When social influence promotes the wisdom of crowds

Abdullah Almaatouq[a,b,*], M. Amin Rahimian[c], and Abdulla Alhajri[b]

[a] Sloan School of Management, Massachusetts Institute of Technology
[b] Center for Computational Engineering, Massachusetts Institute of Technology
[c] Department of Industrial Engineering, University of Pittsburgh
[*] To whom correspondence may be addressed; email: amaatouq@mit.edu.

Whether, and under what conditions, groups exhibit "crowd wisdom" has been a major focus of research across the social and computational sciences. Much of this work has focused on the role of social influence in promoting the wisdom of the crowd versus leading the crowd astray, resulting in conflicting conclusions about how the social network structure determines the impact of social influence. Here, we demonstrate that it is not enough to consider the network structure in isolation. Using theoretical analysis, numerical simulation, and re-analysis of four experimental datasets (totaling $4,002$ human subjects), we find that the wisdom of crowds critically depends on the interaction between (i) the centralization of the social influence network and (ii) the distribution of the initial, individual estimates, i.e., the estimation context. Specifically, we propose a feature of the estimation context that measures the suitability of the crowd to benefit from influence centralization and show its significant predictive powers empirically. By adopting a framework that integrates both the structure of the social influence and the estimation context, we bring previously conflicting results under one theoretical framework and clarify the effects of social influence on the wisdom of crowds.

In its classical definition, the concept of "the wisdom of crowds" refers to the idea that the aggregate estimate of a group of individuals can be superior to that of individual, credentialed experts [1, 2]. Recent applications of this concept include technological, political, and economic forecasting [3], crowdsourcing [4], and public policy design [5]. Conventional statistical accounts of the wisdom of crowds rely on the following two assumptions: (i) the individual errors are uncorrelated or negatively correlated [6], and (ii) the individuals are unbiased, i.e., correct, in mean expectations [2].

However, social influence processes, in which people exchange information about their estimates, can cause them to revise their judgment in estimation tasks [7, 8, 9, 10]. Therefore, aggregating the revised (post-influence) estimates is not the same as aggregating the initial (pre-influence) estimates. Prior research yields conflicting findings on the effects of social influence on the wisdom of crowds. For instance, despite the evidence that social influence can significantly benefit group and individual estimates [9, 11, 12, 13, 14], social influence has also been found to induce systematic bias, herding, and groupthink [7, 8].

In response to these inconsistencies, notable reconciliation efforts have focused on investigating how social network theories interact with the process of collective belief formation. The results of these efforts, including seminal theoretical works [10, 15] and laboratory experiments [9], have established that the wisdom of crowds is preserved only if the influence of the most influential individual vanishes, i.e., becomes negligible, as the group size grows [10]. This condition is satisfied in decentralized influence structures, i.e., structures where everyone has an equal voice, as opposed to centralized structures where one or more individuals have disproportionate influence. Intuitively, the wisdom of crowds benefits from larger group sizes, but centralized influence diminishes this benefit by reducing the collective estimate into the "wisdom of the few".

While these results appear to broadly suggest the superiority of decentralized influence, their conclusions rest on the premise that the distribution of the initial estimates is centered around the truth. In such situations, there are no opportunities for the crowd to improve with social influence [10]. However, empirical distributions of numerical estimates tend to be right-skewed with excess kurtosis, where most estimates are low, with a minority falling on a fat right tail [8, 16, 17]. The skewness of the distribution could emerge due to systematic bias (a tendency to over- or underestimate the actual value [16, 18, 19]) or dispersion (the spread of estimates) in the population. Therefore, it is when the crowd is not initially centered around the truth, as observed in many empirical settings, that centralized influence could present an opportunity to promote crowd wisdom.

A

↓ Systematic bias ↓ Dispersion ↑ Systematic bias ↓ Dispersion ↓ Systematic bias ↑ Dispersion ↑ Systematic bias ↑ Dispersion

True value   True value   True value   True value
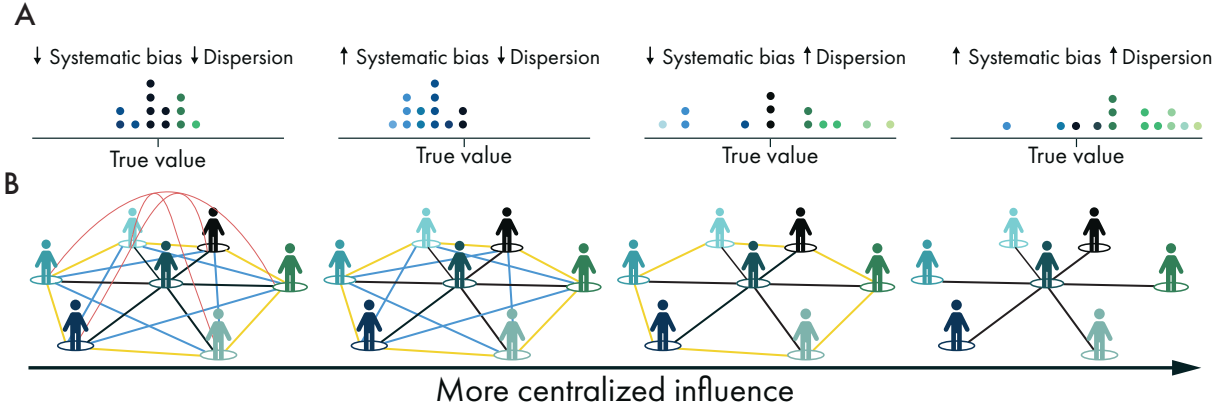
B

More centralized influence

Figure 1: **This schematic illustrates our framework for analyzing the role of estimation context in determining how social influence shapes the wisdom of crowds.** Panel **A** illustrates four types of estimation contexts by varying the distribution of the initial estimates. Panel **B** provides examples of different influence structures arranged in the order of increasing centralization — from a fully decentralized structure, where everyone has an equal voice, to a highly centralized structure, where there is one highly influential individual. In our model, we represent centralized structures by weighted averages that give higher weights to more influential individuals. We point out that when the distribution of the initial estimates is heavy-tailed or right-skewed, then a weighted average with more centralized influence has a high likelihood of falling closer to the truth than the simple, unweighted, decentralized average. Subsequently, we propose a feature of the estimation context that predicts whether the collective estimation performance improves as a result of group interactions, based only on the empirically measurable distribution of the individual estimates.

In this study, we ask *when centralized influence structures improve or hinder the wisdom of crowds in estimation tasks*. Our results demonstrate that the effect of social influence varies systematically with the distribution of the initial estimates, and, therefore, it is more heterogeneous than previously suggested. Specifically, we propose and analyze — theoretically, numerically, and empirically — a feature of the distribution of the initial estimates that measures the suitability of a crowd to benefit from influence centralization.

To illustrate this, we consider a group of n agents tasked to estimate or forecast, with maximal accuracy, some unknown positive quantity such as the unemployment rate in the next quarter, life expectancy of an ill patient, amount of calories in a meal, prevalence of global influenza infections in two weeks, or the number of jellybeans in a jar. To model the population of the agents performing a particular estimation task, we endow each agent with a biased and noisy signal about the truth that constitutes her initial estimate. The initial estimates are

3

independent and identically distributed with the location parameter μ, indicating the center of the distribution that biases the estimates with respect to the truth, and the shape parameter σ, determining the variation and tail shape.

The extent of the systematic bias, μ, and dispersion, σ, of the initial estimates can be viewed as intrinsic properties of the *estimation context*: a population of agents performing a particular estimation task. Different populations of agents, e.g., experts vs. novices, might have different biases and dispersions for similar tasks. Conversely, the same population can vary in terms of their bias and dispersion across different types of tasks. For brevity and to abstract the agents and the estimation task, we refer to the distribution of the initial estimates as the *estimation context*. Figure 1.A shows four estimation contexts with varying levels of bias and dispersion.

Agents frequently have access to the opinions or estimates of other agents. In many common models of social influence [10, 15, 20], as well as in other aggregation mechanisms [6, 21, 22], the collective estimate of the group of agents can be expressed as a linear combination (weighted average) of the initial estimates. This definition of collective estimation contains the simple average of the initial estimates as a special case, i.e., the typical "wisdom of crowds".

We introduce an influence *centralization* parameter, ω, to interpolate between a collective estimate produced by a fully decentralized influence setup where every agent has an equal voice, $\omega = 0$, and a dictatorial setup with a single influential agent, $\omega = 1$. Our definition of ω coincides with Freeman's centralization [23] for a class of network typologies that encompass cases of practical and empirical interest, such as fully connected networks, star networks, empty graphs (isolated individuals), and circular lattices, among others (see Materials and Methods and SI section S1.1). Figure 1.B shows four influence network structures in this class (see SI section S1.1 for the calculation of ω for different networks).

We measure the collective performance of the agents in terms of the proximity of the collective estimate to the truth. Given the estimation context (distribution of the initial estimates), we compute the probability that the collective estimate produced by a centralized influence structure outperforms a decentralized baseline. We denote this probability by Ω_n. Notably, Ω_n captures a critical *feature* of the estimation context, namely, *its suitability to benefit from centralization*. For instance, when $\Omega_n < 1/2$, the initial estimates are better suited for decentralized influence structures; conversely, when $\Omega_n > 1/2$, they are better suited for centralized influence structures.

Our theoretical analysis of Ω_n verifies that for heavy-tailed or right-skewed distributions, the performance of the collective estimate in a centralized structure where a single agent has a
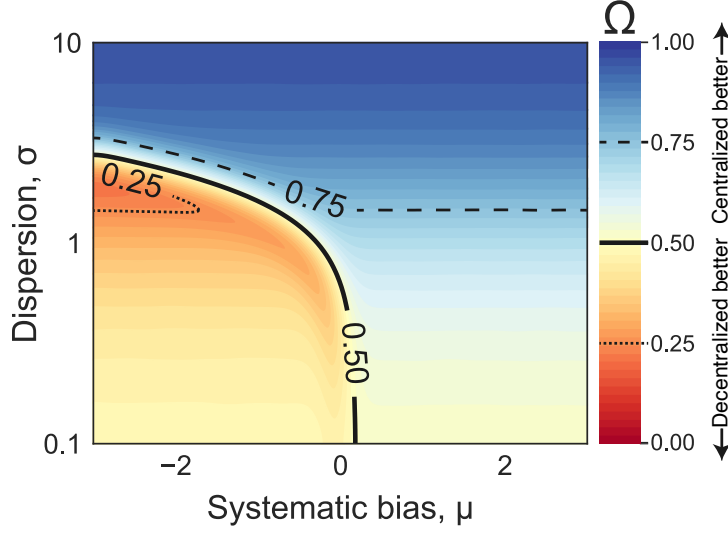
4

Figure 2: **Our simulations illustrate the link between the distribution of the initial esti-mates and the proposed estimation context feature, Ω_n.** Our proposed feature of the esti-mation context, Ω_n, measures the likelihood that a weighted average falls closer to the truth than an unweighted average. Hence, when $\Omega_n < 1/2$, the estimation context is better suited for decentralized (unweighted) influence structures; conversely, when $\Omega_n > 1/2$, it is better suited for centralized (weighted) influence structures. In our theoretical model, we give a lower bound on Ω_n as a function of the estimation context and the network structure (see Materials and Methods). For heavy-tailed distributions (e.g., Pareto, log-normal, and log-Laplace), we identify phase transition behaviors, whereby the lower bound's limiting value transitions from 0 to 1 or $1/2$, as the distribution parameter, σ, crosses a critical value (see SI section S2.1). In this figure, the initial estimates are sampled from a log-normal distribution for a fixed number of agents ($n = 50$) and influence-centralization level ($\omega = 1/3$). See SI Figures S1 and S3 for other distributions and parameter choices.

non-vanishing influence (her contribution to the collective estimate does not become negligible as $n \rightarrow \infty$) is superior to that of the decentralized baseline. Intuitively, this is due to the fact that the sample mean of a heavy-tailed distribution is dominated by its excess tail risk (the egregious errors of a few individuals). On the other hand, using weighted averages in centralized structures, we can guarantee that some random individuals exert enough influence to prevent the group aggregate from being swayed too far by the egregious errors of the few (see SI section S2.1). Notably, in this model, centralized structures violates the vanishing influence condition for the wisdom of crowds, cf. [10] and SI section S2.1.5. This underscores the importance of the distributional assumptions, which are context dependent, when studying the effect of social influence on the wisdom of crowds.

5

In Figure 2, we illustrate the behavior of our proposed feature, Ω_n, for a log-normal estimation context, as reported in several empirical studies [8, 9, 17]. In this case, Ω_n predicts that centralized influence structures improve collective estimates over decentralized influence structures if the distribution of the initial estimates is characterized by a log-normal distribution with an overestimation bias or large dispersion (see SI section S2.2.1 and Figure S2 for the effect of the systematic bias). However, this relationship is reversed when the distribution is characterized by low dispersion and underestimation bias (see SI section S2.2 and Figure S1 for simulation details and other distributional classes).

In order to empirically illustrate the explanatory power of our proposed estimation context feature, Ω_n, we use the data from four published experiments [8, 9, 12, 13]. In these experiments, a total of 4,002 participants, organized into 131 independent groups, completed a total of 57 estimation tasks, generating a total of 20,030 individual estimations and 815 collective estimations (see Figure 3.A).

All experiments followed a similar procedure that involved the following three steps: (1) the participants simultaneously and independently completed numeric estimation tasks on a range of topics, e.g., visual estimation, trivia questions, political facts, and economic forecasts; (2) within groups of varying sizes, the participants in the social interaction condition communicated information about their estimates with each other; and (3) the participants had one or more opportunities to revise their estimates. One trial consisted of a single group of participants answering a single task.

