UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Extremizing Judgements Produces More Inaccurate Individuals but Wiser Crowds

Permalink
https://escholarship.org/uc/item/7dq0095f

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Barrera-Lemarchand, Federico
Navajas, Joaquin

Publication Date
2023

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Extremizing judgements produces more inaccurate individuals but wiser crowds

Federico Barrera-Lemarchand (fedex192@gmail.com)
Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Av. Figueroa Alcorta 7350,
Ciudad de Buenos Aires, Argentina
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad de Buenos Aires, Argentina
Departamento de Física, Universidad de Buenos Aires, Pabellón I, Ciudad Universitaria,
Ciudad de Buenos Aires, Argentina

Joaquín Navajas (joaquin.navajas@utdt.edu)
Laboratorio de Neurociencia, Universidad Torcuato Di Tella, Av. Figueroa Alcorta 7350,
Ciudad de Buenos Aires, Argentina
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Ciudad de Buenos Aires, Argentina
Escuela de Negocios, Universidad Torcuato Di Tella, Av. Figueroa Alcorta 7350,
Ciudad de Buenos Aires, Argentina

Abstract

The crowd wisdom effect is a well-established phenomenon that is widely employed for predicting and estimating variables across various domains. Previous research has focused on enhancing the wisdom of crowds by improving individual estimates while maintaining some of the initial opinion diversity. However, at least theoretically, it is possible to increase collective accuracy by largely increasing the diversity in a crowd while concurrently increasing, to a lesser extent, the individual error. In this study, we propose a method that leverages the anchoring effect to extremize individual judgments and thus increase the diversity of opinions in a crowd. This is achieved by dividing the crowd into two groups, anchoring each group to either a low or high value, and aggregating all estimates. We use a mathematical model of the anchoring effect to determine when this strategy is expected to outperform the crowd wisdom effect. Results from three experiments provide converging evidence that the proposed approach outperforms traditional methods in estimating and forecasting unknown quantities. This research presents a novel approach to reduce collective error by maximizing diversity at the cost of extremizing individual judgements.

Keywords: wisdom of crowds; anchoring bias; predictive diversity; mathematical model

Introduction

The aggregation of many lay estimates often outperforms the expert individual judgement (De Condorcet, 1785; Galton, 1907). This phenomenon, popularly known as the "wisdom of crowds" (Surowiecki, 2005), has been applied to a wide range of domains such as improving medical diagnoses (Kurvers et al., 2016) forecasting geopolitical events (Mellers et al., 2014), predicting financial markets (Ray, 2006), and fact-checking news (Allen, 2021), among many others. Given its practical relevance, understanding the conditions under which crowds produce accurate estimates has become a relevant issue in social and psychological science (Kameda, Toyokawa, & Tindale, 2022; Karachiwalla & Pinkow, 2021; Navajas et al., 2018; Kao & Couzin, 2014).

Previous studies have suggested that an important driver of collective accuracy is the diversity of opinions in the crowd (Hong & Page, 2004; Page, 2008; Becker, Porter & Centola, 2019; Shi et al., 2019; Jönsson, Hahn & Olsson, 2015). A simple intuition underlies this result: when crowds produce diverse estimates, it is likely that some individuals will underestimate the correct answer while others will overestimate it. Therefore, in diverse crowds there is a higher chance that individual errors will cancel out by the process of aggregating their opinions. This intuitive idea can be formalized in a mathematical identity called the "Diversity Prediction Theorem" (Page, 2007), which states that the crowd's error (E) can be expressed as the mean individual error (ε) minus the crowd's predictive diversity (δ, also known as the population variance):

$$E = \varepsilon - \delta \qquad [1]$$

One implication of this theorem is that, in principle, the crowds' accuracy could be increased by reducing the crowd's individual error (ε) or, equivalently, by increasing the predictive diversity (δ). However, while these two alternatives seem equally valid in theory, most research aiming at increasing the wisdom of the crowd has extensively focused on the former strategy. For example, previous studies have proposed to aggregate information from "select" crowds composed by individuals who are more accurate across estimation problems (Mannes, Soll & Larrick, 2014). Other studies have shown that individual error can be reduced by counteracting individual biases (Kao et al., 2018) or by exposing individuals to social information (Jayles et al., 2017; Frey & Van de Rijt, 2021; Madirolas & de Polavieja, 2015; Lorenz et al., 2011). All these different approaches share a common feature: they decrease collective error by increasing accuracy at the individual level, while preserving some of the initial diversity of opinions. Therefore, whether one could effectively reduce the crowd's error by increasing its predictive diversity, even at the cost of slightly reducing individual accuracy, remains unknown.

