phi \in \Phi_R$ as in the proof of (ii) implies (i) of Theorem 1. The profile $\phi = (\phi_i^T)_{i=1}^n$ has strictly increasing shifts and is such that $T^\phi = T$ as well as

$$\phi_i^T(z) = \min_{w \in \Gamma_i(z)} \sum_{j=1}^n w_j z_j^2 \quad \forall z \in \mathbb{R}^n, \forall i \in N. \tag{46}$$

The set $\Gamma_i(z)$ is the set $\{w \in C_i : T_i(z) = w \cdot z\}$ where, with an abuse of notation, T denoted an extension of T from $\mathbb{R}^n$ to $\mathbb{R}^n$. By point 2 of Lemma 6 and since T is constant affine, it follows that such an extension can be assumed to be constant affine as well. This implies that $T : \mathbb{R}^n \to \mathbb{R}^n$ is such that $T(\lambda z) = \lambda T(z)$ for all $\lambda \in \mathbb{R}_+$ and for all $z \in \mathbb{R}^n$. Fix $i \in N$. This yields that $\Gamma_i(\lambda z) = \Gamma_i(z)$ for all $\lambda \in \mathbb{R}_{++}$ and for all $z \in \mathbb{R}^n$. By (46), we have that

$$\phi_i^T(\lambda z) = \min_{w \in \Gamma_i(\lambda z)} \sum_{j=1}^n w_j (\lambda z_j)^2 = \lambda^2 \min_{w \in \Gamma_i(z)} \sum_{j=1}^n w_j z_j^2 = \lambda^2 \phi_i^T(z) \quad \forall z \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}_{++}.$$

Since i was arbitrarily chosen and $\phi_i^T(0) = 0$ for all $i \in N$, it follows that ϕ is positively homogeneous where $\eta_i(\lambda) = \lambda^2$ for all $\lambda \in \mathbb{R}_+$.

2. By Proposition 12 and since $\phi \in \Phi_R$ has strictly increasing shifts, then $\mathbf{T}^\phi = T^\phi$ is single-valued. By point 2 of Lemma 7 and since ϕ is also symmetric, we can conclude that $T = T^\phi$ is an odd robust opinion aggregator. ∎

Proof of Lemma 2. Fix $i \in N$. It is well known that order statistics are functionals from $\mathbb{R}^n$ to $\mathbb{R}$ which are normalized, monotone, and comonotonic additive (cf. the definitions at the beginning of Appendix B). This implies that order statistics are also translation invariant. Since convex linear combinations maintain these properties and i was arbitrarily chosen, the statement follows. ∎

Proof of Proposition 2. In this proof, given $W \in \mathcal{W}$ and a profile of lower semi-continuous functions $\rho = (\rho_i : \mathbb{R} \to \mathbb{R}_+)_{i=1}^n$, for each $i \in N$ we denote by $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ the function defined by

$$\phi_i(z) = \sum_{j=1}^n w_{ij} \rho_i(z_j) \qquad \forall z \in \mathbb{R}^n.$$

We denote by ϕ the profile $\phi = (\phi_i)_{i=1}^n$.

(i) implies (iii). Since $\phi = (W, \rho) \in \Phi_A$, we have that $\phi \in \Phi_R$. Fix $i \in N$. We begin by showing that ρ_i is convex. Consider $g_{i,0}$ as in the proof of Lemma 5. Observe that for each $c \in \mathbb{R}$

$$g_{i,0}(c) = \phi_i(0 + ce) = \sum_{j=1}^n w_{ij} \rho_i(c) = \rho_i(c) \qquad \forall c \in \mathbb{R}.$$

By the proof of Lemma 5 and since $\phi \in \Phi_R$, we have that $g_{i,0} = \rho_i$ is convex. Since ϕ is sensitive, it follows that ρ_i admits a unique minimizer at 0. This implies that ρ_i is strictly decreasing on $\mathbb{R}_-$ and strictly increasing on $\mathbb{R}_+$.[52]

---

[52] Since 0 is a minimizer, $0 \in \partial \rho_i(0)$ and, in particular, the left- and right-derivatives at 0 are such that $\rho'_{i,-}(0) \le 0 \le \rho'_{i,+}(0)$. Since ρ_i is convex, we have that

$$\rho'_{i,+}(c) \le \rho'_{i,-}(0) \le 0 \le \rho'_{i,+}(0) \le \rho'_{i,-}(d) \quad \forall c, d \in \mathbb{R} \text{ s.t. } c < 0 < d.$$

(iii) implies (ii). Fix $i \in N$. Since $\rho_i : \mathbb{R} \to \mathbb{R}_+$ is convex, it follows that ρ_i is continuous and convex. This implies that $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ is continuous and convex. Next, consider $z, v \in \mathbb{R}^n$. It follows that

$$\rho_i (z_j \wedge v_j) + \rho_i (z_j \vee v_j) = \rho_i (v_j) + \rho_i (z_j) \quad \forall j \in N.$$

We can conclude that

$$\phi_i (z \wedge v) + \phi_i (z \vee v) = \sum_{j=1}^{n} w_{ij} \rho_i (z_j \wedge v_j) + \sum_{j=1}^{n} w_{ij} \rho_i (z_j \vee v_j)$$
$$= \sum_{j=1}^{n} w_{ij} \rho_i (v_j) + \sum_{j=1}^{n} w_{ij} \rho_i (z_j) = \phi_i (v) + \phi_i (z),$$

proving that ϕ_i is modular and, in particular, supermodular. Finally, consider $h \in \mathbb{R} \setminus \{0\}$. Since ρ_i is strictly decreasing on $\mathbb{R}_-$ and strictly increasing on $\mathbb{R}_+$, we have that

$$\phi_i (he) = \sum_{j=1}^{n} w_{ij} \rho_i (h) = \rho_i (h) > \rho_i (0) = \sum_{j=1}^{n} w_{ij} \rho_i (0) = \phi_i (0).$$

Since i was arbitrarily chosen, it follows that $\phi = (\phi_i)_{i=1}^{n} \in \Phi_R^*$. By Lemma 1, we have that ϕ is robust, that is, $\phi \in \Phi_R$. By construction, ϕ is additively separable, it follows that $\phi \in \Phi_A$, proving the statement.

(ii) implies (i). It is trivial. ∎

Proof of Proposition 3. We omit the proof of point 2 which follows from well known facts.[53] Recall that for each $i \in N$

$$\phi_i^{\lambda} (z) = \sum_{j=1}^{n} w_{ij} \rho_i (z_j) \qquad \forall z \in \mathbb{R}^n$$

where $\lambda \in \mathbb{R} \setminus \{0\}$ and $\rho_i : \mathbb{R} \to \mathbb{R}_+$ is defined by $\rho_i (s) = e^{\lambda s} - \lambda s$ for all $s \in \mathbb{R}$. It is easy to see that ρ_i satisfies the properties of point (iii) of Proposition 2 for all $i \in N$. This implies that $\phi \in \Phi_A \subseteq \Phi_R$. Since $\rho_i'' > 0$ for all $i \in N$, ρ_i is strictly convex for all $i \in N$. By the same techniques of the second part of the proof of Lemma 1, this yields that ϕ has strictly increasing shifts. By Proposition 12, it follows that $\mathbf{T}^{\phi} = T^{\phi} = T^{\lambda}$ is single-valued and is a robust opinion aggregator. We are only left to compute it.

Since 0 is the unique minimizer and ρ_i is convex, $\rho_{i,+}' (c) < 0 < \rho_{i,-}' (d)$ for all $c, d \in \mathbb{R}$ such that $c < 0 < d$. By the Mean Value Theorem for convex functions, the desired monotonicity properties follow.

[53] The result for $\hat{\lambda} = \infty$ is also known as Laplace's method (see, e.g., [21, Theorem 4.1]). The case for $\hat{\lambda} = -\infty$ is instead obtained from the previous one and by observing that $\lambda x_j = -\lambda (-x_j)$ and $\lambda \to -\infty$ yields $-\lambda \to \infty$. The case of $\hat{\lambda} = 0$ is a standard result in risk theory.

1. Fix $i \in N$. Consider $x \in B$. Since the function $c \mapsto \phi_i^\lambda (x - ce)$ is strictly convex and differentiable. We compute the first order conditions where $c^\star$ is the optimal value.

$$-\sum_{j=1}^{n} w_{ij} \left[\lambda \exp \left(\lambda \left(x_j - c^\star \right) \right) - \lambda \right] = 0 \implies \sum_{j=1}^{n} w_{ij} \left[\exp \left(\lambda \left(x_j - c^\star \right) \right) - 1 \right] = 0 \implies$$

$$\exp \left(-\lambda c^\star \right) \sum_{j=1}^{n} w_{ij} \exp \left(\lambda x_j \right) = 1 \implies \sum_{j=1}^{n} w_{ij} \exp \left(\lambda x_j \right) = \exp \left(\lambda c^\star \right)$$

$$\implies c^\star = \frac{1}{\lambda} \ln \left(\sum_{j=1}^{n} w_{ij} \exp \left(\lambda x_j \right) \right),$$

proving the statement.

3. In order to prove this point, we need to introduce two continuous operators which will be useful. The first one is $S : \mathbb{R}^n \to \mathbb{R}_{++}^n$ defined by $S_i (x) = \exp (\lambda x_i)$ for all $i \in N$ and for all $x \in \mathbb{R}^n$. The second one is $Z : \mathbb{R}_{++}^n \to \mathbb{R}^n$ defined by $Z_i (x) = \frac{1}{\lambda} \ln (x_i)$ for all $i \in N$ and for all $x \in \mathbb{R}_{++}^n$. Clearly, we have that Z is the inverse of S. Define $\hat{T} : \mathbb{R}^n \to \mathbb{R}^n$ by $\hat{T} (x) = W x$ for all $x \in \mathbb{R}^n$. We next show that

$$\left(T^\lambda \right)^t = S^{-1} \hat{T}^t S \qquad \forall t \in \mathbb{N}. \tag{47}$$

By definition of T^λ, if $t = 1$, then $T^\lambda (x) = S^{-1} (W S (x))$ for all $x \in \mathbb{R}^n$, yielding (47). Next, assume that (47) holds for t. We have that

$$\left(T^\lambda \right)^{t+1} = T^\lambda \left(T^\lambda \right)^t = S^{-1} \hat{T} S S^{-1} \hat{T}^t S = S^{-1} \hat{T}^{t+1} S,$$

proving that (47) holds for $t+1$. By induction, (47) follows. Consider $x \in B$. By (12), it follows that

$$\lim_t \hat{T}^t (S (x)) = \lim_t W^t S (x) = \left(\sum_{i=1}^{n} s_i \exp (\lambda x_i) \right) e \in \mathbb{R}_{++}^n.$$

By (47) and since S^{-1} is continuous, we have that

$$\lim_t \left(T^\lambda \right)^t (x) = \left(\frac{1}{\lambda} \ln \left(\sum_{i=1}^{n} s_i \exp (\lambda x_i) \right) \right) e = \bar{T}^\lambda (x).$$

Since x was arbitrarily chosen, subpoint a follows. Subpoint b follows from standard computations. ∎

Proof of Proposition 5. Before starting, we make few observations about strong convexity (see, e.g., [75, p. 268]). Since each ρ_i is strongly convex and twice continuously differentiable, we have that for each $i \in N$ there exists $\alpha_i > 0$ such that $\rho_i'' (s) \geq \alpha_i$ for all $s \in \mathbb{R}$. Moreover, we have that for each $i \in N$

$$\left(\rho_i' (s_1) - \rho_i' (s_2) \right) (s_1 - s_2) \geq \alpha_i (s_1 - s_2)^2 \quad \forall s_1, s_2 \in \mathbb{R}. \tag{48}$$

Finally, since each ρ_i is twice continuously differentiable and I is compact, for each $i \in N$ we have that there exists $L_i > 0$ such that

$$|\rho_i'(s_1) - \rho_i'(s_2)| \le L_i |s_1 - s_2| \quad \forall s_1, s_2 \in [\min I - \max I, \max I - \min I]. \quad (49)$$

Recall that $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ is defined by $\phi_i(z) = \sum_{j=1}^n w_{ij}\rho_i(z_j)$ for all $z \in \mathbb{R}^n$ and for all $i \in N$. By assumption, $\phi \in \Phi_A \subseteq \Phi_R$. Since $\rho_i'' \ge \alpha_i > 0$, this implies that ρ_i is strictly convex for all $i \in N$. By the same techniques of the second part of the proof of Lemma 1, this yields that ϕ has strictly increasing shifts. By Proposition 12, we have that $\mathbf{T}^\phi = T^\phi$ is single-valued and a robust opinion aggregator from B to B. Moreover, $T_i^\phi(x)$ is the unique solution of

$$\min_{c \in \mathbb{R}} \phi_i(x - ce) = \min_{c \in I} \phi_i(x - ce) \quad \forall i \in N, \forall x \in B. \quad (50)$$

Fix $i \in N$. Since ρ_i is differentiable and convex, so is the map $c \mapsto \phi_i(x - ce)$ for all $x \in B$. The solution of (50) is then given by the first order condition

$$\sum_{j=1}^n w_{ij}\rho_i'\left(x_j - T_i^\phi(x)\right) = 0 \quad \forall x \in B.$$

Consider $x \in B$, $h > 0$, and $l \in N$ such that $x + he^l \in B$. We have that

$$\sum_{j=1}^n w_{ij}\rho_i'\left(x_j - T_i^\phi(x)\right) = 0 \text{ and } \sum_{j=1}^n w_{ij}\rho_i'\left(x_j + he_j^l - T_i^\phi(x + he^l)\right) = 0. \quad (51)$$

Note that if $w_{il} = 0$, then $\sum_{j=1}^n w_{ij}\rho_i'\left(x_j + he_j^l - c\right) = \sum_{j=1}^n w_{ij}\rho_i'(x_j - c)$ for all $c \in \mathbb{R}$, proving that $T_i^\phi(x + he^l) = T_i^\phi(x)$ and that l does not strongly influence i. Next, assume that $w_{il} > 0$. By (49), (51), and (48) and since T^ϕ is monotone and $h > 0$, we can conclude that

$$
\begin{aligned}
& L_i \left(T_i^\phi(x + he^l) - T_i^\phi(x)\right) \\
& \ge \sum_{j=1}^n w_{ij}\rho_i'\left(x_j + he_j^l - T_i^\phi(x)\right) - \sum_{j=1}^n w_{ij}\rho_i'\left(x_j + he_j^l - T_i^\phi(x + he^l)\right) \\
& = \sum_{j=1}^n w_{ij}\rho_i'\left(x_j + he_j^l - T_i^\phi(x)\right) - \sum_{j=1}^n w_{ij}\rho_i'\left(x_j - T_i^\phi(x)\right) \\
& = w_{il}\left[\rho_i'\left(x_l + h - T_i^\phi(x)\right) - \rho_i'\left(x_l - T_i^\phi(x)\right)\right] \\
& \ge w_{il}\alpha_i h,
\end{aligned}
$$

proving that $T_i^\phi(x + he^l) - T_i^\phi(x) \ge \varepsilon_{il}h$ where $\varepsilon_{il} = L_i^{-1}w_{il}\alpha_i$. Since x and h were arbitrarily chosen, we have that l strongly influences i. ∎

B Appendix: Convergence

Before starting, we introduce some notation which will prove useful later on. Given $\varepsilon \in (0,1]$, we denote by $\mathcal{W}_{\varepsilon,k}$ the subset of stochastic matrices $W \in \mathcal{W}$ such that the k-th column has all the entries greater than or equal to ε, that is, $w_{ik} \geq \varepsilon$ for all $i \in N$. Let $\mathcal{W}_{\varepsilon} = \cup_{k \in N} \mathcal{W}_{\varepsilon,k}$.[54] Given a functional, $f : B \to \mathbb{R}$, creating a small abuse of terminology with the definitions of Section 2, we say that f is:

1. *normalized* if and only if $f(ke) = k$ for all $k \in I$;

2. *monotone* if and only if $x \geq y$ implies $f(x) \geq f(y)$;

3. *translation invariant* if and only if $f(x + ke) = f(x) + k$ for all $x \in B$ and for all $k \in \mathbb{R}$ such that $x + ke \in B$.

Lemma 8 *If T is a robust opinion aggregator, then T^t is nonexpansive (i.e., Lipschitz continuous of order 1) for all $t \in \mathbb{N}$. In particular, T is nonexpansive.*

Proof. Since T is a robust opinion aggregator, T_i is normalized, monotone, and translation invariant for all $i \in N$. By [22, Theorem 4], it follows that T_i is a niveloid for all $i \in N$. By [22, p. 346], it follows that $|T_i(x) - T_i(y)| \leq \|x - y\|_\infty$ for all $x, y \in B$ and for all $i \in N$. This implies that

$$\|T(x) - T(y)\|_\infty = \max_{i \in N} |T_i(x) - T_i(y)| \leq \|x - y\|_\infty \qquad \forall x, y \in B,$$

proving that T is nonexpansive.

By induction, we next show that T^t is nonexpansive for all $t \in \mathbb{N}$. Since we have shown that T is nonexpansive, T^t is nonexpansive for $t = 1$, proving the initial step. By the induction hypothesis, assume that T^t is nonexpansive, we have that for each $x, y \in B$

$$\left\|T^{t+1}(x) - T^{t+1}(y)\right\|_\infty = \left\|T\left(T^t(x)\right) - T\left(T^t(y)\right)\right\|_\infty \leq \left\|T^t(x) - T^t(y)\right\|_\infty \leq \|x - y\|,$$

proving the inductive step. The statement follows by induction. ∎

Given an opinion aggregator T and $x \in B$, in what follows we first study the limit of the time averages of $\{T^t(x)\}_{t \in \mathbb{N}}$, that is, the Cesaro limit of $\{T^t(x)\}_{t \in \mathbb{N}}$. The next ancillary lemma (proved in the Online Appendix) highlights the properties of the limiting operator $\bar{T}$, whenever it exists. Theorem 5 below provides a sufficient condition for existence: nonexpansivity.

---

[54] The matrices in $\bigcup_{\varepsilon \in (0,1]} \mathcal{W}_{\varepsilon}$ are also said to be Markov's matrices or that they satisfy "Doeblin's condition" (see, respectively, Seneta [80, Definition 4.7] and Stroock [81, p. 32]).

Lemma 9 *Let T be an opinion aggregator. If T is such that*

$$C - \lim_t T^t (x) \ \ exists \qquad \forall x \in B,$$

then $\bar{T} : B \to B$, defined by $\bar{T} (x) = C - \lim_t T^t (x)$ for all $x \in B$, is well defined and $\bar{T} \circ T = \bar{T}$. Moreover,

1. *If T is nonexpansive, so is $\bar{T}$. In particular, $\bar{T}$ is continuous.*

2. *If T is normalized and monotone, so is $\bar{T}$.*

3. *If T is robust, so is $\bar{T}$.*

4. *If T is constant affine, so is $\bar{T}$.*

5. *If T is odd, so is $\bar{T}$, provided I is a symmetric interval, that is, $k \in I$ if and only if $-k \in I$.*

The next result proves that the sequences of updates of a nonexpansive opinion aggregator converge a la Cesaro, provided B is compact. This in turn will yield the same result for robust opinion aggregators (proof of Theorem 2 below) whether B is bounded or not.

Theorem 5 *Let T be an opinion aggregator. If B is compact and T is nonexpansive, then*

$$C - \lim_t T^t (x) \ \ exists \qquad \forall x \in B. \tag{52}$$

Moreover, if $\bar{T} : B \to B$ is defined by

$$\bar{T} (x) = C - \lim_t T^t (x) \qquad \forall x \in B,$$

then $\bar{T}$ is nonexpansive and such that $\bar{T} \circ T = \bar{T}$ as well as

$$\lim_\tau \left(\sup_{x \in B} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t (x) - \bar{T} (x) \right\|_\infty \right) = 0. \tag{53}$$

Proof. By the same inductive argument contained in the proof of Lemma 8, we have that for each $t \in \mathbb{N}$ the maps $T^t : B \to B$ are nonexpansive. Consider the space of continuous functions over B: $C(B)$. We endow this space with the supnorm. With a small abuse of notation, we will denote by $\| \ \|_\infty$ also the supnorm of $C(B)$ where $\|f\|_\infty = \sup_{x \in B} |f(x)|$ for all $f \in C(B)$. Define $S : C(B) \to C(B)$ by

$$S(f) = f \circ T \qquad \forall f \in C(B).$$

Note that S is a positive linear selfmap on $C(B)$. Moreover, $S^t(f) = f \circ T^t$ for all $f \in C(B)$ and for all $t \in \mathbb{N}$. Fix $f \in C(B)$. Since B is compact and f is continuous, it follows that f is uniformly continuous (see, e.g., [3, Corollary 3.31]), that is, for each $\varepsilon > 0$ there exists $\delta > 0$ such that

$$x, y \in B \text{ and } \|x - y\|_\infty < \delta \implies |f(x) - f(y)| < \varepsilon.$$

Since T^t is nonexpansive for all $t \in \mathbb{N}$, this implies that for each $t \in \mathbb{N}$ and $x, y \in B$

$$\|x - y\|_\infty < \delta \implies \left\|T^t(x) - T^t(y)\right\|_\infty < \delta \implies \left|f\left(T^t(x)\right) - f\left(T^t(y)\right)\right| < \varepsilon.$$

We have that $\{S^t(f)\}_{t \in \mathbb{N}}$ is a sequence of equicontinuous functions. Moreover, $|S^t(f)(y)| = |f(T^t(y))| \leq \|f\|_\infty$ for all $t \in \mathbb{N}$ and for all $y \in B$. It follows that $\|S^t(f)\|_\infty \leq \|f\|_\infty$ for all $t \in \mathbb{N}$, that is $\{S^t(f)\}_{t \in \mathbb{N}}$ is bounded. By setting $t = 1$ and since f was arbitrarily chosen, it also follows that S is a bounded operator. We can conclude that S is a positive equicontinuous operator as defined in [77]. For each $\tau \in \mathbb{N}$ also define the operator $S_\tau : C(B) \to C(B)$ by

$$S_\tau = \frac{1}{\tau} \sum_{t=1}^{\tau} S^t.$$

By the Ergodic Theorem in Rosenblatt [77, Theorem 1 p. 134], it follows that $S_\tau(f) \overset{\|\ \|_\infty}{\to} \bar{S}(f)$ for all $f \in C(B)$ where $\bar{S} : C(B) \to C(B)$. It is immediate to see that $\bar{S}$ is linear and bounded as well (see, e.g., [3, Corollary 6.18]).