Each task induces a different distribution on the initial estimates that are measured empirically. Therefore, each task constitutes an estimation context, giving a total of 57 estimation contexts. For each estimation context, we compute our proposed feature, Ω, directly from the data based on the empirical initial estimates, without the need for making any distributional assumptions. To this end, we sample initial estimates with replacement from the data, and numerically estimate the probability that a weighted average (centralized structure) is closer to the true value than an unweighted average (decentralized structure), under the empirically measured distribution of the initial estimates; see SI section S3. Figure 3.B shows the distribution of Ω in these studies. We use the numerically measured Ω for each estimation context, as an independent variable in our regression analysis.

We refer to the average of the individuals in each group, *before* and *after* their interactions, as their collective *initial* and *revised* estimates, respectively. For each trial, we compare the absolute errors of the collective initial and revised estimates. In our regressions, we use the

following two outcome metrics: (1) the standardized (z-score) absolute error of the revised collective estimate for all groups (with or without social interaction), and (2) whether the collective revised estimate is more accurate than the collective initial estimate in groups with social interaction. We use a linear regression for the former, and a logistic regression for the latter.

Our empirical analysis relies on the following premise: the collective initial estimate corresponds to the most decentralized influence structure ($\omega = 0$), and social interactions can only increase the influence centralization ($\omega > 0$). For example, even in social interactions where everybody is equally connected in terms of the communication structure, some group members may become more influential than others, by virtue of being more talkative, more persuasive, or more resistant to social influence [9, 14].

We begin by testing the main hypothesis predicted by our theory; namely, that the effect of social influence centralization on the performance of groups is moderated by our proposed feature of the estimation context, Ω. As shown in Figure 3.C, we find that the interaction between the centralization of the influence and the estimation context feature, Ω, significantly affects the wisdom of crowds ($\beta = -3.88$; t-statistic $= -7.20$; $p < 0.001$). Critically, the results of this analysis show that variation in the estimation context feature, Ω, can completely reverse the effects of social influence centralization: specifically, when $\Omega < 1/2$, the error of the revised collective estimate is lower in decentralized influence structures; whereas when $\Omega > 1/2$, the error of the revised collective estimate is lower in centralized structures.

To find further support for this finding, we limit our attention to the groups in the centralized influence conditions. As shown in Figure 3.D, we find that whether or not a group improved after social interaction is substantially explained by Ω, our proposed feature of the estimation context (odds ratio $= 15.00$; z-statistic $= 5.66$; $p < 0.001$).

The critical implication of the above results is that our proposed feature of the estimation context moderates the effect of influence centralization. Therefore, we find no support to the hypothesis that decentralized influence structures would be preferred over centralized ones independently of the estimation context.

Therefore, the effect of network structure on the collective estimation performance should be reconceptualized under a context-dependent framework, i.e., with respect to the population of individuals performing the particular task. There is no single influence structure that is better than others in all contexts. Such a context-dependent framework can unify previously conflicting findings on crowd wisdom under a single theoretical framework and explain the effects of the influence network structure on the quality of the collective estimates.
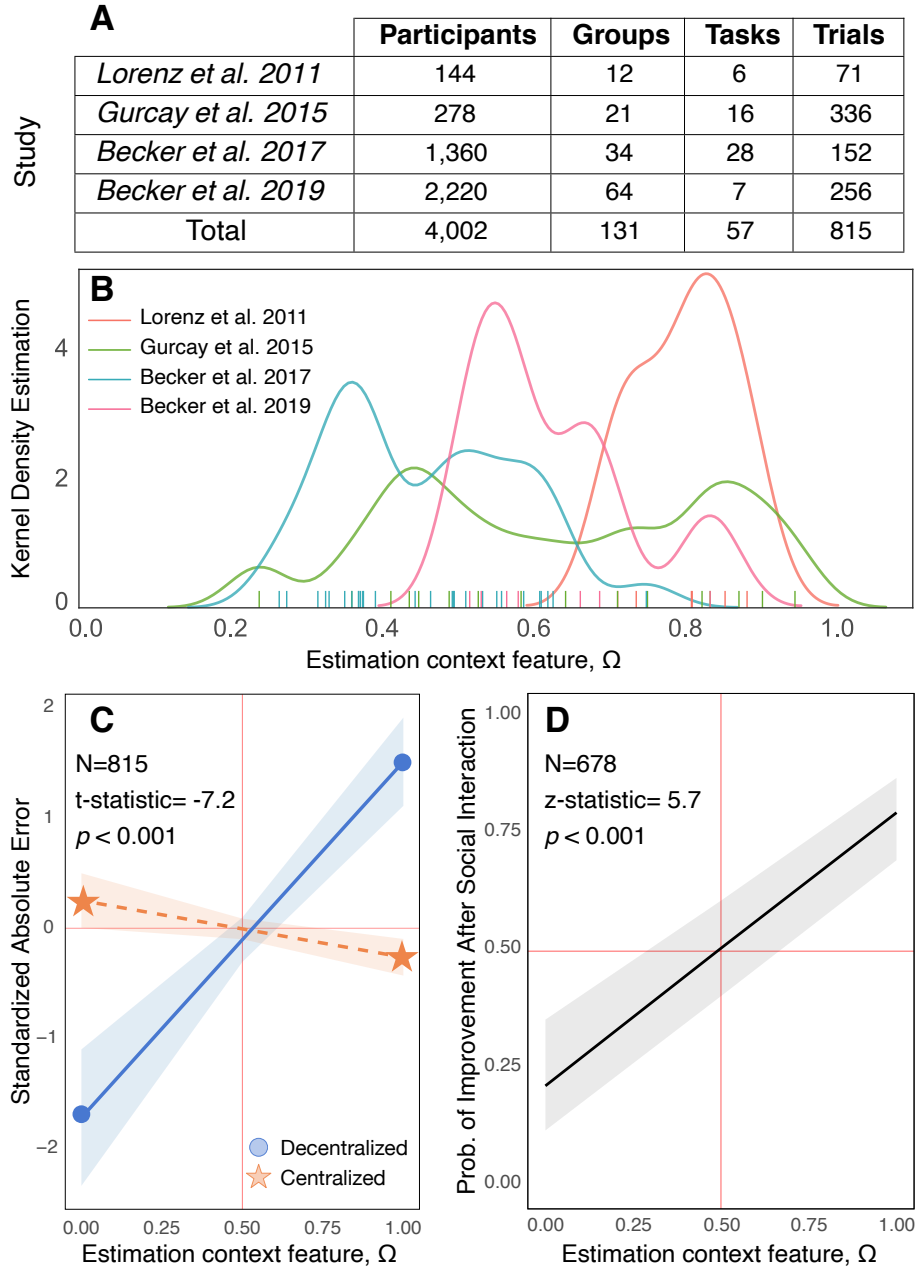
7

Study		Participants	Groups	Tasks	Trials
	Lorenz et al. 2011	144	12	6	71
	Gurcay et al. 2015	278	21	16	336
	Becker et al. 2017	1,360	34	28	152
	Becker et al. 2019	2,220	64	7	256
	Total	4,002	131	57	815

Figure 3: **Reanalysis of previously published experiments indicates that our proposed feature, Ω, has significant predictive powers for determining when the group performance improves as a result of social interactions.** Panel **A** shows the number of participants, groups, tasks, and trials in the reanalyzed experiments. Panel **B** displays the distribution of the estimation context feature, Ω, across these studies. Panel **C** shows the marginal effect of the interaction term between influence centralization and the estimation context feature, Ω: as Ω increases, the group performance improves in the centralized influence conditions, and degrades in the decentralized influence conditions. Panel **D** shows that the probability of groups improving their performance after social interaction is substantially explained by our proposed feature of the estimation context, Ω. The bands are the 95% confidence intervals.

Admittedly, the estimation context is only one of several potential sources of inconsistency in previous studies. For instance, vagueness or ambiguity of some theoretical constructs, e.g., influence, can result in different studies of, seemingly, the same phenomenon, to measure different things. Furthermore, we acknowledge that our proposed feature of the estimation context, Ω, is concerned only with the probability of the following event: the collective estimate generated by the agents interacting in a centralized influence structure is closer to the truth than the collective estimate generated by the agents in a decentralized structure. This is not the same as comparing mean squared error or other expected loss functions for these collective estimates (see Figure S4 for examples of other loss functions). Finally, we note that we only studied one class of tasks: numerical estimation with a non-negative, objective truth. Relevant research on other classes of tasks has similarly demonstrated that variation in context features, such as complexity [24, 25, 26, 27], fundamentally alter collective problem solving outcomes.

Our theoretical and empirical analysis has demonstrated that conclusions about the role of the social influence can be inconsistent unless the estimation context is explicitly accounted for. Many research extensions are warranted from this framework. For example, unlike what is assumed in most available work, including ours, the social networks we live in are not random, nor are they imposed by external forces. Rather, these social networks emerge under the influence of endogenous social processes and gradually evolve within a potentially non-stationary context. A truly context-dependent view on crowd wisdom should open connections with diverse research fields and help advance an interdisciplinary understanding of the design of social systems and their information outcomes.

Materials and Methods

Collective estimation and influence centralization, ω**.** We consider a group of n agents indexed by $i = 1, \ldots, n$, and assume that each agent is endowed with an independent and identically distributed initial estimate $\mathbf{a}_{i,0}$. The distribution of the initial estimates, $\mathcal{F}^{\theta}_{\mu,\sigma}$, is parametrized by the unknown truth, θ, the systematic bias, μ, and the dispersion, σ. In many common models of social influence, the collective estimate, $\mathbf{a}^n$, can be expressed as a convex combination of the initial estimates: $\mathbf{a}^n(\bar{w}) = \sum_{i=1}^{n} w_i \mathbf{a}_{i,0}$, where $w_1, \ldots, w_n$ are positive real weights summing to one. These weights represent the influence of individual agents on shaping the collective estimate. Without loss of generality, we assume that the agents are ordered in the decreasing order of their influence, i.e., $w_1 \geq w_2 \geq \ldots \geq w_n$. In order to investigate the role

of network centralization, $0 \leq \omega \leq 1$, we consider a class of influence structures indexed by ω such that (see SI section S1.1 for more details),

$$\mathbf{a}^n(\omega) = \omega \mathbf{a}_{1,0} + (1 - \omega)\frac{1}{n}\sum_{i=1}^{n} \mathbf{a}_{i,0}.$$

Using ω, we interpolate between a dictatorial setup with a single influential voice (i.e., $w_1 = \omega = 1$ and $w_2 = \ldots = w_n = 0$) and a fully decentralized setup where everyone has an equal voice (i.e., $\omega = 0$ and $w_1 = w_2 = \ldots = w_n = 1/n$).

Proposed feature for the estimation context, Ω. We measure the probability that the collective estimate produced by a centralized influence structure, $\mathbf{a}^n(\omega)$, $\omega > 0$, outperforms the decentralized baseline, $\mathbf{a}^n(0)$. We denote this probability by $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) := \mathbb{P}^\theta_{\mu,\sigma}[|\mathbf{a}^n(\omega) - \theta| < |\mathbf{a}^n(0) - \theta|]$. To compute Ω_n in Figure 2, we have fixed $n = 50$, $\theta = 2$, and $\omega = 1/3$. Therefore, Ω is entirely determined by the distribution of the initial estimates (μ and σ). Similarly, in Figure 3, we have fixed $n = 50$ (sampled randomly and repeatedly from the study's dataset), $\theta = 2$, and $\omega = 1/3$, so that Ω is entirely determined by the empirical distribution of the initial estimates. Figure S3 and Tables S2-S3 replicate our simulation and empirical results for a range of n and ω values. For distributions $\mathcal{F}^\theta_{\mu,\sigma}$, supported over positive reals, with cumulative function $F^\theta_{\mu,\sigma}$, we propose the following lower bound (proved in SI section S2.1):

$$\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) \geq \sup_{\beta > \theta/(1-\omega)} \left\{ F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1}) \right\}.$$

In SI section S2.1, we show how to limit the rate of tail decay for different classes of distributions, to produce a non-trivial (non-zero) lower bound as $n \to \infty$. For heavy-tailed distributions, such as Pareto, log-Laplace, and log-normal (see SI subsections S2.1.1 to S2.1.3), we identify phase transition behaviors, whereby the proposed lower bound's limiting value transitions from 0 to 1 or $1/2$, as σ crosses a critical value.

Statistical tests. All statistics were two-tailed and based on mixed-effects models that included random effects to account for the nested structure of the data. In particular, the regression

equation for Figure 3.C is:

$$y_{ij} = \beta_0 + \beta_1 \Omega_{(j)} + \beta_2 I_i + \beta_3 I_i \Omega_{(j)} + v_i + \epsilon_{ij},$$

where y_{ij} is the standardized (z-score) absolute error of the revised collective estimate for the i-th group in the j-th estimation context, $\Omega_{(j)}$; β_0 is the fixed intercept for the regression model; β_1 is the fixed coefficient for the estimation context feature, Ω; $I_i \in \{0, 1\}$ is an indicator variable of whether or not social interaction has occurred; β_2 is the fixed coefficient for the social influence centralization; β_3 is the fixed coefficient for the interaction term between the estimation context feature, Ω, and influence centralization (shown in Figure 3.C); v_i is the random coefficient for the i-th group; and ϵ_{ij} is a Gaussian error term. The absolute error of the revised collective estimate has been standardized, i.e., z-scored, in order to compare errors across different tasks (the correct answer for different tasks can differ by orders of magnitude). The analysis was conducted on 815 observations; 678 groups with social influence (centralized), and 137 groups without social influence (decentralized).

The logistic regression equation for Figure 3.D is:

$$y_{ij} = \frac{1}{1 + \exp(\beta_0 + \beta_1 \Omega_{(j)} + v_i + \epsilon_{ij})},$$

where y_{ij} is a binary indicator for whether or not the i-th group in the j-th estimation context improved the accuracy of its collective estimate after social interaction; β_0 is the fixed intercept for the regression model; β_1 is the fixed coefficient for the estimation context feature, Ω; v_i is the random coefficient for the i-th group; and ϵ_{ij} is a Gaussian error term. The analysis was conducted on 678 observations (groups with social influence).

