Discussion Paper Series – CRC TR 224

Discussion Paper No. 166
Project C 04

Collective Intelligence: Crowd Wisdom versus Herding

Andreas Engert*

April 2020

*Professor of law at Freie Universität Berlin, Freie Universität Empirical Legal Studies (FUELS) and European Corporate Governance Institute (ECGI)

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Andreas Engert[1]

Collective intelligence: crowd wisdom versus herding

I. Introduction

Vox populi—when Francis Galton, an English statistician and polymath, published an article under this title in a 1907 issue of Nature, he left it to the reader to complete the second half of the famous phrase. Galton was less shy in drawing a bold connection between his finding and the constitutional order of the state, introducing his article with the words: 'In these democratic days, any investigation into the trustworthiness and peculiarities of popular judgments is of interest.'[2] What Galton had to report came not from the lofty spheres of government but from the West of England Fat Stock and Poultry Exhibition at Plymouth. 800 visitors of the show had participated in a competition to submit estimates of the weight that a 'fat ox' would bring to bear after slaughtering and preparation. Tickets were sold at 6 pence, an amount that 'deterred practical joking'. To his pleasant surprise, Galton found that the 'middlemost' estimate was 1207 pounds (about 547 kilograms), less than one percent off the actual weight of 1198 pounds (543 kilograms). The median estimate, in Galton's view, represents the democratic *vox populi*: It is the only amount that is not 'condemned as too low or too high by a majority of the voters'.[3] Galton considered this result as 'more creditable to the trustworthiness of a democratic judgment than might have been expected.'[4]

The Plymouth judging contest is a foundational tale of a line of research in the social sciences that one can group under the heading 'collective intelligence' or the more catchy 'wisdom of crowds', a phrase that the journalist James Sourowiecki coined

---

1 Professor of law at Freie Universität Berlin, Freie Universität Empirical Legal Studies (FUELS) and European Corporate Governance Institute (ECGI). The author wishes to thank the EPoS Collaborative Research Center TR 224 (Project C04) of Deutsche Forschungsgemeinschaft for valuable financial support.

2 Francis Galton, 'Vox populi' (1907) 75 Nature 450, 450.

3 Galton (n 2) 450.

4 Galton (n 2) 451.

and popularized in 2004.[5] Galton's short article—less than one page in length—also encapsulates the peculiar mix of ideas that seem to permeate the collective intelligence literature: a great confidence in, and appeal to, the public as arbiter of truth as reflected in the *vox populi, vox dei* formula;[6] more specifically, a belief in public reason that thrives on the independent thinking and free expression of individuals, but then transcends them so that the collective wisdom ultimately surpasses that of any single person or political leader. These views of liberal enlightenment are entwined with formal models from statistics and probability theory, lending them an air of mathematical inevitability[7] but also of marvel and perhaps magical power.[8] Eventually, the formal theories are put to work in ambitious reform proposals for collective decision making in the public and private sphere.[9]

The chapter attempts not to give in to the considerable rhetorical allure, but nonetheless to take advantage of the analytical rigor, theoretical insights, and empirical evidence that the literature offers to the practitioner and researcher, not least the legal scholar. The law has to provide rules under which decisions are made in the various branches and embodiments of government as well as in private organizations. By the binding force of the law, the choices made under these rules can affect many people without their consent. With this power comes the responsibility to foster well-reasoned, rational decision making. Law and legal

---

5 James Surowiecki, *The Wisdom of Crowds* (Doubleday, 2004). The title mirrors that of Charles Mackay, *Memoirs of Extraordinary Popular Delusions and the Madness of Crowds* (Bentley, 1841).

6 A reference by Machiavelli in the Discourses on Livy (1517/1531) resonates strikingly well with the modern notion of collective intelligence: '[A]nd it is not without good reason that it is said, "The voice of the people is the voice of God"; for we see popular opinion prognosticate events in such a wonderful manner that it would almost seem as if the people had some occult virtue, which enables them to foresee the good and the evil.' Niccoló Machiavelli, *The Historical, Political, and Diplomatic Writings*, vol 2 (Christian Detmold tr, Osgood 1882) 217.

7 E.g., in labeling crowd wisdom a 'truism', John McCoy and Drazden Prelec, *A statistical model for aggregating judgments by incorporating peer predictions* (Working Paper, 2017) 1.

8 '[O]ccult virtue' in the words of Machiavelli (n 6). The Condorcet jury theorem described in II.1 below provides an example.

9 Among others Michael Abramowicz, *Predictocracy, Market Mechanisms for Public and Private Decision Making* (Yale UP 2007); Scott E Page, *The Difference, How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies* (Princeton UP, 2007); Cass R Sunstein, *Infotopia, How Many Minds Produce Knowledge* (OUP, 2006); Cass R Sunstein and Reid Hastie, *Wiser, Getting Beyond Groupthink to Make Groups Smarter* (Harvard Business Review Press, 2015).

scholarship, therefore, must take a natural interest in the art and science of collective choice. The chapter seeks to serve this interest by introducing the reader to core findings from research into collective intelligence. Their value is seen not so much in providing a ready blueprint for ground-breaking change but in better understanding the working and weaknesses of institutions—obviously with the hope of improving them.

The research on collective intelligence confines itself to only one aspect of decision making: It studies social cognition, that is, collective judgments on questions that aim at a true or correct answer. This is not to say that the respective questions have an answer that can be determined with certainty, only that they are properly approached as a matter of cognition, not of interests or preferences. Legal judgment is an example: Courts are supposed to decide the case on the merits, not according to judges' whims or preferences. But whether the court has succeeded in determinug the actual facts often remains a secret, and a critique of the court's legal argument cannot be proven wrong in any strict sense. Within this broader meaning, any collective choice raises cognitive issues. Another example that will be used in the following is that of a corporation contemplating the acquisition of another firm. To the board of directors, the decision will usually present itself as one in favor or against making an offer at particular terms. One can, however, also frame it as determining a decision threshold, for instance, the maximum price that the corporation should offer in negotiating with the target firm. The optimal reservation price is that which leaves shareholders of the corporation indifferent. The question is one of corporate valuation—an estimate of the expectation value of the stock as a function of the price paid in the acquisition. Such estimates are notoriously uncertain but they are nonetheless a matter of cognition, not volition.

In addition, the literature on collective intelligence implicitly assumes that preferences in the outcome do not affect decision makers. This is a very unrealistic assumption for collective choice in politics, where the preferences of voters and politicians likely also affect and distort their opinions and stated views about the cognitive aspects of a decision. The problem should be less pronounced, and the assumption more defensible, in other areas, such as regulation, law, or business.

The chapter is divided in two main parts. The first introduces key ideas in a form that hopes to be accessible and, more importantly, useful to the legal reader. It concludes with a summary discussion of whether and when collective intelligence is likely to confer advantages over individual judgment (section II). The second part examines the methods of distilling collective intelligence from the separate views of individuals. It starts with the formal techniques that are at the focus of the literature. One approach is surveying—asking group members for standardized responses. Voting is a familiar example (subsection III.1). The second general approach are 'prediction markets', with betting markets as a longstanding practice (subsection III.2). Finally, opinion aggregation also occurs in the course of ordinary, non-standardized communication among people. Suspected to be a breeding place of animal spirits and 'herding', informal deliberation turns out to have advantages of its own (subsection III.3). Section IV concludes with observations about complementarities between aggregation methods.

II. Statistics of collective intelligence

The fundamental claim of collective intelligence is unsurprising: Combining the knowledge of different individuals can lead to better judgment than that of any single individual. As usual, the squabble is in the conditions and qualifications as well as ultimately the empirical success of the theory. It is nonetheless worthwhile to start from the most basic idea, the Marquis de Condorcet's jury theorem (subsection 1). A simple model of individual judgment (subsection 2) lays the ground for a closer analysis of collective intelligence and its statistical determinants (subsection 3) before attempting a preliminary conclusion about the benefits of collective judgment (subsection 4).