Showing that it is possible to increase collective performance while concurrently reducing individual accuracy is relevant for at least three different reasons. First,

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*. ©2023 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

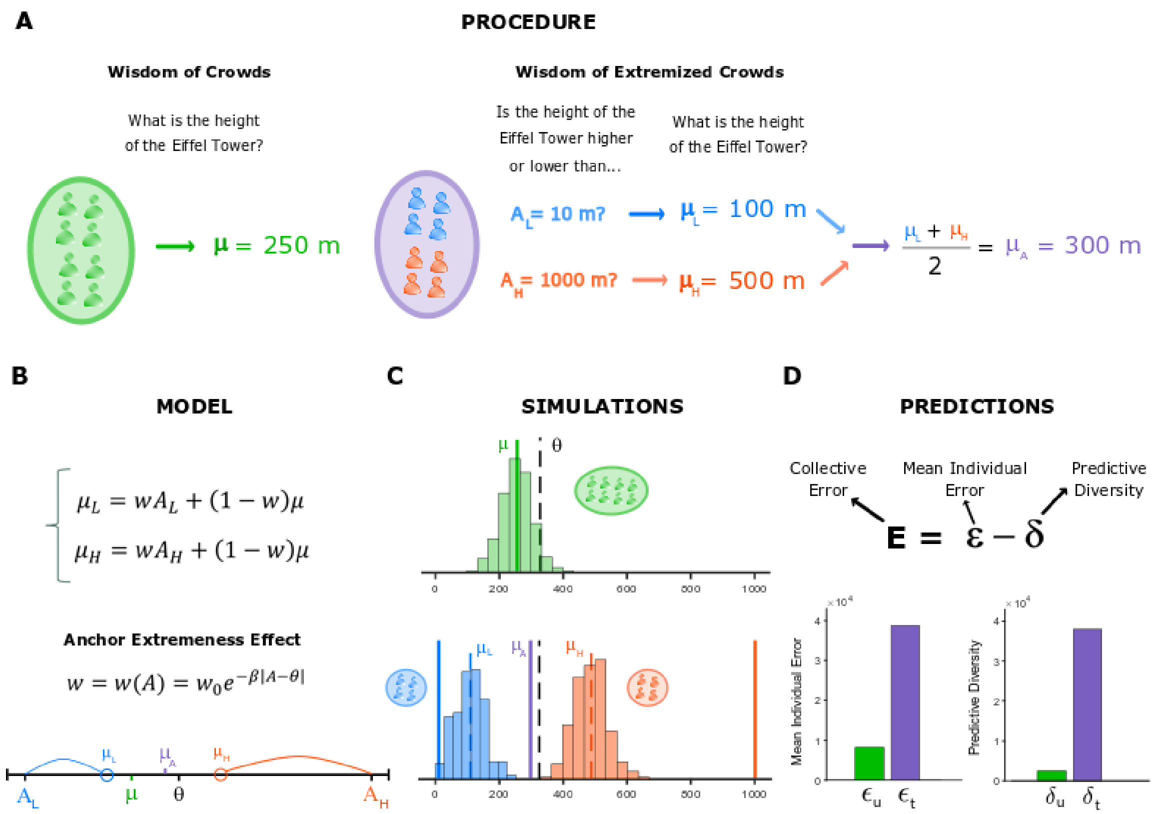

Figure 1: A. The method for estimating a variable through the wisdom of crowds consists in asking a crowd of individuals to estimate a given quantity, and averaging the answers. The method consists in extremizing a crowd by dividing it in two halves, asking an anchoring question with either a low or a high value to each half, and then averaging the answers from both halves. B. Model of the anchoring effect. The anchored mean is a weighted average of the anchor and the wisdom-of-crowds value. The weight *w* depends on the difference between the anchor and the correct answer, reflecting an internal sensitivity to the correct answer. C. Simulations performed using the proposed model show that the method can potentially outperform wisdom of crowds. D. Mean individual error and predictive diversity on the previous simulations, which show an increase in both, but a bigger increase in predictive diversity, resulting in an overall reduction in collective error.

from a theoretical point of view, it would provide empirical evidence that individual error and predictive diversity are indeed two different constructs that independently drive collective error, as proposed by the Diversity Prediction Theorem (Equation 1). Second, from a practical standpoint, it would provide practitioners with a novel approach to increase collective estimation accuracy. Third, given that the wisdom of crowds has been previously interpreted as empirical evidence for the epistemic value of democratic judgements, this putative dissociation between individual and collective accuracy should mitigate concerns about the increase of misinformed voters in recent elections.