Further details of the regression analysis are provided in SI section S3.1, Table S1. Robustness checks for the regression results are presented in Tables S2-S3.

Data and code availability. Replication data and code are available at https://github.com/amaatouq/task-dependence.

References

[1] F. Galton, Vox populi, *Nature* **75**, 450 (1907).

[2] J. Surowiecki, *The Wisdom of Crowds* (Knopf Doubleday Publishing Group, 2005).

[3] J. Wolfers, E. Zitzewitz, Prediction markets, *Journal of economic perspectives* **18**, 107 (2004).

[4] A. Vermeule, Many-minds arguments in legal theory, *J. Legal Analysis* **1**, 1 (2009).

[5] M. G. Morgan, Use (and abuse) of expert elicitation in support of decision making for public policy, *Proceedings of the National academy of Sciences* **111**, 7176 (2014).

[6] C. P. Davis-Stober, D. V. Budescu, J. Dana, S. B. Broomell, When is a crowd wise?, *Decision* **1**, 79 (2014).

[7] L. Muchnik, S. Aral, S. J. Taylor, Social influence bias: A randomized experiment, *Science* **341**, 647 (2013).

[8] J. Lorenz, H. Rauhut, F. Schweitzer, D. Helbing, How social influence can undermine the wisdom of crowd effect, *Proc. of the National Academy of Sciences* **108**, 9020 (2011).

[9] J. Becker, D. Brackbill, D. Centola, Network dynamics of social influence in the wisdom of crowds, *Proceedings of the national academy of sciences* **114**, E5070 (2017).

[10] B. Golub, M. O. Jackson, Naive learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics* **2**, 112 (2010).

[11] B. Bahrami, *et al.*, Optimally interacting minds, *Science* **329**, 1081 (2010).

[12] B. Gürçay, B. A. Mellers, J. Baron, The power of social influence on estimation accuracy, *Journal of Behavioral Decision Making* **28**, 250 (2015).

[13] J. Becker, E. Porter, D. Centola, The wisdom of partisan crowds, *Proceedings of the National Academy of Sciences* **116**, 10717 (2019).

[14] A. Almaatouq, *et al.*, Adaptive social networks promote the wisdom of crowds, *Proceedings of the National Academy of Sciences* **117**, 11379 (2020).

[15] P. M. DeMarzo, D. Vayanos, J. Zwiebel, Persuasion bias, social influence, and unidimensional opinions, *The Quarterly journal of economics* **118**, 909 (2003).

[16] B. Jayles, *et al.*, How social information can improve estimation accuracy in human groups, *Proceedings of the National Academy of Sciences* **114**, 12620 (2017).

[17] A. B. Kao, *et al.*, Counteracting estimation bias and social influence to improve the wisdom of crowds, *Journal of The Royal Society Interface* **15**, 20180130 (2018).

[18] T. Indow, M. Ida, Scaling of dot numerosity, *Perception & Psychophysics* **22**, 265 (1977).

[19] J. P. Simmons, L. D. Nelson, J. Galak, S. Frederick, Intuitive biases in choice versus estimation: Implications for the wisdom of crowds, *Journal of Consumer Research* **38**, 1 (2011).

[20] M. H. DeGroot, Reaching a consensus, *Journal of American Statistical Association* **69**, 118 (1974).

[21] N. H. Anderson, *A functional theory of cognition* (Psychology Press, 2014).

[22] D. Prelec, H. S. Seung, J. McCoy, A solution to the single-question crowd wisdom problem, *Nature* **541**, 532 (2017).

[23] L. C. Freeman, Centrality in social networks conceptual clarification, *Social networks* **1**, 215 (1978).

[24] D. Barkoczi, M. Galesic, Social learning strategies modify the effect of network structure on group performance, *Nature communications* **7**, 1 (2016).

[25] J. Shore, E. Bernstein, D. Lazer, Facts and figuring: An experimental investigation of network structure and performance in information and solution spaces, *Organization Science* **26**, 1432 (2015).

[26] W. Mason, D. J. Watts, Collaborative learning in networks, *Proceedings of the National Academy of Sciences* **109**, 764 (2012).

[27] D. Lazer, A. Friedman, The network structure of exploration and exploitation, *Administrative Science Quarterly* **52**, 667 (2007).

Supplementary Information

This Supplementary Information is organized in five sections. In section S1, we present network models of collective estimation that facilitate our study of the collective estimate, $\mathbf{a}^n(\omega)$, as a function of the centralization, ω, and in relation to its network structure. In section S2, we give the proof of our main proposition (lower bound on Ω_n) and relate it to the rate of tail decay for many common distributions. We study these relationships theoretically and through numerical analysis as well. In section S3, we develop empirical measures to analyze prior experiments in terms of the features of their estimation context. In section S4, we provide robustness checks. Supplementary references are listed in section S5.

Contents

Preliminary notation. For the convenience of the reader, we collect here some notation that will be used throughout. We study collective estimation by a group of agents and use n to denote the group size. For sequences of real numbers f_n and g_n, indexed by integers n, we use the asymptotic notations $f_n \asymp g_n$ to signify that $\lim_{n \to \infty} f_n/g_n = 1$. We consider agents that are indexed by $i = 1, \ldots, n$. We use bold fonts to represent random variables. We use $\mathbf{a}_{1,0}$, ..., $\mathbf{a}_{n,0}$ to denote the individual estimates in the absence of any social interactions. The collective estimate is denoted by $\mathbf{a}^n$; it is determined in terms of the individual estimates, in a manner that involves a centralization parameter ω: $\mathbf{a}^n(\omega) = \omega \mathbf{a}_{1,0} + (1 - \omega)\frac{1}{n}\sum_{i=1}^n \mathbf{a}_{i,0}$. Vectors are denoted by a bar on top of their letters and we use superscript ^{T} to denote matrix transpose.

S1 Network models of collective estimation

Let θ be an unknown state of the world. Consider n agents indexed by $i = 1, \ldots, n$, each endowed with a biased and noisy signal about θ. The signals are independent and identically distributed across the agents and constitute their initial estimates of the unknown θ:

$$\mathbf{a}_{i,0} \sim \mathcal{F}^{\theta}_{\mu,\sigma}. \tag{S.1}$$

The the distribution of the initial estimates in (S.1), $\mathcal{F}^{\theta}_{\mu,\sigma}$, is parametrized by θ, μ, and σ. We think of μ as a location parameter (the center of the distribution) that biases the individual estimates against θ. This captures the level of systematic bias in the population. We think of σ as a variance-proxy/shape parameter that determines the variation and tail-fatness of $\mathcal{F}^{\theta}_{\mu,\sigma}$. In other words, σ can be interpreted as the amount of prior information a group has about the quantity and represents the level of *demonstrability* of the estimation task.

The agents interact in a group. Their group interactions can be modeled in a variety of ways leading to a group aggregate $\mathbf{a}^{n}(\bar{w})$ that is a convex combination of the initial estimates:

$$\mathbf{a}^{n}(\bar{w}) := \bar{w}^{T} \bar{\mathbf{a}}_{0} = \sum_{i=1}^{n} w_{i} \mathbf{a}_{i,0}, \tag{S.2}$$

where $\bar{w}$ is an entry-wise non-negative vector satisfying $\bar{w}^{T} \mathbb{1} = 1$.

In general, different agents' initial estimates will receive different weights in the collective estimate. A common method of modeling group interactions is through DeGroot-style iterated averaging, which has a long history in mathematical sociology and social psychology [1]. The origins of iterated averaging models can be traced to French's seminal work on "A Formal Theory of Social Power" [2], followed up by Harary's investigation of the mathematical properties of the averaging model, including the consensus criteria, and its relations to Markov chain theory [3]. This model was further popularized by DeGroot's seminal work [4] on linear opinion pools and belief exchange dynamics. In a typical iterated averaging setup, an agent's estimate at time t is given by a weighted average of her neighboring estimates at time $t-1$:

$$\mathbf{a}_{i,t} = \sum_{j=1}^{n} W_{ij} \mathbf{a}_{j,t-1},$$

In matrix notation, we have:

$$\bar{\mathbf{a}}_t = W\bar{\mathbf{a}}_{t-1} = W^t\bar{\mathbf{a}}_0, \tag{S.3}$$

where $\bar{\mathbf{a}}_t = (\mathbf{a}_{1,t}, \mathbf{a}_{2,t}, \ldots, \mathbf{a}_{n,t})^T$ and $W = [W_{ij}]$ is the matrix of weights. We refer to matrix W as the social influence matrix. For a strongly connected social network with $W_{ii} > 0$ for all i, the Perron-Frobenius theory [5, Theorems 1.5 and 1.7] implies that W has a simple positive real eigenvalue equal to 1. Moreover, the left and right eigenspaces associated with the unit eigenvalue are both one-dimensional with the corresponding eigenvectors $\overline{w} = (w_1, \ldots, w_n)^T$ and $\mathbb{1} = (1, \ldots, 1)^T$. The magnitude of any other eigenvalue of W is strictly less than 1; therefore, we have:

$$\lim_{t\to\infty} \bar{\mathbf{a}}_t = \lim_{t\to\infty} W^t\bar{\mathbf{a}}_0 = \mathbb{1}\bar{w}^T\bar{\mathbf{a}}_0 = \mathbb{1}\mathbf{a}^n(\bar{w}), \tag{S.4}$$

which implies a consensus on the collective estimate (S.2).

In several experimental settings [6, 7, 8, 9], human participants get to revise their numerical estimates a few times only, and the collective estimate is then calculated by averaging the revised estimates. Let us denote the number of communication rounds in such a scenario by τ. Using (S.3) to model the revision of the numerical estimates, we again arrive at a model that gives the collective estimate as a convex combination of the initial estimates, $\mathbf{a}^n(\bar{w}_\tau) = \bar{w}_\tau^T\bar{\mathbf{a}}_0$, where the transposed vector of weights, $\bar{w}_\tau^T$, is given by:

$$\bar{w}_\tau^T = \frac{1}{n}\mathbb{1}^T W^\tau.$$

S1.1 ω: Parameterizing a class of networks by their centralization

Motivated by our interest in comparing the collective estimation performance of centralized and decentralized networks, we focus our attention on a class of social influence network structures for which the collective estimate can be written as follows:

$$\mathbf{a}^n(\omega) = \omega\mathbf{a}_{1,0} + (1 - \omega)\frac{1}{n}\sum_{i=1}^{n} \mathbf{a}_{i,0}, \tag{S.5}$$

where ω is a measure of influence centralization, with $\omega = 1$ representing a fully centralized social influence structure ($w_1 = 1$, and $w_2 = \ldots = w_n = 0$) and $\omega = 0$ corresponding to a fully decentralized social influence structure ($w_1 = \ldots = w_n = 1/n$). Indeed, with $\bar{w}$ as the centrality vector, the parameter ω corresponds to the Freeman centralization of the underlying network with social influence matrix W. By varying $\omega \in [0, 1]$, we can interpolate between the two extremes: full centralization, $\omega = 1$, and complete decentralization, $\omega = 0$. This class, although not encompassing, is ideal for addressing the central question of interest in this work. The networks in this class are such that one agent, $i = 1$, is distinguished with a higher influence and all others, $i > 1$, have an equal but lower influence. Networks in this class include cases of practical and empirical interest, such as star networks and circular lattices. All networks in Figure 1.B, in the main text, belong to this class.

Here, we consider the special cases of star and cycle networks with bidirectional edges, and equal weights on all edges that are incoming to the same node. Their respective social influence matrices and associated left eigenvectors are given by:

$$
W_\star = \begin{bmatrix}
1/n & 1/n & 1/n & \cdots & 1/n \\
1/2 & 1/2 & 0 & \cdots & 0 \\
1/2 & 0 & 1/2 & 0\cdots & 0 \\
\vdots & \vdots & \ddots & \vdots & 0 \\
1/2 & 0 & \cdots & 0 & 1/2
\end{bmatrix}, \quad
W_o = \begin{bmatrix}
1/3 & 1/3 & 0 & \cdots 0 & 1/3 \\
1/3 & 1/3 & 1/3 & 0\ldots & 0 \\
0 & \ddots & \ddots & & 0 \\
\vdots & & & & \\
0 & \ddots 0 & 1/3 & 1/3 & 1/3 \\
1/3 & \ldots & 0 & 1/3 & 1/3
\end{bmatrix},
$$

$$\bar{w}_\star = (n/(3n-2), 2/(3n-2), \ldots, 2/(3n-2))^T, \quad \bar{w}_o = (1/n, 1/n, \ldots, 1/n)^T.$$

Subsequently, the network centralization parameter ω in (S.5) for the star and cycle networks are given by $\omega_\star = (n-2)/(3n-2)$ and $\omega_o = 0$. Note that as $n \to \infty$, $\omega_\star \to 1/3$. This, together with the fact that experimental studies have used the star topology to test collective estimation in centralized structures [7], motivates our choice of $\omega = 1/3$ in our numerical simulations and empirical analysis. In section S4, we show that our results are robust to this choice of ω.

Although our main proposition in section S2.1 gives a lower bound that is valid for any n, most of the subsequent analysis concerns the limiting behavior of the collective estimates as $n \to \infty$. Our results extend to networks with a finite (non-increasing) collection of influ-

ential agents. To accommodate such cases, one would replace $\mathbf{a}_{1,0}$ in (S.5) with the (possibly weighted) average of the initial estimates of the k influential agents, for some fixed constant k. One can similarly consider generalizations where the remaining, non-influential agents have unequal — but all vanishing — weights (going to zero as $n \to \infty$).