1. Condorcet jury theorem

The Condorcet jury theorem considers a judgment between two mutually exclusive but jointly exhaustive—binary—alternatives, such as the truth of a statement or the liability of a defendant. Provided that each juror has the same independent probability greater than .5 of reaching a correct verdict, the jury theorem implies that the probability of a correct judgment by a majority increases with group size.[10] Table 1 demonstrates how larger groups can lever up individual cognitive ability under this simple logic. For advocates of collective intelligence, a tempting reading is that sufficiently large crowds of laypeople (with an error rate of, say, 40%) can trump single experts (e.g., with an error rate of 20%).

	5	**11**	**21**	**51**	**101**	**501**	**1001**
51%	52%	53%	54%	56%	58%	67%	74%
55%	59%	63%	68%	76%	84%	99%	100%
60%	68%	75%	83%	93%	98%	100%	100%
70%	84%	92%	97%	100%	100%	100%	100%
80%	94%	99%	100%	100%	100%	100%	100%
90%	99%	100%	100%	100%	100%	100%	100%

Table 1: Probability of correct majority judgment depending on probability of correct judgment by individual juror (row) and number of jurors (column)

The Condorcet argument also extends to continuous judgment. The maximum price a corporation should offer for the acquisition of a target is a judgment on a continuous scale. In the spirit of the jury theorem, all directors on the board are assumed to provide an independent and unbiased guess of the acquirer's optimal reservation price. Harnessing the power of collective intelligence could then mean to average the individual estimates. The benefit comes from a reduction in estimation error. If the directors are unbiased, this translates in a narrower confidence interval around the collective estimate. Figure 1 demonstrates this effect. It assumes that all individual directors have the same normal distribution of errors.[11] The bell curves can be interpreted roughly as indicating the probability of estimating a value on the

---

10 While never stated as an explicit theorem, the result and supporting theory is laid out in Nicolas de Condorcet, *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix* (1785); see, e.g., ibid xxiii–xxiv ('si la probabilité de la voix de chaque Votant est plus grand que ½, c'est-à-dire, s'il est plus probable qu'il jugera conformément à la verité, plus le nombre des Votans augmentera, plus la probabilité de la vérité de la décision sera grande: la limite de cette probabilité sera la certitude').

11 Assuming a normal distribution is relatively innocuous in the present context. The central limit theorem ensures that the mean of individual estimates approximates a normal distribution of errors as the number of estimates increases.

horizontal axis when the true value is marked by the vertical line. An individual director's judgment can stray as far away from the optimal reservation price as the first, wide bell curve implies. Averaging over three directors narrows the distribution notably. With a board composed of twenty directors, the probability of arriving at a collective estimate close to the true value increases considerably.[12]

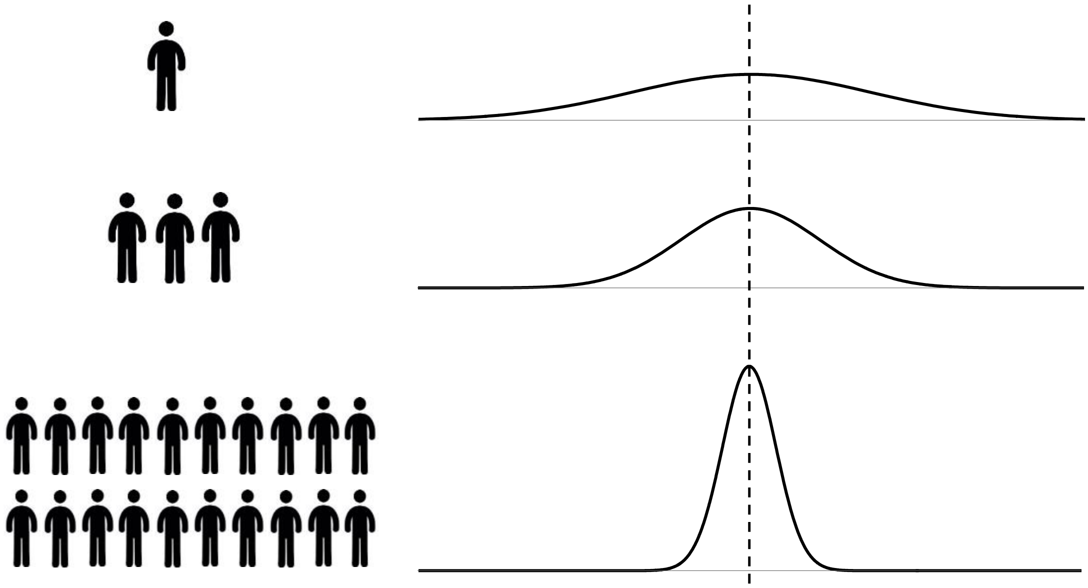

Figure 1: Normal probability distributions of errors for an individual judgment and collective (average) judgments by three and twenty individuals

2. Individual judgment as basis

At first, collective intelligence is an application of statistical sampling—learning about the distribution of human judgment in a given population. The judgment of the population, however, is of interest not in its own right but because it promises information about a relevant question, such as the maximum profitable offer price or the liability of a defendant. Collective intelligence thus seeks to capitalize on individual cognition. While being prone to aberrations and distortions, human cognition has the immense advantage of being able to render judgments about all kinds of variables of interest. Statistical models, by contrast, would need repeated observations in homogenous settings to obtain parameter estimates to enable them to

---

12 Specifically, the standard error declines compared to a single director's estimate by a factor of $\frac{1}{\sqrt{n}}$ where n is the number of directors.

produce predictions. In addition, parameter estimation also requires data on actual outcomes of the dependent, predicted variable. Roughly speaking, one needs to feed the model with correct answers before it can predict unknown answers. Yet many settings never reveal a correct answer—for instance, whether the defendant has in fact committed the tort or ought to be held liable as a matter of law. In such cases, people still make choices or provide intuitive guesses while statistical models can at best predict the outcome of fallible human judgment.

Although explicit statistical models cannot replace it, human judgment itself often implicitly consists of statistical prediction or estimation. When asked to adjudicate a case or to assess the value of an acquisition target, jurors will, consciously or unconsciously, evaluate information available to them and draw inferences from it. The information can be diverse, non-standardized, highly specific to a particular issue and situation, soft or hard; it can range from media reports, official documents, websites, statistical data, personal communication in formal or informal settings, rumors and hearsay to the perceived sentiments of other individuals or a relevant group. It also includes facts, experience, and intuition that bear on the evidentiary weight an individual attaches to any given information and its source. Because evaluators process some—however limited—amount of information their judgment gives at least some indication of the correct answer. Asking individuals for their views, therefore, is a versatile way of obtaining at least some insight.

For any but the most straightforward questions, an individual's judgment will not provide a perfectly reliable 'signal' of the 'true state of the world'. The relationship between the individual assessment and the correct answer can itself be expressed as a statistical model. If the task aims at an estimate on a continuous scale, such as the acquirer's optimal reservation price, a model could be captured in the following expression:

$$\text{individual judgment} = \text{true value} + \text{individual bias} + \text{error term}$$

In the model, individual judgment is linked to the correct answer through the first term of the expression. It can deviate from it, possibly by a large amount, because of random noise from mistakes or haphazard choices the individual evaluator makes on the particular occasion. Such randomness is commonly referred to as the 'error term' with an expectation value of zero. The error term creates variation in judgments but does not drive them in a particular direction away from the true value; the evaluator overshoots as much as she undershoots.[13]

At the same time, individuals also have general tendencies to pick a value that is either too high or too low given a particular question.[14] This introduces bias to their judgments. It can take a positive or negative value indicating a directional deviation from the true value that persists even when the evaluator forms a new judgment about the same question. Bias can reflect individual characteristics, such as psychological dispositions, reliance on heuristics, or intellectual capability to process certain information. A second source of bias is the information an individual is exposed to. Education, profession, personal predilections, ideological leanings, media consumption, and social ties all lead to different patterns of available knowledge from both conscious search and incidental discovery. An individual's information set will hardly be representative of the comprehensive knowledge an ideal evaluator would use in answering the question. A marketing expert, financial analyst, labor representative, lawyer, or engineer will rely on different information when called upon to assess the optimal reservation price in a corporate acquisition. As a consequence, their judgments will diverge in predictable ways. Note that individual bias already threatens to vitiate the optimistic outlook of the Condorcet jury theorem: The expectation value of individual predictions in Figure 1 equals the true value. The wide spread of the bell curve only represents unbiased random error.