Next, for each $i \in N$ define $f_i : B \to \mathbb{R}$ by $f_i(x) = x_i$ for all $x \in B$. Note that f_i is affine and $f_i \in C(B)$ for all $i \in N$. By the previous part of the proof, we have that $S_\tau(f_i) \overset{\|\ \|_\infty}{\to} \bar{S}(f_i)$ for all $i \in N$.

Define $\bar{T} : B \to B$ by $\bar{T}_i(x) = \bar{S}(f_i)(x)$ for all $i \in N$ and for all $x \in B$. Note that $\bar{T}$ is continuous.[55] By definition of $\bar{S}$ and $\bar{T}$, we have that for each $i \in N$ for each $\varepsilon > 0$ there exists $\tau_i(\varepsilon) \in \mathbb{N}$ such that $\tau \geq \tau_i(\varepsilon)$ yields that

$$
\begin{aligned}
\sup_{x \in B} \left| f_i\left(\frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) \right) - \bar{T}_i(x) \right| &= \sup_{x \in B} \left| \frac{1}{\tau} \sum_{t=1}^{\tau} f_i\left(T^t(x)\right) - \bar{S}(f_i)(x) \right| \\
&= \sup_{x \in B} \left| \frac{1}{\tau} \sum_{t=1}^{\tau} S^t(f_i)(x) - \bar{S}(f_i)(x) \right| \\
&= \sup_{x \in B} \left| S_\tau(f_i)(x) - \bar{S}(f_i)(x) \right| \\
&= \left\| S_\tau(f_i) - \bar{S}(f_i) \right\|_\infty < \varepsilon.
\end{aligned}
$$

---

[55]In what follows, inter alia, we will show that $\bar{T}$ coincides with the operator $\bar{T}$ defined in (14), justifying the choice of notation.

For each $\varepsilon > 0$ define $\tau(\varepsilon) = \max_{i \in N} \tau_i(\varepsilon)$. In particular, we have that for each $\varepsilon > 0$ and for each $\tau \geq \tau(\varepsilon)$

$$\varepsilon > \sup_{i \in N} \sup_{x \in B} \left| f_i \left(\frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) \right) - \bar{T}_i(x) \right| = \sup_{x \in B} \sup_{i \in N} \left| f_i \left(\frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) \right) - \bar{T}_i(x) \right|$$

$$= \sup_{x \in B} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_{\infty},$$

proving that $\bar{T}(x) = C - \lim_t T^t(x)$ for all $x \in B$, that is, (52) as well as (53) holds. By Lemma 9, $\bar{T}$ is well defined, nonexpansive, and such that $\bar{T} \circ T = \bar{T}$. ∎

Remark 5 Theorem 5 could be seen as a version of the classic nonlinear ergodic theorem of Baillon (see Aubin and Ekeland [5, p. 253] as well as Krengel [57, Section 9.3]). In this literature, the assumption of finite dimensionality does not seem to play a major role, while the properties of the norm do (e.g., differentiability, strict convexity, etc). In fact, on the one hand, our selfmap is nonexpansive when B is endowed with the $\| \ \|_{\infty}$ norm. On the other hand, in the original version of Baillon's result, T must be nonexpansive with respect to the Euclidean norm $\| \ \|_2$.[56] This is not a mere technical choice, but rather a fundamental one driven by our opinion aggregators and their properties. For example, when T is as in Example 4, T is *not* nonexpansive with respect to $\| \ \|_2$ while it is so for $\| \ \|_{\infty}$. At the same time, generalizations of Baillon's Theorem allow for more general norms (e.g., $\| \ \|_p$ with $p \in (1, \infty)$), but to the best of our knowledge the only one that encompasses the case $\| \ \|_{\infty}$ is the one contained in Baillon, Bruck, and Reich [7, Theorem 3.2 and Corollary 3.1]. Compared to our version, to the best of our knowledge, the part that would be missing is the one contained in (53). Observe that (53), not only guarantees uniform Cesaro convergence of $\{T^t(x)\}_{t \in \mathbb{N}}$ (present in [7] too), but also the independence from the initial condition of the rate of such convergence (cf. also point (iii) of Proposition 4). This latter property might play an important role in applications and is missing in the aforementioned works. ▲

Proof of Theorem 2. Consider $x \in B$. By point 4 of Lemma 6, it follows that there exists a compact subinterval $\tilde{I} \subseteq I$ with nonempty interior such that $x \in \tilde{I}^n \subseteq B$. Define $\tilde{B} = \tilde{I}^n$. Consider the restriction $\tilde{T} = T_{|\tilde{B}}$. By point 4 of Lemma 6, $\tilde{T} : \tilde{B} \to \tilde{B}$ is a robust opinion aggregator and, in particular, nonexpansive. By Theorem 5 and since $\tilde{B}$ is compact, we have that

$$\lim_{\tau} \frac{1}{\tau} \sum_{t=1}^{\tau} \tilde{T}^t(x) \ \text{ exists.}$$

---

[56]Recall that $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$.

Since $\tilde{T}^t(x) = T^t(x)$ for all $t \in \mathbb{N}$ and x was arbitrarily chosen, (13) follows. By Lemma 9, if we define $\bar{T}$ as in (14), $\bar{T}$ is a well defined robust opinion aggregator such that $\bar{T} \circ T = \bar{T}$. Finally, let $\hat{B}$ be a bounded subset of B. Note that there exists a compact interval with nonempty interior $\tilde{I} \subseteq I$ such that $\hat{B} \subseteq \tilde{B}$ where $\tilde{B} = \tilde{I}^n$. Define $\tilde{T}$ as before. By (53) applied to $\tilde{T}$ and since $\bar{\tilde{T}}(x) = \bar{T}(x)$ and $\tilde{T}^t(x) = T^t(x)$ for all $t \in \mathbb{N}$ and for all $x \in \tilde{B}$, we have that

$$\sup_{x \in \hat{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_\infty \leq \sup_{x \in \tilde{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_\infty$$

$$= \sup_{x \in \tilde{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} \tilde{T}^t(x) - \bar{\tilde{T}}(x) \right\|_\infty \to 0,$$

proving (15). $\blacksquare$

Up to now, we proved that robust opinion aggregators generate dynamics which Cesaro convergence. The next result is going to prove useful in transforming the convergence of the updates' time averages into standard convergence.[57]

Theorem 6 (Lorentz) *Let $\{x^t\}_{t \in \mathbb{N}} \subseteq \mathbb{R}^n$ be a bounded sequence. The following statements are equivalent:*

(i) *There exists $\bar{x} \in \mathbb{R}^n$ such that*

$$\forall \varepsilon > 0 \ \exists \bar{\tau} \in \mathbb{N} \ \forall m \in \mathbb{N} \ s.t. \ \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} - \bar{x} \right\|_\infty < \varepsilon \qquad \forall \tau \geq \bar{\tau} \qquad (54)$$

and $\lim_t \| x^{t+1} - x^t \|_\infty = 0$;

(ii) *$\lim_t x^t = \bar{x}$.*

Proof of Proposition 4. By Theorem 2 and since T is robust, we have that if $\hat{B}$ is a bounded subset of B, then

$$\lim_{\tau} \left(\sup_{x \in \hat{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_\infty \right) = 0 \qquad (55)$$

where $\bar{T} : B \to B$ is a robust opinion aggregator such that $\bar{T} \circ T = \bar{T}$. Since $\bar{T}(T(x)) = \bar{T}(x)$ for all $x \in B$, by induction, we have that $\bar{T}(T^m(x)) = \bar{T}(x)$ for all $m \in \mathbb{N}$ and for all $x \in B$.

---

[57] Theorem 6 was proved by Lorentz [61]. His result is stated for a sequence $\{x^t\}_{t \in \mathbb{N}}$ in $\mathbb{R}$ where each x^t is the partial sum up to t of another sequence $\{a^s\}_{s \in \mathbb{N}}$. In other words, Lorentz's result is a Tauberian theorem for series. Nonetheless, the techniques used to prove Theorem 6 are the same elementary ones discovered by Lorentz with the extra caveat of setting $a^1 = x^1$ and $a^s = x^s - x^{s-1}$ for all $s \geq 2$.

(i) implies (ii). Fix $x \in B$. Define the sequence $x^t = T^t(x)$ for all $t \in \mathbb{N}$. By point 3 of Lemma 6, we have that $\{x^t\}_{t \in \mathbb{N}}$ is bounded. Set $\hat{B} = \{x^t\}_{t \in \mathbb{N}}$. Note that for each $\tau \in \mathbb{N}$ and for each $m \in \mathbb{N}$

$$\frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} = \frac{1}{\tau} \sum_{t=1}^{\tau} T^{m+t}(x) = \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(T^m(x)).$$

Since (55) holds, if we define $\bar{x} = \bar{T}(x)$, then we have that for each $m \in \mathbb{N}$

$$\lim_{\tau} \frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} = \lim_{\tau} \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(T^m(x)) = \bar{T}(T^m(x)) = \bar{T}(x) = \bar{x}.$$

It follows that

$$\sup_{m \in \mathbb{N}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} x^{m+t} - \bar{x} \right\|_\infty = \sup_{m \in \mathbb{N}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(T^m(x)) - \bar{T}(T^m(x)) \right\|_\infty$$

$$\leq \sup_{x \in \hat{B}} \left\| \frac{1}{\tau} \sum_{t=1}^{\tau} T^t(x) - \bar{T}(x) \right\|_\infty.$$

Since (55) holds and T is asymptotically regular, we have that $\{x^t\}_{t \in \mathbb{N}}$ satisfies (54) in (i) of Theorem 6. By Theorem 6, we have that $\lim_t T^t(x) = \lim_t x^t$ exists. Since x was arbitrarily chosen, the implication follows.

(i) implies (ii). Fix $x \in B$. Define $x^t = T^t(x)$ for all $t \in \mathbb{N}$. Since T is convergent, we have that $\{x^t\}_{t \in \mathbb{N}}$ converges and, in particular, is bounded. By Theorem 6, we have that $\lim_t \|T^{t+1}(x) - T^t(x)\|_\infty = \lim_t \|x^{t+1} - x^t\|_\infty = 0$. Since x was arbitrarily chosen, the implication follows.

Next, we assume that, in addition, T is also constant affine.

(ii) implies (iii). By point 2 of Lemma 6 and since T is a robust and constant affine opinion aggregator, it admits a unique extension $S : \mathbb{R}^n \to \mathbb{R}^n$ which is also robust and constant affine. By induction, S^t is robust and constant affine for all $t \in \mathbb{N}$. Moreover, by Theorem 2 and Lemma 9, we have that the limiting operator $\bar{S}$ has the same properties and, in particular, it is continuous. Observe also that if the domain is $\mathbb{R}^n$, then constant affinity yields positive homogeneity. Define $\tilde{I} = [-1, 1]$ and $\tilde{B} = \tilde{I}^n$. Let $\tilde{S}$ be the restriction of S to $\tilde{B}$. Consider the space $C\left(\tilde{B}, \mathbb{R}^n\right)$: the space of continuous functions with $\tilde{B}$ as domain and $\mathbb{R}^n$ as target space. The space $C\left(\tilde{B}, \mathbb{R}^n\right)$ is a Banach space once endowed with the supnorm: $\|f\|_* = \sup_{x \in \tilde{B}} \|f(x)\|_\infty$ for all $f \in C\left(\tilde{B}, \mathbb{R}^n\right)$. Note that $\left\{\tilde{S}^t\right\}_{t \in \mathbb{N}} \subseteq C\left(\tilde{B}, \mathbb{R}^n\right)$. Since T is convergent, so is the extension S and we have that $\lim_t \tilde{S}^t(x) = \lim_t S^t(x) = \bar{S}(x)$ for all $x \in \tilde{B}$.[58] This

---

[58] Observe that for each $x \in \mathbb{R}^n$ there exists $\lambda > 0$ and $k \in \mathbb{R}$ such that $\tilde{x} = \lambda x + ke \in B$. This implies that

$$\lim_t S^t(x) = \lim_t S^t\left(\frac{1}{\lambda}\tilde{x} - \frac{k}{\lambda}e\right) = \lim_t \frac{1}{\lambda} S^t(\tilde{x}) - \frac{k}{\lambda}e = \lim_t \frac{1}{\lambda} T^t(\tilde{x}) - \frac{k}{\lambda}e,$$

implies that $\left\{\tilde{S}^t (x)\right\}_{t \in \mathbb{N}} \subseteq \tilde{B} \subseteq \mathbb{R}^n$ is bounded for all $x \in \tilde{B}$. Since $\bar{S} : \mathbb{R}^n \to \mathbb{R}^n$ is continuous, so is its restriction to $\tilde{B}$ which we still denote by $\bar{S}$. By Lemma 8 and since $\tilde{S}$ is a robust opinion aggregator, it follows that $\tilde{S}^t$ is nonexpansive for all $t \in \mathbb{N}$. By [29, pp. 135–136], this implies that the sequence $\left\{\tilde{S}^t\right\}_{t \in \mathbb{N}} \subseteq C\left(\tilde{B}, \mathbb{R}^n\right)$ is equicontinuous. By contradiction, assume that $\lim_t \left\|\tilde{S}^t - \bar{S}\right\|_* \neq 0$. This would imply that there exists $\varepsilon > 0$ and a subsequence $\left\{\tilde{S}^{t_m}\right\}_{m \in \mathbb{N}} \subseteq \left\{\tilde{S}^t\right\}_{t \in \mathbb{N}}$ such that $\left\|\tilde{S}^{t_m} - \bar{S}\right\|_* \geq \varepsilon$ for all $m \in \mathbb{N}$. By Arzela-Ascoli Theorem (see, e.g., [29, Theorem 7.5.7]) and since $\left\{\tilde{S}^{t_m}\right\}_{m \in \mathbb{N}}$ is equicontinuous and $\left\{\tilde{S}^{t_m} (x)\right\}_{m \in \mathbb{N}} \subseteq \mathbb{R}^n$ is bounded for all $x \in \tilde{B}$, this would imply that there exists a subsequence $\left\{\tilde{S}^{t_{m(l)}}\right\}_{l \in \mathbb{N}}$ and a function $\hat{S} \in C\left(\tilde{B}, \mathbb{R}^n\right)$ such that $\lim_l \left\|\tilde{S}^{t_{m(l)}} - \hat{S}\right\|_* = 0$. By definition of $\| \ \|_*$, it would follow that $\bar{S} (x) = \lim_l \tilde{S}^{t_{m(l)}} (x) = \hat{S} (x)$ for all $x \in \tilde{B}$, that is, $\hat{S} = \bar{S}$ on $\tilde{B}$. This would imply that $0 < \varepsilon \leq \lim_l \left\|\tilde{S}^{t_{m(l)}} - \bar{S}\right\|_* = 0$, a contradiction. We can conclude that

$$\lim_t \left(\sup_{x \in \tilde{B}} \left\|\tilde{S}^t (x) - \bar{S} (x)\right\|_\infty\right) = \lim_t \left\|\tilde{S}^t - \bar{S}\right\|_* = 0.$$

By point 4 of Lemma 6, recall that $\tilde{S}^t (x) = S^t (x)$ for all $x \in \tilde{B}$ and for all $t \in \mathbb{N}$. Note also that $\bar{S} (x) \in \tilde{B}$ for all $x \in \tilde{B}$. Consequently, define $\{c_t\}_{t \in \mathbb{N}} \subseteq [0, \infty)$ by

$$c_t = \sup_{x \in \tilde{B}} \left\|\tilde{S}^t (x) - \bar{S} (x)\right\|_\infty = \sup_{x \in \tilde{B}} \left\|S^t (x) - \bar{S} (x)\right\|_\infty \qquad \forall t \in \mathbb{N}. \qquad (56)$$

By the previous part of the proof, note that $c_t \to 0$. Consider $y \in \mathbb{R}^n \backslash \{0\}$ and $t \in \mathbb{N}$. By (56) and since $y/ \|y\|_\infty \in \tilde{B}$ and $\bar{S}$ and S^t are positively homogeneous for all $t \in \mathbb{N}$, it follows that

$$\frac{1}{\|y\|_\infty} \left\|S^t (y) - \bar{S} (y)\right\|_\infty = \left\|\frac{1}{\|y\|_\infty} S^t (y) - \frac{1}{\|y\|_\infty} \bar{S} (y)\right\|_\infty$$
$$= \left\|S^t \left(\frac{y}{\|y\|_\infty}\right) - \bar{S} \left(\frac{y}{\|y\|_\infty}\right)\right\|_\infty \leq c_t.$$

Since y was arbitrarily chosen in $\mathbb{R}^n \backslash \{0\}$ and $S^t (0) = 0 = \bar{S} (0)$ for all $t \in \mathbb{N}$, we have that

$$\left\|S^t (y) - \bar{S} (y)\right\|_\infty \leq c_t \|y\|_\infty \qquad \forall t \in \mathbb{N}, \forall y \in \mathbb{R}^n.$$

Since S is the extension of T, we can conclude that

$$\left\|\bar{T} (x) - T^t (x)\right\|_\infty = \left\|\bar{S} (x) - S^t (x)\right\|_\infty \leq c_t \|x\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B,$$

proving (17).

(iii) implies (ii). Since T satisfies (17), we clearly have that T is convergent. ∎

yielding that $\lim_t S^t (x)$ exists.

Remark 6 To the best of our knowledge, the use of Lorentz's Theorem to transform Cesaro convergence of the orbits of T into standard convergence seems to have first appeared in Bruck [17]. Compared to our result, he assumes that T is nonexpansive with respect to $\| \ \|_2$ (cf. Remark 5) and he does not elaborate on the rate of convergence. This latter feature is mainly due to the property of constant affinity of T. Finally, proving that asymptotic regularity is equivalent to convergence can also be obtained with the techniques of Browder and Petryshyn [16, Theorem 2].　　　▲

We next move to the proof of our main result on standard convergence: Theorem 3. In order to do so, we will show how the property of having a nontrivial network (see Definition 12) is equivalent to T having a useful decomposition.

Proposition 13 *Let T be a robust opinion aggregator. The following statements are equivalent:*

(i) T has a nontrivial network;

(ii) There exist $W \in \mathcal{W}$ and $\varepsilon \in (0,1)$ such that

$$T(x) = \varepsilon W x + (1 - \varepsilon) S(x) \qquad \forall x \in B \tag{57}$$

where S is a robust opinion aggregator.

Moreover, we have that W in (ii) can be chosen to be such that $A(W) = A(T)$.

Proof. (i) implies (ii). For each $i, j \in N$ if j strongly influences i, consider $\varepsilon_{ij} \in (0,1)$ as in (18) otherwise let $\varepsilon_{ij} = 1/2$. Define $\tilde{W}$ to be such that $\tilde{w}_{ij} = a_{ij}\varepsilon_{ij}$ for all $i, j \in N$ where a_{ij} is the ij-th entry of $A(T)$. Since each row of $A(T)$ is not null, for each $i \in N$ there exists $j \in N$ such that $a_{ij} = 1$ and, in particular, $\tilde{w}_{ij} > 0$. This implies that $\sum_{l=1}^{n} \tilde{w}_{il} > 0$ for all $i \in N$. Define also $\varepsilon = \min \{\min_{i \in N} \sum_{l=1}^{n} \tilde{w}_{il}, 1/2\} \in (0,1)$. Define $W \in \mathcal{W}$ to be such that $w_{ij} = \tilde{w}_{ij}/\sum_{l=1}^{n} \tilde{w}_{il}$ for all $i, j \in N$. Clearly, we have that for each $i, j \in N$

$$w_{ij} > 0 \iff \tilde{w}_{ij} > 0 \iff a_{ij} = 1. \tag{58}$$

This yields that $A(W) = A(T)$. Next, consider $x, y \in B$ such that $x \geq y$. Define $y^0 = y$. For each $t \in \{1, ..., n-1\}$ define $y^t \in B$ to be such that $y_i^t = x_i$ for all $i \leq t$ and $y_i^t = y_i$ for all $i \geq t + 1$. Define $y^n = x$. Note that $x = y^n \geq ...y^1 \geq y^0 = y$. It

follows that

$$T_i\left(x\right) - T_i\left(y\right) = \sum_{j=1}^n \left[T_i\left(y^j\right) - T_i\left(y^{j-1}\right)\right] \geq \sum_{j=1}^n a_{ij}\varepsilon_{ij}\left(y_j^j - y_j^{j-1}\right)$$

$$= \sum_{j=1}^n \tilde{w}_{ij}\left(x_j - y_j\right) = \left(\sum_{l=1}^n \tilde{w}_{il}\right)\left(\sum_{j=1}^n \frac{\tilde{w}_{ij}}{\sum_{l=1}^n \tilde{w}_{il}}\left(x_j - y_j\right)\right)$$

$$= \left(\sum_{l=1}^n \tilde{w}_{il}\right)\left(\sum_{j=1}^n w_{ij}\left(x_j - y_j\right)\right)$$

$$\geq \varepsilon \sum_{j=1}^n w_{ij}\left(x_j - y_j\right) \qquad \forall i \in N.$$

It follows that

$$x \geq y \implies T\left(x\right) - T\left(y\right) \geq \varepsilon W\left(x - y\right) = \varepsilon\left(Wx - Wy\right). \tag{59}$$

Define $S : B \to \mathbb{R}^n$ by

$$S\left(x\right) = \frac{T\left(x\right) - \varepsilon Wx}{1 - \varepsilon} \qquad \forall x \in B. \tag{60}$$

By definition of S and since $W \in \mathcal{W}$ and T is normalized and translation invariant, it is immediate to see that $S\left(ke\right) = ke$ for all $k \in I$ and that S is translation invariant. Since (59) holds and $\varepsilon \in (0, 1)$, routine computations yield that S is monotone. Since S is normalized and monotone, then S is a selfmap, that is, $S\left(B\right) \subseteq B$ and, in particular, S is a robust opinion aggregator. By rearranging (60), (57) follows.