S1.2 Ω: A proposed feature for the estimation context

We consider a case where agents are randomly placed in the social influence network. This is typical of many experimental setups [8, 7, 6]. Let $\mathbb{E}^\theta_{\mu,\sigma}$ be the expectation with respect to the random draws of the n i.i.d. initial estimates, $\mathbf{a}_{1,0}, \ldots, \mathbf{a}_{n,0} \sim \mathcal{F}^\theta_{\mu,\sigma}$, and let $\mathbb{P}^\theta_{\mu,\sigma}$ be the corresponding probability measure. The expected mean square root error, mean absolute error, and mean squared error of the collective estimate are given by:

$$\text{MSRE}_n(\bar{w}, \mathcal{F}^\theta_{\mu,\sigma}) := \mathbb{E}^\theta_{\mu,\sigma}\left[|\mathbf{a}^n(\bar{w}) - \theta|^{1/2}\right],$$

$$\text{MAE}_n(\bar{w}, \mathcal{F}^\theta_{\mu,\sigma}) := \mathbb{E}^\theta_{\mu,\sigma}\left[|\mathbf{a}^n(\bar{w}) - \theta|\right],$$

$$\text{MSE}_n(\bar{w}, \mathcal{F}^\theta_{\mu,\sigma}) := \mathbb{E}^\theta_{\mu,\sigma}\left[(\mathbf{a}^n(\bar{w}) - \theta)^2\right].$$

In order to investigate the interaction between the network structure and the distribution of the initial estimates, (i.e. the *estimation context*: a population of agents performing a particular estimation task), we propose the following measure of how the collective estimate $\mathbf{a}^n(\bar{w}) = \bar{w}^T \bar{\mathbf{a}}_0$ performs against a fully decentralized aggregate $\mathbf{a}^n(0)$:

$$\Omega_n(\overline{w}, \mathcal{F}^\theta_{\mu,\sigma}) := \mathbb{P}^\theta_{\mu,\sigma}[|\mathbf{a}^n(\overline{w}) - \theta| < |\mathbf{a}^n(0) - \theta|],$$

where $\overline{w}$ is the centrality vector in (S.4). Restricting attention to the class of networks in subsection S1.1, with $\mathbf{a}^n(\omega) = \omega \mathbf{a}_{1,0} + (1 - \omega)\frac{1}{n}\sum_{i=1}^n \mathbf{a}_{i,0}$, we can write:

$$\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) = \mathbb{P}^\theta_{\mu,\sigma}[|\mathbf{a}^n(\omega) - \theta| < |\mathbf{a}^n(0) - \theta|].$$

This measure, $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$, corresponds to the probability that a network with social influence centralization $\omega > 0$ outperforms a decentralized network with $\omega = 0$, in absolute error perfor-

mance. Similarly, for other performance measures, we can write:

$$\text{MSRE}_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta) := \mathbb{E}_{\mu,\sigma}^\theta \left[|\mathbf{a}^n(\omega) - \theta|^{1/2} \right], \tag{S.6}$$

$$\text{MAE}_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta) := \mathbb{E}_{\mu,\sigma}^\theta \left[|\mathbf{a}^n(\omega) - \theta| \right],$$

$$\text{MSE}_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta) := \mathbb{E}_{\mu,\sigma}^\theta \left[(\mathbf{a}^n(\omega) - \theta)^2 \right].$$

Our focus throughout the paper will be on Ω_n, which we propose as a critical feature of the estimation context, to capture its suitability to benefit from centralization.

In section S2, we present a theoretical and numerical analysis of the properties of our proposed feature, Ω_n. In particular, we show how the behavior of Ω_n varies with the estimation context, i.e. the distribution of the initial estimates. In section S2.1, we present a theoretical lower bound on $\Omega_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta)$ and analyze its behavior for various classes of distributions, $\mathcal{F}_{\mu,\sigma}^\theta$. In section S2.2, we supplement these findings by numerical analysis and simulations.

In section S3, we demonstrate the explanatory power of Ω_n in determining how and when social influence improves the collective estimation accuracy in prior empirical studies. In particular, in experimental conditions with no social interactions, an external observers polls each of the participants for their opinions. Therefore, in the absence of social influence, the aggregate is given by $\mathbf{a}^n(0) = (1/n) \sum_{i=1}^n \mathbf{a}_{i,0}$, which is equivalent to a fully decentralized influence structure. On the other hand, in the presence of social influence, the participants revise their estimates as a result of their social interactions, thus leading to an aggregate that is a weighted average of the initial estimate, $\mathbf{a}^n(\bar{w}) = \sum_{i=1}^n w_i \mathbf{a}_{i,0}$. Hence, social interaction leads to a collective estimate that is less decentralized. In our model, we capture this case by $\mathbf{a}^n(\omega), \omega > 0$. We present our empirical results in the main text for $\omega = 1/3$ and $n = 50$. In section S4 we show that our results are robust to our choices of ω and n.

Our proposed feature of the estimation context, Ω_n, is concerned only with the probability of the following event: the collective estimate generated by the agents interacting in a centralized influence structure will be closer to the truth than the collective estimate generated by the agents in a decentralized structure. This is not the same as comparing the expected loss or error magnitudes. In section S4, Figure S4, we show the results for various loss function choices.

S2 Theoretical analysis of Ω

In this section, we first propose a lower bound on $\Omega_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta)$ in subsection S2.1, followed by analyses of its behavior for different distributions in five sub-subsections: these are the Pareto, S2.1.1, log-Laplace, S2.1.2, log-normal, S2.1.3, other heavy-tailed, S2.1.4, and thin-tailed distributions, S2.1.5. In subsection S2.2, we describe the procedure for direct numerical simulations of Ω for these various classes and offer additional numerical insights.

S2.1 The main proposition

Motivated by empirical literature that pose estimation questions to human participants, we focus on $0 < \theta$ and distributions $\mathcal{F}_{\mu,\sigma}^\theta$ with support over positive reals. Fix $0 < \theta$, $0 < \omega < 1$, $\theta/(1-\omega) < \beta$, and consider the event $\mathcal{E}_1 = \{\mathbf{a}_{1,0} < \beta\}$. Note that for many distributions we can make $\mathbb{P}_{\mu,\sigma}^\theta(\mathcal{E}_1)$ arbitrarily close to one by taking β large enough. Next consider the event $\mathcal{E}_n = \{\mathbf{a}^n(0) > \beta + \mathbf{a}_{1,0}/n\}$. Note that $\mathcal{E}_n$ implies $\mathbf{a}^n(0) > \beta$; furthermore $\mathcal{E}_1$ and $\mathcal{E}_n$ are independent events. On the other hand, conditioned on the events $\mathcal{E}_1$ and $\mathcal{E}_n$, we have:

$$
\begin{aligned}
|\mathbf{a}^n(\omega) - \theta| = |\omega\mathbf{a}_{1,0} + (1-\omega)\mathbf{a}^n(0) - \theta| &< \omega\mathbf{a}_{1,0} + |(1-\omega)\mathbf{a}^n(0) - \theta| \\
&< \omega\beta + |(1-\omega)\mathbf{a}^n(0) - \theta| \\
&= \omega\beta + (1-\omega)\mathbf{a}^n(0) - \theta < \omega\mathbf{a}^n(0) + (1-\omega)\mathbf{a}^n(0) - \theta = |\mathbf{a}^n(0) - \theta|,
\end{aligned}
$$

where in the second line we have used $\beta > \mathbf{a}_{1,0}$, and in the third line we have used $\mathbf{a}^n(0) > \beta$ and $(1-\omega)\mathbf{a}^n(0) > (1-\omega)\beta > \theta$. Hence, conditioned on $\mathcal{E}_1 \cap \mathcal{E}_n$, we have $|\mathbf{a}^n(\omega) - \theta| < |\mathbf{a}^n(0) - \theta|$, i.e., centralized networks outperform decentralized ones. We can thus bound $\Omega_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta)$, the probability that a social influence network with centralization ω outperforms a decentralized one ($\omega = 0$) in absolute error measure:

$$
\Omega_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta) \geq \mathbb{P}_{\mu,\sigma}^\theta[\mathcal{E}_1 \cap \mathcal{E}_n] = \mathbb{P}_{\mu,\sigma}^\theta[\mathcal{E}_1]\mathbb{P}_{\mu,\sigma}^\theta[\mathcal{E}_n]. \tag{S.7}
$$

Denoting the cumulative distribution $F^\theta_{\mu,\sigma}(x) := \mathbb{P}^\theta_{\mu,\sigma}[\mathbf{a}_{1,0} \le x]$ and the tail probability $\bar{F}^\theta_{\mu,\sigma}(x) := \mathbb{P}^\theta_{\mu,\sigma}[\mathbf{a}_{1,0} > x]$, we can write $\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_1] = \mathbb{P}^\theta_{\mu,\sigma}[\mathbf{a}_{1,0} \le \beta] = F^\theta_{\mu,\sigma}(\beta)$, and

$$\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_n] = \mathbb{P}^\theta_{\mu,\sigma}[\textstyle\sum_2^n \mathbf{a}_{i,0} > n\beta] \ge \mathbb{P}^\theta_{\mu,\sigma}[\max_{2,\dots,n} \mathbf{a}_{i,0} > n\beta] = 1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1}. \quad \text{(S.8)}$$

Using $\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_1]\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_n] \ge F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1})$, together with (S.7), we arrive at our main proposition:

Proposition. $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) \ge \sup_{\beta > \theta/(1-\omega)} \left\{ F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1}) \right\}.$

To proceed, let us denote the lower bound:

$$\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) := \sup_{\beta > \theta/(1-\omega)} \left\{ F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1}) \right\}. \quad \text{(S.9)}$$

Some observations are now in order regarding the behavior of the lower bound, $\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$. First note that $\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$ is decreasing in θ and ω. Secondly, the asymptotic behavior of $\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$, as $n \to \infty$, is determined by $F^\theta_{\mu,\sigma}(n\beta)^{n-1}$. Note that $F^\theta_{\mu,\sigma}(n\beta) \le 1$ and $F^\theta_{\mu,\sigma}(\beta) \to 1$ as $n \to \infty$ for any $\beta > 0$. Therefore, if $F^\theta_{\mu,\sigma}(n\beta) \to 1$ at a slow enough rate such that $F^\theta_{\mu,\sigma}(n\beta)^{n-1}$ is bounded away from one for all n, then $F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1})$ is bounded away from zero as $n \to \infty$. Subsequently, for various classes of distributions with slow enough tail decay, we can establish nontrivial lower bounds $\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) > 0$.

In subsections S2.1.1 to S2.1.3, we give conditions on the estimation context, i.e., the distribution of the initial estimates (parameterized by μ, σ, and θ), of well-known heavy-tailed distributions such that $\underline{\Omega}_n, \Omega_n \to 1$ as $n \to \infty$. In subsection S2.1.4, we discuss the general properties of heavy-tailed distributions that make them relevant to our proposition. Finally, in subsection S2.1.5 we present countervailing arguments for thin-tailed distributions.

S2.1.1 Pareto (power-law)

Pareto or power-law distributions are archetypal, heavy-tailed distributions characterized by their polynomial tail decay. Consider a Pareto distribution with location parameter θe^μ and

shape parameter σ, defined as follows:

$$\bar{F}_{\mu,\sigma}^{\theta}(x) = (\theta e^{\mu}/x)^{1/\sigma}, \text{ for } x \geq \theta e^{\mu}. \tag{S.10}$$

For Pareto distributions we have:

$$F_{\mu,\sigma}^{\theta}(\beta)(1 - F_{\mu,\sigma}^{\theta}(n\beta)^{n-1}) = \left(1 - \left(\frac{\theta e^{\mu}}{\beta}\right)^{1/\sigma}\right)\left(1 - \left(1 - \left(\frac{\theta e^{\mu}}{n\beta}\right)^{1/\sigma}\right)^{n-1}\right)$$

$$\asymp \left(1 - \left(\frac{\theta e^{\mu}}{\beta}\right)^{1/\sigma}\right)\left(1 - e^{-n(\theta e^{\mu}/n\beta)^{1/\sigma}}\right), \tag{S.11}$$

where we used the $n \to \infty$ asymptotic equality

$$e^{-n(\theta e^{\mu}/n\beta)^{1/\sigma}} \asymp \left(1 - \left(\frac{\theta e^{\mu}}{n\beta}\right)^{1/\sigma}\right)^{n-1}. \tag{S.12}$$

We now consider the three distinct $n \to \infty$ limiting behavior that arise for $\sigma > 1$, $\sigma = 1$, and $\sigma < 1$:

- For $\sigma > 1$, as $n \to \infty$ we get

$$\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,\sigma}^{\theta}) \geq 1 - (\theta e^{\mu}/\beta)^{1/\sigma} \text{ for all } \beta > \theta/(1 - \omega),$$

  and letting $\beta \to \infty$ we conclude that

$$\Omega_n(\omega, \mathcal{F}_{\mu,\sigma}^{\theta}) \asymp \underline{\Omega}_n(\omega, \mathcal{F}_{\mu,\sigma}^{\theta}) \asymp 1, \text{ for } \sigma > 1, \text{ and any } 0 < \theta, 0 < \omega < 1, \text{ and real } \mu.$$

- Replacing $\sigma = 1$ in (S.11), the lower bound $\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,1}^{\theta})$ can be calculated as follow:

$$\sup_{\beta > \theta/(1-\omega)} \left\{1 - \frac{\theta e^{\mu}}{\beta} - e^{-\theta e^{\mu}/\beta} + \frac{\theta e^{\mu}}{\beta}e^{-\theta e^{\mu}/\beta}\right\}$$

$$= \begin{cases} W_0(e^2) + 1/W_0(e^2) - 2 \approx 0.199, & \text{if } \mu \geq \log\left(\frac{2 - W_0(e^2)}{1-\omega}\right), \\ 1 - e^{\mu}(1-\omega) - e^{-e^{\mu}(1-\omega)} + e^{\mu}(1-\omega)e^{-e^{\mu}(1-\omega)}, & \text{if } \mu < \log\left(\frac{2 - W_0(e^2)}{1-\omega}\right). \end{cases}$$

S9

To see why, denote $x = \theta e^\mu / \beta$. The maximum of $1 - x - e^{-x} + x e^{-x}$ occurs at $x^\star$ satisfying $-1 + 2e^{-x^\star} - x^\star e^{-x^\star} = 0$. The latter has a unique solution over the positive reals given by $x^\star = 2 - W_0(e^2) \approx 0.443$, where W_0 is the principal branch Lambert W function, uniquely satisfying $W_0(e^2) e^{W_0(e^2)} = e^2$ on positive reals. Replacing $x = 2 - W_0(e^2)$ in $1 - x - e^{-x} + x e^{-x}$ gives the maximum value $W_0(e^2) + 1/W_0(e^2) - 2$ realized at $\beta = \theta e^\mu / (2 - W_0(e^2))$ if $e^\mu \geq (2 - W_0(e^2))/(1 - \omega)$. If $e^\mu < (2 - W_0(e^2))/(1 - \omega)$, then the supremum is achieved with $\beta = \theta/(1 - \omega)$ at a value that is strictly less than $W_0(e^2) + 1/W_0(e^2) - 2 \approx 0.199$.