3. Statistical determinants of collective intelligence

The goal of utilizing collective intelligence is to mitigate the shortcomings of individual judgments—to reduce or eliminate random error and bias. Moving from individual to collective judgment promises progress on both fronts. Perhaps the more fundamental improvement is that individual bias partly cancels out, making

13 The error term is a 'random variable'. It takes a new, random value ('realization') every time a judgment is made.

14 For a very condensed overview of causes of individual bias, see Clintin P Davis-Stober, David V Budescu, Stephen B Broomell, and Jason Dana, 'The Composition of Optimally Wise Crowds' (2015) 12 Decision Analysis 130, 131.

collective intelligence less biased than a randomly chosen individual (subsection a)). In addition, aggregating individual judgments reduces the undirected noise from individual randomness and different individual biases. This lower variance represented the main lesson obtained earlier from the Condorcet jury theorem (subsection b)).

a) Bias

Errors in judgment are deviations from the true value. They can result from persistent deviation (bias) or from random variation (the 'error term' in the expression above). Knowing an evaluator's ability and available information in theory permits an observer to predict a positive or negative, greater or smaller, error in judgment—the individual bias. Unfortunately, such a reliable error forecast is usually not at hand.[15] If it were, the observer and perhaps even the individual herself could correct her bias by simple addition or subtraction. Yet even without knowledge of the sign and extent of individual bias, the use of collective intelligence offers at least a partial repair. For the organizer of collective choice, ignorance of individual biases makes them a random variable of group members. When collective judgment is formed through averaging of individual judgments, the individual biases add up. As bad as this sounds, it mitigates the bias in collective judgment if individual biases have opposite signs and cancel out. Averaging them reduces collective bias compared to the average individual bias.[16] Fittingly, the idea has been described as 'bracketing' the true value between individual judgments.[17] With much luck, individual biases could reduce to zero so that collective judgment exhibits no bias at all.

---

15 In a specified statistical model of individual judgment, the bias would be a function of observable explanatory variables reflecting the juror's characteristics and information environment. The parameters of the function would have to be estimated based on data from the past on the explanatory variables, judgments, and the true value (!) in the same or sufficiently similar settings.

16 To be more precise, the absolute value of the average bias is less than the average of absolute biases. With b_i denoting individual bias of individual i in a group of n jurors and equally weighted average, it must be that $\left|\sum_{i=1}^{n} \frac{b_i}{n}\right| < \sum_{i=1}^{n} \left|\frac{b_i}{n}\right|$ if there is any pair of b_i, b_j that have opposite sign.

17 Richard P Larrick and Jack B Soll, 'Intuitions About Combining Opinions: Misappreciation of the Averaging Principle' (2006) Management Science 111, 111–112.

The above presentation of the Condorcet jury theorem assumed individual judgments to be unbiased. This could be a shorthand way of saying that in collective cognition, individual biases wash out and can be ignored on average. But even at the collective level, this seems an extreme and unlikely case. Empirically, whenever the collective judgment of a large group misses the mark by a large margin, this almost surely reflects aggregate, collective bias; undirected individual errors cannot explain large deviations as they tend to disappear in larger groups.[18] Major errors in collective judgment therefore attest to group bias. The voluminous body of literature in psychology and behavioral economics about biases in individual judgment is, in fact, also a collection of errors that fail to average out in large subject pools; it is this property that makes them systematic biases of general interest, rather than just individual gaffes.

We consider only one example, an experiment conducted specifically about collective judgment: Participants were asked to bet on American football games based on 'point spreads'. A point spread is set by bookmakers and added to the score of the home team to designate a game's winner for the purpose of the bet. The aim is to balance the odds of the teams by eliminating the greater probability for the favorite team; historical data suggests that point spreads accomplish this almost perfectly.[19] In the experiment, however, the point spreads were manipulated against the respective favorite teams. Without additional information about the particular games or teams, a rational gambler would have to bet against the favorite. Nonetheless, the majority of participants opted to bet on the favorite in 80–90% of 226 games.[20] A majority vote would have led the group to lose significantly more than half of its bets. The result bodes ill for the Condorcet jury theorem, especially in

---

18 This remains the main takeaway from Figure 1 above and is elaborated further in the following subsection b).

19 For a description and discussion, see Joseph P Simmons, Leif D Nelson, Jeff Galak, and Shane Frederick, 'Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds' (2011) 38 Journal of Consumer Research 1, 2–3.

20 The result persisted even when participants were warned of the manipulation. Asking instead for a prediction of the score difference between teams effectively de-biased participants and their collective judgment, Simmons et al (n 19) 7–12.

its original domain of majority voting: the probability of reaching a correct judgment could be below .5.[21]

b) Variance

What collective intelligence reliably delivers is a reduction in variance. This concerns the ‘error term’ in the model of individual judgment, representing random mistakes, arbitrary choices, and the whims of the moment. The individual error terms ‘drops out’ in very large groups and is substantially reduced even in smaller groups as positive and negative deviations cancel each other out. This is precisely the effect showcased in Table 1 and Figure 1 above. Pooling evaluators also reduces the variance in bias across group members, compared to picking a single evaluator at chance and provided that biases are less than perfectly correlated.

The flip side is that the lower variance also reduces chance corrections of collective bias. Whatever systematic judgment distortion exists will come out more reliably in the aggregate. Collective intelligence is more persistent also in its mistakes. Because of this, certain decision making procedures—most prominently the majority principle—can make groups look worse than individuals even if they are in aggregate less biased than the average individual, as one should expect based on the previous analysis. In fact, in the experiment about betting against rigged point spreads, the collective of participants, deciding by the majority rule, performed worse than 93% of its members on an individual basis.[22] Collective intelligence thus seems at a disadvantage to individual judgment. This, however, is an artefact of the majority rule, which translates the group’s reduced variance into more consistent implementation of its bias. Based on a continuous measure, the group in the experiment performed better at avoiding the favorites. Knowing the size and direction of the bias, a suitable supermajority requirement could have restored the advantage of collective intelligence.[23]

---

[21] Condorcet did anticipate this possibility, especially for large assemblies that, in his view, necessarily had to include less enlightened voters, see Condorcet (n 10) xxiv.

[22] Absent the de-biasing procedure described in n 20, Simmons et al (n 19) Table 4.

[23] Clintin P Davis-Stober, David V Budescu, Jason Dana, and Stephen B Broomell, ‘When Is a Crowd Wise’ (2014) 1 Decision 79, 94–96.

4. Are groups wiser than individuals?

The contention of 'wise crowds' raises the question whether and when groups have an advantage over single evaluators. We leave aside the cost of group judgment; it is most likely higher and can only be justified if collective intelligence promises better outcomes. We also assume that individual judgment remains equally accurate—in terms of bias and variance—when it is delivered as sole evaluator compared to as a member of a potentially large group. These aspects are kept for later, as is the possibility of amplified bias from interaction within a group or 'herding'.

With these complications out of the way, one still needs to be explicit about the available knowledge when choosing between collective intelligence and individual judgment. The decision would be trivial and pointless if the social planner already knew the correct answer to the question under consideration. Therefore, one should conceive of her as being susceptible to bias and error as much as any other evaluator. This prevents her from hand-picking a team with zero aggregate bias. But without making her task trivial, the social planner can have—and plausibly has, as other observers—some information about the 'general ability' of evaluators, including her own, for particular classes of judgments. In the above expression describing the generation of individual judgment, such (noisy) information could relate to the probability distributions of individual biases, relative to the unknown true value, and of the error terms.

To illustrate, the social planner may be able to rank potential evaluators by their expertise on the subject matter. If expertise is marked by greater investment in knowledge and frequent validation of past judgments from social feedback and real-world outcomes, one would expect an expert's judgment to reflect more information and more different aspects than that of a layperson. Such considerations suggest that expert judgments tend to suffer from less individual bias. Conversely, if non-experts are more prone to using a limited number of cues, this could entail not only greater variance but also higher correlation of individual biases because the available and salient information could come from shared public sources.

Knowledge of differences in cognitive characteristics can guide the choice between collective and individual judgment. If distributions of individual bias and random error among candidates tend to be similar, collective intelligence safely beats a single evaluator— keep in mind that we abstract from decision-making cost, individual effort, and 'herding'. Under these conditions, aggregating the judgments of cognitively similar individuals reduces random variance and has a fair chance of mitigating bias while never increasing it.