(ii) implies (i). Consider $i \in N$. Since W is a stochastic matrix, there exists $j \in N$ such that $w_{ij} > 0$. Let $x \in B$ and $h > 0$ be such that $x + he^j \in B$. By (57) and since S is monotone, we have that

$$T_i\left(x + he^j\right) - T_i\left(x\right) = \varepsilon w_{ij}h + \left(1 - \varepsilon\right) S_i\left(x + he^j\right) - \left(1 - \varepsilon\right) S_i\left(x\right)$$

$$\geq \varepsilon w_{ij}h,$$

proving that j strongly influences i and $a_{ij} = 1$. It follows that the i-th row of $A\left(T\right)$ is not null. Since i was arbitrarily chosen, the statement follows.

Finally, by (58), note that W in (ii) can be chosen to be such that $A\left(W\right) = A\left(T\right)$. ∎

Our standard convergence theorem (Theorem 3) builds on two assumptions: a) the adjacency matrix $A\left(T\right)$ has no null row and b) each closed group of $\left(N, A\left(T\right)\right)$ is aperiodic. On the one hand, the first assumption allows for a decomposition of T into a convex linear combination of a linear opinion aggregator with matrix W and a

robust opinion aggregator S (cf. Proposition 13). We next show that if W takes a very particular shape, which we dub partition matrix, then T is asymptotically regular and, in particular, convergent (see Lemma 10 and Proposition 14). On the other hand, the second assumption yields that W can be always chosen such that W^t is eventually a partition matrix. This will prove Theorem 3.

Definition 20 *Let $J : B \to B$ be an opinion aggregator. We say that J is a* partition *operator/matrix if and only if there exists a family of disjoint nonempty subsets $\{N_l\}_{l=1}^m$ of N such that $\cup_{l=1}^m N_l = N$ and for each $l \in \{1, ..., m\}$ there exists $k_l \in N_l$ such that $J_i(x) = x_{k_l}$ for all $i \in N_l$.*

Note that a partition operator is linear. With a small abuse of notation, we will denote the matrix and the operator by the same symbol. In particular, $J_{k_l}(x) = x_{k_l}$ for all $l \in \{1, ..., m\}$ and for all $x \in B$.

Lemma 10 *Let T be a robust opinion aggregator such that $T = \varepsilon J + (1 - \varepsilon) S$ where $\varepsilon \in (0, 1)$, J is a partition operator, and $S : B \to B$ is a robust opinion aggregator. Let A be a nonempty subset of B such that there exists $k > 0$ satisfying*

$$\|T(x) - x\|_\infty < k \qquad \forall x \in A. \tag{61}$$

If there exists $\delta > 0$ such that for each $t \in \mathbb{N}_0$ there exists $x \in A$ satisfying

$$\left\|T^{t+1}(x) - T^t(x)\right\|_\infty \geq \delta, \tag{62}$$

then $\{T^t(x) : x \in A \text{ and } t \in \mathbb{N}_0\}$ is unbounded.

The previous lemma, proved in the Online Appendix, and next proposition are related to the work of Edelstein and O'Brien [30, Lemma 1 and Theorem 1]. In Remark 9, before proving Lemma 10, we elaborate on the differences.

Proposition 14 *Let T be a robust opinion aggregator. If T is such that $T = \varepsilon J + (1 - \varepsilon) S$ where $\varepsilon \in (0, 1)$, J is a partition operator, and S is a robust opinion aggregator, then T is asymptotically regular and, in particular, convergent.*

Proof. Fix $x \in B$. In Lemma 10, set $A = \{x\}$. Clearly, there exists $k > 0$ that satisfies $\|T(x) - x\|_\infty < k$. By point 3 of Lemma 6 and since T is a robust opinion aggregator, it follows that $\{T^t(x)\}_{t \in \mathbb{N}_0}$ is bounded. By Lemma 10, we have that for each $\delta > 0$ there exists $\bar{t} \in \mathbb{N}$ such that

$$\left\|T^{\bar{t}+1}(x) - T^{\bar{t}}(x)\right\|_\infty < \delta. \tag{63}$$

Since T is nonexpansive, $\{\|T^{t+1}(x) - T^t(x)\|_\infty\}_{t \in \mathbb{N}_0}$ is a decreasing sequence. By (63) and since $\{\|T^{t+1}(x) - T^t(x)\|_\infty\}_{t \in \mathbb{N}_0}$ is a decreasing sequence, we have that for each $\delta > 0$ there exists $\bar{t} \in \mathbb{N}$ such that $\|T^{t+1}(x) - T^t(x)\|_\infty < \delta$ for all $t \geq \bar{t}$, that is, $\lim_t \|T^{t+1}(x) - T^t(x)\|_\infty = 0$. Since x was arbitrarily chosen, it follows that T is asymptotically regular. By Proposition 4, this implies that T is convergent. ∎

Next, we show that if T is strongly aperiodic and has a nontrivial network, then there exists $\bar{t} \in \mathbb{N}$ such that $T^{\bar{t}} = \gamma J + (1-\gamma)S$ (resp., $T^{\bar{t}+1} = \gamma J + (1-\gamma)S$) where J is a partition operator, $\gamma \in (0,1)$, and S is a robust opinion aggregator. The operator J only depends on $A(T)$ while γ and S both depend on $\bar{t}$ (resp., $\bar{t}+1$). In turn, Proposition 14 yields that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent. This will be sufficient to imply the convergence of T.

Lemma 11 *Let T be a robust opinion aggregator. If T is strongly aperiodic and has a nontrivial network, then there exists $\bar{t} \in \mathbb{N}$ such that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent.*

Proof. By Proposition 13 and since T has a nontrivial network, we have that there exists $W \in \mathcal{W}$, $\varepsilon \in (0,1)$, and a robust opinion aggregator $S : B \to B$ such that

$$T(x) = \varepsilon W x + (1-\varepsilon)S(x) \quad \forall x \in B. \tag{64}$$

Moreover, W can be chosen to be such that $A(W) = A(T)$. By [42, Theorems 2 and 3] and since T is strongly aperiodic, this implies that there exist $\bar{t} \in \mathbb{N}$ and a partition $\{N_l\}_{l=1}^m$ of N such that for each $l \in \{1, ..., m\}$ there exists $k_l \in N_l$ satisfying $w_{ik_l}^{(\bar{t})}, w_{ik_l}^{(\bar{t}+1)} > 0$ for all $i \in N_l$.[59] It follows that

$$W^{\bar{t}} = \delta_{\bar{t}} J + (1 - \delta_{\bar{t}}) \tilde{W}_{\bar{t}} \text{ and } W^{\bar{t}+1} = \delta_{\bar{t}+1} J + (1 - \delta_{\bar{t}+1}) \tilde{W}_{\bar{t}+1} \tag{65}$$

where $\delta_{\bar{t}}, \delta_{\bar{t}+1} \in (0,1)$, J is a partition operator,[60] and $\tilde{W}_{\bar{t}}$ as well as $\tilde{W}_{\bar{t}+1}$ are stochastic matrices. By (64) and induction, we also have that

$$T^{\bar{t}}(x) = \varepsilon^{\bar{t}} W^{\bar{t}} x + \left(1 - \varepsilon^{\bar{t}}\right) \tilde{S}_{\bar{t}}(x) \qquad \forall x \in B$$

and

$$T^{\bar{t}+1}(x) = \varepsilon^{\bar{t}+1} W^{\bar{t}+1} x + \left(1 - \varepsilon^{\bar{t}+1}\right) \tilde{S}_{\bar{t}+1}(x) \qquad \forall x \in B$$

where $\tilde{S}_{\bar{t}}$ and $\tilde{S}_{\bar{t}+1}$ are robust opinion aggregators. By (65), it follows that

$$T^{\bar{t}} = \gamma_{\bar{t}} J + (1 - \gamma_{\bar{t}}) \hat{S}_{\bar{t}} \text{ and } T^{\bar{t}+1} = \gamma_{\bar{t}+1} J + (1 - \gamma_{\bar{t}+1}) \hat{S}_{\bar{t}+1}$$

---

[59] As usual, we denote by $w_{ik_l}^{(\bar{t})}$ (resp., $w_{ik_l}^{(\bar{t}+1)}$) the entry in the i-th row and k_l-th column of the matrix $W^{\bar{t}}$ (resp., $W^{\bar{t}+1}$).

[60] That is, $J_i(x) = x_{k_l}$ for all $i \in N_l$ and for all $l \in \{1, ..., m\}$ where $\{N_l\}_{l=1}^m$ and $\{k_l\}_{l=1}^m$ have been defined above.

where $\gamma_{\bar{t}} = \varepsilon^{\bar{t}}\delta_{\bar{t}}$ (resp., $\gamma_{\bar{t}+1} = \varepsilon^{\bar{t}+1}\delta_{\bar{t}+1}$) and $\hat{S}_{\bar{t}}(x) = \frac{\varepsilon^{\bar{t}}(1-\delta_{\bar{t}})}{1-\varepsilon^{\bar{t}}\delta_{\bar{t}}}\tilde{W}_{\bar{t}}x + \frac{1-\varepsilon^{\bar{t}}}{1-\varepsilon^{\bar{t}}\delta_{\bar{t}}}\tilde{S}_{\bar{t}}(x)$ (resp.,

$\hat{S}_{\bar{t}+1}(x) = \frac{\varepsilon^{\bar{t}+1}(1-\delta_{\bar{t}+1})}{1-\varepsilon^{\bar{t}+1}\delta_{\bar{t}+1}}\tilde{W}_{\bar{t}}x + \frac{1-\varepsilon^{\bar{t}+1}}{1-\varepsilon^{\bar{t}+1}\delta_{\bar{t}+1}}\tilde{S}_{\bar{t}+1}(x)$) for all $x \in B$. It follows that $\gamma_{\bar{t}}, \gamma_{\bar{t}+1} \in (0,1)$ and $\hat{S}_{\bar{t}}$ as well as $\hat{S}_{\bar{t}+1}$ are robust opinion aggregators. By Proposition 14, this implies that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent. ∎

Proof of Theorem 3. We adopt the usual convention $T^0(x) = x$ for all $x \in B$. By Lemma 11 and since T is strongly aperiodic and has a nontrivial network, there exists $\bar{t} \in \mathbb{N}$ such that $T^{\bar{t}}$ and $T^{\bar{t}+1}$ are convergent. We next show that this implies that T is convergent. Fix $x \in B$. Since $T^{\bar{t}}$ is convergent, we can conclude that $\lim_k T^{k\bar{t}}(x)$ exists. Denote $\bar{x} = \lim_k T^{k\bar{t}}(x)$. Since T is continuous and so is $T^{\bar{t}}$, it is plain that $T^{\bar{t}}(\bar{x}) = \bar{x}$. This implies that

$$T^{\bar{t}}(T^s(\bar{x})) = T^{\bar{t}+s}(\bar{x}) = T^{s+\bar{t}}(\bar{x}) = T^s\left(T^{\bar{t}}(\bar{x})\right) = T^s(\bar{x}) \qquad \forall s \in \mathbb{N}_0.$$

By induction on k, this yields that for each $s \in \mathbb{N}_0$

$$T^{(k+1)\bar{t}}(T^s(\bar{x})) = T^{k\bar{t}}\left(T^{\bar{t}}(T^s(\bar{x}))\right) = T^{k\bar{t}}(T^s(\bar{x})) = T^s(\bar{x}) \quad \forall k \in \mathbb{N}.$$

In particular, by setting $k = s$, we obtain that for each $s \in \mathbb{N}$

$$T^{s(\bar{t}+1)}(\bar{x}) = T^{s\bar{t}}(T^s(\bar{x})) = T^s(\bar{x}). \tag{66}$$

Since $T^{\bar{t}+1}$ is convergent, we have that $\lim_s T^{s(\bar{t}+1)}(\bar{x})$ exists. By (66), this implies that $\lim_s T^s(\bar{x})$ exists. Denote $\hat{x} = \lim_s T^s(\bar{x})$. Since T is continuous, it is plain that $T(\hat{x}) = \hat{x}$. Since $\{T^{k\bar{t}}(\bar{x})\}_{k\in\mathbb{N}} \subseteq \{T^s(\bar{x})\}_{s\in\mathbb{N}}$ and $T^{k\bar{t}}(\bar{x}) = \bar{x}$ for all $k \in \mathbb{N}$, we have that

$$\bar{x} = \lim_k T^{k\bar{t}}(\bar{x}) = \lim_s T^s(\bar{x}) = \hat{x} \text{ and } T(\hat{x}) = \hat{x}. \tag{67}$$

We can now prove that $\{T^t(x)\}_{t\in\mathbb{N}}$ converges too. By (67) and since T is nonexpansive, we have that

$$\left\|\bar{x} - T^{t+1}(x)\right\|_\infty = \left\|T(\bar{x}) - T(T^t(x))\right\|_\infty \leq \left\|\bar{x} - T^t(x)\right\|_\infty \qquad \forall t \in \mathbb{N},$$

yielding that $\{\|\bar{x} - T^t(x)\|_\infty\}_{t\in\mathbb{N}}$ is a decreasing sequence. Moreover, since $\bar{x} = \lim_k T^{k\bar{t}}(x)$, we have that the subsequence $\{\|\bar{x} - T^{k\bar{t}}(x)\|_\infty\}_{k\in\mathbb{N}} \subseteq \{\|\bar{x} - T^t(x)\|_\infty\}_{t\in\mathbb{N}}$ converges to 0. This implies that $\lim_t T^t(x) = \bar{x}$ and, in particular, that T is convergent at x. Since x was arbitrarily chosen, the statement follows. ∎

Remark 7 Most of our results on convergence rely on the notion of adjacency matrix induced by T (cf. Definition 10). We are aware of three other notions of directed graph/network associated to an operator $T : \mathbb{R}^n \to \mathbb{R}^n$, coming from the mathematical literature. They are due respectively to Gaubert and Gunawardena [40, p. 4943],

Amghibech and Dellacherie [4], and Nussbaum [73].[61] These three notions are too weak to yield our convergence results. In fact, given any of these three notions, we can provide examples of nonconvergent robust opinion aggregators T such that the corresponding graph features each agent being connected to any other agent via a path of length one (that is each ij-th entry is 1). ▲

Proof of Corollary 1. Since T is self-influential, it follows that each row of $A(T)$ is not null, yielding that T has a nontrivial network. Moreover, since there is a simple cycle of length 1 from i to i for all $i \in N$, each closed group is trivially aperiodic. By Theorem 3, the statement follows.

Proof of Lemma 3. By Lemma 8 and since T is robust, T is nonexpansive and, in particular, continuous. By Theorems 2 and 6 and since T is robust and convergent, we have that

$$\bar{T}(x) = C - \lim_t T^t(x) = \lim_t T^t(x) \qquad \forall x \in B.$$

For ease of notation, for each $x \in B$ define $\bar{T}(x) = \bar{x}$. Since T is continuous, we have that $T(\bar{x}) = T(\lim_t T^t(x)) = \lim_t T(T^t(x)) = \lim_t T^{t+1}(x) = \bar{x}$, proving the statement. ■

Proof of Proposition 6. Before starting, observe that if T has the pairwise common influencer property (resp., is strongly connected), then T has a nontrivial network. By Proposition 13, there exist $W \in \mathcal{W}$ and $\varepsilon \in (0, 1)$ such that

$$T(x) = \varepsilon W x + (1 - \varepsilon) S(x) \qquad \forall x \in B \qquad (68)$$

where $S : B \to B$ is a robust opinion aggregator. Moreover, W can be chosen to be such that $A(W) = A(T)$ and so, in particular, W is scrambling (resp., irreducible). Finally, by induction and (68), we have that if $t \in \mathbb{N}$, then there exist $\gamma \in (0, 1)$ and a robust opinion aggregator $\tilde{S} : B \to B$ (which both depend on t) such that

$$T^t(x) = \gamma W^t x + (1 - \gamma) \tilde{S}(x) \qquad \forall x \in B. \qquad (69)$$

a. Assume that T has the pairwise common influencer property. By contradiction, assume that there exists $x \in B \backslash D$ such that $T(x) = x$. Define $x_i = \min_{l \in N} x_l$ and $x_j = \max_{l \in N} x_l$. It follows that $x_j > x_i$ and $i \neq j$. Since W is scrambling, there exists $k = k(i, j) \in N$ such that $w_{ik} > 0$ and $w_{jk} > 0$. We have two cases:

---

[61] The notion of Nussbaum pertains homogeneous operators defined over $\mathbb{R}^n_+$. We consider the natural corresponding notion for translation invariant operator which is obtained via the usual log-exp transformation.

1. $x_k < x_j$. It follows that

$$0 = \|T(x) - x\|_\infty \geq |T_j(x) - x_j| = \left| \varepsilon \sum_{l=1}^n w_{jl} x_l + (1 - \varepsilon) S_j(x) - x_j \right|$$

$$= \left| \varepsilon \sum_{l=1}^n w_{jl}(x_l - x_j) + (1 - \varepsilon)(S_j(x) - x_j) \right|$$

$$= \varepsilon \sum_{l=1}^n w_{jl}(x_j - x_l) + (1 - \varepsilon)(x_j - S_j(x)) \geq \varepsilon w_{jk}(x_j - x_k) > 0,$$

a contradiction.

2. $x_k > x_i$. It follows that

$$0 = \|T(x) - x\|_\infty \geq |T_i(x) - x_i| = \left| \varepsilon \sum_{l=1}^n w_{il} x_l + (1 - \varepsilon) S_i(x) - x_i \right|$$

$$= \left| \varepsilon \sum_{l=1}^n w_{il}(x_l - x_i) + (1 - \varepsilon)(S_i(x) - x_i) \right|$$

$$= \varepsilon \sum_{l=1}^n w_{il}(x_l - x_i) + (1 - \varepsilon)(S_i(x) - x_i) \geq \varepsilon w_{ik}(x_k - x_i) > 0,$$

a contradiction.

Case 1 and 2 prove that the only equilibria of T are the constant vectors in B, proving the statement.

b. Assume that T has the uniform common influencer property. It follows that T has the pairwise common influencer property. By point a, we have that T is a consensus operator.

c. We start by making one simple observation. By induction, if $t \in \mathbb{N}$, then the equilibria of T are a subset of the ones of $T^t : B \to B$. By contradiction, assume that there exists $x \in B \backslash D$ such that $T(x) = x$. Define $x_i = \min_{l \in N} x_l$ and $x_j = \max_{l \in N} x_l$. Clearly, $x_j > x_i$. Since $A(W) = A(T)$ and T is strongly connected, there exists $\bar{t} \in \mathbb{N}$ such that $w_{ij}^{(\bar{t})} > 0$. Since $x \in E(T)$, we have that $T^{\bar{t}}(x) = x$. By (69), this implies that

$$0 = \left\| T^{\bar{t}}(x) - x \right\|_\infty \geq \left| T_i^{\bar{t}}(x) - x_i \right| = \left| \gamma \sum_{l=1}^n w_{il}^{(\bar{t})} x_l + (1 - \gamma) \tilde{S}_i(x) - x_i \right|$$

$$= \left| \gamma \sum_{l=1}^n w_{il}^{(\bar{t})}(x_l - x_i) + (1 - \gamma)\left(\tilde{S}_i(x) - x_i \right) \right|$$

$$= \gamma \sum_{l=1}^n w_{il}^{(\bar{t})}(x_l - x_i) + (1 - \gamma)\left(\tilde{S}_i(x) - x_i \right) \geq \gamma w_{ij}^{(\bar{t})}(x_j - x_i) > 0,$$

a contradiction. ■

Our convergence results in Section 5.3.3 (Corollaries 2–4) can all be reconducted to the following convergence result which generalizes one of the findings of DeGroot (see, e.g., [26, Theorem 1]).