In this paper, we present a new approach to increase collective accuracy by boosting the crowd's predictive diversity. We do so by extremizing estimates using a cognitive bias known as the anchoring effect (Tversky & Kahneman, 1974). The method consists in anchoring one half of the crowd to a small value (low anchor) and the other half to a large value (high anchor), and then averaging all estimates. We hypothesized that this simple technique should lead to an increase in the predictive diversity of the crowd that surpasses the increase in mean individual error, thus leading to lower collective error. Using a simple mathematical model, we showed that this method is expected to enhance collective accuracy across a wide range of model parameters. We then empirically demonstrated that the procedure reduces collective error across three behavioral experiments involving the estimation and forecasting of different quantities.

Results

Procedure

Let us consider the scenario where someone needs to estimate a numerical variable unknown to them; for example, the height of the Eiffel Tower. Based on the standard wisdom-of-crowds effect, one could obtain an approximate value by asking a large number of individuals to provide an estimate. Then, to estimate the height of the Eiffel Tower, the person would aggregate these values, for example, by averaging them (Figure 1A). In this work, we propose an alternative approach that consists in dividing the crowd into two halves and extremizing each half in opposite directions (Figure 1B). We suggest doing so by using the anchoring effect: before estimating the relevant variable, individuals are first asked to consider either an extremely low or high value. In the previous example, half of the individuals would be asked to consider if the height of the Eiffel Tower is greater or less than 10 meters (low anchor, A_L) and the other half, if it is greater or less than 1000 meters (high anchor, A_H). After providing a categorical answer to this initial question, all individuals would then be asked to provide their best-guess estimate. An extensive literature has shown that these estimates should then be consistently biased towards the initially considered values (Furnham & Boo, 2011; Röseler et al., 2022). Because these anchors are extreme in opposite directions, this procedure should extremize the estimates produced by the crowd as a whole, leading to an increase in predictive diversity. We therefore propose to average all numbers, across both halves of the crowd.

While this procedure requires pre-defining two extreme values that will be used as anchors, the strategy does not require knowing the correct answer. However, reasonably, its accuracy will depend on the specific choice of anchors. Therefore, to better understand the conditions under which the proposed approach is expected to increase collective accuracy, we developed a simple mathematical model of the anchoring effect.

Model

Based on the Diversity Prediction Theorem, we reasoned that this procedure should reduce collective error by means of increasing the diversity of opinions in the crowd. However, we noted that this claim is true if and only if the method produces an increase in diversity that surpasses the increase in individual error (Equation 1).

We consider a set of individuals who, being asked to estimate the variable θ, produce a distribution of values with mean μ. The model assumes that, when those individuals are anchored to a low value A_L, they produce a set of estimates with a different mean

$$\mu_L = w_L\, A_L + (1 - w_L)\, \mu \qquad [2]$$

where $0 \leq w_L \leq 1$ is an "anchoring index" reflecting the strength of the anchoring procedure given by the anchor A_L. Similarly, a population anchored to a high value A_H would produce a distribution of estimates given by

$$\mu_H = w_H\, A_H + (1 - w_H)\, \mu \qquad [3]$$

where $0 \leq w_H \leq 1$ is the corresponding "anchoring index" given by anchor A_H. Equations 2 and 3 imply that the anchored mean is a weighted average of the anchor and the mean of the original distribution of values μ.

Following a variety of empirical findings linking the anchoring effect to the plausibility of the anchor (Mussweiler & Strack, 2001; Wegener et al., 2001), we propose that w is given by

$$w_j = w_0 e^{-\beta|A_j - \theta|} \qquad [4]$$

where j indicates whether the given weight corresponds to the low or high anchor (w_L or w_H, respectively) and w_0 is a parameter reflecting the anchoring index when individuals are anchored to the correct value θ. The parameter β is an "inverse temperature" encoding the sensitivity of the individuals to the distance between the anchor and θ. Thus, the value of β modulates the strength of the "anchoring extremeness effect" (Röseler et al., 2022).