To find advantage in using only a single decision-maker, there must be a substantial difference in judgment quality. Even then, the argument is not straightforward. As regards variance, more evaluators still improve performance provided that one can adjust the weight of individual judgments to differences in variance.[24] Without proper adjustment, however, adding less able evaluators spoils the broth. Expanding the group thus requires knowledge of variances in individual biases and error terms; the social planner would also have to make further adjustments for uncertainty in her own assessment of variances. More difficulties arise if additional evaluators are susceptible to aggregate bias as can be suspected in certain instances of lay judgment. While one can try to correct for such systematic bias, this opens up new opportunities for serious mistakes and may not be worth the rather modest gains. Intuitively, one should not meddle with a construction engineer's assessment of the carrying capacity of a building structure based on a population survey. The best response to a cardiac arrest is to call for a doctor, not to conduct a poll among bystanders.[25]

By way of a preliminary conclusion,[26] the collective intelligence of groups has much to commend it. The default advice rarely is to restrict decision making to a single mind. The principal reason to exclude evaluators, perhaps even down to only one person, is major differences in the quality of individual judgment, specifically in variance and systematic, directional bias, the model example being expertise in a field of specialized knowledge or practice. With noticeable differences in evaluator

---

24 A higher-variance evaluator optimally receives a lower weight in averaging but should never be completely disregarded.

25 Albert E Mannes, Jack B Soll, and Richard P Larrick, 'The Wisdom of Select Crowds' (2014) 107 Journal of Personality and Social Psychology 276, 278.

26 Davis-Stober et al (n 23) provide a more rigorous and comprehensive assessment. See also Mannes et al (n 25) 278–279 for a discussion of the relevant settings.

quality, limiting the scope of the group to the more able individuals seems advisable.[27]

III. Generation of collective intelligence

Gathering and combining the dispersed knowledge of individuals is not a recent idea—it pertains to human nature as a social animal. What fuels the excitement over collective intelligence are quantitative techniques of extracting or representing the common judgment of a group, starting from the very basic mechanism of majority voting. Besides giving the group's judgment separate expression from that of its members, these formal mechanisms lend themselves to mathematical modeling and empirical evaluation. The resulting framework allows more rigorous analysis and learning about the performance of collective intelligence and about conditions conducive to it. We first consider the more conventional approach of eliciting assessments by asking or 'surveying' individuals (subsection 1) before turning to the more fanciful contrivance of a prediction market (subsection 2). These two general types of formal mechanisms are then contrasted with the naturally occurring formation of collective views in informal exchanges or 'deliberation' (subsection 3).

1. Surveying

Asking participants to answer a question in standardized form and then aggregating the responses is a common technique to arrive at a collective judgment. A prime example is voting in politics, the law, and all kinds of organizations. Voting is set apart, however, because it aims at producing not just a cognitive assessment but a collective choice between alternatives. It often involves divergent preferences and power struggles. Even with uniform preferences, joint decision making through voting is more than just extracting the maximum amount of information from a group. The number of options and the rules to establish a decision—such as majority requirements—reflect considerations about organizational needs, the expected quality of collective judgment, and risk trade-offs. The above example of betting decisions illustrates the difference: the group's informationally superior collective

[27] For a simulation of different settings, see Mannes et al (n 25) 279–284.

judgment ultimately produced inferior choices under a simple-majority decision rule.[28]

Turning to purely informational surveying, a thriving research branch explores strategies for improving the aggregate estimates of groups. In line with the informal arguments presented above, it has been shown that limiting the group to the most able individuals or giving their assessments greater weight enhances collective judgment.[29] Ability can be measured on a stand-alone basis—having more accuracy individually—or relative to other jurors, either by having less bias or a consistently corrective one.[30] The latter approach overlaps with the notion that the group benefits from using more diverse information.[31] Besides including *individuals* with different cognitive approaches and information backgrounds, researchers have invented algorithms that seek to detect and overweight *judgments* containing novel or divergent information that is under-represented in the group. The common idea behind these algorithms is asking evaluators about not only their own judgment but also the judgment distribution among others. The latter answer reflects what individuals believe to be the group's shared information base while their own judgment in addition incorporates the respective evaluator's own, private information. Simple averaging underweights private information, which shows only

---

28 For strategic voting in spite of homogenous preferences, see J M M Goertz, 'Inefficient committees: small elections with three alternatives' (2014) 43 Social Choice and Welfare 357, including the short literature overview ibid 358–360.

29 Regarding individual differences in ability (outside the group composition literature), a large-scale experiment in forecasting political events over three years showed persistently superior performance of participants with the top 2% record in the respectively previous year, Barbara Mellers et al, 'Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions' (2015) 10 Perspectives on Psychological Science 267; Barbara Mellers et al, 'Psychological Strategies for Winning a Geopolitical Forecasting Tournament' (2014) 25 Psychological Science 1106, 1109, 1111. Note that measuring ability requires information about the correct answer (which in the case of forecasting is provided by future development); for crowd wisdom itself as the target value, see Eyal Baharad, Jacob Goldberger, Moshe Koppel, and Shumel Nitzan, 'Beyond Condorcet: optimal aggregation rules using voting records' (2012) 72 Theory and Decision 113; for crowd wisdom as indicator of correct answers even when they are in principle observable R H J M Kurvers et al, 'How to detect high-performing individuals and groups: Decision similarity predicts accuracy' (2019) 5 Science Advances eaaw9011.

30 Evidence for improvements based on stand-alone criteria: Mannes et al (n 25) 284–286; on contribution to the group: David V Budescu and Eva Chen, 'Identifying Expertise to Extract the Wisdom of Crowds' (2015) 61 Management Science 267, 270–277.

31 The trade-off between diversity and accuracy is analyzed in Davis-Stober et al (n 14), qualifying the strong emphasis put on diversity by P J Lamberson and Scott E Page, 'Optimal Forecasting Groups' (2012) 58 Management Science 805.

in a single judgment (or few judgments) while there is no a priori reason to assume that it is less relevant than shared information.[32] An improved aggregation rule then strengthens the effect of private information on group judgment.[33]

The accuracy of individual judgments hinges not only on expertise and ability but also on effort. Genuine interest in the question often motivates effort but explicit incentives also matter. Rewards or penalties can attach at the level of the group, such as when committees are held accountable for their decisions. Small and stable groups also are more effective at policing individual shirking. Larger crowds depend more on self-motivated individuals, which can introduce a self-selection bias from participation.[34] Monetary rewards would often be too costly. A different kind of explicit incentives are social 'bragging rights'[35] tied to tournament-style competition over the most accurate judgments.[36] Yet such contests can induce evaluators to report extreme views away from their own best guess to maximize the chance of finishing first.[37]

---

32 See Oliver Kim, Steve C Lim, and Kenneth W Shaw, 'The Inefficiency of the Mean Analyst Forecast as a Summary Forecast of Earnings' (2001) 39 Journal of Accounting Research 329, 330–333; empirical evidence regarding economic forecasts in Christopher Crowe, *Consensus Forecasts and Inefficient Information Aggregation* (IMF Working Paper 10/178, 2010).

33 Asa B Palley and Jack B Soll, 'Extracting the Wisdom of Crowds When Information is Shared' (2019) 65 Management Science 2291 ('pivoting' procedure); Dražen Prelec, H. Sebastian Seung, and John McCoy, 'A solution to the single-question crowd wisdom problem' (2017) 541 Nature 532 ('surprisingly popular' rule for binary choice); McCoy and Prelec (n 7) (providing a generalization of Prelec et al).

34 A well known example are online reviews where dissatisfied or moderately satisfied customers tend not to participate, see the overview in Steven Tadelis, 'Reputation and Feedback Systems in Online Platform Markets' (2016) 8 Annual Review of Economics 321, 333–336; for improvements from explicit incentives at a platform for employer reviews Ioana Marinescu, Nadav Klein, Andrew Chamberlain, and Morgan Smart, *Incentives Can Reduce Bias in Online Reviews* (NBER Working Paper 24372, 2018).