Theorem 7 *Let T be a robust opinion aggregator. If there exists $\hat{t} \in \mathbb{N}$ such that $T^{\hat{t}}$ has the uniform common influencer property, then $\bar{T}(x) = \lim_t T^t(x) \in D$ for all $x \in B$. Moreover, there exists $\varepsilon \in (0,1)$ such that*

$$\left\|\bar{T}(x) - T^t(x)\right\|_\infty \le 2(1-\varepsilon)^{\left\lfloor \frac{t}{\hat{t}} \right\rfloor} \|x\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B. \tag{70}$$

Proof. Before proving the main statement, we need to state and prove an ancillary claim. We just introduce some notation. Given a sequence of stochastic matrices, $\{W_l\}_{l \in \mathbb{N}} \subseteq \mathcal{W}$, we denote by $\Pi_{l=1}^{t+1} W_l$ the backward product of the first $t+1$ elements, that is, $\Pi_{l=1}^{t+1} W_l = W_{t+1} \Pi_{l=1}^t W_l = W_{t+1}...W_1$ for all $t \in \mathbb{N}$.

Claim: *If $\{W_l\}_{l \in \mathbb{N}} \subseteq \mathcal{W}_\varepsilon$ for some $\varepsilon \in (0,1)$, then for each $t, m \in \mathbb{N}$ such that $m \ge t$*

$$\left\|\left(\Pi_{l=1}^m W_l\right) x - \left(\Pi_{l=1}^t W_l\right) x\right\|_\infty \le 2(1-\varepsilon)^t \|x\|_\infty \qquad \forall x \in B.$$

Proof of the Claim. Recall that the product of stochastic matrices is a stochastic matrix, thus $\Pi_{l=1}^t W_l \in \mathcal{W}$ for all $t \in \mathbb{N}$. Next, define $V_0 = \{y \in \mathbb{R}^n : \sum_{i=1}^n y_i = 0\}$. By [81, p. 28], note that for each $y \in V_0$ and $l \in \mathbb{N}$

$$y^T W_l \in V_0 \text{ and } \left\|y^T W_l\right\|_1 \le (1-\varepsilon) \|y\|_1$$

where $\|y\|_1 = \sum_{i=1}^n |y_i|$. By induction, this yields that

$$\left\|y^T \Pi_{l=1}^t W_l\right\|_1 \le (1-\varepsilon)^t \|y\|_1 \qquad \forall y \in V_0. \tag{71}$$

Finally, consider $x \in B$ and $m > t$. By the definition of backward product, it follows that $\left(\Pi_{l=1}^m W_l\right) x = \left(\Pi_{l=t+1}^m W_l\right) \left(\left(\Pi_{l=1}^t W_l\right) x\right)$. Observe that

$$\left(\Pi_{l=t+1}^m W_l\right) \left(\left(\Pi_{l=1}^t W_l\right) x\right) - \left(\Pi_{l=1}^t W_l\right) x = \left(\left(\Pi_{l=t+1}^m W_l\right) - I\right) \left(\left(\Pi_{l=1}^t W_l\right) x\right) = Z(Rx)$$

where $Z = \Pi_{l=t+1}^m W_l - I$ and $R = \Pi_{l=1}^t W_l$. Note that $R, \Pi_{l=t+1}^m W_l \in \mathcal{W}$ and, in particular, the entries of each row of Z sum up to 0. Denote by z^i the column vector whose transpose is exactly the i-th row of Z. It is immediate to see that it is the difference of two probability vectors, thus $z^i \in V_0$ and $\|z^i\|_1 \le 2$. The quantities $Z(Rx)$ and Rx are vectors of $\mathbb{R}^n$, and the i-th component of the former is exactly $(z^i)^T Rx$. By (71) and since $z^i \in V_0$, we can conclude that

$$\left|(z^i)^T (Rx)\right| = \left|\left((z^i)^T R\right) x\right| \le \left\|(z^i)^T R\right\|_1 \|x\|_\infty \le (1-\varepsilon)^t \|z^i\|_1 \|x\|_\infty$$
$$\le 2(1-\varepsilon)^t \|x\|_\infty.$$

Since $(\Pi_{l=1}^m W_l)\, x - (\Pi_{l=1}^t W_l)\, x = Z\,(Rx)$ and t, m, x, and i were arbitrarily chosen, we have that

$$\left|\left((\Pi_{l=1}^m W_l)\, x - (\Pi_{l=1}^t W_l)\, x\right)_i\right| = \left|(z^i)^{\mathrm{T}}\,(Rx)\right| \leq 2\,(1-\varepsilon)^t\,\|x\|_\infty \qquad \forall x \in B, \forall i \in N, \forall m > t.$$

We can conclude that

$$\left\|(\Pi_{l=1}^m W_l)\, x - (\Pi_{l=1}^t W_l)\, x\right\|_\infty \leq 2\,(1-\varepsilon)^t\,\|x\|_\infty \qquad \forall x \in B, \forall t, m \in \mathbb{N} \text{ s.t. } m \geq t,$$

proving the claim. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We start by proving the statement for $\hat{t} = 1$. Since $T = T^{\hat{t}}$ has the uniform common influencer property, it follows that T has a nontrivial network. By Proposition 13, we have that there exist $W \in \mathcal{W}$ and $\gamma \in (0,1)$ such that

$$T\,(x) = \gamma W x + (1-\gamma)\, S\,(x) \qquad \forall x \in B \tag{72}$$

where S is a robust opinion aggregator. Moreover, W can be chosen to be such that $A\,(W) = A\,(T)$ and so, in particular, $W \in \mathcal{W}_\delta$ for some $\delta \in (0,1)$. This implies that for each $x \in B$ there exists a matrix $W\,(x) \in \mathcal{W}_{\gamma\delta}$ such that $T\,(x) = W\,(x)\,x$ for all $x \in B$.[62] Fix $x \in B$. Define $W_l = W\left(T^{l-1}\,(x)\right)$ for all $l \in \mathbb{N}$ with the usual convention $T^0\,(x) = x$ for all $x \in B$. By induction, we have that for each $m \in \mathbb{N}$

$$T^m\,(x) = T\left(T^{m-1}\,(x)\right) = W\left(T^{m-1}\,(x)\right) T^{m-1}\,(x) = (\Pi_{l=1}^m W_l)\, x.$$

By the previous claim, we have that

$$\left\|T^m\,(x) - T^t\,(x)\right\|_\infty = \left\|(\Pi_{l=1}^m W_l)\, x - (\Pi_{l=1}^t W_l)\, x\right\|_\infty \leq 2\,(1-\varepsilon)^t\,\|x\|_\infty \qquad \forall t, m \in \mathbb{N} \text{ s.t. } m \geq t \tag{73}$$

$\varepsilon = \gamma\delta > 0$. By setting $m = t+1$ and since x was arbitrarily chosen, this implies that T is asymptotically regular. By Proposition 4, it follows that T is convergent. By Lemma 3 and since T is convergent, we have that $\bar{T}\,(x) = \lim_t T^t\,(x) \in E\,(T)$ for all $x \in B$. By taking the limit in m in (73), (70) immediately follows for the case $\hat{t} = 1$. By Proposition 6 and since T has the uniform common influencer property and $\bar{T}\,(x) = \lim_t T^t\,(x) \in E\,(T)$ for all $x \in B$, we have that T is a consensus operator and $\bar{T}\,(x) = \lim_t T^t\,(x) \in D$ for all $x \in B$.

---

[62]Observe that $S_i : B \to \mathbb{R}$ is normalized, monotone, and translation invariant for all $i \in N$. By [20] and since S_i is continuous, for each $i \in N$ there exists a convex and compact subset C_i of Δ such that

$$S_i\,(x) = \alpha\,(x) \min_{w \in C_i} w \cdot x + [1 - \alpha\,(x)] \max_{w \in C_i} w \cdot x \qquad \forall x \in B$$

where $\alpha : B \to [0,1]$. It follows that for each $i \in N$ and for each $x \in B$ there exists $w^i\,(x) \in \Delta$ such that $S_i\,(x) = w^i\,(x) \cdot x$. If we define $\hat{W}\,(x) \in \mathcal{W}$ to be such that its i-th row is the transpose of $w^i\,(x)$, then we have that $S\,(x) = \hat{W}\,(x)\,x$. By (72), we can conclude that $T\,(x) = W\,(x)\,x$ where $W\,(x) = \gamma W + (1-\gamma)\,\hat{W}\,(x)$.

If $\hat{t} > 1$, define $U : B \to B$ to be such that $U = T^{\hat{t}}$. Note that U is a robust opinion aggregator. By assumption, it follows that U has the uniform common influencer property. By the previous part of the proof, we have that

$$\left\| \bar{U}(x) - U^t(x) \right\|_\infty \leq 2(1 - \varepsilon)^t \|x\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B$$

and $\bar{U}(x) = \lim_t U^t(x) \in D$ for all $x \in B$. Consider $x \in B$ and, for ease of notation, denote $\bar{x} = \bar{U}(x)$. Let $t \in \mathbb{N}$. We have two cases:

1. $t < \hat{t}$. In this case, it follows that $\lfloor \frac{t}{\hat{t}} \rfloor = 0$ and

$$\left\| \bar{x} - T^t(x) \right\|_\infty \leq \left\| \bar{U}(x) \right\|_\infty + \left\| T^t(x) \right\|_\infty \leq \|x\|_\infty + \|x\|_\infty = 2(1 - \varepsilon)^{\lfloor \frac{t}{\hat{t}} \rfloor} \|x\|_\infty .$$

2. $t \geq \hat{t}$. In this case, it follows that there exists $n \in \mathbb{N}$ such that $n\hat{t} \leq t < (n+1)\hat{t}$. This implies that $n \leq \frac{t}{\hat{t}} < (n+1)$ and $n = \lfloor \frac{t}{\hat{t}} \rfloor$. We have two subcases:

   (a) $n = \frac{t}{\hat{t}}$. In this case, we have that

   $$\left\| \bar{x} - T^t(x) \right\|_\infty = \left\| \bar{x} - T^{n\hat{t}}(x) \right\|_\infty = \left\| \bar{x} - U^n(x) \right\|_\infty \leq 2(1 - \varepsilon)^n \|x\|_\infty .$$

   (b) $n < \frac{t}{\hat{t}}$. In this case, define $j = t - n\hat{t}$. Since $\bar{x} \in D$ and T is nonexpansive, we have that

   $$\begin{aligned}
   \left\| \bar{x} - T^t(x) \right\|_\infty = \left\| T(\bar{x}) - T^t(x) \right\|_\infty &\leq \left\| \bar{x} - T^{t-1}(x) \right\|_\infty \\
   &= \left\| T(\bar{x}) - T^{t-1}(x) \right\|_\infty \leq \dots \\
   &\leq \left\| \bar{x} - T^{t-j}(x) \right\|_\infty = \left\| \bar{x} - T^{n\hat{t}}(x) \right\|_\infty \\
   &= \left\| \bar{x} - U^n(x) \right\|_\infty \leq 2(1 - \varepsilon)^n \|x\|_\infty .
   \end{aligned}$$

Since t and x were arbitrarily chosen, points 1 and 2 prove that

$$\left\| \bar{U}(x) - T^t(x) \right\|_\infty \leq 2(1 - \varepsilon)^{\lfloor \frac{t}{\hat{t}} \rfloor} \|x\|_\infty \qquad \forall t \in \mathbb{N}, \forall x \in B. \qquad (74)$$

It follows that $\{T^t(x)\}_{t \in \mathbb{N}}$ converges for all $x \in B$. By taking the limit in t and Lemma 3, this implies that $\bar{T}(x) = \lim_t T^t(x) = \bar{U}(x) \in D$ for all $x \in B$. Thus, (74) implies (70). ∎

Proof of Corollary 2. By hypothesis, in Theorem 7, we can set $\hat{t} = 1$. By Theorem 7, the statement follows. ∎

Proof of Corollary 3. By repeating the same arguments contained in the beginning of the proof of Proposition 6 up to (69), we have that if $t \in \mathbb{N}$, then there exist $\gamma \in (0,1)$, a stochastic matrix $W \in \mathcal{W}$, and a robust opinion aggregator $\tilde{S} : B \to B$ such that

$$T^t(x) = \gamma W^t x + (1 - \gamma)\tilde{S}(x) \qquad \forall x \in B$$

where γ and $\tilde{S}$ depend on T as well as t and W can be chosen to be such that $A(W) = A(T)$. Since T has the pairwise common influencer property and $A(W) = A(T)$, it follows that W is scrambling. By [80, Theorem 4.11 and Exercise 4.13] and since W is scrambling, we have that there exists $\hat{t} \in \mathbb{N}$ such that $W^{\hat{t}}$ has a column whose entries are all strictly positive. By [49, Theorem 2], it follows that $\hat{t}$ is such that $\hat{t} \leq n^2 - 3n + 3$. Since $W^{\hat{t}}$ has a column whose entries are all strictly positive, note that $W^{\hat{t}} \in \mathcal{W}_\delta$ for some $\delta \in (0, 1)$. Denote the strictly positive column by $k \in N$. It follows that if $x \in B$ and $h > 0$ are such that $x + he^k \in B$, then for each $i \in N$

$$T_i^{\hat{t}}\left(x + he^k\right) - T_i^{\hat{t}}(x) \geq \gamma \sum_{l=1}^n w_{il}^{(\hat{t})} he_l^k \geq \gamma w_{ik}^{(\hat{t})} h \geq \gamma \delta h,$$

proving that k strongly influences i for all $i \in N$ and, in particular, that $T^{\hat{t}}$ has the uniform common influencer property. By Theorem 7, $\bar{T}(x) = \lim_t T^t(x) \in D$ for all $x \in B$ and there exists $\varepsilon \in (0, 1)$ such that (19) holds. ∎

Proof of Corollary 4. By repeating the same arguments contained in the beginning of the proof of Proposition 6 up to (69), we have that if $t \in \mathbb{N}$, then there exist $\gamma \in (0, 1)$, a stochastic matrix $W \in \mathcal{W}$, and a robust opinion aggregator $\tilde{S} : B \to B$ such that

$$T^t(x) = \gamma W^t x + (1 - \gamma)\tilde{S}(x) \qquad \forall x \in B$$

where γ and $\tilde{S}$ depend on T as well as t and W can be chosen to be such that $A(W) = A(T)$. Since T is strongly connected and strongly aperiodic and $A(W) = A(T)$, it follows that W is primitive. By [47, Corollary 8.5.8] and since W is primitive, we have that there exists $\hat{t} \in \mathbb{N}$ such that $W^{\hat{t}}$ has a column whose entries are all strictly positive and $\hat{t}$ is such that $\hat{t} \leq (n-1)^2 + 1$. Since $W^{\hat{t}}$ has a column whose entries are all strictly positive, note that $W^{\hat{t}} \in \mathcal{W}_\delta$ for some $\delta \in (0, 1)$. Denote the strictly positive column by $k \in N$. It follows that if $x \in B$ and $h > 0$ are such that $x + he^k \in B$, then for each $i \in N$

$$T_i^{\hat{t}}\left(x + he^k\right) - T_i^{\hat{t}}(x) \geq \gamma \sum_{l=1}^n w_{il}^{(\hat{t})} he_l^k \geq \gamma w_{ik}^{(\hat{t})} h \geq \gamma \delta h,$$

proving that k strongly influences i for all $i \in N$ and, in particular, that $T^{\hat{t}}$ has the uniform common influencer property. By Theorem 7, $\bar{T}(x) = \lim_t T^t(x) \in D$ for all $x \in B$ and there exists $\varepsilon \in (0, 1)$ such that (20) holds.

Finally, since $A(W) = A(T)$, if T is self-influential, then W is such that $w_{ii} > 0$ for all $i \in N$. By [47, Lemma 8.5.4] and since W is primitive, we have that the above index $\hat{t}$ is such that $\hat{t} \leq n - 1$. ∎

Proof of Theorem 4. Let $x \in B$. Call V the set of values the components of x take: $V = \{x_1, ..., x_n\}$. Define U to be the subset of vectors y in B such that each component of y coincides in value to the value of some component of x, formally,

$$U = \{y \in B : y_i \in V \qquad \forall i \in \{1, ..., n\}\}.$$

Since the components of x might be distinct, note that the cardinality of U is at most n^n. Since ν_i is $\{0, 1\}$-valued for all $i \in \{1, ..., n\}$, note that $T_i(y) \in V$ for all $y \in U$ and for all $i \in \{1, ..., n\}$. This implies that $T(x) \in U$. By induction, it follows that $T^t(x) \in U$ for all $t \in \mathbb{N}$. This implies that the sequence $\{T^t(x)\}_{t \in \mathbb{N}}$ can take at most a finite number of values. We have two cases:

1. $\{T^t(x)\}_{t \in \mathbb{N}}$ converges. If $\{T^t(x)\}_{t \in \mathbb{N}}$ converges, then the previous part implies that $\{T^t(x)\}_{t \in \mathbb{N}}$ becomes constant, that is, there exists $\tilde{t} \in \mathbb{N}$ such that

$$T^t(x) = T^{\tilde{t}}(x) \in U \qquad \forall t \geq \tilde{t}. \tag{75}$$

Call $\bar{x}$ the limit of $\{T^t(x)\}_{t \in \mathbb{N}}$. Note that $\bar{x} = T^{\tilde{t}}(x)$ and $T^t(\bar{x}) = \bar{x}$ for all $t \in \mathbb{N}$. In particular, we have that

$$T(\bar{x}) = \bar{x}. \tag{76}$$

Define now $\bar{t} \in \mathbb{N}$ to be such that $\bar{t} = \min\{t \in \mathbb{N} : T^t(x) = \bar{x}\}$. By (75), $\bar{t}$ is well defined. By (76), we have that $T^t(x) = \bar{x}$ for all $t \geq \bar{t}$. If $\bar{t} = 1$, then $\{T^t(x)\}_{t \in \mathbb{N}}$ is constant to begin with and so it becomes constant after at most n^n periods. Assume $\bar{t} > 1$. We next show that $T^t(x) \neq T^m(x)$ for all $m, t < \bar{t}$ such that $m \neq t$. By contradiction, assume that there exist $m, t < \bar{t}$ such that $m \neq t$ and $T^t(x) = T^m(x)$. Without loss of generality, we assume that $m > t$. This would imply that $T^{t+n}(x) = T^n(T^t(x)) = T^n(T^m(x)) = T^{m+n}(x)$ for all $n \in \mathbb{N}$. In particular, by setting $n = \bar{t} - m > 0$, we would have that $T^{t+n}(x) = T^{m+n}(x) = T^{\bar{t}}(x) = \bar{x}$. Note that $\hat{t} = t + n < m + n = \bar{t}$. Thus, this would imply that

$$T^{\hat{t}}(x) = \bar{x} \text{ and } \hat{t} < \bar{t},$$

a contradiction with the minimality of $\bar{t}$. By definition of $\bar{t}$, we can also conclude that $T^t(x) \neq \bar{x}$ for all $t < \bar{t}$. This implies that $\{T^t(x)\}_{t=1}^{\bar{t}-1}$ is contained in $U \backslash \{\bar{x}\}$. Since U contains at most n^n elements and the elements of $\{T^t(x)\}_{t=1}^{\bar{t}-1}$ are pairwise distinct, it follows that $\bar{t} - 1 \leq n^n - 1$, proving that $\{T^t(x)\}_{t \in \mathbb{N}}$ converges only if it becomes constant after at most n^n periods.

2. $\{T^t(x)\}_{t \in \mathbb{N}}$ does not converge. Define $\tilde{n} = n^n$. Recall that $\{T^t(x)\}_{t=1}^{\tilde{n}+1} \subseteq U$ where the latter set has cardinality at most $\tilde{n}$. This implies that there exist $\hat{m}, \hat{t} \leq \tilde{n} + 1$ such that $T^{\hat{m}}(x) = T^{\hat{t}}(x)$ and $\hat{m} \neq \hat{t}$. Without loss of generality, we assume that $\hat{m} > \hat{t}$. It follows that

$$T^{\hat{t}+n}(x) = T^n\left(T^{\hat{t}}(x)\right) = T^n\left(T^{\hat{m}}(x)\right) = T^{\hat{m}+n}(x) \qquad \forall n \in \mathbb{N}_0.$$

Define $p = \hat{m} - \hat{t} > 0$. Since $\hat{t} \geq 1$ and $\hat{m} \leq \tilde{n} + 1$, note that $\hat{m} - \hat{t} \leq \tilde{n}$ and $\hat{t} \leq \tilde{n}$. We have that $T^{\hat{t}+n}(x) = T^{\hat{t}+n+p}(x)$ for all $n \in \mathbb{N}_0$, proving that $T^t(x) = T^{t+p}(x)$ for all $t \geq \hat{t}$.

Points 1 and 2 prove the first part of the statement as well as the "only if" of the second part. The "if" part is trivial. ∎

Proof of Proposition 7. Before starting, define $B_{\text{inj}} = \{x \in \text{int } B : x_l \neq x_k \text{ if } l \neq k\}$. It is routine to check that B_{inj} is dense in $\text{int } B$, and so, in B. Since T is continuous and B_{inj} is dense, if $T_i(x) = x_j$ for all $x \in B_{\text{inj}}$, then $T_i(x) = x_j$ for all $x \in B$. Thus, $T_i(x) \neq x_j$ for some $x \in B$ if and only if there exists $y \in B_{\text{inj}}$ such that $T_i(y) \neq y_j$.