In this work, we propose averaging estimates from two populations of individuals, each of which is anchored to either a low or a high value (A_L or A_H). Following the previous model, assuming both populations are equally sized, we can estimate the mean of estimates from both populations as

$$\mu_A = \frac{\mu_L + \mu_H}{2} \qquad [5]$$

Simulations (for a set of fixed parameters, with θ=324, μ=250, A_L=10, A_H=1000, w_0=1, and β=0.0017) show that this model can produce a reduction in collective error (Figure 1C), which is due to an increase in predictive diversity that surpasses the increase in mean individual error (Figure 1D). Using this simple model of the anchoring effect, we analytically derived the general conditions under which the proposed method leads to an increase in collective accuracy. We found that the key variable determining the success of the approach is the mid-point of the anchors, defined as $\bar{A} = \frac{A_L + A_H}{2}$. Following a simple mathematical procedure (omitted here for, brevity), we found that the range of $\bar{A}$ values where the method outperforms the wisdom of crowds (Δ) is

$$\Delta = \frac{4|\mu - \theta|}{w_L + w_H} \qquad [6]$$

The expression derived in Equation 6 implies that the range of values where the method outperforms the wisdom of crowds is always equal to or larger than two times the collective error. This can be shown by examining two opposite extreme scenarios. If the sensitivity β is small (i.e., when the anchoring procedure does not depend on the distance between the anchor and the truth), the range of values where the method outperforms the wisdom of crowds converges to twice the collective error (i.e., if $\beta \to 0$, then $w_j \to w_0$, and therefore $\Delta \to \frac{2|\mu - \theta|}{w_0}$). In the opposite case, when the sensitivity β is large (i.e., when individuals are more attracted to anchors that are closer to the correct answer), then this method is always better than the classic wisdom of crowds (i.e., if $\beta \to \infty$, then $w_j \to 0$, and thus $\Delta \to \infty$).

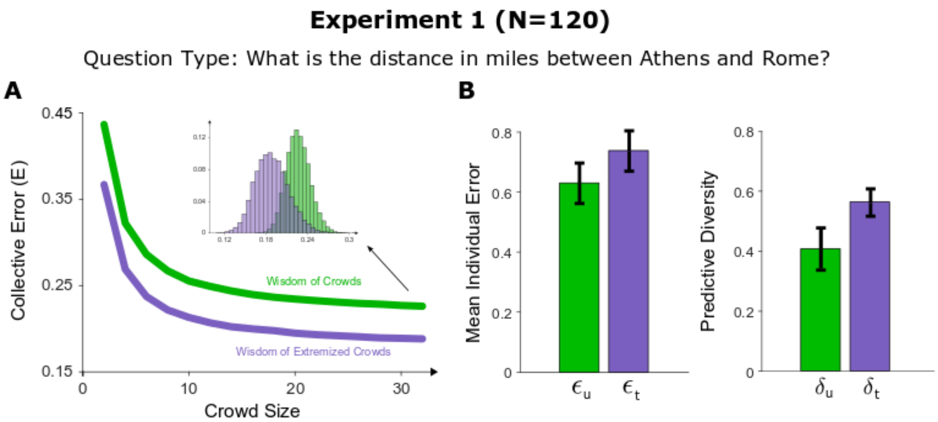

Figure 2: empirical results for Experiment 1. A. Change in collective error when changing the crowd size, for the classic wisdom of crowds (green) and the extremized crowd (purple). The standard error is contained within the line width. We also show the distribution of values from the resampling method for the biggest crowd size. B. Mean individual error and predictive diversity for both the classic wisdom of crowds (green) and the extremized crowd (purple). The error bars are the standard deviation of the means.