35 Jens Witkowski et al, 'Incentive-Compatible Forecasting Competitions' (2018) 32 AAAI Conference on Artificial Intelligence 1282, 1282.

36 The top 2% participants in the experiment referred to in n 29 showed far greater effort than other participants; besides explaining their success it also attests to the very lopsided effect of tournament incentives, see Mellers et al, 'Identifying and Cultivating' (n 29) 277; Mellers et al, 'Psychological Strategies' (n 29) 1111.

37 For a complicated remedy, see Witkowski et al (n 35). But see also Kenneth C Lichtendahl Jr, Yael Grushka-Cockayne, and Phillip E Pfeifer, 'The Wisdom of Competitive Crowds' (2013) 61 Operations Research 1383 (discussing the desirable effect that tournament incentives tilt individual judgments away from shared towards private information).

Incentives—whether they be explicit or implicit—require a measure of success. In many important real-world applications, the correct answer remains subject to debate; think again of the acquirer's optimal reservation price or the judgment rendered by a court. Even without a clear indicator of truth it is possible, at least theoretically, to provide incentives that induce effort. The key idea is to link rewards or penalties to the reported judgments of others.[38] Implementing such a scheme requires much knowledge about the probability distributions mapping different possible truths to individual judgments.[39] The difficulty explains why the law offers no explicit incentives to judges to promote accurate adjudication.

A final aspect of surveying is the optimal size of the group, about which a lot has already been said. If evaluators differ in ability, expanding the group helps little and could even do harm without proper adjustment of aggregation weights.[40] Effort motivation and incentives likewise counsel against large groups. Smaller groups also tend to be cheaper to survey and administer. All of these considerations caution against enthusiasm for large crowds as collective arbiter.

2. Prediction markets

Asking individuals is straightforward but not particularly inventive. The vision of a collective intelligence might seem to call for something more congenial to spontaneous order than a centrally administered collection and summation of responses. It thus has caught the imagination of many researchers that collective judgment, rather than being established by an intermediating agent, could emerge

---

38 The incentive scheme interprets an evaluator's judgment as a prediction of others' judgments and apply a 'proper scoring rule' that induces truthful reporting of the prediction, see Nolan Miller, Paul Resnick, and Richard Zeckhauser, 'Eliciting Informative Feedback: The Peer-Prediction Method' (2005) 51 Management Science 1359, 1360–1364; see also Radu Jurca and Boi Faltings, 'Mechanisms for Making Crowds Truthful' (2009) 34 Journal of Artificial Intelligence 209. For a simple explanation of scoring rules, see Abramowicz (n 9) 111–114. On the related notion of evaluating individual ability based on others' behavior see the last references in n 29.

39 Miller et al (n 38) rightly point out that participants themselves need not possess this information as long as they believe the administrator to apply the scheme properly, ibid 1363. They argue that internet rating platforms could have sufficient information, ibid 1368–1369.

40 For limiting groups to the most able individuals, see again Mannes et al (n 25). For the gain in accuracy depending on group size, see, e.g., Stefano DellaVigna and Devin Poper, 'Predicting Experimental Results: Who Knows What?' (2018) 126 Journal of Political Economy 2410, 2425–2426.

naturally from the uncoordinated choices of its constituents. Such a procedure in fact exists in the form of 'information' or 'prediction markets'. Hayek is usually cited for the notion that markets collect dispersed information and incorporate it in a single price.[41] The Hayekian theory, however, related to individual knowledge of needs and resources,[42] that is, to the divergent private values of goods. To become a medium of collective intelligence, the market exchange has to be in goods of a common value reflecting the question of interest. After explaining the design of such markets (subsection a)) their limitations will be considered (subsection b)).

a) Design

To turn market prices into oracles of collective intelligence, the traded good must be imbued with meaning, and that meaning should determine the value of the good for all market participants.[43] The means to this end is a derivative contract or—more simply—a bet: One party promises to pay an amount contingent on an event that the market is meant to predict. A prominent example are bets on the outcome of an election under which the promisor has to pay one Euro if a particular candidate wins or loses. In exchange, the promisee pays a contract price for receiving the promise. The good or asset consists of the contractual promise.

Figure 2 provides a simplified description how a market in such contracts can aggregate the views of few individual traders. Each of the horizontal lines represents a trader's order: the length of the lines depicts the quantity of contracts the trader is willing to buy or sell (by taking the position of promisee or promisor), the vertical position shows her reservation price, assumed to be equal for buying or selling. In the example of the one-Euro bet on the election of a candidate, a risk-neutral bettor

---

41 Professor Sunstein eloquently characterizes prediction markets as 'Hayek's Challenge to Habermas', Cass R Sunstein, 'Deliberating Groups versus Prediction Markets (or Hayek's Challenge to Habermas)' [2006] Episteme 192; see also Sunstein and Hastie (n 9) 181–183; Robert Forsythe, Forrest Nelson, George R Neumann, and Jack Wright, 'Anatomy of an Experimental Political Stock Market' (1992) 82 American Economic Review 1142, 1143 ('Hayek hypothesis').

42 The 'particular circumstances of time and place', see F A Hayek, 'The Use of Knowledge in Society' (1945) 35 American Economic Review 519, 521–522.

43 That is, the fundamental value of the good—from holding it to maturity—should be the same for all potential owners. In the parlance of bargaining and auction theory, it should be a common, not private value.

might be willing to pay any amount smaller than her probability estimate times one Euro for buying the contract (being the promisee), and take the position of seller (promisor) for any greater amount. For instance, if she tallies the candidate's winning probability at .37, she would buy at any price less than 37 Cents and sell at any greater price. Competitive bidding—such as through the order book of a stock exchange[44]—should lead the market to 'clear', that is, to realize all trading opportunities from matching orders with different reservation prices. Market clearing results in a price closely around the reservation price of the median bid as shown in Figure 2.[45]

44 For a handy explanation of the operation of order books, see Oxera, *The design of equity trading markets in Europe* (2019), 19–22 <https://www.oxera.com/wp-content/uploads/2019/03/design-of-equity-trading-markets-1.pdf > accessed 25 March 2020.

45 The median bid itself can be matched only in part. The market clearing price can be anywhere in the range between the median bid and the next higher or lower resevation price, depending on whether the median trader acts as seller or buyer in the contracts that she can conclude.