(i) implies (ii). By assumption, there exists $\varepsilon_{ij} \in (0,1)$ such that for each $y \in B$ and for each $h > 0$ such that $y + he^j \in B$

$$T_i\left(y + he^j\right) - T_i(y) \geq \varepsilon_{ij} h.$$

By contradiction, assume that there exists $y \in B_{\text{inj}}$ such that $T_i(y) \neq y_j$. Since $B_{\text{inj}} \subseteq \text{int } B$, there exists $\bar{\varepsilon} > 0$ such that $z \in \mathbb{R}^n$ and $\|z - y\|_\infty < \bar{\varepsilon}$ yields that $z \in \text{int } B$. Since $T_i(x) = \int_N x d\nu_i$ for all $x \in B$ and ν_i is $\{0,1\}$-valued, it follows that there exists $m \in N$ such that $T_i(y) = y_m$. Since $T_i(y) \neq y_j$, this implies that $j \neq m$. Define $\delta = \min\left\{\min_{i',i'' \in N : i' \neq i''} |y_{i'} - y_{i''}|, \bar{\varepsilon}\right\}/2$. Since $y \in B_{\text{inj}}$ and $\bar{\varepsilon} > 0$, we have that $\delta > 0$ and $\delta < \bar{\varepsilon}$. Consider $h = \delta$. It is immediate to see that $y + he^j \in B$. At the same time, we have that $T_i\left(y + he^j\right) = \int\left(y + he^j\right) d\nu_i = y_m$, yielding that

$$0 = T_i\left(y + he^j\right) - T_i(y) \geq \varepsilon_{ij} h = \varepsilon_{ij}\delta > 0,$$

a contradiction.

(ii) implies (i). It is trivial.

Note that T is self-influential if and only if i strongly influences i for all $i \in N$. By the previous part, this happens if and only if $T_i(x) = x_i$ for all $i \in N$, that is, $T(x) = x$ for all $x \in B$. The second part of the statement follows in a similar fashion. ∎

Proof of Proposition 8. Before proving the statement, we introduce some terminology and some useful facts. Fix $i \in N$. Recall that

$$\nu_i(A) = f_i\left(\sum_{l \in A} w_{il}\right) \qquad \forall A \subseteq N.$$

Fix $j \in N$ and $A \in 2^N$ such that $j \notin A$. Since f_i is strictly increasing, we have that

$$\nu_i(A \cup \{j\}) - \nu_i(A) > 0 \Leftrightarrow f_i\left(\sum_{l \in A} w_{il} + w_{ij}\right) - f_i\left(\sum_{l \in A} w_{il}\right) > 0 \Leftrightarrow w_{ij} > 0. \quad (77)$$

Call Π the collection of all permutations of N, that is, the collection of all bijections $\pi : \{1, ..., n\} \to \{1, ..., n\}$. Given $\pi \in \Pi$, consider $p \in \Delta$ as defined in (22), to signal the dependence on π and i we here denote p by $p^{\pi,i}$. Define by D_i the set $\{p^{\pi,i} : \pi \in \Pi\}$. By definition of $p^{\pi,i}$, we have that

$$p^{\pi,i}_{\pi(j)} = \nu_i \left(\cup_{l=1}^j \{\pi(l)\} \right) - \nu_i \left(\cup_{l=1}^{j-1} \{\pi(l)\} \right) \qquad \forall j \in N, \forall \pi \in \Pi \qquad (78)$$

with the assumption that $\nu_i \left(\cup_{l=1}^0 \{\pi(l)\} \right) = 0$. By (77) and (78), we have that for each $\pi \in \Pi$ and for each $j \in N$

$$p^{\pi,i}_{\pi(j)} > 0 \iff w_{i\pi(j)} > 0. \qquad (79)$$

Finally, since each ν_i induces a functional from $\mathbb{R}^n$ to $\mathbb{R}$, we can assume without loss of generality that $B = \mathbb{R}^n$. By [41, Theorem 14 and Example 17] and (79), we have that the Clarke's differential of T_i at 0, $\partial T_i(0)$, coincides with the convex hull of D_i. This yields that

$$p_j > 0 \quad \forall p \in \partial T_i(0) = \text{co}(D_i) \iff w_{ij} > 0. \qquad (80)$$

(i) implies (ii). By counterpositive, we prove that if $w_{ij} \not> 0$, then j does not strongly influence i. Since $w_{ij} \geq 0$ and $w_{ij} \not> 0$, we have that $w_{ij} = 0$. Fix $x = 0$. This implies that

$$T_i \left(x + he^j \right) - T_i(x) = T_i \left(he^j \right) = hT_i \left(e^j \right) = hf_i(w_{ij}) = 0 \quad \forall h > 0,$$

proving that j cannot strongly influence i.

(ii) implies (i). Define $\varepsilon_{ij} = \min_{\pi \in \Pi} p^{\pi,i}_j / 2$. By (79) and since $w_{ij} > 0$, we have that $\varepsilon_{ij} \in (0,1)$. Consider $x \in B$ and $h > 0$. Define $y = x + he^j$. By Lebourg's Mean Value Theorem and (80) and since T_i is a Choquet integral, there exist $\gamma \in (0,1)$ and $p \in \partial T_i(\gamma y + (1 - \gamma) x) \subseteq \partial T_i(0)$ such that

$$T_i \left(x + he^j \right) - T_i(x) = T_i(y) - T_i(x) = \sum_{l=1}^n p_l (y_l - x_l) = p_j h \geq \varepsilon_{ij} h,$$

proving the implication.

Finally, by the previous part of the proof and since the ij-th entry of $A(T)$ is 1 if and only if j strongly influences i if and only if $w_{ij} > 0$ if and only if the ij-th entry of $A(W)$ is 1, the final part of the statement follows. ∎

Proof of Proposition 9. Consider $x \in B$. Let $\pi \in \Pi$ be such that $x \in B_\pi$. We first show that if $z \in B_\pi$, then $T(z) \in B_\pi$. Since T_i is a Choquet average, it is well known that for each $i \in N$

$$T_i(z) = \int_N z d\nu_i = \int_0^\infty \nu_i (\{j \in N : z_j \geq t\}) dt + \int_{-\infty}^0 [\nu_i (\{j \in N : z_j \geq t\}) - 1] dt. \qquad (81)$$

Since T is assortative, we have that

$$i' \geq i \implies \nu_{\pi(i)}\left(\{j \in N : z_j \geq t\}\right) \geq \nu_{\pi(i')}\left(\{j \in N : z_j \geq t\}\right) \qquad \forall t \in \mathbb{R}.$$

By (81), this implies that

$$T_{\pi(1)}(z) \geq T_{\pi(2)}(z) \ldots \geq T_{\pi(n)}(z),$$

proving that $T(z) \in B_\pi$. By induction, this implies that $T^t(z) \in B_\pi$ for all $z \in B_\pi$ and for all $t \in \mathbb{N}$. Next, for each $i \in N$, given ν_i, construct $p^{\pi,i}$ as in the proof of Proposition 8. Define W_π where the i-th row is the transpose of the column vector $p^{\pi,i}$. By (23), we have that $T(z) = W_\pi z$ for all $z \in B_\pi$. Since $x \in B_\pi$, $T(z) = W_\pi z$, and $T^t(z) \in B_\pi$ for all $z \in B_\pi$ and for all $t \in \mathbb{N}$, we can conclude that

$$T(x) = W_\pi x \text{ as well as } T^{t+1}(x) = T\left(T^t(x)\right) = W_\pi T^t(x) \qquad \forall t \in \mathbb{N}. \qquad (82)$$

By (82) and using induction, this proves the statement. ∎

C   Appendix: Vox populi, vox Dei

Proof of Proposition 10. Before starting, define $\hat{B} = \hat{I}^n$. We proceed by proving two intermediate steps. First, we prove that $\bar{T}_i(n)$ is an unbiased estimator for all $i \in N$ and for all $n \in \mathbb{N}$. Second, we show that (29) yields that $\bar{T}_i(n)$ is not extremely sensitive to changes coming from a single observation. Finally, by applying McDiarmid's inequality, these two steps will yield (27). Before starting, we make a few observations. By Theorem 2 and by points 3 and 5 of Lemma 9 and since $I = \mathbb{R}$, we have that $\bar{T}(n)$ is a well defined odd robust opinion aggregator for all $n \in \mathbb{N}$. By Assumption 3, we have that $\bar{T}_i(n) = \bar{T}_j(n)$ for all $i, j \in N$ and for all $n \in \mathbb{N}$. Since the random variables $\{X_i(n)\}_{i \in N, n \in \mathbb{N}}$ are uniformly bounded and measurable and $\bar{T}_i(n)$ is continuous for all $i \in N$ and for all $n \in \mathbb{N}$, it follows that $\omega \mapsto \bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega))$ is integrable for all $i \in N$ and for all $n \in \mathbb{N}$.

Step 1. For each $i \in N$ and for each $n \in \mathbb{N}$

$$\mathbb{E}\left(\bar{T}_i(n)(X_1(n), ..., X_n(n))\right) = \int_\Omega \bar{T}_i(n)(X_1(n), ..., X_n(n)) \, dP = \mu. \qquad (83)$$

Proof of the Step. Recall that $X_i(n) = \mu + \varepsilon_i(n)$ for all $i \in N$ and for all $n \in \mathbb{N}$ where $\{\varepsilon_i(n)\}_{i \in N, n \in \mathbb{N}}$ is a collection of uniformly bounded, symmetric, and independent random variables. Since $\bar{T}_i(n)$ is continuous for all $i \in N$ and for all $n \in \mathbb{N}$, it follows that $\omega \mapsto \bar{T}_i(n)(\varepsilon_1(n)(\omega), ..., \varepsilon_n(n)(\omega))$ is integrable for all $i \in N$ and for all $n \in \mathbb{N}$.

Since $\bar{T}(n)$ is odd for all $n \in \mathbb{N}$, this implies that for each $i \in N$ and for each $n \in \mathbb{N}$

$$\int_\Omega \bar{T}_i(n)\left(\varepsilon_1(n),...,\varepsilon_n(n)\right)dP = \int_\Omega \bar{T}_i(n)\left(-\varepsilon_1(n),...,-\varepsilon_n(n)\right)dP$$
$$= -\int_\Omega \bar{T}_i(n)\left(\varepsilon_1(n),...,\varepsilon_n(n)\right)dP.$$

It follows that for each $i \in N$ and for each $n \in \mathbb{N}$

$$2\int_\Omega \bar{T}_i(n)\left(\varepsilon_1(n),...,\varepsilon_n(n)\right)dP = 0.$$

Since $\bar{T}(n)$ is translation invariant, we can conclude that for each $i \in N$ and for each $n \in \mathbb{N}$

$$\mathbb{E}\left(\bar{T}_i(n)\left(X_1(n),...,X_n(n)\right)\right) = \int_\Omega \bar{T}_i(n)\left(X_1(n),...,X_n(n)\right)dP$$
$$= \int_\Omega \bar{T}_i(n)\left(\mu+\varepsilon_1(n),...,\mu+\varepsilon_n(n)\right)dP$$
$$= \mu + \int_\Omega \bar{T}_i(n)\left(\varepsilon_1(n),...,\varepsilon_n(n)\right)dP = \mu,$$

proving (83). □

Step 2. For each $i,j \in N$ and for each $n \in \mathbb{N}$

$$\sup_{\left\{(x,t)\in\hat{B}\times\mathbb{R}:x+te^j\in\hat{B}\right\}} \left|\bar{T}_i(n)\left(x+te^j\right)-\bar{T}_i(n)(x)\right| \leq \ell c(n)\,w_j(n).$$

Proof of the Step. Fix $i \in N$ and $n \in \mathbb{N}$. By Rademacher's Theorem and since $\bar{T}(n)$ is nonexpansive, this implies that $\bar{T}(n)$ is almost everywhere Frechet differentiable. Let $\mathcal{D}\left(\bar{T}(n)\right) \subseteq \hat{I}^n = \hat{B}$ be the subset of $\hat{B}$ where $\bar{T}(n)$ is Frechet differentiable. Clearly, $\bar{T}_i(n)$ is Frechet differentiable on $\mathcal{D}\left(\bar{T}(n)\right)$ and, in particular, Clarke differentiable. Since $\bar{T}_i(n)$ is monotone and translation invariant, note that $\nabla\bar{T}_i(n)(x) \in \Delta_n$ for all $x \in \mathcal{D}\left(\bar{T}(n)\right)$. Consider $\bar{x} \in \hat{B}$. Recall that Clarke's differential is the set (see, e.g., [25, Theorem 2.5.1]):

$$\partial\bar{T}_i(n)(\bar{x}) = \text{co}\left\{p \in \Delta_n : p = \lim_k \nabla\bar{T}_i(n)\left(x^k\right) \text{ s.t. } x^k \to \bar{x} \text{ and } x^k \in \mathcal{D}\left(\bar{T}(n)\right)\right\}. \tag{84}$$

By Definition 19 and (84) and since $\bar{T}_1(n) = \bar{T}_i(n)$, note that

$$0 \leq p_j \leq s_j(T(n)) \quad \forall p \in \partial\bar{T}_i(n)(x), \forall x \in \hat{B}, \forall j \in N. \tag{85}$$

Consider $j \in N$, $x \in \hat{B}$, and $t \in \mathbb{R}$ such that $x+te^j \in \hat{B}$. Define $y = x+te^j$. By Lebourg's Mean Value Theorem, we have that there exist $\lambda \in (0,1)$ and $\bar{p} \in \partial\bar{T}_i(n)(z)$ where $z = \lambda y + (1-\lambda)x \in \hat{B}$ such that

$$\bar{T}_i(n)\left(x+te^j\right)-\bar{T}_i(n)(x) = \bar{T}_i(n)(y)-\bar{T}_i(n)(x) = \sum_{l=1}^n \bar{p}_l\left(y_l - x_l\right).$$

It follows that $\left|\bar{T}_i(n)(x + te^j) - \bar{T}_i(n)(x)\right| = |\bar{p}_j(y_j - x_j)| = \bar{p}_j|y_j - x_j| \leq \ell\bar{p}_j$. By (85), this implies that

$$\left|\bar{T}_i(n)(x + te^j) - \bar{T}_i(n)(x)\right| \leq \ell\bar{p}_j \leq \ell s_j(T(n)) \leq \ell c(n) w_j(n).$$

Since x and t were arbitrarily chosen, it follows that

$$\sup_{\{(x,t)\in\hat{B}\times\mathbb{R}:x+te^j\in\hat{B}\}} \left|\bar{T}_i(n)(x + te^j) - \bar{T}_i(n)(x)\right| \leq \ell c(n) w_j(n).$$

Since i, n, and j were also arbitrarily chosen, the statement follows. $\qquad\square$

By McDiarmid's inequality as well as Steps 1 and 2, we can conclude that for each $\delta > 0$

$$P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu\right| \geq \delta\right\}\right)$$
$$= P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mathbb{E}\left(\bar{T}_i(n)(X_1(n), ..., X_n(n))\right)\right| \geq \delta\right\}\right)$$
$$\leq 2\exp\left(-\frac{2\delta^2}{\sum_{j=1}^n(\ell c(n) w_j(n))^2}\right) = 2\exp\left(-\frac{2\delta^2}{\ell^2 c(n)^2 \sum_{j=1}^n w_j(n)^2}\right)$$
$$\leq 2\exp\left(-\frac{2\delta^2}{\ell^2 c(n)^2 \max_{k\in N} w_k(n) \sum_{j=1}^n w_j(n)}\right)$$
$$= 2\exp\left(-\frac{2\delta^2}{\ell^2 c(n)^2 \max_{k\in N} w_k(n)}\right) \to 0 \text{ as } n \to \infty,$$

proving the statement. $\qquad\blacksquare$

Proof of Corollary 5. For each $i \in N$ and for each $n \in \mathbb{N}$ set

$$w_i(n) = \frac{1}{n} \text{ and } c(n) = n \max_{k\in N} s_k(T(n)).$$

It is immediate to see that $c(n) \in \mathbb{R}$ and $w(n) \in \Delta_n$ for all $n \in \mathbb{N}$. Moreover, we have that for each $n \in \mathbb{N}$

$$s_i(T(n)) \leq \max_{k\in N} s_k(T(n)) = c(n) w_i(n) \qquad \forall i \in N. \tag{86}$$

Since $\max_{k\in N} s_k(T(n)) = o\left(\frac{1}{\sqrt{n}}\right)$, we have that

$$c(n)^2 \max_{k\in N} w_k(n) = n\left(\max_{k\in N} s_k(T(n))\right)^2 \to 0 \text{ as } n \to \infty. \tag{87}$$

By (86) and (87), it follows that (29) holds. By Proposition 10 and since $\{T(n)\}_{n\in\mathbb{N}}$ is a sequence of odd robust opinion aggregators, this implies the statement. $\qquad\blacksquare$

Proof of Remark 4. Let $i \in N$ and $n \in \mathbb{N}$. By the last part of the proof of Proposition 10, we have that

$$P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu\right|^2 \geq \delta\right\}\right)$$
$$= P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu\right| \geq \sqrt{\delta}\right\}\right)$$
$$\leq 2\exp\left(-\frac{2\delta}{\ell^2 c^2 \max_{k \in N} w_k(n)}\right) \qquad \forall \delta > 0.$$

By [12, Equation 21.9], observe that

$$\text{Var}\left(\bar{T}_i(n)(X_1(n), ..., X_n(n))\right)$$
$$= \mathbb{E}\left(\left(\bar{T}_i(n)(X_1(n), ..., X_n(n)) - \mu\right)^2\right)$$
$$= \int_0^\infty P\left(\left\{\omega \in \Omega : \left(\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu\right)^2 \geq t\right\}\right) dt$$
$$= \int_0^{\ell^2} P\left(\left\{\omega \in \Omega : \left|\bar{T}_i(n)(X_1(n)(\omega), ..., X_n(n)(\omega)) - \mu\right|^2 \geq t\right\}\right) dt$$
$$\leq \int_0^{\ell^2} 2\exp\left(-\frac{2t}{\ell^2 c^2 \max_{k \in N} w_k(n)}\right) dt$$
$$= \ell^2 c^2 \max_{k \in N} w_k(n) \left[1 - \exp\left(-\frac{2}{c^2 \max_{k \in N} w_k(n)}\right)\right],$$

proving (32). ∎

Proof of Proposition 11. Fix $n \in \mathbb{N}$. Since $T(n)$ is a robust opinion aggregator, we have that $T(n)$ is Lipschitz continuous. By Rademacher's Theorem, this implies that $T(n)$ is almost everywhere Frechet differentiable and, in particular, Clarke differentiable. Let $\mathcal{D}(T(n)) \subseteq \hat{I}^n = \hat{B}$ be the subset of $\hat{B}$ where $T(n)$ is Frechet differentiable. Clearly, $T_h(n)$ is Frechet differentiable on the same set for all $h \in N$. Since $T_h(n)$ is monotone and translation invariant, note that $\nabla T_h(n)(x) \in \Delta_n$ for all $x \in \mathcal{D}(T(n))$. Consider $\bar{x} \in \hat{B}$. Recall that Clarke's differential is the set (see, e.g., [25, Theorem 2.5.1]):

$$\partial T_h(n)(\bar{x}) = \text{co}\left\{p \in \Delta_n : p = \lim_k \nabla T_h(n)(x^k) \text{ s.t. } x^k \to \bar{x} \text{ and } x^k \in \mathcal{D}(T(n))\right\}. \tag{88}$$

Similarly, we have that

$$\partial \bar{T}_1(n)(\bar{x}) = \text{co}\left\{p \in \Delta_n : p = \lim_k \nabla \bar{T}_1(n)(x^k) \text{ s.t. } x^k \to \bar{x} \text{ and } x^k \in \mathcal{D}(\bar{T}(n))\right\}.$$

By Theorem 2, recall that $\bar{T}(n) \circ T(n) = \bar{T}(n)$, yielding that $\bar{T}_1(n) \circ T(n) = \bar{T}_1(n)$. Fix $\bar{x} \in \hat{B}$. Define by $\Pi_{h=1}^n \partial T_h(n)(\bar{x})$ the collection of all $n \times n$ square matrices whose

h-th row is an element of $\partial T_h(n)(\bar{x})$. From the previous part of the proof, we have that $\Pi_{h=1}^n \partial T_h(n)(\bar{x}) \subseteq \mathcal{W}$. Define

$$\partial \bar{T}_1(n)(T(n)(\bar{x}))\,\Pi_{h=1}^n \partial T_h(n)(\bar{x})$$
$$= \left\{\tilde{w} \in \Delta_n : \exists p \in \partial \bar{T}_1(n)(T(n)(\bar{x})), \exists W \in \Pi_{h=1}^n \partial T_h(n)(\bar{x}) \text{ s.t. } p^{\mathrm{T}} W = \tilde{w}^{\mathrm{T}}\right\}.$$

In words, $\tilde{w} \in \partial \bar{T}_1(n)(T(n)(\bar{x}))\,\Pi_{h=1}^n \partial T_h(n)(\bar{x})$ only if it is a stochastic vector which is a convex linear combination of the rows of some matrix $W \in \Pi_{h=1}^n \partial T_h(n)(\bar{x})$. By the Chain Rule (see, e.g., [25, Theorem 2.6.6 and point e of Proposition 2.6.2]) and since $\bar{T}_1(n) = \bar{T}_1(n) \circ T(n)$, we have that

$$\partial \bar{T}_1(n)(\bar{x}) \subseteq \mathrm{co}\left\{\partial \bar{T}_1(n)(T(n)(\bar{x}))\,\Pi_{h=1}^n \partial T_h(n)(\bar{x})\right\} \tag{89}$$

By assumption, we have that for each $h \in N$

$$\sup_{x \in \mathcal{D}(T(n))} \frac{\partial T_h(n)}{\partial x_i}(x) \le c(n)\,w_i(n) \text{ and } c(n)^2 \max_{k \in N} w_k(n) \to 0 \text{ as } n \to \infty \tag{90}$$

By (88) and (90), we have that

$$0 \le p_i \le c(n)\,w_i(n) \quad \forall p \in \partial T_h(n)(\bar{x}), \forall i, h \in N.$$

By (89), we can conclude that

$$0 \le p_i \le c(n)\,w_i(n) \quad \forall p \in \partial \bar{T}_1(n)(\bar{x}), \forall i \in N.$$

Since $\bar{x}$ was arbitrarily chosen, this implies that

$$0 \le p_i \le c(n)\,w_i(n) \quad \forall p \in \partial \bar{T}_1(n)(x), \forall x \in \hat{B}, \forall i \in N.$$

Finally, observe that if $x \in \mathcal{D}\left(\bar{T}(n)\right)$, we have that $\nabla \bar{T}_1(n)(x) \in \partial \bar{T}_1(n)(x)$ and, in particular, $\frac{\partial \bar{T}_1(n)}{\partial x_i}(x) \le c(n)\,w_i(n)$ for all $i \in N$. This yields that

$$s_i(T(n)) = \sup_{x \in \mathcal{D}\left(\bar{T}(n)\right)} \frac{\partial \bar{T}_1(n)}{\partial x_i}(x) \le c(n)\,w_i(n),$$

proving the statement. ∎

References

[1] D. Acemoglu and A. Ozdaglar, Opinion dynamics and learning in social networks, *Dynamic Games and Applications*, 1, 3–49, 2011.