Empirical Results

To evaluate the effectiveness of the proposed method, we performed three experiments. In Experiment 1, we recruited 120 participants (from the USA, recruited online on Amazon Mechanical Turk), and asked them to estimate 14 general-knowledge quantities. All variables involved positive unbounded quantities, like the example of used in Figure 1 (e.g., "What is the height of the Eiffel Tower?"). Participants had monetary incentives to estimate them as accurately as possible. One third of the sample was randomly assigned to a control condition where they simply estimated the variable. The other two thirds of the participants were randomly assigned to a condition where, before estimating the quantity, they considered either an extremely low or extremely high value. Half of the anchored participants considered a low value, and the other half considered a high value. These extreme values were set as the 5 and 95 percentiles of the empirical distribution of the non-anchored condition. By employing a simple bootstrapping resampling method, we estimated the collective error of differently-sized groups for both the wisdom of crowds, and the wisdom of extremized crowds (Figure 2A). We observed that the average collective error of the latter was always smaller than the former for all crowd sizes. Specifically, the error is vastly reduced for the largest crowd size (34 individuals, unpaired t-test: $t(998)=29.7$, $p=2x10^{-139}$; effect size: cohen's D = 1.84). This collective error reduction was due to an increase in predictive diversity that was higher than the increase in mean individual error (Figure 2B). One limitation of Experiment 1 is that anchors were defined after collecting the data of the non-anchored population. Because this procedure may be inconvenient from a practical point of view, we performed a second pre-registered (https://aspredicted.org/RYC_4Y5) experiment where anchors were pre-defined and fixed across all questions.

In Experiment 2, we recruited 396 Argentinian participants online, and asked them 30 general-knowledge questions that involved the estimation of a percentage. Therefore, their answers were bounded in the range [0,100] (e.g. what percentage of the population of Argentina is under 15 years old?). Unlike the previous study, here we used the same anchors for all questions, always set at either 5% (low anchor A_L) or 95% (high anchor A_H). We observed very similar results to Experiment 1 (Figure 3A and 3B). Collective error was lower for the extremized crowd (crowd size N=50, unpaired t-test: t(998)=19.1, p=2x10-69; effect size: cohen's D = 1.21). We also found this to be accompanied by an increase in mean individual error, and by a greater increase in predictive diversity.

Given that this experiment used fixed anchors across all questions, it allowed us to test the key element of the model, i.e., the anchoring extremeness effect (Equation 4). We did so by performing two separate analyses. First, we examined the biases associated with each experimental condition. We reasoned that, if participants anchoring effects were sensitive to the distance to the correct answer, then the effectiveness of the anchoring procedure should be higher when the correct answer is close to the anchor. For example, we should see that participants considering a low value (5%) should be more attracted to the anchor when the correct answer is low (below 50%) compared to when the correct answer is high (above 50%). Consistent with this idea, when the correct answer was below 50% (17 questions), we observed that the population considering the "low anchor" provided a distribution of estimates that was statistically indistinguishable from the correct answer (paired t-test, t(16)=0.95, p=.36). In turn, both the non-anchored (paired t-test, $t(16)=4.16$, $p=3x10^{-4}$) and the population anchored to a high value (paired t-test, $t(16)=5.99$, $p=2x10^{-5}$) provided a distribution of estimates that significantly overestimated the correct answer (Figure 3C). Conversely, when the correct answer was above 50% (13 questions), we observed the opposite pattern: the population