- For $\sigma < 1$, as $n \to \infty$ we get

$$\left(1 - \left(\frac{\theta e^\mu}{\beta}\right)^{1/\sigma}\right)\left(1 - e^{-n(\theta e^\mu / n\beta)^{1/\sigma}}\right) \to 0.$$

  for any β.

We thus obtain the following asymptotic characterization of the lower bound for Pareto distributions (indicating a phase transition at $\sigma = 1$):

$$\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) \asymp \begin{cases} 0, & \text{if } \sigma < 1, \\ (1 - e^\mu(1 - \omega))(1 - e^{-e^\mu(1-\omega)}), & \text{if } \sigma = 1 \text{ and } \mu < \log\left(\frac{2 - W_0(e^2)}{1 - \omega}\right), \\ W_0(e^2) + 1/W_0(e^2) - 2 \approx 0.199, & \text{if } \sigma = 1 \text{ and } \mu \geq \log\left(\frac{2 - W_0(e^2)}{1 - \omega}\right), \\ 1, & \text{if } \sigma > 1. \end{cases}$$
(S.13)

In Figure S1.A, top, we have plotted (S.13) for $\omega = 1/3$, $\theta = 2$, and $n = 50$. Comparing with the direct numerical simulation in Figure S1.A, bottom, shows how the bound gets tighter for large σ.

S2.1.2   Log-Laplace

Jayles et al. [10] point out that log-Laplace provides a better fit to the empirically measured distribution of the initial estimates, compared to log-Cauchy [11], or log-normal. Here, we analyze the asymptotic behavior of the proposed lower bound as $n \to \infty$, when the initial estimates are distributed according to a log-Laplace distribution with parameters $\log \theta + \mu$ and

σ:

$$F_{\mu,\sigma}^{\theta}(x) = \begin{cases} \frac{1}{2}\exp\left(\frac{\log(x/\theta)-\mu}{\sigma}\right), & \text{if } \log(x/\theta) < \mu, \\ 1 - \frac{1}{2}\exp\left(\frac{\mu-\log(x/\theta)}{\sigma}\right), & \text{if } \log(x/\theta) \geq \mu. \end{cases} \tag{S.14}$$

For n large enough, we have $\log(n\beta/\theta) \geq \mu$, and using the same asymptotics as in (S.12) we get:

$$\begin{aligned} F_{\mu,\sigma}^{\theta}(n\beta)^{n-1} &= \left(1 - \frac{1}{2}\exp\left(\frac{\mu - \log(n\beta/\theta)}{\sigma}\right)\right)^{n-1} \\ &= \left(1 - \left(\frac{1}{n}\right)^{1/\sigma}\exp\left(\frac{\mu + \log\theta - \log\beta}{\sigma} - \log(2)\right)\right)^{n-1} \\ &\asymp \exp\left(-n\left(\frac{1}{n}\right)^{1/\sigma}\exp\left(\frac{\mu + \log\theta - \log\beta}{\sigma} - \log(2)\right)\right) \\ &\asymp \begin{cases} 1, & \text{if } \sigma < 1, \\ \exp\left(\frac{-e^{\mu}\theta}{2\beta}\right), & \text{if } \sigma = 1, \\ 0, & \text{if } \sigma > 1. \end{cases} \end{aligned}$$

Subsequently, for $\sigma < 1$ we have $F_{\mu,\sigma}^{\theta}(\beta)(1 - F_{\mu,\sigma}^{\theta}(n\beta)^{n-1}) \asymp 0$, for any β. On the other hand, for $\sigma > 1$ we have $F_{\mu,\sigma}^{\theta}(\beta)(1 - F_{\mu,\sigma}^{\theta}(n\beta)^{n-1}) \asymp F_{\mu,\sigma}^{\theta}(\beta)$, which is increasing in β and goes to one as β increases to ∞. Hence,

$$\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,\sigma}^{\theta}) = \sup_{\beta > \theta/(1-\omega)} \left\{F_{\mu,\sigma}^{\theta}(\beta)(1 - F_{\mu,\sigma}^{\theta}(n\beta)^{n-1})\right\} \asymp 1, \text{ for } \sigma > 1.$$

Finally, for $\sigma = 1$, we have:

$$F_{\mu,1}^{\theta}(\beta)(1 - F_{\mu,1}^{\theta}(n\beta)^{n-1}) \asymp \begin{cases} \left(1 - \exp\left(\frac{-e^{\mu}\theta}{2\beta}\right)\right)\frac{\beta}{2\theta e^{\mu}}, & \text{if } \frac{\beta}{e^{\mu}\theta} < 1, \\ \left(1 - \exp\left(\frac{-e^{\mu}\theta}{2\beta}\right)\right)\left(1 - \frac{e^{\mu}\theta}{2\beta}\right), & \text{if } \frac{\beta}{e^{\mu}\theta} \geq 1. \end{cases}$$

Optimizing β gives:

$$\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,1}^\theta) \asymp \begin{cases} \frac{1-1/\sqrt{e}}{2} \approx 0.1967, & \text{if } \mu \geq -\log(1-\omega), \\ \left(1 - e^{-e^\mu(1-\omega)/2}\right)(1 - e^\mu(1-\omega)/2) < \frac{1-1/\sqrt{e}}{2}, & \text{if } \mu < -\log(1-\omega). \end{cases}$$

We summarize the above results in the following asymptotic characterization of the lower bound for log-Laplace distributions, with a phase transition at $\sigma = 1$:

$$\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,\sigma}^\theta) \asymp \begin{cases} 0, & \text{if } \sigma < 1, \\ \left(1 - e^{-e^\mu(1-\omega)/2}\right)(1 - e^\mu(1-\omega)/2), & \text{if } \sigma = 1 \text{ and } \mu < -\log(1-\omega), \\ \frac{1-1/\sqrt{e}}{2} \approx 0.1967, & \text{if } \sigma = 1 \text{ and } \mu \geq -\log(1-\omega), \\ 1, & \text{if } \sigma > 1. \end{cases} \tag{S.15}$$

In Figure S1.B, top, we have plotted (S.15) for $\omega = 1/3$, $\theta = 2$, and $n = 50$. Comparing with the direct numerical simulation in Figure S1.B, bottom, shows how the bound gets tighter for large σ.

S2.1.3 Log-normal

Several empirical studies report a log-normal distribution for the initial estimates [7, 8, 12]. Here, we analyze the case where the initial estimates are distributed according to a log-normal distribution with parameters $\log \theta + \mu$ and σ:

$$F_{\mu,\sigma}^\theta(x) = \Phi\left(\frac{\log(x/\theta) - \mu}{\sigma}\right), x > 0, \tag{S.16}$$

where Φ is the standard normal distribution. We next apply the following control over the Gaussian tail:

$$1 - \frac{1}{\sqrt{2\pi}t}e^{-t^2/2} \leq \Phi(t) \leq 1 - \frac{t}{\sqrt{2\pi}(t^2+1)}e^{-t^2/2}$$

to obtain:

$$\left(1 - \frac{\sigma}{\sqrt{2\pi}(\log(n\beta/\theta) - \mu)} \exp\left(-\frac{(\log(n\beta/\theta) - \mu)^2}{2\sigma^2}\right)\right)^{n-1} \leq F_{\mu,\sigma}^{\theta}(n\beta)^{n-1}$$

$$\leq \left(1 - \frac{(\log(n\beta/\theta) - \mu)\sigma}{\sqrt{2\pi}((\log(n\beta/\theta) - \mu)^2 + \sigma^2)} \exp\left(-\frac{(\log(n\beta/\theta) - \mu)^2}{2\sigma^2}\right)\right)^{n-1}. \qquad \text{(S.17)}$$

We next choose $\sigma = \sigma_n$, with β and θ, fixed such that $(\log(n\beta/\theta) - \mu)/\sigma_n \to \infty$ as $n \to \infty$. Taking the limit $n \to \infty$, with $\sigma = \sigma_n$ in (S.17), we get:

$$F_{\mu,\sigma_n}^{\theta}(n\beta)^{n-1} \qquad \text{(S.18)}$$

$$\asymp \exp\left(-\frac{n\sigma_n}{\sqrt{2\pi}\log n} \exp\left(-\frac{(\log(n\beta/\theta) - \mu)^2}{2\sigma_n^2}\right)\right)$$

$$= \exp\left(-\frac{1}{\sqrt{2\pi}\log n} \exp\left(\log n + \log \sigma_n - \frac{(\log(n\beta/\theta) - \mu)^2}{2\sigma_n^2}\right)\right)$$

Focusing on the second exponent, we obtain :

$$f_n := \log n + \log \sigma_n - \frac{(\log(n\beta/\theta) - \mu)^2}{2\sigma_n^2} \qquad \text{(S.19)}$$

$$= \log n + \log \sigma_n - \frac{\log(n\beta/\theta)^2}{2\sigma_n^2} - \frac{\mu^2}{2\sigma_n^2} + \frac{\mu\log(n\beta/\theta)}{\sigma_n^2}$$

$$= \log n + \log \sigma_n - \frac{(\log n)^2}{2\sigma_n^2} - \frac{\log(\beta/\theta)\log n}{\sigma_n^2} - \frac{\log(\beta/\theta)^2}{2\sigma_n^2} - \frac{\mu^2}{2\sigma_n^2} + \frac{\mu\log(n\beta/\theta)}{\sigma_n^2}.$$

Setting $\sigma_n = \sqrt{\log n/k_n}$ in (S.19) yields:

$$f_n = (1 - k_n/2)\log n + (1/2)\log\log n + (\mu - \log(\beta/\theta))\,k_n - (1/2)\log k_n$$

$$+ (-(1/2)\log(\beta/\theta)^2 - \mu^2/2 + \mu\log(\beta/\theta))\frac{k_n}{\log n}$$

$$\asymp \begin{cases} \frac{(2-k_n)}{2}\log n, & \text{if } k_n < 2, \\ (1/2)\log\log n, & \text{if } k_n = 2, \\ \frac{(2-k_n)}{2}\log n, & \text{if } k_n > 2. \end{cases}$$

S13

Finally, substituting in (S.18), we get:

$$F_{\mu,\sigma_n}^{\theta}(n\beta)^{n-1} \asymp \exp\left(-e^{f_n}/(\sqrt{2\pi}\log n)\right) \asymp \begin{cases} 0, & \text{if } \sigma_n > \sqrt{\log n/2}, \\ 1, & \text{if } \sigma_n \leq \sqrt{\log n/2}. \end{cases}$$

The latter together with the fact that $F_{\mu,\sigma_n}^{\theta}(\beta) = \Phi\left((\log(\beta/\theta) - \mu)/\sigma_n\right) \asymp \Phi(0) = 1/2$, for $\sigma_n > \sqrt{\log n/2}$, leads to the following asymptotic characterization for log-normal distributions (indicating a phase transition at $\sigma_n = \sqrt{\log n/2}$):

$$F_{\mu,\sigma_n}^{\theta}(\beta)(1 - F_{\mu,\sigma_n}^{\theta}(n\beta)^{n-1}) \asymp \begin{cases} 0, & \text{if } \sigma_n \leq \sqrt{\log n/2}, \\ 1/2, & \text{if } \sigma_n > \sqrt{\log n/2}, \end{cases}$$

which is true for all β, and in particular yields:

$$\begin{aligned}
\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,\sigma}^{\theta}) &= \sup_{\beta > \theta/(1-\omega)} \left\{ F_{\mu,\sigma}^{\theta}(\beta)(1 - F_{\mu,\sigma}^{\theta}(n\beta)^{n-1}) \right\} \\
&\asymp \begin{cases} 0, & \text{if } \sigma_n \leq \sqrt{\log n/2}, \\ 1/2, & \text{if } \sigma_n > \sqrt{\log n/2}. \end{cases}
\end{aligned} \tag{S.20}$$

In Figure S1.C, top, we have plotted (S.20) for $\omega = 1/3$, $\theta = 2$, and $n = 50$. Comparing with the result of the direct numerical simulation, Figure S1.C, bottom, shows how the bound gets tighter for larger σ.

S2.1.4 Other heavy-tailed distributions

Many empirical studies [8, 11, 12, 13, 14] point out a heavy-tailed distribution for the numerical estimates (with a few estimates that fall on a fat right tail). Following the proof of the main proposition, we pointed out that for heavy-tailed distributions where $\bar{F}_{\mu,\sigma}^{\theta}(n\beta)$ decreases slowly, we can provide non-trivial lower bounds on Ω_n that remain bounded away from zero, $\underline{\Omega}_n(\omega, \mathcal{F}_{\mu,\sigma}^{\theta}) > 0$, even as $n \to \infty$. In fact, if $\bar{F}_{\mu,\sigma}^{\theta}(n\beta)$ decreases at a rate that is slower than $1/n$, i.e., $n\bar{F}_{\mu,\sigma}^{\theta}(n\beta) \to \infty$, then $F_{\mu,\sigma}^{\theta}(n\beta)^{n-1} \to 0$ as $n \to \infty$. For such slowly decaying tails, the supremum in (S.9) is achieved as $\beta \to \infty$, and we can guarantee that $\underline{\Omega}_n \asymp \Omega_n \asymp 1$; hence, the proposed lower bound is asymptotically tight.