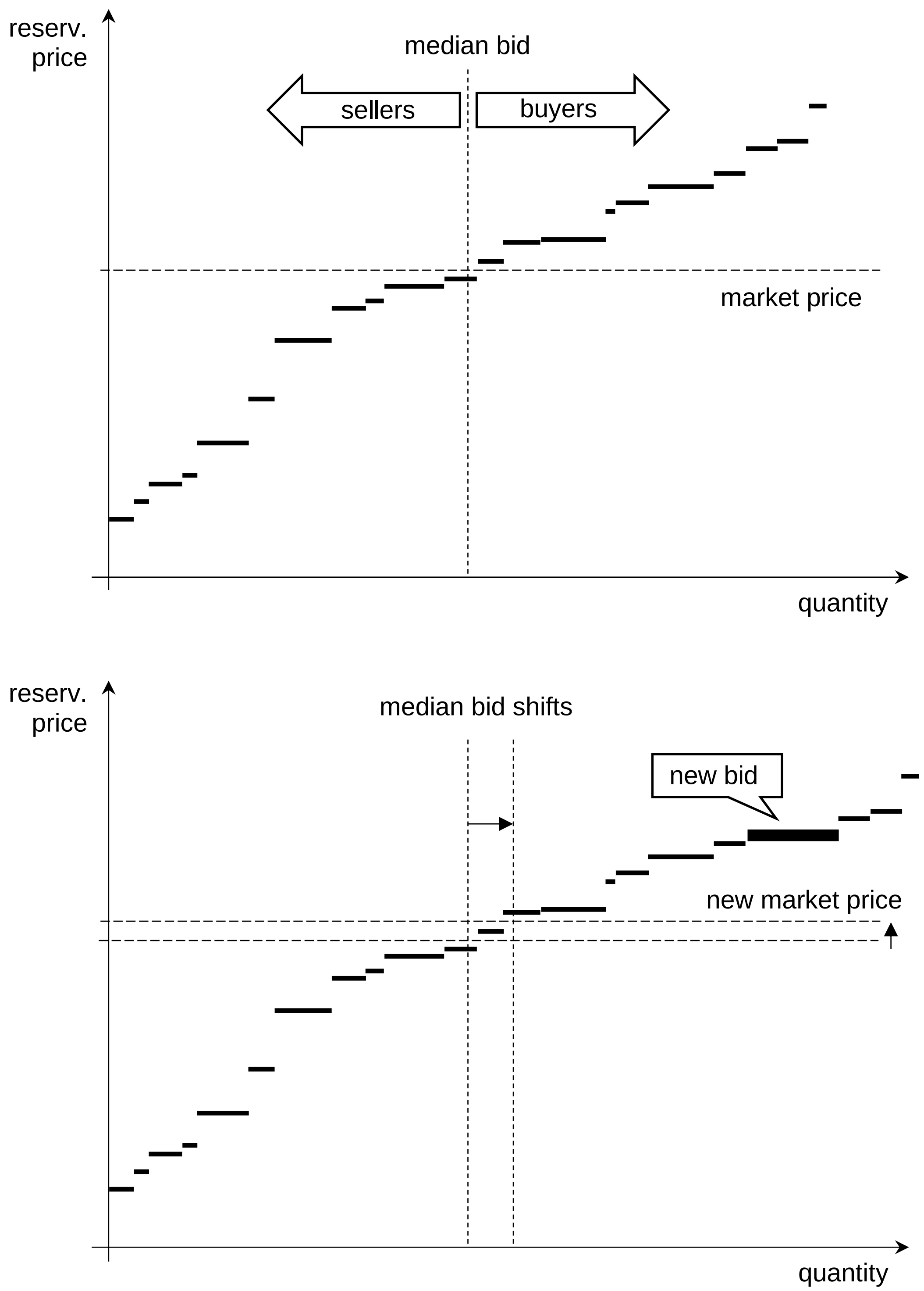

Figure 2: Market price depending on buy/sell orders with quantities and truthfully revealed reservation prices (with willingness to pay equal to willingness to accept). Price is set to match as many orders as possible, that is, to clear the market. The below graph depicts the price change in response to the arrival of a new order.

The lower graph of Figure 2 demonstrates how adding an individual judgment—in the form of a new order—affects the market price: The new order shifts the median to the right or left, depending on whether it is above or below the original median bid. The resulting market price reflects the change in the median bid. No central agent attaches weights to individual judgments. In fact, the information contained in the distribution of reservation prices is entirely lost.[46] Order size—the amount a trader is willing to wager—determines how much her judgment counts, not whether it is deemed valuable by a survey organizer. Incentives naturally align as market participants decide how much they invest in the accuracy of their own assessment. As only the more able and diligent traders have positive expected earnings, prediction markets also tend to self-select better evaluators without a need to actively manage the composition of the group.

b) Limitations

The self-governing features of prediction markets have ignited much excitement. Markets have been hailed as an information panacea for all domains of decision making and credited with staggering forecasting successes.[47] It should be kept in mind, however, that markets are not a divine or natural order that human intervention can only destroy or taint. They are subject to, and often even the creation of, human design.[48] This is particularly true of prediction markets. With the sole purpose of aggregating the dispersed information or judgments of individuals, prediction markets are up against a fundamental result of economic theory: The 'no-trade theorem' essentially states that market participants are unwilling to trade on valuation differences that reflect only private information.[49] The underlying intuition

---

46 The key role of the marginal trader has even been considered the major driver of prediction market accuracy, Forsythe et al (n 41), 1157–1160.

47 See the call for a regulatory safe harbor from 22 leading social scientists in Kenneth J Arrow et al, 'The Promise of Prediction Markets' (2008) 320 Science 877; Sunstein (n 41) ('In countless domains, their forecasts have proved extremely accurate.'); Robin Hanson, 'Shall We Vote on Values, But Bet on Beliefs' (2013) 21 Journal of Political Philosophy 151, 154 ('While government policy may often suffer info failures, speculative markets show striking info successes.'); Jason Dana, Pavel Atanasov, Philip Tetlock, and Barbara Mellers, 'Are markets more accurate than polls? The surprising informational value of "just asking"' (2019) 14 Judgment and Decision Making 135, 135 ('resounding success', 'impressively accurate predictions').

48 For a more comprehensive summary of the limitations, see Justin Wolfowitz and Eric Zitzewitz, *Five Open Questions About Prediction Markets* (NBER Working Paper 2006).

49 Paul Milgrom and Nancy Stokey, 'Information, Trade and Common Knowledge' (1982) 26 Journal of Economic Theory 17, 21–24; Jean Tirole, 'On the Possibility of Speculation under Rational Expectations' (1982) 50 Econometrica 1163, 1166–1168.

is that rational traders agree on an asset's valuation when they all share the same information.[50] Another trader's willingness to accept (or pay) less (more) than one's own reservation price reveals that the other party has different information and forces a correction of one's own valuation, eliminating the bargaining range. More simply put, without a money inflow from outside every trader knows the market to be a zero-sum game. Whatever one market participant hopes to gain must come at the expense of another. Trading on valuation differences would be no more than a gamble that risk-averse people usually dislike. It can only arise between parties with inconsistent, self-serving beliefs about their own chances to outsmart each other.

The emergence of prediction markets becomes even less likely when participation costs are taken into account. These are the costs of entering into transactions but also—more significantly—the effort needed to gather information to avoid losing money against better informed traders. When these effort costs are considered, entering the market becomes a losing proposition. Ironically, while they are supposed to feed on the rational self-interest of traders, many real world prediction markets depend on non-financial motives such as gambling preferences or a genuine interest in the subject matter (making judgment formation effortless and even a source of utility); examples are betting markets on sports or politics events.[51] Information processing can also be a byproduct of markets that exist for other purposes. The prime examples are financial markets that offer investment and hedging (insurance) services. Investors and hedgers trade in these markets to match their complementing time and risk preferences—for instance, to invest when others want to liquidate. These complementarities create so many benefits that they can sustain positive expected returns for rational information traders (arbitrageurs, speculators) from exploiting mispricings and thereby enhancing the predictive accuracy of market prices.

---

50 More specifically, the theorem assumes traders to have 'common knowledge' of each other's rationality and common Bayesian priors, Milgrom and Stokey (n 49) 19–24.

51 For betting markets on the outcomes of U.S. presidential elections 1884–1928, see Paul W Rhode and Koleman S Strumpf, 'Historical Presidential Betting Markets' (2004) 18 Journal of Economic Perspectives 127, 128–129.

Designated prediction markets thus face a similar motivational challenge as do survey techniques. Sustaining them in principle requires an outside subsidy, turning the market into a positive-sum game for traders at the expense of the sponsor.[52] Apart from infusing money, a social planner can try to encourage participation by appealing to curiosity in the question, commitment to a cause or group—that the market is meant to serve—or again to preferences for gambling or competitive contests.[53] Once sufficiently many traders have been attracted, their behavior can still fall in line with the logic of self-interested market trading—even if the market uses play money that cannot be exchanged for cash.[54] Market incentives offer distinctive advantages for information aggregation. It was observed above that averaging of survey responses should attach greater weight to private information to prevent it from being swamped in the many judgments reflecting only the same widely shared information. Prediction markets provide an inbuilt incentive to behave accordingly. Profits can be made in expectation only based on information that the market price does not already incorporate. Rational traders therefore should act only on information that they believe not to be shared by many others.[55] The very promise of prediction markets rests on this incentive to seek out new information instead of reaffirming the conventional view.

---

[52] Designing a subsidy can be tricky. For instance, simply rewarding trading activities could induce profit-maximizing participants to place random orders. An intriguing albeit untested proposal are information-forcing scoring rules (see n 38) directed at the public to induce market activity, see Robin Hanson, 'Combinatorial Information Market Design' (2003) 5 Information Systems Frontiers 107.

[53] Prediction markets have been used for purposes of public policy, within firms, and for academic research, see, respectively, Robin D Hanson, 'Designing real terrorism futures' (2006) 128 Public Choice 257, 258–260; Bo Cowgill and Eric Zitzewitz, 'Corporate Prediction Markets: Evidence from Google, Ford, and firm X' (2015) 82 Review of Economic Studies 1309, 1314–1319; Colin F Camerer et al, 'Evaluating replicability of laboratory experiments in economics' (2016) 351 Science 1433, 1434–1435.