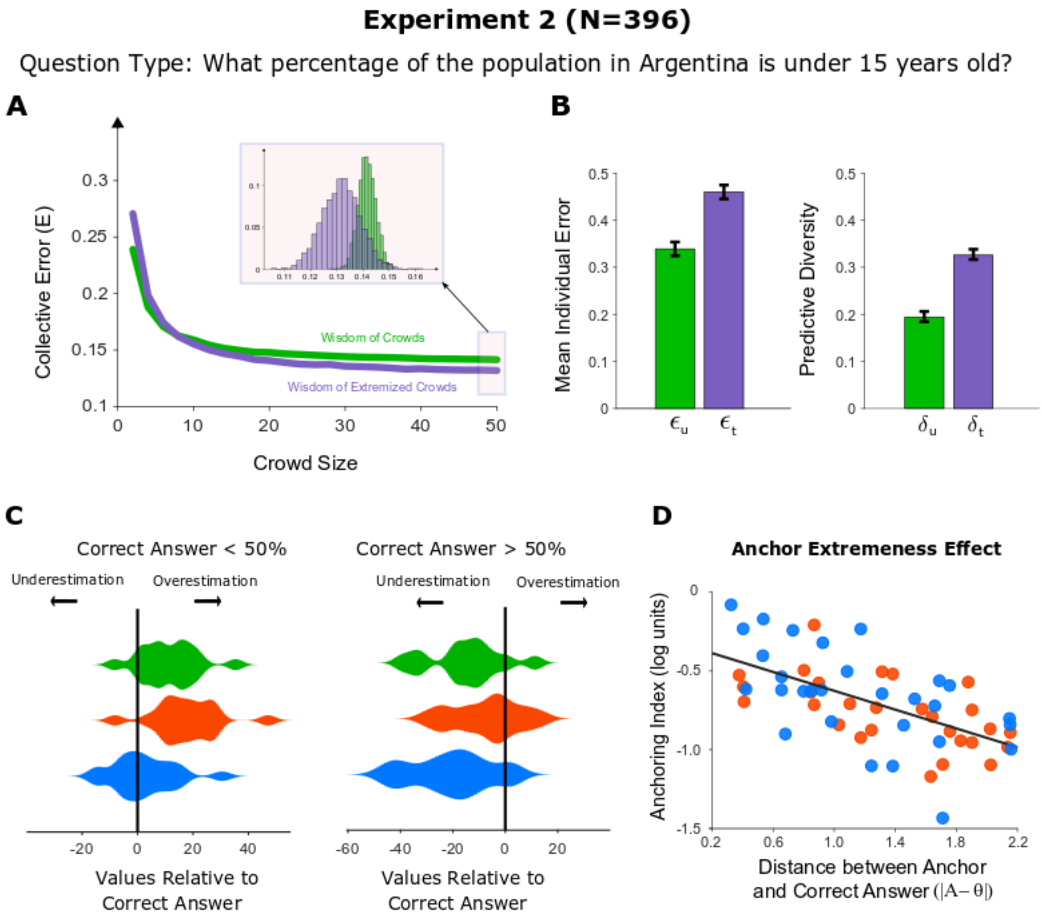

Figure 3: empirical results for Experiment 2. A. Change in collective error when changing the crowd size, for the classic wisdom of crowds (green) and the extremized crowd (purple). The standard error is contained within the line width. We also show the distribution of values from the resampling method for the biggest crowd size. B. Mean individual error and predictive diversity for both the classic wisdom of crowds (green) and the extremized crowd (purple). The error bars are the standard deviation of the means. C. Distributions of values corresponding to the difference between the mean answers and the correct answer for each question, for the classic wisdom of crowds (green), the crowd extremized with a high anchor (red) and the crowd extremized with a low anchor (blue). We separate the cases where the correct answer is above and below 50%. The black line represents the case where the mean value is equal to the correct answer. D. Empirical anchoring index (*w*) for each question, for the case of low anchor (blue) and high anchor (red), against the distance between the corresponding anchor and the correct answer (logarithmic units on anchoring index). The black line represents the best linear fit of the data.

considering the "high anchor" provided a distribution of estimates that was statistically indistinguishable from the correct answer (paired t-test, $t(12)=2.01$, $p=0.07$). Also, both the non-anchored (paired t-test, $t(16)=3.90$, $p=0.002$) and the population anchored to a low value (paired t-test, $t(16)=4.90$, $p=4x10^{-4}$) provided a distribution of estimates that significantly underestimated the correct answer.

Second, we directly examined the anchoring extremeness effect by studying the association between the strength of the anchoring procedure and the distance between the anchor and the correct answer (Equation 4). We estimated the strength of the anchoring effect by calculating the "anchoring index" as in previous literature (Jacowitz & Kahneman, 1995). The index takes a value of 0 when the anchoring procedure does not affect the estimates, and a value of 1 when the estimates are equal to the anchor. Consistent with the existence of the anchoring extremeness effect (Equation 4), we observed a significantly negative correlation between the anchoring index in log units and the distance between the anchor and the correct answer (Pearson correlation coefficient, $r=-0.56$, $p=3x10^{-6}$). Finally, we asked whether the proposed method could outperform the wisdom of crowds for forecasting tasks. To answer this question, we performed a third pre-registered experiment (https://aspredicted.org/HZC_PTH).