S14

Here, we identify a second way, in which, our proposed lower bound is tighter for heavy-tailed distributions. To this end, let us revisit (S.8) — a critical step in deriving the proposed lower bound:

$$\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_n] = \mathbb{P}^\theta_{\mu,\sigma}\big[\textstyle\sum_{i=2}^n \mathbf{a}_{i,0} > n\beta\,\big] > \mathbb{P}^\theta_{\mu,\sigma}\big[\max_{2,\ldots,n} \mathbf{a}_{i,0} > n\beta\,\big].$$

This inequality is at the heart of the so-called "catastrophe principle" [15, Chapter 3] that applies to many heavy-tailed distributions. Intuitively, this principle entails that when one observes a larger than expected average value for a collection of heavy-tailed random variables, then this observation is most likely explained by the existence of a very large sample in the collection, i.e. a "catastrophe". On the other hand, the countervailing explanation in the case of light-tailed random variables is that "most" of the samples in the collection happen to be larger than expected. Formally, the distribution $\mathcal{F}^\theta_{\mu,\sigma}$ of the initial estimates is said to satisfy the catastrophe principle [15, Definition 3.1], if for any n:

$$\lim_{t\to\infty} \frac{\mathbb{P}^\theta_{\mu,\sigma}\big[\max_{1,\ldots,n} \mathbf{a}_{i,0} > t\,\big]}{\mathbb{P}^\theta_{\mu,\sigma}\big[\sum_{i=1}^n \mathbf{a}_{i,0} > t\,\big]} = 1.$$

The preceding condition is equivalent to having:

$$\lim_{t\to\infty} \frac{n\mathbb{P}^\theta_{\mu,\sigma}[\mathbf{a}_{1,0} > t]}{\mathbb{P}^\theta_{\mu,\sigma}[\sum_{i=1}^n \mathbf{a}_{i,0} > t]} = 1,\ \text{ for all } n \geq 2.$$

The latter is the defining property for the subexponential family of distributions, which include many common classes of heavy-tailed distributions such as those considered in subsections S2.1.1 to S2.1.3. Setting $t = n\beta$ and letting $n \to \infty$, we obtain that if $\mathcal{F}^\theta_{\mu,\sigma}$ is a member of the subexponential family, then

$$\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_n] = \mathbb{P}^\theta_{\mu,\sigma}\big[\textstyle\sum_{i=2}^n \mathbf{a}_{i,0} > n\beta\,\big] \asymp \mathbb{P}^\theta_{\mu,\sigma}\big[\max_{2,\ldots,n} \mathbf{a}_{i,0} > n\beta\,\big].$$

Hence, for such distributions belonging to the subexponential family our proposed lower bound is asymptotically tight in as much as $\mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_1 \cap \mathcal{E}_n] \asymp F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1})$, and the only way in which our lower bound may be loose is through (S.7), i.e. if $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) > \mathbb{P}^\theta_{\mu,\sigma}[\mathcal{E}_1 \cap \mathcal{E}_n]$ for all $\beta > \theta/(1-\omega)$ as $n \to \infty$.

It is worth noting that many light-tailed distributions portray an opposite picture, referred to as "conspiracy principle" in [15, Definition 3.2]; formally defined as follows:

$$\lim_{t \to \infty} \frac{\mathbb{P}^{\theta}_{\mu,\sigma}[\max_{1,\dots,n} \mathbf{a}_{i,0} > t]}{\mathbb{P}^{\theta}_{\mu,\sigma}[\sum_{i=1}^{n} \mathbf{a}_{i,0} > t]} = 0, \text{ for all } n \geq 2.$$

As an example, suppose that the initial estimates are exponentially distributed with mean θe^{μ} and the following tail probability:

$$\bar{F}^{\theta}_{\mu,\sigma}(x) = e^{-x/\theta e^{\mu}}, x > 0.$$

Then their sum follows an Erlang distribution, satisfying:

$$\mathbb{P}^{\theta}_{\mu,\sigma}[\sum_{2}^{n} \mathbf{a}_{i,0} > n\beta] = \sum_{k=0}^{n-2} \frac{e^{-n\beta/(\theta e^{\mu})}}{k!} \left(\frac{n\beta}{\theta e^{\mu}}\right)^{k},$$

such that

$$\mathbb{P}^{\theta}_{\mu,\sigma}[\max_{2,\dots,n} \mathbf{a}_{i,0} > n\beta] \asymp (n-1)\mathbb{P}^{\theta}_{\mu,\sigma}[\mathbf{a}_{1,0} > n\beta] = (n-1)e^{-n\beta/(\theta e^{\mu})}$$

$$\ll \sum_{k=0}^{n-2} \frac{e^{-n\beta/(\theta e^{\mu})}}{k!} \left(\frac{n\beta}{\theta e^{\mu}}\right)^{k} = \mathbb{P}^{\theta}_{\mu,\sigma}[\sum_{i=2}^{n} \mathbf{a}_{i,0} > n\beta].$$

and

$$\frac{\mathbb{P}^{\theta}_{\mu,\sigma}[\max_{2,\dots,n} \mathbf{a}_{i,0} > n\beta]}{\mathbb{P}^{\theta}_{\mu,\sigma}[\sum_{i=2}^{n} \mathbf{a}_{i,0} > n\beta]} \asymp 0.$$

S2.1.5 Distributions with strong tail decay and classical accounts of the wisdom of crowds

It is instructive to investigate the behavior of the lower bound for light-tailed distributions as well. Sub-Gaussian distributions are a class of probability distribution with strong tail decay (at least as fast as a Gaussian). Suppose $\mathbf{x}$ is a random variable with mean $\mu + \theta$ and cumulative distribution $F^{\theta}_{\mu,\sigma}$. Furthermore, suppose that $\mathbf{x} - \mu - \theta$ is sub-Gaussain with variance-proxy

parameter σ, thereby, satisfying:

$$\bar{F}^\theta_{\mu,\sigma}(n\beta) = \mathbb{P}^\theta_{\mu,\sigma}[\mathbf{x} > n\beta] \leq e^{-(n\beta - \mu - \theta)^2/2\sigma^2}$$

On the other hand, we have $(1 + F^\theta_{\mu,\sigma}(n\beta) + F^\theta_{\mu,\sigma}(n\beta)^2 + \ldots + F^\theta_{\mu,\sigma}(n\beta)^{n-2}) \leq n - 1$, which we can combine with the above to get that

$$F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta)^{n-1})$$
$$= F^\theta_{\mu,\sigma}(\beta)(1 - F^\theta_{\mu,\sigma}(n\beta))(1 + F^\theta_{\mu,\sigma}(n\beta) + F^\theta_{\mu,\sigma}(n\beta)^2 + \ldots + F^\theta_{\mu,\sigma}(n\beta)^{n-2})$$
$$\leq nF^\theta_{\mu,\sigma}(\beta) \exp\left(-\frac{(n\beta - \mu - \theta)^2}{2\sigma^2}\right) \to 0, \text{ as } n \to \infty, \text{ for any } \beta > 0.$$

Therefore, there are no set of parameters μ and σ that lead to a non-trivial, asymptotic lower bound on Ω_n for random variables with sub-Gaussian tails: $\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) \asymp 0$, for all θ, μ, σ. As an example, consider the folded Gaussian distribution, which is defined as the absolute value of a normally distributed random variable with mean θe^μ and variance σ:

$$F^\theta_{\mu,\sigma}(x) = \Phi\left(\frac{x - \theta e^\mu}{\sigma}\right) + \Phi\left(\frac{x + \theta e^\mu}{\sigma}\right) - 1, x > 0, \tag{S.21}$$

where Φ is the standard normal distribution. In Figure S1.D, bottom, we have plotted $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$ with $\omega = 1/3$, $\theta = 2$, $n = 50$, and initial estimates following a folded-Gaussian distribution. There are no range of distribution parameters, μ and σ, for which Ω_n increases above 0.6. Indeed, for such light-tailed distributions, admitting finite first and second moments, we can show that the limiting expected absolute error of the collective estimate with centralization ω, $\mathbf{a}^n(\omega)$, is higher than the decentralized baseline, $\mathbf{a}^n(0)$.

Consider the case where $\mathcal{F}^\theta_{\mu,\sigma}$ admits the following finite first and second moments: $\mathbb{E}^\theta_{\mu,\sigma}[\mathbf{a}_{1,0}] = \theta e^\mu$, and $\mathbb{E}^\theta_{\mu,\sigma}[\mathbf{a}^2_{1,0}] = \theta^2 e^{2\mu} + \sigma^2$. Then $\mathbf{a}^n(0) \to \theta e^\mu$, and $\mathbf{a}^n(\omega) \to \omega \mathbf{a}_{1,0} + (1 - \omega)\theta e^\mu$, both almost surely, as $n \to \infty$. Hence,

$$\mathbb{E}^\theta_{\mu,\sigma}[|\mathbf{a}^n(\omega) - \theta|] > |\mathbb{E}^\theta_{\mu,\sigma}[\mathbf{a}^n(\omega) - \theta]| = |\theta(1 - e^\mu)| \asymp \mathbb{E}^\theta_{\mu,\sigma}[|\mathbf{a}^n(0) - \theta|].$$

We can repeat the same calculations for the expected mean squared errors as well:

$$\mathbb{E}^{\theta}_{\mu,\sigma}[(\mathbf{a}^n(0) - \theta)^2] \to \theta^2(1 - e^\mu)^2, \text{ and}$$

$$\mathbb{E}^{\theta}_{\mu,\sigma}[(\mathbf{a}^n(\omega) - \theta)^2] \to \theta^2(1 - e^\mu)^2 + \omega^2\sigma^2 > \theta^2(1 - e^\mu)^2 \asymp \mathbb{E}^{\theta}_{\mu,\sigma}[(\mathbf{a}^n(0) - \theta)^2].$$

For distributions with light tails decentralized networks outperform centralized ones, in expectation for absolute and squared errors, for any choice of parameters μ and σ. This verifies the classical accounts of the wisdom of crowds, whereby the law of large numbers guarantees almost sure convergence of the collective estimate for structures with vanishing individual influences [16, Proposition 2].

Indeed, among all convex combinations the simple average, $\mathbf{a}^n(0)$, has the minimum variance. Since all estimators in our class, $\mathbf{a}^n(\omega), \omega \geq 0$, have the same expected value, the simple average, $\mathbf{a}^n(0)$, is, in fact, the mean squared error (MSE) minimizer in this class. In this subsection, we point out that even though the variance/MSE of $\mathbf{a}^n(\omega)$ is minimum when $\omega = 0$, $\mathbf{a}^n(\omega)$, with $\omega > 0$ fixed, can "often" outperform $\mathbf{a}^n(0)$, i.e., fall closer to the truth, θ. We show this by lower bounding the probability, Ω_n, that $|\mathbf{a}^n(\omega) - \theta| < |\mathbf{a}^n(0) - \theta|$. Subsequently, we identify heavy-tailedness conditions that make this event likely and Ω_n large.

It is worth highlighting that although for $\mu = 0$, the sample mean is a minimum-variance, unbiased estimator (MVUE) of the truth, it is well-known in statistics that simple average is not "optimal" for closeness in many situations. For instance, in the presence of outliers or extreme values the sample median is preferred to sample mean due to its robustness properties [17].

In subsection S2.2, we discuss direct numerical simulation of the value of Ω_n for various distribution classes. In sub-subsection S2.2.1, we identify other comparable right-skewness conditions for making Ω_n large, e.g., greater than $1/2$, by analyzing the location of the medians of the two estimators, $\mathbf{a}^n(\omega), \omega > 0$ and $\mathbf{a}^n(0)$. In section S3, we show that the feature, Ω, that we identify from this theory has significant explanatory power for determining whether experimentally measured collective estimation outcomes improve, after group members interact with each other.

S2.2 Numerical simulations

For numerical simulations, we have fixed $\theta = 2$, $\omega = 1/3$, and $n = 50$. The choice of $\omega = 1/3$ is arbitrary and our conclusions remain valid for $\omega > 0$, as verified by the robustness

checks in section S4. This choice is motivated by our observation in subsection S1.1 that ω for a star network converges to $1/3$ as $n \to \infty$. This also allows us to juxtapose our simulations with common experimental setups that use the star topology as archetypal of centralized structures [7].

Note that with $\omega = 1/3$ fixed, the dependence of Ω_n on the network structure is removed. Therefore, $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$ is entirely determined by the distribution of the initial estimates, $\mathcal{F}^\theta_{\mu,\sigma}$, i.e. the estimation context. Here, we study our proposed task feature, Ω_n, numerically for a palette of empirically relevant distributions.

For any distribution of the initial estimates, $\mathcal{F}^\theta_{\mu,\sigma}$, and number of agents, n, we calculate Ω_n using a Monte Carlo method. We sample n initial estimates and calculate the collective estimates, $\mathbf{a}^n(1/3)$ and $\mathbf{a}^n(0)$, using equation (S.5). If $\mathbf{a}^n(1/3)$ is closer to the truth, θ, than $\mathbf{a}^n(0)$, implying that a centralized network performed better than a decentralized network, then we add to our tally of Ω_n. We repeat this procedure N times, where N is large enough to allow for the value of Ω_n to converge (see the simulation procedure 1). The results in Figure S1 are obtained in this manner with $\theta = 2$, $n = 50$, and $N = 10,000$ for four different distributions: Pareto (S.10), log-Laplace (S.14), log-normal (S.16), and folded-Gaussian (S.21).