[2] V. Alatas, A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken, Network structure and the aggregation of information: Theory and evidence from Indonesia, *American Economic Review*, 106, 1663–1704, 2016.

[3] C. D. Aliprantis and K. C. Border, *Infinite Dimensional Analysis*, 3rd ed., Springer-Verlag, Berlin, 2006.

[4] S. Amghibech and C. Dellacherie, Une version non-lineaire, d'apres G. J. Olsder, du theoreme d'existence et d'unicite d'une mesure invariante pour une transition sur un espace fini, *Seminaire de Probabilites de Rouen*, 1994.

[5] J-P. Aubin and I. Ekeland, *Applied Nonlinear Analysis*, Dover, New York, 1984.

[6] Y. Babichenko and O. Tamuz, Graphical potential games, *Journal of Economic Theory*, 163, 889–899, 2016.

[7] J. B. Baillon, R. E. Bruck, and S. Reich, On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces, *Houston Journal of Mathematics*, 4, 1–9, 1978.

[8] A. Banerjee, E. Breza, A. G. Chandrasekhar, and M. Mobius, Naive learning with uninformed agents, *National Bureau of Economic Research*, 25497, 2019.

[9] A. Banerjee and D. Fudenberg, Word-of-mouth learning, *Games and Economic Behavior*, 46, 1–22, 2004.

[10] D. J. Benjamin, Errors in probabilistic reasoning and judgment biases, in *Handbook of Behavioral Economics: Applications and Foundations 1* (B. D. Bernheim, S. DellaVigna, and D. Laibson, eds.), North-Holland, Amsterdam, 2019.

[11] V. Berinde, *Iterative Approximation of Fixed Points*, 2nd ed., Springer-Verlag, Berlin, 2007.

[12] P. Billingsley, *Probability and Measure*, 3rd ed., John Wiley & Sons, New York, 1995.

[13] U. Bolletta and P. Pin, Polarization when people choose their peers, mimeo, 2019.

[14] E. Breza, A. G. Chandrasekhar, B. Golub, and A. Parvathaneni, Networks in economic development, *Oxford Review of Economic Policy*, 35, 678–721, 2019.

[15] E. Breza, A. G. Chandrasekhar, and A. Tahbaz-Salehi, Seeing the forest for the trees? An investigation of network knowledge, *National Bureau of Economic Research*, 24359, 2018.

[16] F. E. Browder and W. V. Petryshyn, The solution by iteration of nonlinear functional equations in Banach spaces, *Bulletin of the American Mathematical Society*, 72, 571–575, 1966.

[17] R. E. Bruck, On the almost-convergence of iterates of a nonexpansive mapping in Hilbert space and the structure of the weak ω-limit set, *Israel Journal of Mathematics*, 29, 1–16, 1979.

[18] N. L. Carothers, *Real Analysis*, Cambridge University Press, Cambridge, 2000.

[19] S. Cerreia-Vioglio, R. Corrao, and G. Lanzani, Adaptation, coordination, and inertia in network games, mimeo, 2020.

[20] S. Cerreia-Vioglio, P. Ghirardato, F. Maccheroni, M. Marinacci, and M. Siniscalchi, Rational preferences under ambiguity, *Economic Theory*, 48, 341–375, 2011.

[21] S. Cerreia-Vioglio, F. Maccheroni, and M. Marinacci, A characterization of probabilities with full support and the Laplace method, *Journal of Optimization Theory and Applications*, 181, 470–478, 2019.

[22] S. Cerreia-Vioglio, F. Maccheroni, M. Marinacci, and A. Rustichini, Niveloids and their extensions: Risk measures on small domains, *Journal of Mathematical Analysis and Applications*, 413, 343–360, 2014.

[23] A. G. Chandrasekhar, H. Larreguy, and J. P. Xandri, Testing models of social learning on networks: Evidence from two experiments, *Econometrica*, 88, 1–32, 2020.

[24] S. Chatterjee and E. Seneta, Towards consensus: Some convergence theorems on repeated averaging, *Journal of Applied Probability*, 14, 89–97, 1977.

[25] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM, Philadelphia, 1990.

[26] M. H. DeGroot, Reaching a consensus, *Journal of the American Statistical Association*, 69, 118–121, 1974.

[27] P. M. DeMarzo, D. Vayanos, and J. Zwiebel, Persuasion bias, social influence, and unidimensional opinions, *Quarterly Journal of Economics*, 118, 909–968, 2003.

[28] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd ed., Addison-Wesley, New York, 1989.

[29] J. Dieudonne, *Foundations of Modern Analysis*, Academic Press, New York, 1960.

[30] M. Edelstein and R. C. O'Brien, Nonexpansive mappings, asymptotic regularity, and successive approximations, *Journal of the London Mathematical Society*, 17, 547–554, 1978.

[31] D. M. Frankel, S. Morris, and A. Pauzner, Equilibrium selection in global games with strategic complementarities, *Journal of Economic Theory*, 108, 1–44, 2003.

[32] M. Frick, R. Iijima, and Y. Ishii, Dispersed behavior and perceptions in assortative societies, Cowles Foundation Discussion Paper, 2128R, 2018.

[33] N. E. Friedkin and E. C. Johnsen, Social influence and opinions, *Journal of Mathematical Sociology*, 15, 193–205, 1990.

[34] N. E. Friedkin and E. C. Johnsen, Social influence networks and opinion change, *Advances in Group Processes*, 16, 1–29, 1999.

[35] D. Fudenberg and D. K. Levine, Conditional universal consistency, *Games and Economic Behavior*, 29, 104–130, 1999.

[36] F. Galton, Vox populi, *Nature*, 75, 450–451, 1907.

[37] A. Galeotti, B. Golub, and S. Goyal, Targeting interventions in networks, mimeo, 2019.

[38] A. Galeotti, S. Goyal, M. O. Jackson, F. Vega-Redondo, and L. Yariv, Network games, *The Review of Economic Studies*, 77, 218–244, 2010.

[39] S. Galperti and J. Perego, Belief meddling in social networks: An information-design approach, mimeo, 2019.

[40] S. Gaubert and J. Gunawardena, The Perron-Frobenius Theorem for homogeneous, monotone functions, *Transactions of the American Mathematical Society*, 356, 4931–4950, 2004.

[41] P. Ghirardato, F. Maccheroni, and M. Marinacci, Differentiating ambiguity and ambiguity attitude, *Journal of Economic Theory*, 118, 133–173, 2004.

[42] B. Golub and M. O. Jackson, Naïve learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics*, 2, 112–149, 2010.

[43] B. Golub and M. O. Jackson, How homophily affects the speed of learning and best-response dynamics, *Quarterly Journal of Economics*, 127, 1287–1338, 2012.

[44] B. Golub and S. Morris, Expectations, networks, and conventions, mimeo, 2017.

[45] B. Golub and E. Sadler, Learning in social networks, in *The Oxford Handbook of the Economics of Networks* (Y. Bramoullé, A. Galeotti, and B. Rogers, eds.), Oxford University Press, Oxford, 2016.

[46] M. J. Granovetter, The strength of weak ties, *American Journal of Sociology*, 78, 1360–1380, 1973.

[47] R. A. Horn and C. R. Johnson, Matrix Analysis, 2nd ed., Cambridge University Press, Cambridge, 2013.

[48] P. J. Huber, Robust estimation of a location parameter, *Annals of Mathematical Statistics*, 35, 73–101, 1964.

[49] D. Isaacson and R. Madsen, Positive columns for stochastic matrices, *Journal of Applied Probability*, 11, 829–835, 1974.

[50] S. Ishikawa, Fixed points and iteration of a nonexpansive mapping in a Banach space, *Proceedings of the American Mathematical Society*, 59, 65–71, 1976.

[51] M. O. Jackson, *Social and Economic Networks*, Princeton University Press, Princeton, 2008.

[52] A. Jadbabaie, P. Molavi, A. Sandroni, and A. Tahbaz-Salehi, Non-Bayesian social learning, *Games and Economic Behavior*, 76, 210–225, 2012.

[53] D. Kempe, J. Kleinberg, and É. Tardos, Maximizing the spread of influence through a social network, *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 137–146, 2003.

[54] J. Korevaar, *Tauberian Theory: A Century of Developments*, Springer-Verlag, Berlin, 2004.

[55] S. Kotz, T. J. Kozubowski, and K. Podgorski, *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*, Springer, New York, 2001.

[56] U. Krause, *Positive Dynamical Systems in Discrete Time: Theory, Models, and Applications*, de Gruyter, Berlin, 2015.

[57] U. Krengel, *Ergodic Theorems*, de Gruyter, Berlin, 1985.

[58] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed., Springer-Verlag, New York, 1998.

[59] A. Liang and X. Mu, Complementary information and learning traps, *Quarterly Journal of Economics*, 135, 389–448, 2020.

[60] C. G. Lord, L. Ross, and M. R. Lepper, Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence, *Journal of Personality and Social Psychology*, 37, 2098–2109, 1979.

[61] G. G. Lorentz, A contribution to the theory of divergent sequences, *Acta Mathematica*, 80, 167–190, 1948.

[62] W. R. Mann, Mean value methods in iteration, *Proceedings of the American Mathematical Society*, 4, 506–510, 1953.

[63] F. Maccheroni, M. Marinacci, and A. Rustichini, Ambiguity aversion, robustness, and the variational representation of preferences, 74, *Econometrica*, 1447–1498, 2006.

[64] M. Marinacci and L. Montrucchio, Ultramodular functions, *Mathematics of Operations Research*, 30, 311–332, 2005.

[65] W. Mei, F. Bullo, G. Chen, and F. Dorfler, Occam's razor in opinion dynamics: The weighted-median influence process, mimeo, 2019.

[66] P. R. Milgrom, Good news and bad news: Representation theorems and applications, *The Bell Journal of Economics*, 12, 380–391, 1981.

[67] P. Milgrom and C. Shannon, Monotone comparative statics, *Econometrica*, 62, 157–180, 1994.

[68] M. Mobius and T. Rosenblat, Social learning in economics, *Annual Review of Economics*, 6, 827–847, 2014.

[69] P. Molavi, A. Tahbaz-Salehi, and A. Jadbabaie, A theory of non-Bayesian social learning, *Econometrica*, 86, 445–490, 2018.

[70] E. Mossel, N. Olsman, and O. Tamuz, Efficient Bayesian learning in social networks with Gaussian estimators, *Proceedings of the 54th Annual Allerton Conference on Communication, Control, and Computing*, 425–432, 2016.

[71] M. Muller-Frank, Robust non-Bayesian learning, mimeo, 2017.

[72] S. W. Neufeld, A diameter bound on the exponent of a primitive directed graph, *Linear Algebra and its Applications*, 245, 27–47, 1996.

[73] R. D. Nussbaum, Convexity and log convexity for the spectral radius, *Linear Algebra and its Applications*, 73, 59–122, 1986.

[74] D. Prelec, The probability weighting function, *Econometrica*, 66, 497–527, 1998.

[75] A. W. Roberts and D. E. Varberg, *Convex Functions*, Academic Press, New York, 1973.

[76] C. Robinson, *Dynamical Systems: Stability, Symbolic Dynamics, and Chaos*, CRC Press, Boca Raton, 1995.

[77] M. Rosenblatt, *Markov Processes. Structure and Asymptotic Behavior*, Springer-Verlag, Berlin, 1971.

[78] D. Schmeidler, Integral representation without additivity, *Proceedings of the American Mathematical Society*, 97, 255–261, 1986.

[79] D. Schmeidler, Subjective probability and expected utility without additivity, *Econometrica*, 57, 571–587, 1989.

[80] E. Seneta, *Non-negative Matrices and Markov Chains*, 2nd ed., Springer, New York, 1981.

[81] D. W. Stroock, *An Introduction to Markov Processes*, 2nd ed., Springer-Verlag, Berlin, 2014.

[82] D. van Dolder and M. J. van den Assem, The wisdom of the inner crowd in three large natural experiments, *Nature Human Behavior*, 2, 21–26, 2018.

[83] M. West, On scale mixtures of normal distributions, *Biometrika*, 74, 646–648, 1987.

[84] L. Yariv, I'll see it when I believe it - A simple model of cognitive consistency, mimeo, 2005.

D   Online Appendix

D.1   Additional statistical procedures

We saw that robust estimation provides a tight foundation for our robust opinion aggregators. At the same time, robust estimation is only one of several statistical procedures that agents might use to estimate the underlying parameter μ, particularly when in a network. In what follows, we discuss two other common estimation procedures and show that they still lead to robust opinion aggregators. Before doing so, we make an assumption which we will keep throughout this section.

Assumption Each agent believes errors $\varepsilon = (\varepsilon_i)_{i=1}^n$ are independently distributed across agents.

Maximum likelihood estimation   The approach followed by Huber [48], that we combine to a network structure (N, A), is completely non-parametric, in the sense that Huber did not postulate any functional form for the probability distribution of the error terms. An alternative approach for this network framework is the one proposed by DeMarzo, Vayanos, and Zwiebel [27]. They consider agents who may entertain incorrect point beliefs about the errors' distribution. Each agent $i \in N$ believes that the errors across the population are independently distributed according to a pdf f_{ij}, which is normal with 0 mean and variance $\sigma_{ij}^2 > 0$ for all $j \in N$. Thus, agent i believes that the errors' vector ε in (1) is distributed according to $N(0, \Sigma_i)$, where Σ_i is a diagonal matrix with diagonal entries $(\sigma_{ij}^2)_{j=1}^n$. According to [27], in period 0 there is a signals' realization $(x_j^0)_{j=1}^n$, but each agent i only observes the signals' realizations of her neighbors: $(x_j^0)_{j \in N_i}$.[63] In period 1 agent i performs a maximum likelihood estimation of μ obtaining

$$x_i^1 = T_i(x^0) = \sum_{j=1}^n w_{ij} x_j^0 = \sum_{j \in N_i} w_{ij} x_j^0 \qquad \forall i \in N$$

where

$$w_{ij} = a_{ij} \frac{1/\sigma_{ij}^2}{\sum_{l \in N_i} 1/\sigma_{il}^2} \qquad \forall i, j \in N.$$

In other words, as well known, in this case the maximum likelihood estimator is a weighted average of the realizations where the weight given by agent i to j is inversely proportional to the variance of j's error, provided j is in the neighborhood of i, and it is 0 otherwise. In the following periods $t \in \mathbb{N}$, in the most basic version of the model, agents naively keep updating their opinions by combining their neighbors' estimates with the weights just defined, that is, $x_i^t = T_i(x^{t-1})$ (see Section D.2 below).

---

[63] Clearly, we assume that $N_i \neq \emptyset$ for all $i \in N$.

In this way, DeMarzo, Vayanos, and Zwiebel provide a foundation for the linear De-Groot's model based on optimal information acquisition with a behavioral component: persuasion bias as discussed above. At the same time, their approach relies heavily on the normality of errors which in turn yields the linearity of the opinion aggregator T.

In what follows, we maintain the same maximum likelihood estimation procedure, but we relax the normality assumption. This naturally leads us again toward robust opinion aggregators.

Lemma 12 *If for each $i, j \in N$ the marginal pdf $f_{ij} : \mathbb{R} \to \mathbb{R}_{++}$ is log-concave with a unique local maximum in 0, then the agents' maximum likelihood estimators admit a selection which is a robust opinion aggregator.*[64]

Proof. Each agent $i \in N$, given $x \in B$, solves

$$\max_{c \in \mathbb{R}} \prod_{j \in N_i} f_{ij} (x_j - c).$$

Since the set of maximizers is not affected by a strictly increasing transformation of the objective function and $f_{ij} > 0$ for all $i, j \in N$, this problem is equivalent to

$$\max_{c \in \mathbb{R}} \sum_{j \in N_i} \ln f_{ij} (x_j - c) = -\min_{c \in \mathbb{R}} \left(-\sum_{j \in N_i} \ln f_{ij} (x_j - c) \right). \tag{91}$$

For each $i \in N$ define $\phi_i : \mathbb{R}^n \to \mathbb{R}_+$ by $\phi_i (z) = \sum_{j=1}^n a_{ij} \rho_{ij} (z_j)$ for all $z \in \mathbb{R}^n$ where $\rho_{ij} = -\ln f_{ij} + \ln f_{ij} (0)$ for all $i, j \in N$. By the same arguments of (iii) implies (ii) of Proposition 2, it follows that $\phi = (\phi_i)_{i=1}^n \in \Phi_R^* \subseteq \Phi_R$. For each $i \in N$ define $\mathbf{T}_i^\phi : B \rightrightarrows I$ by

$$\mathbf{T}_i^\phi (x) = \mathrm{argmin}_{c \in \mathbb{R}} \phi_i (x - ce) \qquad \forall x \in B.$$

For each $i \in N$ and $x \in B$ the solutions of (91) coincide with the set $\mathbf{T}_i^\phi (x)$. Define $\mathbf{T}^\phi : B \rightrightarrows B$ by $\mathbf{T}^\phi (x) = \prod_{i=1}^n \mathbf{T}_i^\phi (x)$ for all $x \in B$. By Proposition 12, the statement follows. ∎

These properties are quite mild and satisfied by several parametric families: e.g., normal, logistic, Gumbel, and Laplace with mode 0. Also, some of the aggregators studied in the main text can be mapped into these families. For example, if the f_{ij}s are normal, then the resulting opinion aggregator T is linear; if they are Laplace, then each component T_i is a weighted median; if they are Gumbel with uniform scale parameter across the j indexes, then the aggregator is a quasi-arithmetic one as in (11).

---

[64] As the maximum likelihood estimator is the result of a maximization problem, a priori, the solution might fail to be unique. In applications, this is not a concern since typically the f_{ij} pdfs are assumed to be strictly log-concave.

For example, Laplacian errors arise when agents still believe that the errors of the others are normally distributed, but they are uncertain about their variance. In particular, they have a subjective belief that the variance of the signal of another agent is $\sigma_{ij} = \sqrt{2\nu_i}/\tau_{ij}$ where ν_i has an exponential distribution with parameter 1.[65] In other words, agent i believes that agent j error has the shape $\sigma_{ij}Z_i$ where Z_i is a standard normal. It is then well known that this random variable is distributed according to Laplace $\left(0, \frac{1}{\tau_{ij}}\right)$ (see, e.g., Kotz, Kozubowski, and Podgorski [55, Proposition 2.2.1]). This is just an example of a much more general fact: any distribution in the exponential power family is a mixture of normals (see West [83]). This allows us to view some robust opinion aggregators as optimal behavior in face of various forms of uncertainty about the precision of the signals of the other agents.[66]

Bayesian estimation   The updating rule proposed by DeMarzo, Vayanos, and Zwiebel has also been rationalized as a Bayesian procedure with a diffuse prior (see [27, Footnote 17] and [45]). Formally, we keep all the assumptions of above and only change the estimation procedure from maximum likelihood estimation to a Bayesian one, that is, agents have a uniform improper prior Λ over the entire real line. With this procedure, at period 1, agents update their beliefs about μ in the following way:

$$\Lambda\left(\mu \mid (x_j)_{j \in N_i}\right) = \frac{\prod_{j \in N_i} f_{ij}(x_j - \mu)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(x_j - \mu') \, d\Lambda(\mu')}.$$

The posterior expectation of each agent i, given x, is defined by

$$T_i(x) = \frac{\int_{\mathbb{R}} \mu \prod_{j \in N_i} f_{ij}(x_j - \mu) \, d\Lambda(\mu)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(x_j - \mu') \, d\Lambda(\mu')}.$$

Next result, which is an extension of Milgrom [66, Proposition 1], shows that also this procedure yields a robust opinion aggregator.