In Experiment 3, we recruited 620 participants (from the USA, recruited online on Prolific), and asked them to

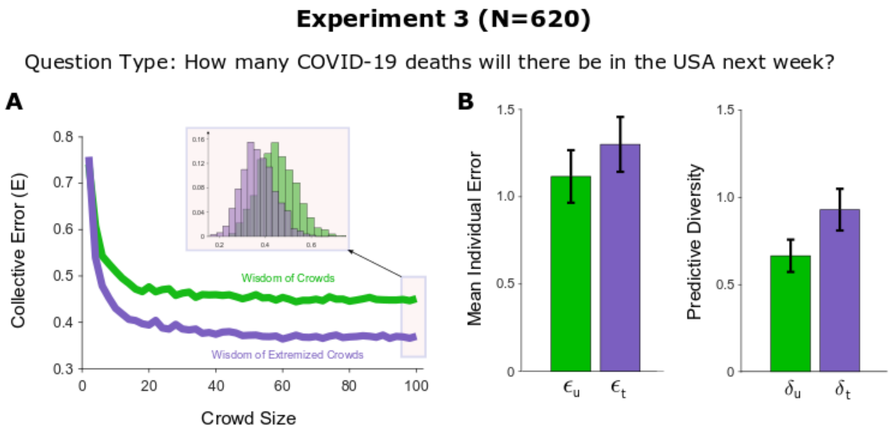

Figure 4: empirical results for Experiment 3. A. Change in collective error when changing the crowd size, for the classic wisdom of crowds (green) and the extremized crowd (purple). The standard error is contained within the line width. We also show the distribution of values from the resampling method for the biggest crowd size. B. Mean individual error and predictive diversity for both the classic wisdom of crowds (green) and the extremized crowd (purple). The error bars are the standard deviation of the means.

estimate the number of COVID-19 cases and deaths in the United States in the following week (from 27 July to 2 August, 2020). Participants had monetary incentives to estimate these variables as accurately as possible. Anchors were selected as extreme values based on historical data, namely, two orders of magnitude less or more the number of COVID-19 cases and deaths reported on the two weeks before the beginning of the experiment. Again, as in Experiments 1 and 2, we observed that the collective error was lower for the extremized crowd for all group sizes compared to the wisdom of crowds (Figure 4A). We found that the decrease in collective error (specifically, for the largest crowd size N=100, unpaired t-test: t(998)=16.4, p=7x10-54; effect size: cohen's D = 1.04) was due to an increase in predictive diversity that was greater than the increase in mean individual error (Figure 4B).

Discussion

In this work, we provide evidence to support the employment of a novel tactic for improving upon the well-known wisdom-of-crowds effect (Surowiecki, 2005). By means of the anchoring bias (Tversky, & Kahneman, 1974), we can extremize crowds and reduce collective error by increasing diversity (Page, 2007).

In previous literature, methods for increasing the wisdom of crowds often involved strategies for reducing individual error (Madirolas & de Polavieja, 2015; Mannes, Soll & Larrick, 2014). However, while theoretical analysis suggested this goal could also be achieved by increasing the predictive diversity (as seen in Equation 1), that possibility remained unexplored. Here, we thoroughly studied this approach, both theoretically, by developing a mathematical model of the anchoring effect, and empirically, by performing three experiments. In all of the experiments, irrespective of their differences (sample size, country of residence, type of anchors, bounded or unbounded variables, estimation or forecasting), we observed very similar results, with collective error being lower for the extremized crowd with respect to the wisdom of crowds, and with an increase in both mean individual error and predictive diversity, but an overall reduction in collective error. We also found compelling evidence that the simple anchoring extremeness effect model we proposed is appropriate to explain why the suggested method works in this way.

One limitation of the proposed method is that the selection of appropriate anchors could potentially prove hard (for example, in forecasting problems). However, the presented model suggests that the range of values where this method improves the wisdom of crowds is large, and that a reduction in error is expected. We also showed that this is empirically feasible, by employing different tactics for anchor selection in all three experiments, reaching very similar results. The third experiment is especially relevant, since it shows a direct application of the proposed method in a real-world forecasting problem, where the answer was unknown at the time. Thus, appropriate anchors can be chosen for problems with unknown answers if the order of magnitude of those answers is not completely indiscernible.

The first study on the wisdom of crowds (Galton, 1907) was regarded as an empirical demonstration that democratic aggregation rules can be trustworthy and efficient. This was counter-intuitive at the time, since it showed that erroneous individuals could make correct choices. Nowadays, when political opinions tend to become extremized, and polarization looks like a threat to democracies, these results imply that democratic decisions can still be surprisingly accurate. Needless to say, we must first understand how and if these results, which hold for factual problems, also extend to moral and political dilemmas, a matter left to be explored in future works.