[54] A comparison of two sports betting markets found no significant difference in accuracy between the one operating with real money and the other using play money, Emile Servan-Schreiber, Justin Wolfers, David M Pennock, and Brian Galebach, 'Prediction markets: Does money matter' (2004) 14 Electronic Markets 243. For the modest incentives in corporate prediction markets, see Cowgill and Zitzewitz (n 53) 1319.

[55] On this anti-herding feature, see Christopher Avery and Peter Zemsky, 'Multidimensional Uncertainty and Herd Behavior in Financial Markets' (1998) 88 American Economic Review 724, 725, 728–730.

Whether the promise is fulfilled depends on how well individuals ascertain the distribution of information in the market.[56] The available evidence confirms the forecasting faculties[57] of markets but also points to certain persistent biases.[58] In a large-scale experiment involving hundreds of probability estimates for political events, prediction markets turned out to be more accurate than an unweighted average of individual estimates from a survey.[59] However, adjusting weights for individual skill and excluding stale forecasts[60] equated performance of the two aggregation methods.[61] Yet again, tweaking the survey predictions requires sufficient data to estimate the parameters of the algorithm[62]—a benefit that the market mechanism provides at no cost.[63]

For all their advantages and allure, prediction markets have one major prerequisite that limits their application: They need something to predict. A derivative contract must relate to a well-defined 'underlying' variable that at some point assumes a

---

56 Avery and Zemsky (n 55) 735–737 offer an explanation of herding in financial markets along these lines.

57 A well-known success story is elections forecasting where prediction markets tend to perform better than opinion polls, see Forsythe et al (n 41); Joyce E Berg, Forrest D Nelson, and Thomas A Rietz, 'Prediction market accuracy in the long run' (2008) 24 International Journal of Forecasting 285, 290–298. For corporate prediction markets Cowgill and Zitzewitz (n 53) 1320–1332. Further performance results in Justin Wolfers and Eric Zitzewitz, 'Prediction Markets', in Steven N Durlauf and Lawrence E Blume (eds), New Palgrave Dictionary of Economics (2nd ed, Palgrave, 2008). A classic early reference regarding horserace betting is Stephen Figlewski, 'Subjective Information and Market Efficiency in a Betting Market' (1979) 87 Journal of Political Economy 75.

58 The 'favorite-longshot' bias leads markets to underrate likely winners and to overvalue very low-probability events, see Lionel Page and Robert T Clemen, 'Do Prediction Markets Produce Well-Calibrated Probability Forecasts' (2013) 123 Economic Journal 491; see also the brief summary of explanations, ibid, 492–494. For over-optimism and other biases in corporate prediction markets, see Cowgill and Zitzewitz (n 53) 1327–1332.

59 Pavel Atanasov et al, 'Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls' (2017) 63 Management Science 691, 696–697; Dana et al (n 47) 138–139.

60 The market's prediction was always taken from the most recent price. Excluding older responses from the survey can be seen as a special case of correcting the bias in favor of shared information, see n 32. A third correction consisted of 'extremizing' the aggregate estimates. See Atanasov et al (n 59) 694, 696 for a summary of the algorithm and its justifications.

61 Atanasov et al (n 59) 696–697, 702–703; Dana et al (n 47) 138–139.

62 Atanasov et al (n 59) 696–698 (also providing a robustness analysis).

63 As a final twist of the argument, survey predictions appeared to contain information beyond market prices and to perform better than markets on longer horizons, Dana et al (n 47) 139–141.

verifiable value,[64] such as the outcome of an election, a firm's sales in a given quarter or the occurrence (or not) of an event within a stipulated timeframe. Many important and much needed judgments lack such clear validation. The problem surfaced already in connection with possible incentives for survey respondents; sharp incentives for traders are the hallmark of prediction markets. They stand or fall by the prospect of an unequivocal resolution of the question. Where collective intelligence is to pass judgments of a different kind, it cannot be obtained from a market.[65]

3. Deliberation

Formal aggregation procedures are fascinating objects of study and, at least in the form of voting, have broad application and importance for the law. Nonetheless, in the greater scheme of things collective judgment is formed predominantly through non-standardized communication—talking with each other. Like surveys or prediction markets, deliberation among group members leads to an aggregation of views, only that it occurs at the individual level by adjusting individual judgments to what one learns from others. People constantly 'survey' each other. If formal aggregation eventually takes place it often ingests the results of informal aggregation within individual minds. Yet unavoidable as it is, one may ask whether deliberation rather helps or hampers collective intelligence, especially when formal aggregation methods could serve as partial substitutes. In fact, the hope that standardized surveys or prediction markets can yield better judgment than conventional discussions accounts for some of the appeal of research in collective intelligence.[66]

a) Information suppression ('herding')

Because merging different views can yield a more accurate and reliable judgment, rational individuals should take the opinions of others into account. There is,

---

64 Justin Wolfers and Eriz Zitzewitz, *Five Open Questions About Prediction Markets* (NBER Working Paper 12060, 2006), 9–10 ('contractability').

65 But see the vision of a 'predictocracy' developed by Abramowicz (n 9) 137 ff where markets are to predict and partially substitute decision makers.

66 See, e.g., Sunstein (n 41) and the references in n 9.

however, a critical condition: Individuals have to act as well-calibrated Bayesian updaters. In incorporating the views of others, they must give adequate weight to the different pieces of information. They should continue to use their own information and, beforehand, produce it through observation and reflection. Skeptics of deliberation question that this condition is often and sufficiently met. They suspect that individuals are poor aggregators of judgments and dissipate much of the potential of collective intelligence. The conclusion would be to substitute scientifically proven surveying methods or prediction markets as much as possible for unguided deliberation.

Critics of deliberation can point to abundant evidence about people's inclination to neglect their own information and overemphasize the views of others. Common rubrics include 'social influence', 'conformity', or 'herding'. While 'herding' evokes an image of impulsive animal behavior, it should be repeated that adjusting one's prior assessment to information from others is a tenet of rationality and, after all, the principle behind collective intelligence. People *ought to* 'follow the herd' cognitively if crowd wisdom has greater evidentiary value than the individual's private view. A telling example are experiments from the 1950s where participants succumbed to the consensus of an orchestrated group over the lengths of graphical lines, even though it manifestly contradicted the readily visible evidence.[67] As much as these findings are astonishing, they need not result from erroneous aggregation. If participants did not suspect that other group members had been staged to lie about their perceptions, the very simplicity of the task—comparing the length of lines—provided strong reason to be confident about the independent assessment of *several* others. The unfortunate victims had excellent reason to question their own senses.[68]

---

67 Solomon E Asch, 'Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority' (1956) 70 Psychological Monographs 1; Solomon E Asch, 'Opinions and Social Pressure' (1955) 193 Scientific American 31; Morton Deutsch and Harold B. Gerard, 'A Study of Normative and Informational Social Influences upon Individual Judgment' (1955) 51 Journal of Abnormal and Social Psychology 51, 629.

68 In line with this rationalization, the propensity to yield to the group depended on its size and on the presence of dissenters, Asch 'Opinions' (n 67) 33–35; the experiment by Deutsch and Gerard (n **Fehler! Textmarke nicht definiert.**) showed that participants gave in even in anonymous settings without social pressure.

The reiterated caveat notwithstanding, judgment aggregation at the individual level is certainly far from optimal. It poses a threat to collective intelligence mostly when it introduces systematic bias in the resulting individual judgments, that is, when all or most group members attribute excessive or too little weight to the same pieces of information. In the model of individual judgment introduced earlier,[69] this adds a uniform bias to the judgments of all group members, preventing individual biases from cancelling out. 'Following the herd' by adopting a generally accepted—potentially biased—view can nonetheless be individually rational if the cost of information is taken into account: When several others have endorsed a prevailing judgment, adding another independent evaluation may not be worth the effort individually as even a piece contradictory evidence would weigh little against the information accumulated in the consensus; collectively, challenging the dominant view could still be worthwhile.[70] In addition, plenty of empirical and theoretical work indicates that people also herd irrationally by using flawed aggregation strategies. For instance, it is both theoretically plausible and supported by experimental evidence that individuals fail to discount properly the weight of identical judgments that reflect the same underlying information.[71] Also credible but less clearly defined is the effect of social norms and 'groupthink' on judgment

[69] II.2 above.