Lemma 13 *If for each $i, j \in N$ the marginal pdf $f_{ij} : \mathbb{R} \to \mathbb{R}_{++}$ is log-concave and symmetric around 0, then the collection of agents' posterior expectations is a robust opinion aggregator.*

---

[65] Clearly, $\tau_{ij} > 0$ for all $i, j \in N$.

[66] Observe that the exponential power densities with power parameter equal or larger than 1 satisfy the assumptions of Lemma 12.

Proof. Fix $i \in N$. Since each $f_{ij} : \mathbb{R} \to \mathbb{R}_{++}$ is log-concave and symmetric around 0, it follows that $T_i(x)$ is well defined for all $x \in B$. We have that for each $k \in \mathbb{R}$

$$
\begin{aligned}
T_i(ke) &= \int_{\mathbb{R}} \mu \frac{\prod_{j \in N_i} f_{ij}(k - \mu)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(k - \mu')\, d\Lambda(\mu')}\, d\Lambda(\mu) \\
&= \int_{\mathbb{R}} \mu \frac{\prod_{j \in N_i} f_{ij}(\mu - k)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(\mu' - k)\, d\Lambda(\mu')}\, d\Lambda(\mu) \\
&= \int_{\mathbb{R}} (s + k) \frac{\prod_{j \in N_i} f_{ij}(s)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s')\, d\Lambda(s')}\, d\Lambda(s) \\
&= \int_{\mathbb{R}} s \frac{\prod_{j \in N_i} f_{ij}(s)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s')\, d\Lambda(s')}\, d\Lambda(s) + k \frac{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s)\, d\Lambda(s)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s')\, d\Lambda(s')} \\
&= k
\end{aligned}
$$

where the first equality follows by definition, the second and the fifth by the symmetry of the f_{ij}s, and the third by the Change of Variable Theorem. Next, consider $x \in B$ and $k \in \mathbb{R}$. By similar arguments, we have that

$$
\begin{aligned}
T_i(x + ke) &= \int_{\mathbb{R}} \mu \frac{\prod_{j \in N_i} f_{ij}(x_j + k - \mu)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(x_j + k - \mu')\, d\Lambda(\mu')}\, d\Lambda(\mu) \\
&= \int_{\mathbb{R}} \mu \frac{\prod_{j \in N_i} f_{ij}(\mu - k - x_j)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(\mu' - k - x_j)\, d\Lambda(\mu')}\, d\Lambda(\mu) \\
&= \int_{\mathbb{R}} (s + k) \frac{\prod_{j \in N_i} f_{ij}(s - x_j)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s' - x_j)\, d\Lambda(s')}\, d\Lambda(s) \\
&= \int_{\mathbb{R}} s \frac{\prod_{j \in N_i} f_{ij}(s - x_j)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s' - x_j)\, d\Lambda(s')}\, d\Lambda(s) + k \frac{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s - x_j)\, d\Lambda(s)}{\int_{\mathbb{R}} \prod_{j \in N_i} f_{ij}(s' - x_j)\, d\Lambda(s')} \\
&= T_i(x) + k.
\end{aligned}
$$

By [66, Proposition 1] and since the f_{ij}s are log-concave, we have that T_i is monotone. Since i was arbitrarily chosen, the statement follows. ∎

D.2 Alternative updating rules

In DeGroot's linear model, given x^0, the updates' dynamics are of the type

$$
x^t = T\left(x^{t-1}\right) \qquad \forall t \in \mathbb{N}
$$

where T is linear. Insofar, what we proposed was the study of the same type of dynamics where the aggregator T was only assumed to be robust, thus not necessarily linear. At the same time, despite keeping the assumption that T is linear, other types of opinion evolution have been studied in the literature (see, e.g., Jackson [51, Chapter

8]). In this section, we focus on two particular examples: the one of DeMarzo, Vayanos, and Zwiebel [27] as well as the one of Friedkin and Johnsen [33] and [34].

We start by considering the procedure of [27]. In this case, DeMarzo, Vayanos, and Zwiebel have agents revise an opinion x at each round with a linear operator T, but they also allow agents to vary the weight they give to their own beliefs. This results in the following revision dynamic:

$$x^t = T_t \left(x^{t-1} \right) \text{ and } T_t = (1 - \lambda_t) I + \lambda_t T \qquad \forall t \in \mathbb{N}. \tag{92}$$

They further assume that $\{\lambda_t\}_{t \in \mathbb{N}} \subseteq (0, 1]$ and $\sum_{t=1}^{\infty} \lambda_t = \infty$. Moreover, in their case, T is assumed to be linear and self-influential. The above condition on the weights λ_t intuitively captures the idea that agents cannot get fixed on their own opinion too quickly. By definition, since T is linear, there exists a stochastic matrix $W \in \mathcal{W}$ such that $T(x) = Wx$ for all $x \in B$. Theorem 1 of [27] shows that if W is irreducible, that is T is strongly connected, then $\{x^t\}_{t \in \mathbb{N}}$ converges and to an equilibrium point $\bar{x}$ of T. It turns out that $\bar{x}$ is also a consensus opinion. In what follows, we generalize this result in two directions. First, we show that if T is robust and self-influential, then the sequence of updates defined as in (92) still converges and to an equilibrium of T. Second, we can also offer a version where T is not necessarily assumed to be self-influential. This relaxation comes at the cost of requiring λ_t to be bounded away from 1, that is, $\lambda_t \leq b < 1$ for all $t \in \mathbb{N}$ and some $b \in (0, 1)$. Intuitively, this means that agents, at each round, are stuck on their own opinion for at least a factor of $1 - b$, which can be small, but must also be strictly positive.

Proposition 15 *Let T be a robust opinion aggregator, $x^0 \in B$, and $\{\lambda_t\}_{t \in \mathbb{N}} \subseteq (0, 1]$ such that $\sum_{t=1}^{\infty} \lambda_t = \infty$. The following statements are true:*

1. *If T is self-influential and $\{x^t\}_{t \in \mathbb{N}}$ is defined as in (92), then $\lim_t x^t$ exists and it is an equilibrium of T.*

2. *If there exists $b \in (0, 1)$ such that $\lambda_t \leq b$ for all $t \in \mathbb{N}$ and $\{x^t\}_{t \in \mathbb{N}}$ is defined as in (92), then $\lim_t x^t$ exists and it is an equilibrium of T.*

Proof. We first prove point 2, then point 1. By Lemma 8, recall that robust opinion aggregators are nonexpansive.

2. Given $x^0 \in B$, the sequence $\{x^t\}_{t \in \mathbb{N}}$, defined as in (92), is a specification of the Mann's iterates of T, using as weights $\{\lambda_t\}_{t \in \mathbb{N}}$. Consider $\tilde{B}$ as in point 4 of Lemma 6. Note that $T\left(\tilde{B}\right) \subseteq \tilde{B}$ and $x^t \in \tilde{B}$ for all $t \in \mathbb{N}_0$. By Ishikawa [50, Theorem 1], we have that $\{x^t\}_{t \in \mathbb{N}}$ converges and its limit is a fixed point of T.

1. By Proposition 13 and since T is self-influential, there exists $\varepsilon \in (0,1)$ such that $T = \varepsilon I + (1 - \varepsilon) S$ where S is a robust opinion aggregator. Note that given $\lambda \in (0,1]$ we have that

$$T_\lambda \overset{\text{def}}{=} (1 - \lambda) I + \lambda T = (1 - \lambda) I + \lambda \varepsilon I + (1 - \varepsilon) \lambda S = (1 - \gamma) I + \gamma S \overset{\text{def}}{=} S_\gamma$$

where $\gamma = (1 - \varepsilon) \lambda \leq (1 - \varepsilon)$. Given $x^0 \in B$, the sequence $\{x^t\}_{t\in\mathbb{N}}$, defined as in (92), can be rewritten as

$$x^t = S_t \left(x^{t-1} \right) \text{ and } S_t = (1 - \gamma_t) I + \gamma_t S \qquad \forall t \in \mathbb{N}$$

where $\gamma_t = (1 - \varepsilon) \lambda_t$ for all $t \in \mathbb{N}$. We have that $0 < \gamma_t \leq (1 - \varepsilon) < 1$ for all $t \in \mathbb{N}$ as well as $\sum_{t=1}^\infty \gamma_t = \infty$. By point 2, we can conclude that $\{x^t\}_{t\in\mathbb{N}}$ converges to a fixed point of S. It is immediate to see that the fixed points of T and S coincide, proving the statement. ∎

Remark 8 On the one hand, the updating process that we studied in Section 5 generates a sequence of updates $\{T^t (x^0)\}_{t\in\mathbb{N}}$. If we define $x^t = T^t (x^0)$ for all $t \in \mathbb{N}$, then the sequence of updates is such that

$$x^t = T \left(x^{t-1} \right) \qquad \forall t \in \mathbb{N}$$

which are also known in the mathematical literature as Picard's iterates. On the other hand, the sequence of updates generated by the updating procedure of DeMarzo, Vayanos, Zwiebel [27] are known as (a version of) Mann's iterates (see Mann [62] and Ishikawa [50]). Both types of iterates are very common in the literature of fixed points approximation which deals with the study of recursive procedures that yield existence and convergence to fixed points of selfmaps (see, e.g., Berinde [11]). ▲

Note that for both results consensus is the only possible limit, provided T is a consensus operator (cf. Definition 14 and Proposition 6). Note also that point 2 generalizes [27] in the linear case too. If each agent has a minimal stickiness to her own opinion, then convergence happens irrespective of the network structure. This fact is remarkable because convergence to a consensus could be obtained by having strong connectedness, but without self-influentiality or, more in general, aperiodicity.

We next consider the procedure of Friedkin and Johnsen [33] and [34] (see also Golub and Sadler [45, p. 16]), which is popular in sociology. In this case, agents are assumed to aggregate an opinion x at each round with a linear operator T (represented by a matrix W), but they are also allowed to hold onto their initial opinion:

$$x^t = \alpha T \left(x^{t-1} \right) + (1 - \alpha) x^0 \qquad \forall t \in \mathbb{N}. \tag{93}$$

Friedkin and Johnsen interpret α as a measure of agents' susceptibilities to personal influence. In other words, the t-th update of agent i is a mixture of her linear $t-1$-th update, x^{t-1}, and her initial opinion x_i^0 where the mixture weight is given by α. The convergence of the corresponding different updating process is derived under the assumption $\alpha \in [0,1)$ for all $i \in N$.[67]

In what follows, we study the Friedkin and Johnsen updating process in (93) when T is only assumed to be robust, but not necessarily linear. Next result shows that if T is robust, then the procedure in (93) still yields convergent updating dynamics.

Fact 1 *Let T be a robust opinion aggregator and $x^0 \in B$. If $\alpha \in (0,1)$ and $\{x^t\}_{t\in\mathbb{N}}$ is defined as in (93), then $\bar{x} = \lim_t x^t \in B$ exists and it is such that*

$$\bar{x} = \alpha T(\bar{x}) + (1-\alpha) x^0.$$

Proof. Fix $x^0 \in B$. Consider $\tilde{B}$ as in point 4 of Lemma 6. Observe that $T(\tilde{B}) \subseteq \tilde{B}$. Define also $R : \tilde{B} \to \tilde{B}$ to be such that $R(x) = \alpha T(x) + (1-\alpha) x^0$ for all $x \in \tilde{B}$. Since $\alpha \in (0,1)$, R is well defined. By Lemma 8 and since T is robust, we have that T is nonexpansive. Since $\alpha \in (0,1)$ and $x^0 \in \tilde{B}$, this immediately implies that R is a contraction. By definition of R, we have that $R(x^0) = x^1$. If $R^t(x^0) = x^t \in \tilde{B}$ when $t \in \mathbb{N}$, then we can conclude that $R^{t+1}(x^0) = R(R^t(x^0)) = R(x^t) = x^{t+1}$. By induction, it follows that $\{R^t(x^0)\}_{t\in\mathbb{N}} = \{x^t\}_{t\in\mathbb{N}} \subseteq \tilde{B}$. By Banach Contraction Principle and since R is a contraction and $\tilde{B}$ is compact, we have that $\bar{x} = \lim_t x^t = \lim_t R^t(x^0)$ exists and $\bar{x}$ is a fixed point of R, that is,

$$\bar{x} = R(\bar{x}) = \alpha T(\bar{x}) + (1-\alpha) x^0,$$

proving the statement. ∎

D.3 Comparison with Molavi, Tahbaz-Salehi, and Jadbabaie

In this section we show that for the issues analyzed in this paper, namely convergence and the wisdom of the crowd, when each agent observes a unique *initial* signal, the log-linear learning rule axiomatized in Molavi, Tahbaz-Salehi, and Jadbabaie [69] can be equivalently analyzed by means of a linear system. However, the equivalence with a linear system may be lost for a problem of learning with repeated signals like the one in [69].

Formally, Molavi, Tahbaz-Salehi, and Jadbabaie consider the case of agents having a full support belief $\mu \in \Delta(\Theta)$ where Θ is a finite set of possible states of the world.

---

[67]For example, in discussing convergence, they require α^{-1} not to be an eigenvalue of W, thus $\alpha \neq 1$.

In particular, there is a bijection between beliefs and the profile of likelihood ratios $\left(x \left(\theta, \hat{\theta} \right) \right)_{(\theta, \hat{\theta}) \in \Theta \times \Theta}$ with

$$x \left(\theta, \hat{\theta} \right) = \frac{\mu(\theta)}{\mu \left(\hat{\theta} \right)} \qquad \forall \left(\theta, \hat{\theta} \right) \in \Theta \times \Theta.$$

Their assumption of IIA allows to study the evolution of $x \left(\theta, \hat{\theta} \right) \in \mathbb{R}_{++}$ independently from the value of the other likelihood ratios. Therefore, we fix a particular $\left(\theta, \hat{\theta} \right)$ and denote the likelihood ratio obtained from the belief of agent i at time t as

$$x_i^t = \frac{\mu_i^t(\theta)}{\mu_i^t \left(\hat{\theta} \right)} \qquad \forall i \in N, \forall t \in \mathbb{N}.$$

When the agents do not observe any additional signal at period t equation (3) in [69] reads as

$$\ln x_i^t = \sum_{j \in N_i} a_{ij,t} \ln x_j^{t-1}$$

where $a_{ij,t} > 0$ for all $i, j \in N$ and for all $t \in \mathbb{N}$. We can explicitly write the law of motion of $\{x^t\}_{t \in \mathbb{N}}$ as

$$x_i^t = \exp \left(\sum_{j \in N_i} a_{ij,t} \ln x_j^{t-1} \right) \qquad \forall i \in N, \forall t \in \mathbb{N}. \tag{94}$$

For each $t \in \mathbb{N}$ define the operator $S_t : \mathbb{R}_{++}^n \to \mathbb{R}_{++}^n$ to be such that the i-th component is $x \mapsto \exp \left(\sum_{j \in N_i} a_{ij,t} \ln x_j \right)$. Thus, we can rewrite (94) as $x^t = S_t \left(x^{t-1} \right)$ for all $t \in \mathbb{N}$. Each operator S_t is *not* a robust opinion aggregator, since it does not satisfy translation invariance.

At the same time, we can analyze an equivalent system whose law of motion is described by time-varying linear aggregation. Toward this end, fix x^0 and let $\{x^t\}_{t \in \mathbb{N}}$ be recursively defined as in equation (94). Define A_t to be the matrix whose ij-th entry is $a_{ij,t}$. Next, define $\{y^t\}_{t \in \mathbb{N}_0}$ by

$$y_i^t = \ln x_i^t \qquad \forall i \in N, \forall t \in \mathbb{N}_0 \tag{95}$$

and note that

$$y^t = A_t y^{t-1} \qquad \forall t \in \mathbb{N}.$$

Note that, whenever each A_t is equal to the same stochastic matrix $W \in \mathcal{W}$, the iterates $\{y^t\}_{t \in \mathbb{N}}$ are described by the standard DeGroot's model with matrix W. Therefore, in this case, it is possible to appeal to the results in [42] to study the limit behavior of $\{y^t\}_{t \in \mathbb{N}}$. In general, whenever each $A_t \in \mathcal{W}$, one can rely on the more general results

for time-varying matrices (see, e.g., Seneta [80] and Krause [56]). Given the continuity of the transformation in (95), we have that

$$\lim_t x^t \text{ exists in } \mathbb{R}^n_{++} \iff \lim_t y^t \text{ exists in } \mathbb{R}^n.$$

The previous simple equivalence shows that $\lim_t y^t$ uniquely pins down $\lim_t x^t$.

D.4 Missing proofs

In this section, we confine all the proofs which involve routine arguments. They appear in the order in which the corresponding statements appear in the text.

Proof of Lemma 6. 1. Since T is robust, we have that $T_i : B \to \mathbb{R}$ is monotone and translation invariant for all $i \in N$ (cf. the definitions at the beginning of Appendix B). By [22, Theorem 4], T_i is a niveloid for all $i \in N$. By [22, Theorem 1], T_i admits an extension $S_i : \mathbb{R}^n \to \mathbb{R}$ which is a niveloid for all $i \in N$. By [22, Theorem 4], S_i is monotone and translation invariant for all $i \in N$. Define $S : \mathbb{R}^n \to \mathbb{R}^n$ to be such that the i-th component of $S(x)$ is $S_i(x)$ for all $i \in N$ and for all $x \in B$. It is immediate to see that S is monotone and translation invariant. Fix $k' \in I$. Since S is translation invariant and T is normalized, it follows that for all $k \in \mathbb{R}$

$$S(ke) = S(k'e + (k - k')e) = S(k'e) + (k - k')e$$
$$= T(k'e) + (k - k')e = k'e + (k - k')e = ke,$$

proving that S is normalized and, in particular, that S is robust.

2. We only prove uniqueness of the extension. The existence of a robust extension $S : \mathbb{R}^n \to \mathbb{R}^n$ which is constant affine follows from routine arguments.[68] Let $R : \mathbb{R}^n \to \mathbb{R}^n$ be a robust and constant affine opinion aggregator such that $R(x) = T(x)$ for all $x \in B$. Let $r \in \text{int } I$. It follows that $re \in \text{int } B$. Consider $x \in \mathbb{R}^n$ and $\alpha \in (0, 1)$ such that $\alpha x + (1 - \alpha) re \in B$. We have that

$$\alpha R(x) + (1 - \alpha) re = R(\alpha x + (1 - \alpha) re) = T(\alpha x + (1 - \alpha) re)$$
$$= S(\alpha x + (1 - \alpha) re) = \alpha S(x) + (1 - \alpha) re,$$

proving that $R(x) = S(x)$. Since x was arbitrarily chosen, it follows that $S = R$.

3. By induction, if T is normalized and monotone, then T^t is normalized and monotone for all $t \in \mathbb{N}$. Consider $x \in B$ and $t \in \mathbb{N}$. Define $k_\star = \min_{i \in N} x_i$ and $k^\star = \max_{i \in N} x_i$. Note that $\|x\|_\infty = \max\{|k_\star|, |k^\star|\}$, $k_\star, k^\star \in I$, and $k_\star e \leq x \leq k^\star e$. Since T^t is normalized and monotone, we have that

$$k_\star e = T^t(k_\star e) \leq T^t(x) \leq T^t(k^\star e) = k^\star e,$$

---

[68] A proof is available upon request.

yielding that $|T^t(x)| \leq \max\{|k_\star|, |k^\star|\} e$ and $\|T^t(x)\|_\infty \leq \|x\|_\infty$. Since t and x were arbitrarily chosen, the statement follows.

4. Let $x \in B$. Define $k_\star = \min_{i \in N} x_i$ and $k^\star = \max_{i \in N} x_i$. We have two cases:

a. $k_\star < k^\star$. Clearly, we have that $k_\star, k^\star \in I$. Note that $\tilde{I} = [k_\star, k^\star] \subseteq I$ is compact and with nonempty interior. Moreover, $x \in \tilde{I}^n = \tilde{B}$.

b. $k_\star = k^\star$. Since I has nonempty interior, there exists $\varepsilon > 0$ such that either $\tilde{I} = [k_\star, k_\star + \varepsilon] \subseteq I$ or $\tilde{I} = [k_\star - \varepsilon, k_\star] \subseteq I$. In all these cases, $\tilde{I}$ is compact and with nonempty interior. Moreover, $x \in \tilde{I}^n = \tilde{B}$.

Consider the restriction $\tilde{T} = T_{|\tilde{B}}$. Note that $T(\tilde{B}) \subseteq \tilde{B}$, yielding that $\tilde{T}$ is a robust opinion aggregator. By induction, we have that $\tilde{T}^t(x) = T^t(x)$ for all $t \in \mathbb{N}$ and for all $x \in \tilde{B}$. It follows that for each $x \in \tilde{B}$

$$\bar{T}(x) = \lim_\tau \frac{1}{\tau} \sum_{t=1}^\tau T^t(x) = \lim_\tau \frac{1}{\tau} \sum_{t=1}^\tau \tilde{T}^t(x) = \bar{\tilde{T}}(x) \qquad \forall x \in \tilde{B},$$

proving the point. ∎

Proof of Lemma 9. Let $x \in B$. Since T is a selfmap, we have that $\{T^t(x)\}_{t \in \mathbb{N}} \subseteq B$. Since B is convex, we have that

$$\frac{1}{\tau} \sum_{t=1}^\tau T^t(x) \in B \qquad \forall \tau \in \mathbb{N}.$$

Since x was arbitrarily chosen, this implies that $A_\tau : B \to B$, defined by $A_\tau(x) = \sum_{t=1}^\tau T^t(x)/\tau$ for all $x \in B$, is well defined for all $\tau \in \mathbb{N}$. Since B is closed, we have that $\bar{T}(x) = \lim_\tau A_\tau(x) = \lim_\tau \frac{1}{\tau} \sum_{t=1}^\tau T^t(x) \in B$ for all $x \in B$, proving that $\bar{T}$ is well defined. By the same computations contained in [3, Lemma 20.12], despite T being nonlinear, one can prove that

$$A_\tau(T(x)) = \frac{\tau+1}{\tau} A_{\tau+1}(x) - \frac{1}{\tau} T(x) \qquad \forall x \in B, \forall \tau \in \mathbb{N}.$$

This implies that

$$\bar{T}(T(x)) = \lim_\tau A_\tau(T(x)) = \lim_\tau \frac{\tau+1}{\tau} \lim_\tau A_{\tau+1}(x) - \lim_\tau \frac{1}{\tau} T(x) = \bar{T}(x) \qquad \forall x \in B,$$

proving that $\bar{T} \circ T = \bar{T}$.