Acknowledgments

This research was supported by a Sponsored Research Agreement between Facebook and Fundación Universidad Torcuato Di Tella (grant #INB2376941).

References

Allen, J., Arechar, A. A., Pennycook, G., & Rand, D. G. (2021). Scaling up fact-checking using the wisdom of crowds. *Science advances*, *7*(36), eabf4393.

Becker, J., Porter, E., & Centola, D. (2019). The wisdom of partisan crowds. *Proceedings of the National Academy of Sciences*, 116(22), 10717-10722.

De Condorcet, N. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'impremerie royale.

Frey, V., & Van de Rijt, A. (2021). Social influence undermines the wisdom of the crowd in sequential decision making. *Management science*, 67(7), 4273-4286.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The journal of socio-economics*, *40*(1), 35-42.

Galton, F. (1907). Vox populi. *Nature 7*, 450–451.

Hong, L., & Page, S. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences*, *101*(46), 16385-16389.

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161-1166.

Jayles, B., Kim, H. R., Escobedo, R., Cezera, S., Blanchet, A., Kameda, T., ... & Theraulaz, G. (2017). How social information can improve estimation accuracy in human groups. *Proceedings of the National Academy of Sciences*, 114(47), 12620-12625.

Jönsson, M. L., Hahn, U., & Olsson, E. J. (2015). The kind of group you want to belong to: Effects of group structure on group accuracy. *Cognition*, 142, 191-204.

Kameda, T., Toyokawa, W., & Tindale, R. S. (2022). Information aggregation and collective intelligence beyond the wisdom of crowds. *Nature Reviews Psychology*, *1*(6), 345-357.

Kao, A. B., & Couzin, I. D. (2014). Decision accuracy in complex environments is often maximized by small group sizes. *Proceedings of the Royal Society B: Biological Sciences*, 281(1784), 20133305.

Kao, A. B., Berdahl, A. M., Hartnett, A. T., Lutz, M. J., Bak-Coleman, J. B., Ioannou, C. C., ... & Couzin, I. D. (2018). Counteracting estimation bias and social influence to improve the wisdom of crowds. *Journal of The Royal Society Interface*, 15(141), 20180130.

Karachiwalla, R., & Pinkow, F. (2021). Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. *Creativity and innovation management*, *30*(3), 563-584.

Kurvers, R. H., Herzog, S. M., Hertwig, R., Krause, J., Carney, P. A., Bogart, A., Zalaudek, I., & Wolf, M. (2016). Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, 113(31), 8777-8782.

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the national academy of sciences*, 108(22), 9020-9025.

Madirolas, G., & de Polavieja, G. G. (2015). Improving collective estimations using resistance to social influence. *PLoS computational biology*, *11*(11), e1004594.

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of personality and social psychology*, *107*(2), 276.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., ... & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological science*, *25*(5), 1106-1115.

Mussweiler, T., & Strack, F. (2001). Considering the impossible: Explaining the effects of implausible anchors. *Social Cognition*, *19*(2), 145-160.

Navajas, J., Niella, T., Garbulsky, G., Bahrami, B., & Sigman, M. (2018). Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour*, 2(2), 126-132.

Page, S. (2007). Making the difference: Applying a logic of diversity. *Academy of Management Perspectives*, *21*(4), 6-20.

Page, S. (2008). The difference. In *The Difference*. Princeton University Press.

Ray, R. (2006). Prediction markets and the financial" wisdom of crowds". *The Journal of Behavioral Finance*, *7*(1), 2-4.

Röseler, L., Weber, L., Helgerth, K., Stich, E., Günther, M., Tegethoff, P., ... & Schütz, A. (2022). The Open Anchoring Quest Dataset: Anchored Estimates from 96 Studies on Anchoring Effects. *Journal of Open Psychology Data*, *10*(1), 16.

Shi, F., Teplitskiy, M., Duede, E., & Evans, J. A. (2019). The wisdom of polarized crowds. *Nature human behaviour*, 3(4), 329-336.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124-1131.

Wegener, D. T., Petty, R. E., Detweiler-Bedell, B. T., & Jarvis, W. B. G. (2001). Implications of attitude change theories for numerical anchoring: Anchor plausibility and the limits of anchor effectiveness. *Journal of Experimental Social Psychology*, *37*(1), 62-69.