[70] The standard references are Sushil Bikhchandani, David Hirshleifer, and Ivo Welch, 'A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades' (1992) 100 Journal of political Economy 992; Abhijit V Banerjee, 'A Simple Model of Herd Behavior' (1992) 107 Quarterly Journal of Economics 797.

[71] See, e.g., Peter M DeMarzo, Dimitri Vayanos, and Jeffrey Zwiebel, 'Persuasion Bias, Social Influence and Uni-Dimensional Opinions' (2003) 118 Quarterly Journal of Economis 909 (theoretical implications of 'persuasion bias' as the failure to adjust to repetitious presentation of the same information); David V Budescu and Hsiu-Ting Yu, 'Aggregation of Opinions Based on Correlated Cues and Advisors' (2007) 20 Journal of Behavioral Decision Making 153, 166–173 (low effect of correlation in the hints of 'advisors' on confidence of 'decision makers' in experiments about a medical diagnostic task); Ilan Yaniv, Shoham Choshen-Hillel, and Maxim Milyavsky, 'Spurious Consensus and Opinion Revision: Why Might People Be More Confident in Their Less Accurate Judgments?' (2009) 35 Journal of Experimental Psychology: Learning, Memory, and Cognition 558, 560–561 (lower confidence in aggregated judgment based on *random* sample of 'advisors' as compared to *selective* sample of advisor estimates close to own original judgment); for a review of the 'hidden profile' literature on group deliberation, see Li Lu, Y Connie Yuan, and Poppie Lauretta McLeod 'Twenty-Five Years of Hidden Profiles in Group Decision Making: A Meta-Analysis' (2012) 16 Personality and Social Psychology Review 54; for the general profile of a 'credulous Bayesian' Edward L Glaeser and Cass R Sunstein, 'Extremism and Social Learning' (2009) 1 Journal of Legal Analysis 263, 275–300.

aggregation at the individual level.[72] A review of the extensive and diverse literature on herding and social influence is beyond the scope of this chapter.[73] Suffice it to say that there is ample evidence for major shortcomings in how individuals factor in the views of others.

b) Information discovery

These flaws suggest that free deliberation poses a hazard to collective intelligence. Formal aggregation then presents itself as a superior alternative. Intuitively, it seems clear that ultimately no common judgment can be formed without informal communication. But there is value in explaining why deliberation is more than just a necessary evil and enjoys an additional advantage over aggregation methods that rely on standardized messages, be it survey responses or market orders. This distinctive feature is the lack of pre-imposed structure—the possibility of every participant to refine the space of available signals by specifying the content of her own message. This allows each member of the deliberating group to highlight information complementarities that in a formal aggregation scheme the central designer would have to anticipate.

To elaborate this claim, consider again the examples of adjudicating a case or deciding on the maximum offer price in a corporate acquisition. To rule in favor of the plaintiff, the court has to convince itself that all elements of the claim have been proven. Suppose that in a panel of three judges two find for the plaintiff and one against her. If the judges disagree on evidence regarding the same element the

---

72 The 'groupthink' literature seems not to have coalesced much since the coinage of the term in 1972, see the reviews of James D Rose, 'Diverse Perspectives on the Groupthink Theory – a Literary Review' (2005) 4 Emerging Leadership Journeys 37, 37 ('Groupthink […] ironically is controversial in itself'); Robert S Baron, 'So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making' (2005) 37 Advances in Experimental Social Psychology 219. For an economic model of groupthink driven by anticipatory utility from future prospects (e.g., disutility from bad news about future losses) Roland Bénabou, 'Groupthink: Collective Delusions in Organizations and Markets' (2013) 80 Review of Economic Studies 429. A recent piece of evidence for the cost of dissent in courts is Felipe de Mendonça Lopes, 'Dissent Aversion and Sequential Voting in the Brazilian Supreme Court' (2019) 16 Journal of Empirical Legal Studies 933.

73 For a non-technical exposition Sunstein and Hastie (n 9) 21–99; for a survey of herding in financial markets David Hirshleifer and Siew Hing Teoh, 'Thought and Behavior Contagion in Capital Markets', in Thorsten Hens and Klaus Reiner Schenk-Hoppé (eds), *Handbook of Financial Markets: Dynamics and Evolution* (Elsevier, 2009) 1.

majority may well be justified to disregard the opposing vote. If, however, the dissenter has identified a flaw in the plaintiff's argument regarding another element that the other judges have overlooked, learning about the dissenter's reason can be vital to avoid a misjudgment. Simple outcome voting would miss this critical difference. In a similar vein, the reservation price of a corporate acquirer is a function of many variables (say, the expected cash-flows of different business divisions of the target, potential cost savings from synergies, legal or reputational risks, the effect of the transaction on the acquirer's cost of capital, etc.). Even if evaluators use the same valuation formula, they will have private information—based on their expertise, research, or random knowledge—on some variables but not on others. Their judgment will reflect updated, posterior estimates for the former and uninformed priors for the latter. Again, information aggregation improves dramatically when evaluators can communicate their information sources and assumptions, instead of averaging only final valuations.

The flexibility of deliberation demonstrates how rigid a structure formal aggregation requires: Firstly, it mechanically limits the effect that any single individual can have on the outcome. Even the clever algorithms for giving greater weight to private information in averaging miss much of what could be distilled from an examination of stated reasons.[74] Similarly in prediction markets, no matter how important a trader's insights, she can invest no more than her capital and what her risk aversion and the market rules permit. In consequence, a group guided by formal aggregation would fatally underweight a single individual's information that the building is on fire. Secondly, formal aggregation is bound by the prespecified task and the messages made available to individuals. To some extent, the scheme can accommodate information complementarities across evaluators, namely by collecting individual judgments on component variables instead of final outcomes. But this would imply that the sponsor of the survey or prediction market dictates the model structure for making the final judgment—a stark departure from the ideal of tapping the intelligence of the many.

---

[74] See n 33 and accompanying text.

IV. Conclusion

The discussion of freewheeling communication highlights not only the potential 'madness' of crowds[75] but also the comparative weakness of formal aggregation: the amount of structure imposed on the judgment problem and its possibly poor adaptation. In consequence, it would hardly ever be advisable to establish a statutory or regulatory rule prescribing, as a default or mandate, a specific formal aggregation scheme for a class of collective choices. The law does, of course, provide formal rules for making decisions, namely by consent or some variant of majority voting. The rules, however, leave blank the proposal that is to be consented or voted upon. The formal mechanism relies on input from free deliberation that precedes it. The complementarity is mutual: The prospect of a binding decision—including the continuation of the status quo—stimulates the exchange of views during deliberation, as each participant with an interest in the outcome will want to feed her information into the process.

Complementarity between deliberation and formal aggregation can also arise in connection with prediction markets. Because these markets are competitive zero-sum games between the traders, they seem at first not to encourage an exchange of information other than through price. However, once a trader has placed her bet, she should be willing to divulge her information and indeed even try to persuade others of her valuation so as to feed it into the market price and make a profit. A remaining impediment to deliberation arises from the fear that other traders could manipulate the market price by misrepresenting their information. On the flip side, the market creates incentives to uncover new evidence that, after it has been used to take a position, can benefit deliberation.

Overall, it seems a fair assessment that the analysis of collective intelligence so far has failed to deliver a break-through technology to revolutionize decision making. The promise of the Condorcet jury theorem dwindles as the underlying assumptions are scrutinized.[76] Prediction markets could hold additional unrealized potential but

---

75 Mackay (n 5).

76 Condorcet, an aspiring social reformer, cannot be blamed for inspiring excessive optimism, see n 21.

their requirements are demanding and prevent them from becoming a universal cure or improvement for the pitfalls of collective choice. The sober conclusion should not be considered a disappointment. The research on collective intelligence provides valuable insights. Its formalizations and experiments should be conceived less as a practical guidebook to reform than as a disciplined way of studying the relevant tradeoffs in shaping collective choice. In this capacity, they contribute to the analysis of legal institutions and eventually to improving them.