1. By the same inductive argument contained in the proof of Lemma 8, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is nonexpansive. Since the convex linear combination of nonexpansive maps is nonexpansive, the map $A_\tau : B \to B$ is nonexpansive for all $\tau \in \mathbb{N}$. We can conclude that for each $x, y \in B$

$$\left\|\bar{T}(x) - \bar{T}(y)\right\|_\infty = \left\|\lim_\tau A_\tau(x) - \lim_\tau A_\tau(y)\right\|_\infty = \lim_\tau \|A_\tau(x) - A_\tau(y)\|_\infty \leq \|x - y\|_\infty,$$

proving that $\bar{T}$ is nonexpansive. Continuity of $\bar{T}$ trivially follows.

2. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is normalized and monotone. Since the convex linear combination of normalized and monotone operators is normalized and monotone, the map $A_\tau : B \to B$ is normalized and monotone for all $\tau \in \mathbb{N}$. We can conclude that $\bar{T}(ke) = \lim_\tau A_\tau(ke) = ke$ for all $k \in I$ as well as

$$x \geq y \implies \bar{T}(x) = \lim_\tau A_\tau(x) \geq \lim_\tau A_\tau(y) = \bar{T}(y),$$

proving that $\bar{T}$ is normalized and monotone.

3. Since T is robust, T is normalized, monotone, and translation invariant. By the previous point, $\bar{T}$ is normalized and monotone. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is translation invariant. Since the convex linear combination of translation invariant operators is translation invariant, the map $A_\tau : B \to B$ is translation invariant for all $\tau \in \mathbb{N}$. We can conclude that for each $x \in B$ and for each $k \in I$ such that $x + ke \in B$

$$\bar{T}(x + ke) = \lim_\tau A_\tau(x + ke) = \lim_\tau [A_\tau(x) + ke] = \bar{T}(x) + ke,$$

proving that $\bar{T}$ is translation invariant and, in particular, robust.

4. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is constant affine. Since the convex linear combination of constant affine maps is constant affine, the map $A_\tau : B \to B$ is constant affine for all $\tau \in \mathbb{N}$. We can conclude that for each $x \in B$, for each $k \in I$, and for each $\lambda \in [0, 1]$

$$\bar{T}(\lambda x + (1 - \lambda) ke) = \lim_\tau A_\tau(\lambda x + (1 - \lambda) ke)$$
$$= \lim_\tau [\lambda A_\tau(x) + (1 - \lambda) ke] = \lambda \bar{T}(x) + (1 - \lambda) ke,$$

proving that $\bar{T}$ is constant affine.

5. By induction, we have that for each $t \in \mathbb{N}$ the map $T^t : B \to B$ is odd. Since the convex linear combination of odd maps is odd, the map $A_\tau : B \to B$ is odd for all $\tau \in \mathbb{N}$. We can conclude that

$$\bar{T}(-x) = \lim_\tau A_\tau(-x) = \lim_\tau [-A_\tau(x)] = -\bar{T}(x) \quad \forall x \in B \text{ s.t. } -x \in B,$$

proving that $\bar{T}$ is odd. ∎

Remark 9 The proof of Lemma 10 (cf. Proposition 14) below is a tedious adaptation of the techniques contained in the proof of Edelstein and O'Brien [30, Lemma 1]. Their case is more general in terms of domain of T in that B can be any convex subset of a normed vector space. Their generality comes at the cost of having J equal to the identity operator which in our case would only cover Corollary 1. ▲

Proof of Lemma 10. We first offer two definitions and make two observations. Define the diameter of $\{T^t(x) : x \in A \text{ and } t \in \mathbb{N}_0\}$ by D.[69] Given $x \in B$, define $x^t = T^t(x)$ as well as $y^t = S(x^t)$ for all $t \in \mathbb{N}_0$. Since T is nonexpansive, recall that $\{\|x^t - x^{t-1}\|_\infty\}_{t \in \mathbb{N}}$ is a decreasing sequence for all $x \in B$.[70] Note that this implies that $\|T(x) - x\|_\infty \geq \|T^{t+1}(x) - T^t(x)\|_\infty$ for all $t \in \mathbb{N}_0$ and for all $x \in B$, yielding that $k > \delta$.

By contradiction, assume that $\{T^t(x) : x \in A \text{ and } t \in \mathbb{N}_0\}$ is bounded. This implies that $D < \infty$. Consider $M \in \mathbb{N}\setminus\{1\}$ and $P \in \mathbb{N}$ to be such that

$$M\delta > D + \delta + 1 \text{ and } \left\lfloor \frac{P}{M} \right\rfloor > \max\left\{1, \frac{k}{(1-\varepsilon)\varepsilon M}\right\}.[71]$$

By (62) and since $P \in \mathbb{N}$, there exists $x \in A$ such that

$$\left\|x^{P+1} - x^P\right\|_\infty = \left\|T^{P+1}(x) - T^P(x)\right\|_\infty \geq \delta.$$

Now, we enlist seven useful facts:

1. By (61) and since $\{\|x^t - x^{t-1}\|_\infty\}_{t \in \mathbb{N}}$ is a decreasing sequence, it follows that

$$k \geq \left\|x^{i+1} - x^i\right\|_\infty \geq \delta \qquad \forall i \in \{1, ..., P\}.$$

2. By definition of $\{y^t\}_{t \in \mathbb{N}_0}$ and since S is nonexpansive, we have that

$$\left\|y^t - y^{t-1}\right\|_\infty = \left\|S(x^t) - S(x^{t-1})\right\|_\infty \leq \left\|x^t - x^{t-1}\right\|_\infty \qquad \forall t \in \mathbb{N}.$$

3. By definition of $\{x^t\}_{t \in \mathbb{N}_0}$ and since $T = \varepsilon J + (1 - \varepsilon) S$, we have that $x^t = T(x^{t-1}) = \varepsilon J(x^{t-1}) + (1-\varepsilon)y^{t-1}$ for all $t \in \mathbb{N}$, that is,

$$y^{t-1} = \frac{1}{1-\varepsilon}x^t - \frac{\varepsilon}{1-\varepsilon}J(x^{t-1}) \qquad \forall t \in \mathbb{N}.$$

---

[69] Recall that the diameter of a subset $\hat{A}$ of B is the quantity

$$\sup\left\{\|x - y\|_\infty : x, y \in \hat{A}\right\}.$$

[70] For, we have that for each $t \in \mathbb{N}$ and for each $x \in B$

$$\left\|x^{t+1} - x^t\right\|_\infty = \left\|T^{t+1}(x) - T^t(x)\right\|_\infty$$
$$= \left\|T(T^t(x)) - T(T^{t-1}(x))\right\|_\infty \leq \left\|T^t(x) - T^{t-1}(x)\right\|_\infty \leq \left\|x^t - x^{t-1}\right\|_\infty.$$

[71] Once M is chosen, one could set $P = \left(\left\lfloor \max\left\{1, \frac{k}{(1-\varepsilon)\varepsilon M}\right\}\right\rfloor + 1\right)M$, so that

$$\left\lfloor \frac{P}{M} \right\rfloor = \left\lfloor \max\left\{1, \frac{k}{(1-\varepsilon)\varepsilon M}\right\}\right\rfloor + 1 > \max\left\{1, \frac{k}{(1-\varepsilon)\varepsilon M}\right\}.$$

By point 2, this yields that for each $t \in \mathbb{N}$

$$\left\| \frac{1}{1-\varepsilon} \left(x^{t+1} - x^t \right) - \frac{\varepsilon}{1-\varepsilon} \left(J\left(x^t \right) - J\left(x^{t-1} \right) \right) \right\|_\infty = \left\| y^t - y^{t-1} \right\|_\infty \leq \left\| x^t - x^{t-1} \right\|_\infty.$$

4. Let L be an integer in $\mathbb{N}$ such that

$$L > \frac{k}{(1-\varepsilon)\,\varepsilon^M}. \tag{96}$$

Define $b_m = \delta + m\,(1-\varepsilon)\,\varepsilon^M$ for all $m \in \{0, ..., L\}$. It follows that the collection of intervals $\{[b_m, b_{m+1}]\}_{m=0}^{L-1}$ contains L elements whose union is a superset of $[\delta, k]$.

5. Note that $\varepsilon^{M-1}\frac{1-\varepsilon^i}{\varepsilon^i} = \varepsilon^{M-i-1} - \varepsilon^{M-1} \leq \varepsilon^{M-i-1}$ for all $i \in \{1, M-1\}$. Since $\varepsilon \in (0,1)$, this implies that

$$(1-\varepsilon)\,\varepsilon^M \sum_{i=1}^{M-1} \frac{1-\varepsilon^i}{\varepsilon^i} \leq (1-\varepsilon)\,\varepsilon \sum_{i=1}^{M-1} \varepsilon^{M-i-1} = (1-\varepsilon)\,\varepsilon \sum_{i=0}^{M-2} \varepsilon^i$$

$$\leq (1-\varepsilon)\,\varepsilon \frac{1}{1-\varepsilon} \leq \varepsilon < 1.$$

6. Let $t \in \mathbb{N}$, $j \in N$, and $b, \kappa, c \geq 0$. If $x_j^{t+1} - x_j^t \geq b - c$ and $\left\| x^t - x^{t-1} \right\|_\infty \leq b + \kappa$, then (by point 3):[72]

$$\frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon} \left(x_{k_l}^t - x_{k_l}^{t-1} \right) = \frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon} \left(J_j\left(x^t \right) - J_j\left(x^{t-1} \right) \right) \leq b + \kappa$$

where l is such that $j \in N_l$. This yields that

$$x_{k_l}^t - x_{k_l}^{t-1} \geq b - \frac{c}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}\kappa. \tag{97}$$

7. Let $t \in \mathbb{N}$, $j \in N$, and $b, \kappa, c \geq 0$. If $x_j^t - x_j^{t+1} \geq b - c$ and $\left\| x^t - x^{t-1} \right\|_\infty \leq b + \kappa$,

---

[72] For, we have that

$$\frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon} \left(x_{k_l}^t - x_{k_l}^{t-1} \right) = \frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon} \left(J_j\left(x^t \right) - J_j\left(x^{t-1} \right) \right)$$

$$\leq \frac{1}{1-\varepsilon} \left(x_j^{t+1} - x_j^t \right) - \frac{\varepsilon}{1-\varepsilon} \left(J_j\left(x^t \right) - J_j\left(x^{t-1} \right) \right)$$

$$\leq \left\| \frac{1}{1-\varepsilon} \left(x^{t+1} - x^t \right) - \frac{\varepsilon}{1-\varepsilon} \left(J\left(x^t \right) - J\left(x^{t-1} \right) \right) \right\|_\infty$$

$$= \left\| y^t - y^{t-1} \right\|_\infty \leq \left\| x^t - x^{t-1} \right\|_\infty.$$

then (by point 3):[73]

$$\frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(x_{k_l}^{t-1} - x_{k_l}^t\right) = \frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(J_j\left(x^{t-1}\right) - J_j\left(x^t\right)\right) \le b + \kappa$$

where l is such that $j \in N_l$. This yields that

$$x_{k_l}^{t-1} - x_{k_l}^t \ge b - \frac{c}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}\kappa. \tag{98}$$

By definition of P, we have that $\lfloor P/M \rfloor$ satisfies (96). By point 4, there exists a collection of intervals $\{[b_m, b_{m+1}]\}_{m=0}^{\lfloor P/M \rfloor - 1}$ which covers $[\delta, k]$. By point 1, $[\delta, k]$ contains $\{\|x^{i+1} - x^i\|_\infty\}_{i=1}^P$. Since we have $\lfloor P/M \rfloor$ intervals and the first P elements (of the sequence $\{\|x^{t+1} - x^t\|_\infty\}_{t \in \mathbb{N}}$) belong to these intervals, we have that there exists one of them, $\hat{I} = [b_{\bar{m}}, b_{\bar{m}+1}]$, which contains at least M elements of $\{\|x^{i+1} - x^i\|_\infty\}_{i=1}^P$.[74] Since $\{\|x^t - x^{t-1}\|_\infty\}_{t \in \mathbb{N}}$ is decreasing, we have that there exists $K \in \mathbb{N}_0$ such that

$$\left\|x^{K+i+1} - x^{K+i}\right\|_\infty \in \hat{I} \qquad \forall i \in \{1, ..., M\}.$$

This implies that there exists $j \in \{1, ..., n\}$ such that $\left|x_j^{K+M+1} - x_j^{K+M}\right| \ge b_{\bar{m}}$ and $\left\|x^{K+M} - x^{K+M-1}\right\|_\infty \le b_{\bar{m}+1} = b_{\bar{m}} + (1-\varepsilon)\varepsilon^M$. We have two cases:

a. $x_j^{K+M+1} - x_j^{K+M} \ge b_{\bar{m}}$. Set $b = b_{\bar{m}}$, $c = 0$, and $\kappa = (1-\varepsilon)\varepsilon^M$. By (97), we can conclude that

$$x_{k_l}^{K+M} - x_{k_l}^{K+M-1} \ge b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon)}{\varepsilon}. \tag{99}$$

By (finite) induction, we next prove that

$$x_{k_l}^{K+M+1-i} - x_{k_l}^{K+M-i} \ge b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i} \qquad \forall i \in \{1, M-1\}, \tag{100}$$

---

[73] For, we have that

$$\begin{aligned}
\frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(x_{k_l}^{t-1} - x_{k_l}^t\right) &= \frac{b-c}{1-\varepsilon} - \frac{\varepsilon}{1-\varepsilon}\left(J_j\left(x^{t-1}\right) - J_j\left(x^t\right)\right) \\
&\le \frac{1}{1-\varepsilon}\left(x_j^t - x_j^{t+1}\right) - \frac{\varepsilon}{1-\varepsilon}\left(J_j\left(x^{t-1}\right) - J_j\left(x^t\right)\right) \\
&\le \left\|\frac{1}{1-\varepsilon}\left(x^t - x^{t+1}\right) - \frac{\varepsilon}{1-\varepsilon}\left(J\left(x^{t-1}\right) - J\left(x^t\right)\right)\right\|_\infty \\
&= \left\|\frac{1}{1-\varepsilon}\left(x^{t+1} - x^t\right) - \frac{\varepsilon}{1-\varepsilon}\left(J\left(x^t\right) - J\left(x^{t-1}\right)\right)\right\|_\infty \\
&= \left\|y^t - y^{t-1}\right\|_\infty \le \left\|x^t - x^{t-1}\right\|_\infty.
\end{aligned}$$

[74] Otherwise, each interval would contain strictly less than M elements. This would yield that $\{\|x^{i+1} - x^i\|_\infty\}_{i=1}^P$ would consist of strictly less than

$$\left\lfloor \frac{P}{M} \right\rfloor M \le \frac{P}{M}M = P$$

elements, a contradiction.

that is,

$$x_{k_l}^{K+M+1-i} \geq x_{k_l}^{K+M-i} + b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i} \qquad \forall i \in \{1, M-1\}. \qquad (101)$$

By (99), the statement is true for $i = 1$. Next, we assume it is true for $i \in \{1, ..., M-1\}$ and prove it is still true for $i+1$ when $i+1 \in \{1, ..., M-1\}$. This implies that $i \leq M-2$. Define $t = K+M-i$. By the induction hypothesis, we have that

$$x_{k_l}^{t+1} - x_{k_l}^t = x_{k_l}^{K+M+1-i} - x_{k_l}^{K+M-i} \geq b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i}.$$

Moreover, we also have that $\|x^t - x^{t-1}\|_\infty = \|x^{K+M-i} - x^{K+M-i-1}\|_\infty \leq b_{\bar{m}} + (1-\varepsilon)\varepsilon^M$. Set $b = b_{\bar{m}}$, $c = (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i}$, and $\kappa = (1-\varepsilon)\varepsilon^M$. By (97), we can conclude that

$$\begin{aligned}
x_{k_l}^{K+M+1-(i+1)} - x_{k_l}^{K+M-(i+1)} &= x_{k_l}^{K+M-i} - x_{k_l}^{K+M-i-1} = x_{k_l}^t - x_{k_l}^{t-1} \\
&\geq b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i} \frac{1}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}(1-\varepsilon)\varepsilon^M \\
&= b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^{i+1})}{\varepsilon^{i+1}},
\end{aligned}$$

proving (100). By (101), repeated substitution, and point 5, this implies that

$$x_{k_l}^{K+M} \geq x_{k_l}^{K+1} + (M-1)b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \sum_{i=1}^{M-1} \frac{1-\varepsilon^i}{\varepsilon^i} \geq x_{k_l}^{K+1} + (M-1)b_{\bar{m}} - 1,$$

that is,
$$\|x^{K+M} - x^{K+1}\|_\infty \geq x_{k_l}^{K+M} - x_{k_l}^{K+1} \geq (M-1)b_{\bar{m}} - 1.$$

Since $b_{\bar{m}} \geq \delta > 0$, we have that $(M-1)b_{\bar{m}} \geq (M-1)\delta > D+1$. We can conclude that

$$D \geq \|x^{K+M} - x^{K+1}\|_\infty \geq (M-1)b_{\bar{m}} - 1 > D,$$

a contradiction.

b. $x_j^{K+M} - x_j^{K+M+1} \geq b_{\bar{m}}$. Set $b = b_{\bar{m}}$, $c = 0$, and $\kappa = (1-\varepsilon)\varepsilon^M$. By (98), we can conclude that
$$x_{k_l}^{K+M-1} - x_{k_l}^{K+M} \geq b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{1-\varepsilon}{\varepsilon}. \qquad (102)$$

By (finite) induction, we next prove that

$$x_{k_l}^{K+M-i} - x_{k_l}^{K+M+1-i} \geq b_{\bar{m}} - (1-\varepsilon)\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i} \qquad \forall i \in \{1, M-1\}, \qquad (103)$$

that is,

$$x_{k_l}^{K+M-i} \geq x_{k_l}^{K+M+1-i} + b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i} \qquad \forall i \in \{1, M-1\}. \quad (104)$$

By (102), the statement is true for $i = 1$. Next, we assume it is true for $i \in \{1, ..., M-1\}$ and prove it is still true for $i+1$ when $i+1 \in \{1, ..., M-1\}$. This implies that $i \leq M-2$. Define $t = K+M-i$. By the induction hypothesis, we have that

$$x_{k_l}^t - x_{k_l}^{t+1} = x_{k_l}^{K+M-i} - x_{k_l}^{K+M+1-i} \geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i}.$$

Moreover, we also have that $\|x^t - x^{t-1}\|_\infty = \|x^{K+M-i} - x^{K+M-i-1}\|_\infty \leq b_{\bar{m}} + (1-\varepsilon)\,\varepsilon^M$. Set $b = b_{\bar{m}}$, $c = (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i}$, and $\kappa = (1-\varepsilon)\,\varepsilon^M$. By (98), we can conclude that

$$\begin{aligned}
x_{k_l}^{K+M-(i+1)} - x_{k_l}^{K+M+1-(i+1)} = x_{k_l}^{K+M-i-1} - x_{k_l}^{K+M-i} &= x_{k_l}^{t-1} - x_{k_l}^t \\
&\geq b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon^i)}{\varepsilon^i}\frac{1}{\varepsilon} - \frac{1-\varepsilon}{\varepsilon}(1-\varepsilon)\,\varepsilon^M \\
&= b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \frac{(1-\varepsilon^{i+1})}{\varepsilon^{i+1}},
\end{aligned}$$

proving (103). By (104), repeated substitution, and point 5, this implies that

$$x_{k_l}^{K+1} \geq x_{k_l}^{K+M} + (M-1)\,b_{\bar{m}} - (1-\varepsilon)\,\varepsilon^M \sum_{i=1}^{M-1} \frac{1-\varepsilon^i}{\varepsilon^i} \geq x_{k_l}^{K+M} + (M-1)\,b_{\bar{m}} - 1,$$

that is,
$$\|x^{K+1} - x^{K+M}\|_\infty \geq x_{k_l}^{K+1} - x_{k_l}^{K+M} \geq (M-1)\,b_{\bar{m}} - 1.$$

Since $b_{\bar{m}} \geq \delta > 0$, we have that $(M-1)\,b_{\bar{m}} \geq (M-1)\,\delta > D+1$. We can conclude that

$$D \geq \|x^{K+1} - x^{K+M}\|_\infty \geq (M-1)\,b_{\bar{m}} - 1 > D,$$

a contradiction.

Points a and b prove the statement. ∎