Procedure 1: $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$ computation

---

Input: $\mathcal{F}_\theta(\mu, \sigma)$, n, ω, N
Output: $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$
Initialize $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) = 0$.
for $j = 1 : N$ **do**
    Sample $\mathbf{a}_{1,0}, \ldots, \mathbf{a}_{n,0} \sim \mathcal{F}_\theta(\mu, \sigma)$.
    Compute $\mathbf{a}^n(0)$ and $\mathbf{a}^n(\omega)$ according to (S.5).
    if $|\mathbf{a}^n(\omega) - \theta| < |\mathbf{a}^n(0) - \theta|$ **then**
        Update $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) = \Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma}) + 1/N$.
    end
end

S2.2.1 The effect of the systematic bias, μ

To see the effect of the log-normal distribution parameters, μ and σ, in a different light, it is instructive to study the behavior of the median of the random variable $\mathbf{a}^n(\omega)$. In particular, we are interested in the location of $\text{Median}[\mathbf{a}^n(\omega)]$ with respect to the truth θ, as the distribution
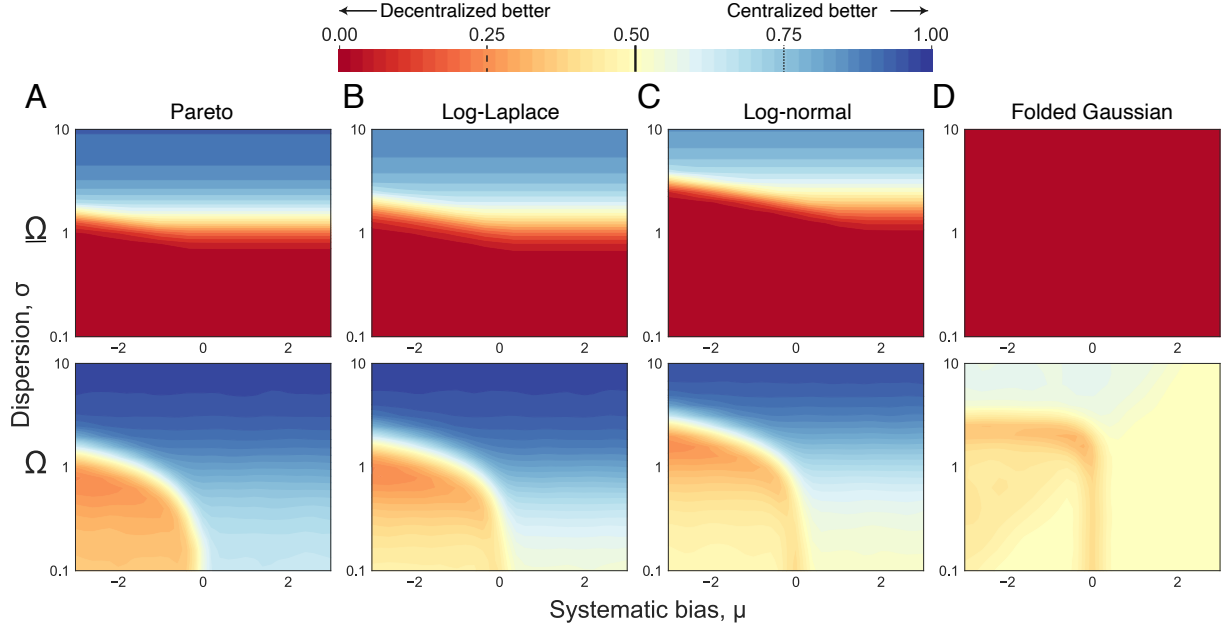
Figure S1: Simulating the lower bound $\underline{\Omega}_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$ and the actual value $\Omega_n(\omega, \mathcal{F}^\theta_{\mu,\sigma})$, under different distributions for the initial estimates: Pareto (S.10), log-Laplace (S.14), log-normal (S.16), and folded Gaussian (S.21). In all the plots, we have fixed $\omega = 1/3$, $\theta = 2$, and $n = 50$.

parameter μ is varied. We do so in the limit of large group sizes, $n \to \infty$. Note that since log-normal distributions have finite moments, the strong law of large numbers applies. Hence, as $n \to \infty$, $\mathbf{a}^n(0)$ converges almost surely to

$$\mathbb{E}[\mathbf{a}_{i,0}] = \exp(\log\theta + \mu + \sigma^2/2)$$

In particular we also have that

$$\lim_{n\to\infty} \text{Median}[\mathbf{a}^n(0)] = \exp(\log\theta + \mu + \sigma^2/2).$$

On the other hand, note that for $\omega > 0$, $\mathbf{a}^n(\omega) = \omega\mathbf{a}_{1,0} + (1-\omega)\mathbf{a}^n(0)$. Hence, as $n \to \infty$, $\mathbf{a}^n(\omega) \to \omega\mathbf{a}_{1,0} + \exp(\log\theta + \mu + \sigma^2/2)$, almost surely. Therefore,

$$\lim_{n\to\infty} \text{Median}[\mathbf{a}^n(\omega)] = \omega\,\text{Median}[\mathbf{a}_{i,0}] + (1-\omega)\exp(\log\theta + \mu + \sigma^2/2)$$

$$= \omega\exp(\log\theta + \mu) + (1-\omega)\exp(\log\theta + \mu + \sigma^2/2),$$

where we have used the fact that $\text{Median}[\mathbf{a}_{i,0}] = \exp(\log\theta + \mu)$. Next, we note that depending on where the distributional parameters, θ, σ, and μ, are located, three cases may arise:

- If $\exp(\log\theta+\mu+\sigma^2/2) < \theta$, or equivalently, $\mu < -\sigma^2/2$, then $\lim_{n\to\infty} \text{Median}[\mathbf{a}^n(\omega)] < \lim_{n\to\infty} \text{Median}[\mathbf{a}^n(0)] < \theta$. In this case, as $n \to \infty$, at least half of the time, $\mathbf{a}^n(0)$ falls closer to θ than $\mathbf{a}^n(\omega)$, hence, $\lim_{n\to\infty} \Omega_n < 1/2$; see Figure S2.A.

- If $\exp(\log\theta + \mu) < \theta < \exp(\log\theta + \mu + \sigma^2/2)$, or equivalently, $-\sigma^2/2 < \mu < 0$, then $\lim_{n\to\infty} \text{Median}[\mathbf{a}^n(\omega)] < \theta < \lim_{n\to\infty} \text{Median}[\mathbf{a}^n(0)]$. In this case, the limiting value of Ω_n may be less then, or greater than, but is close to $1/2$; see Figure S2.B.

- If $\theta < \exp(\log\theta+\mu+\sigma^2/2)$, or equivalently, $0 < \mu$, then $\theta < \lim_{n\to\infty} \text{Median}[\mathbf{a}^n(\omega)] < \lim_{n\to\infty} \text{Median}[\mathbf{a}^n(0)]$. In this case, as $n \to \infty$, at least half of the time, $\mathbf{a}^n(\omega)$ falls closer to θ than $\mathbf{a}^n(0)$, hence, $\lim_{n\to\infty} \Omega_n > 1/2$; see Figure S2.C.
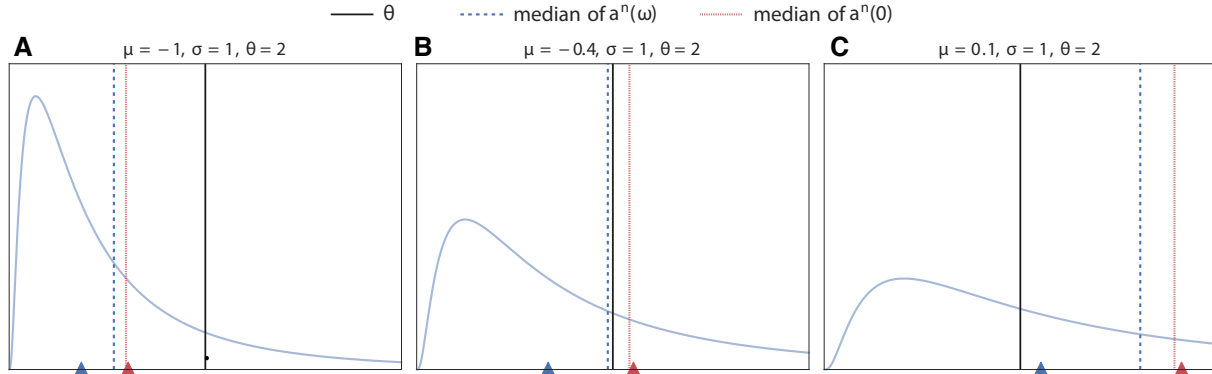

Figure S2: Simulating the medians of $\mathbf{a}^n(\omega)$ and $\mathbf{a}^n(0)$ for three different values of the systematic bias, μ. The distribution median, marked in blue on the x-axis, lags the distribution mean, marked in red. Subsequently, the median of $\mathbf{a}^n(\omega)$ is always less than the median of $\mathbf{a}^n(0)$ for the distributions studied here. In this framing, there are three different levels of bias: panel **A**, when the distribution of initial estimates significantly under-estimates the truth, $\mu < -\sigma^2/2$, then the median of $\mathbf{a}^n(0)$ is closer to the truth, θ, than the median of $\mathbf{a}^n(\omega)$, in this case, $\Omega_n < 1/2$; panel **B**, when the distribution of initial estimates slightly under-estimates the truth, $-\sigma^2/2 < \mu < 0$, then the truth lies between the medians of $\mathbf{a}^n(\omega)$ and $\mathbf{a}^n(0)$, and $\Omega \approx 0.5$; panel **C**, when the distribution of initial estimates over-estimates the truth, $\mu > 0$, then the median $\mathbf{a}^n(\omega)$ is closer to the truth, leading to $\Omega_n > 1/2$. In these simulations, we have fixed $\omega = 1/3$, $\theta = 2$, $n = 50$, and $N = 10{,}000$, where N is the number of samples used to simulate the median values numerically.

Finally, it is worth noting that a similar argument applies to any right-skewed and heavy-tailed distribution, for which the population mean exists and is greater than the population median.

S3 Empirical analysis of estimation contexts in prior work

To empirically illustrate the explanatory power of this theory, we use data from four published experiments [8, 18, 7, 6], in which $4,002$ participants organized into 131 independent groups completed a total of 57 estimation tasks generating $20,030$ individual estimations and 815 collective estimations. Each task induces a different distribution on the initial estimates that are measured empirically. Therefore, each task constitutes an estimation context in our framework and we have a total of 57 estimation contexts. We calculate Ω directly from the data based on the empirical initial estimates without the need for making any distributional assumptions, i.e., sampling initial estimates with replacement from the data. The computational procedure is as follows ($\omega = 1/3$, $n = 50$, $N = 10,000$):

Procedure 2: $\Omega_n(\omega, \text{experimental task})$ computation

Input: experimental initial estimates, n, ω, N

Output: $\Omega_n(\omega, \text{experimental task})$

Initialize $\Omega_n(\omega, \text{experimental task}) = 0$.

for $j = 1 : N$ **do**
    Sample $\mathbf{a}_{1,0}, \ldots, \mathbf{a}_{n,0}$ with replacement from the experimental initial estimates.
    Compute $\mathbf{a}^n(0)$ and $\mathbf{a}^n(\omega)$ according to (S.5).
    if $|\mathbf{a}^n(\omega) - \theta| < |\mathbf{a}^n(0) - \theta|$ **then**
        | Update $\Omega_n(\text{experimental task}) = \Omega_n(\text{experimental task}) + 1/N$.
    end
end

S3.1 Regression Analysis

Our main empirical analyses, shown in Figure 3 and Table S1, are based on a mixed effect model with a random effect to account for the nested structure of the data.

In particular, the regression equation for Figure 3.C (Table S1, Model 1) is:

$$y_{ij} = \beta_0 + \beta_1 \Omega_{(j)} + \beta_2 I_i + \beta_3 I_i \Omega_{(j)} + v_i + \epsilon_{ij}, \qquad \text{(Model 1)}$$

where y_{ij} is the standardized (z-score) absolute error of the revised collective estimate for the i-th group in the j-th estimation context, β_0 is the fixed intercept for the regression model, β_1 is the fixed coefficient for our proposed feature of the estimation context, $I_i \in \{0, 1\}$ is an indicator variable of whether social interaction has occurred or not, β_2 is the fixed coefficient for the social influence centralization, β_3 is the fixed coefficient for the interaction term between our proposed feature of the estimation context and influence centralization, and v_i is the random coefficient for the i-th group. Finally, ϵ_{ij} is a Gaussian error term.

The logistic regression equation for Figure 3.D (Table S1, Model 2) is:

$$y_{ij} = \frac{1}{1 + \exp(\beta_0 + \beta_1 \Omega_{(j)} + v_i + \epsilon_{ij})}, \qquad \text{(Model 2)}$$

where y_{ij} is a binary indicator of whether the i-th group on the j-th task improved its collective estimate after social interaction, β_0 is the fixed intercept, β_1 is the fixed coefficient for our proposed feature of the estimation context, v_i is the random coefficient for the i-th group, and ϵ_{ij} is a Gaussian error term.

Table S1: The main effects of our proposed feature, Ω, and its interaction with social influence (i.e., centralization). Each datapoint is an experimental trial. The results are from a mixed effect model with a random effect for the group. Note that Ω is computed for a fixed number of agents ($n = 50$) and centralization level ($\omega = 1/3$). See Tables S2-S3 for robustness to these choices.

|	Dependent variable:	
	Standardized Absolute Error	Improved after social interaction
	linear mixed-effects	*generalized linear mixed-effects*
	(Model 1)	(Model 2)
Estimation context feature Ω	3.359***	2.708***
	(0.504)	(0.478)
Social Influence $\in \{0,1\}$	2.026***	
	(0.343)	
Ω x Social Influence	-3.882***	
	(0.539)	
Intercept	-1.772***	-1.334***
	(0.320)	(0.293)
Observations	815	678
Log Likelihood	$-1,097.161$	-447.346
Akaike Inf. Crit.	2,206.323	900.692
Bayesian Inf. Crit.	2,234.542	914.249
Note:		*p<0.1; **p<0.05; ***p<0.01

S4 Robustness checks
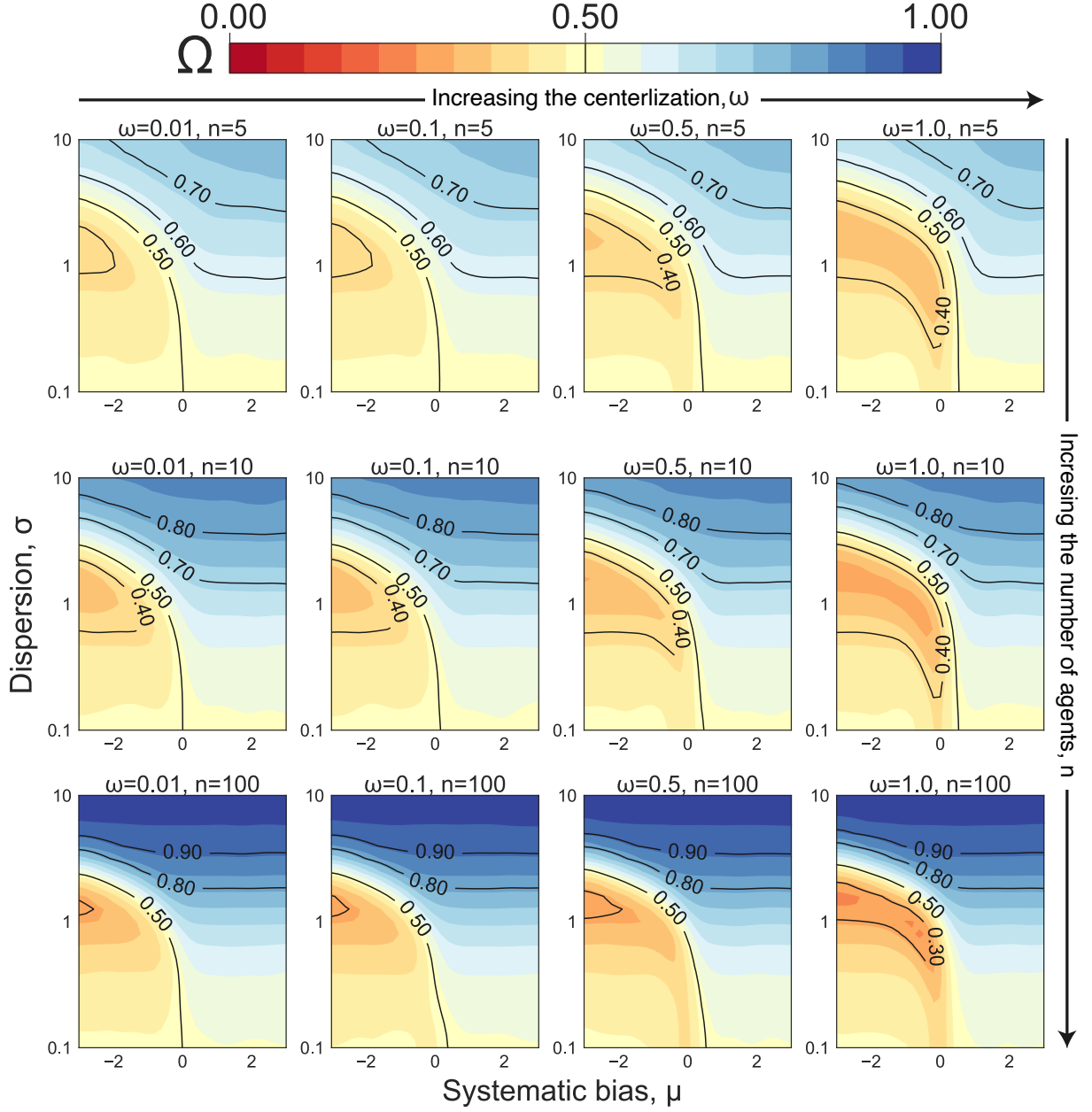
Figure S3: Robustness checks of the simulation results by varying ω and n when calculating our proposed feature of the estimation context, Ω, for the log-normal distribution. We find that the qualitative behavior of the phase diagram is robust to these changes. Increasing n or ω leads to sharper transitions from low Ω to high Ω.

Table S2: Robustness checks for Model 1 (by varying ω and n when calculating our proposed feature of the estimation context, Ω, from the empirical data) for the marginal effect of the interaction term between the centralization of influence and the estimation context feature on group performance—in terms of standardized absolute error. Each datapoint is an experimental trial. The results are from a mixed effect model with a random effect for the group. We find that the nature of the results is robust to alternative parameter choices.

	Dependent variable:		
	Standardized Absolute Error		
	$\omega = 0.1, n = 5$	$\omega = 1/3, n = 50$	$\omega = 1, n = 35$
	(1)	(2)	(3)
Estimation context feature Ω	6.321***	3.046***	2.742***
	(1.256)	(0.460)	(0.413)
Social Influence $\in \{0, 1\}$	3.666***	1.882***	1.315***
	(0.743)	(0.325)	(0.251)
Ω x Social Influence	-7.351***	-3.538***	-3.168***
	(1.349)	(0.492)	(0.442)
Intercept	-3.170***	-1.638***	-1.154***
	(0.692)	(0.302)	(0.233)
Observations	815	815	815
Log Likelihood	$-1,105.754$	$-1,097.441$	$-1,097.743$
Akaike Inf. Crit.	2,223.507	2,206.882	2,207.486
Bayesian Inf. Crit.	2,251.726	2,235.101	2,235.705
Note:			*p<0.1; **p<0.05; ***p<0.01

Table S3: Robustness checks for Model 2 (by varying ω and n when calculating our proposed feature of the estimation context, Ω, from the empirical data) for the effect of the estimation context feature on the group performance after social interaction. Each datapoint is an experimental trial. The results are from a mixed effect model with a random effect for the group. We find that the nature of the results is robust to alternative parameter choices.

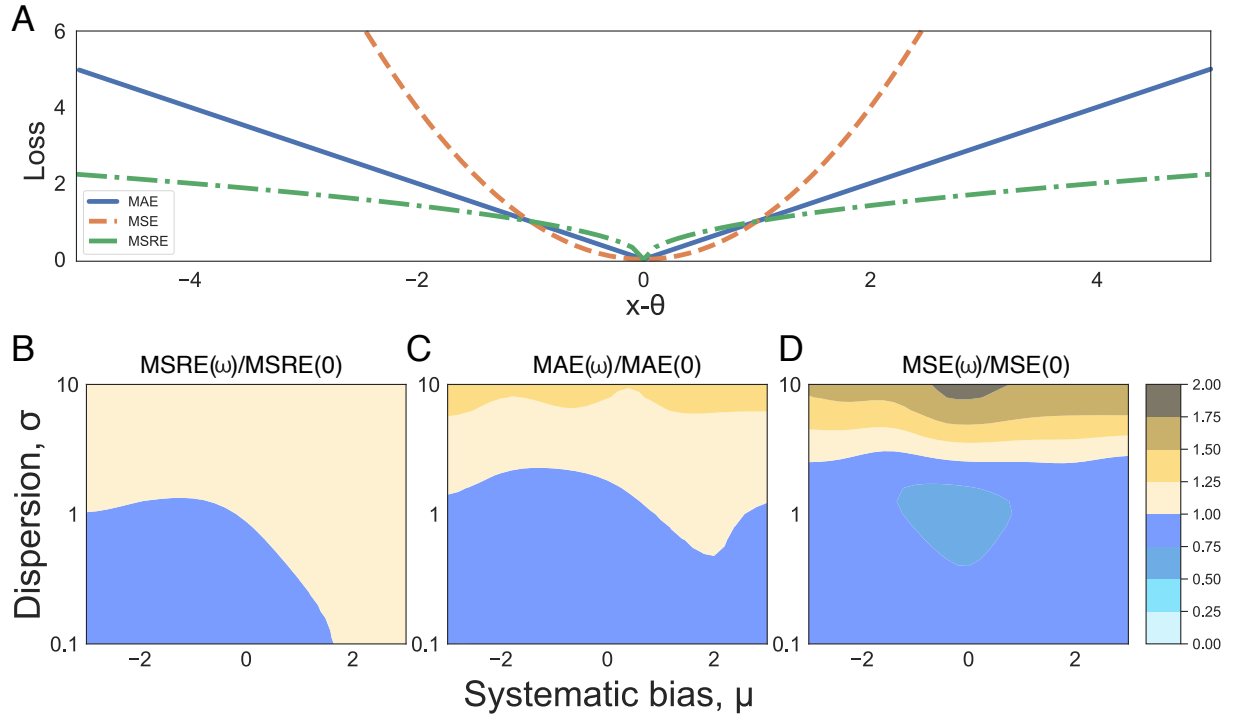|	*Dependent variable:*		
	Whether the group improved after social interaction		
	$\omega = 0.1, n = 5$	$\omega = 1/3, n = 50$	$\omega = 1, n = 35$
	(1)	(2)	(3)
Estimation context feature Ω	5.601***	2.521***	1.993***
	(1.173)	(0.439)	(0.391)
Intercept	−2.759***	−1.264***	−0.743***
	(0.635)	(0.277)	(0.215)
Observations	678	678	678
Log Likelihood	−452.745	−446.900	−450.914
Akaike Inf. Crit.	911.490	899.799	907.829
Bayesian Inf. Crit.	925.047	913.356	921.386
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure S4: Panel **A** shows the loss as a function of error for the Mean Squared Error (MSE), Mean Absolute Error (MAE), and the Mean Square Root Error (MSRE) loss functions — see (S.6). Panels **B**, **C**, and **D** show the three loss functions for different values of μ and σ for a log-normal distribution. In each case, we plot the ratio of the loss function evaluated in a centralized structure, $\omega > 0$, over a decentralized structure $\omega = 0$. A ratio less than 1 indicates that the centralized network performs better than the decentralized network. The performance of the two influence structures can vary significantly as a function of the selected loss function [19]. The choice of the loss function is typically application-dependent. For instance, if the reward for 'getting it right' is greater than the cost of being frequently wrong—as in domains where the loss and payoff are asymmetric, unbounded, or have a remote boundary [20, 21, 13]—then the decentralized influence structure is more desirable when the dispersion is high. The initial estimates in these simulations are sampled from a log-normal distribution for a fixed number of agents ($n = 50$) and centralization level ($\omega = 1/3$).

S5   References

[1] N. E. Friedkin, A formal theory of social power, *Journal of Mathematical Sociology* **12**, 103 (1986).

[2] J. R. French Jr, A formal theory of social power., *Psychological review* **63**, 181 (1956).

[3] F. Harary, *Studies in Social Power* (University of Michigan, 1959), pp. 168–182. D. Cartwright, Ed.

[4] M. H. DeGroot, Reaching a consensus, *Journal of American Statistical Association* **69**, 118 (1974).

[5] E. Seneta, *Non-negative matrices and Markov chains* (Springer, 2006).

[6] J. Becker, E. Porter, D. Centola, The wisdom of partisan crowds, *Proceedings of the National Academy of Sciences* **116**, 10717 (2019).

[7] J. Becker, D. Brackbill, D. Centola, Network dynamics of social influence in the wisdom of crowds, *Proceedings of the national academy of sciences* **114**, E5070 (2017).

[8] J. Lorenz, H. Rauhut, F. Schweitzer, D. Helbing, How social influence can undermine the wisdom of crowd effect, *Proc. of the National Academy of Sciences* **108**, 9020 (2011).

[9] A. Almaatouq, *et al.*, Adaptive social networks promote the wisdom of crowds, *Proceedings of the National Academy of Sciences* **117**, 11379 (2020).

[10] B. Jayles, *et al.*, The impact of incorrect social information on collective wisdom in human groups, *arXiv preprint arXiv:2003.06160* (2020).

[11] B. Jayles, *et al.*, How social information can improve estimation accuracy in human groups, *Proceedings of the National Academy of Sciences* **114**, 12620 (2017).

[12] A. B. Kao, *et al.*, Counteracting estimation bias and social influence to improve the wisdom of crowds, *Journal of The Royal Society Interface* **15**, 20180130 (2018).

[13] M. S. Lobo, D. Yao, Human judgement is heavy tailed: Empirical evidence and implications for the aggregation of estimates and forecasts, *INSEAD Working Paper No. 2010/48/DS. Available at SSRN: 1638811* .

[14] P. Mavrodiev, C. J. Tessone, F. Schweitzer, Quantifying the effects of social influence, *Scientific reports* **3**, 1360 (2013).

[15] J. Nair, A. Wierman, B. Zwart, *The Fundamentals of Heavy-Tails: Properties, Emergence, and Identification* (2020). Preprint.

[16] B. Golub, M. O. Jackson, Naive learning in social networks and the wisdom of crowds, *American Economic Journal: Microeconomics* **2**, 112 (2010).

[17] P. J. Bickel, *et al.*, On some robust estimates of location, *The Annals of Mathematical Statistics* **36**, 847 (1965).

[18] B. Gürçay, B. A. Mellers, J. Baron, The power of social influence on estimation accuracy, *Journal of Behavioral Decision Making* **28**, 250 (2015).

[19] N. N. Taleb, P. E. Tetlock, On the difference between binary prediction and true exposure with implications for forecasting tournaments and decision making research, *Unpublished Manuscript. http://papers. ssrn. com/sol3/papers. cfm* (2013).

[20] L. Fang, A. Yasuda, The effectiveness of reputation as a disciplinary mechanism in sell-side research, *The Review of Financial Studies* **22**, 3735 (2009).

[21] N. N. Taleb, C. Sandis, *Future Perspectives in Risk Models and Finance* (Springer, 2015), pp. 125–136.