INTELLIGENCE · ARTIFICIAL AND COLLECTIVE INTELLIGENCE

Collective Intelligence and the Wisdom of Crowds

Groups can know more than any of their members. This article explains the research on collective intelligence and the wisdom of crowds, why averaging independent judgements works, what makes some groups smarter than others, and how organisations can draw on the knowledge of many people.

ModQu research article · 1,681 words · 18 references (APA 7th edition)

In 1906, the scientist Francis Galton visited a livestock fair in Plymouth where visitors paid to guess the weight of an ox once it had been slaughtered and dressed. Galton collected about 800 tickets afterwards and examined them. The middle value of all the guesses was within 1% of the true weight, closer than most individual entries, including those of butchers and farmers (Galton, 1907). The observation has become the classic illustration of what is now called the **wisdom of crowds**: under the right conditions, the combined judgement of many people can be more accurate than the judgement of almost any individual among them.

Collective intelligence is the broader idea: the capacity of groups, sometimes with the help of technology, to act more intelligently than their members could alone. Malone and Bernstein (2015) define it as groups of individuals acting collectively in ways that seem intelligent. It includes simple averaging of estimates, structured group discussion, markets, open collaboration and combinations of people and machines. For intelligence and research work, it offers a way to draw on knowledge spread across many minds.

Why averaging works

The statistical logic is simple. Each person's estimate contains some truth and some error. If the errors are independent, some people guess too high and others too low, and the errors tend to cancel out when the estimates are averaged. What remains is closer to the truth than a typical individual estimate.

Larrick and Soll (2006) showed that the average of two judgements is guaranteed to be at least as accurate as the typical individual judgement, and is often far better, whenever the two people's errors fall on opposite sides of the truth. Yet in their experiments, people consistently underrated averaging and preferred to pick the judgement they thought was best. Clemen (1989), reviewing decades of forecasting research, reached a similar conclusion: combining forecasts from different methods or people reliably improves accuracy, and simple averages often work as well as more complicated weighting schemes.

Figure 1 shows the effect, and its main limit, in a simulation.

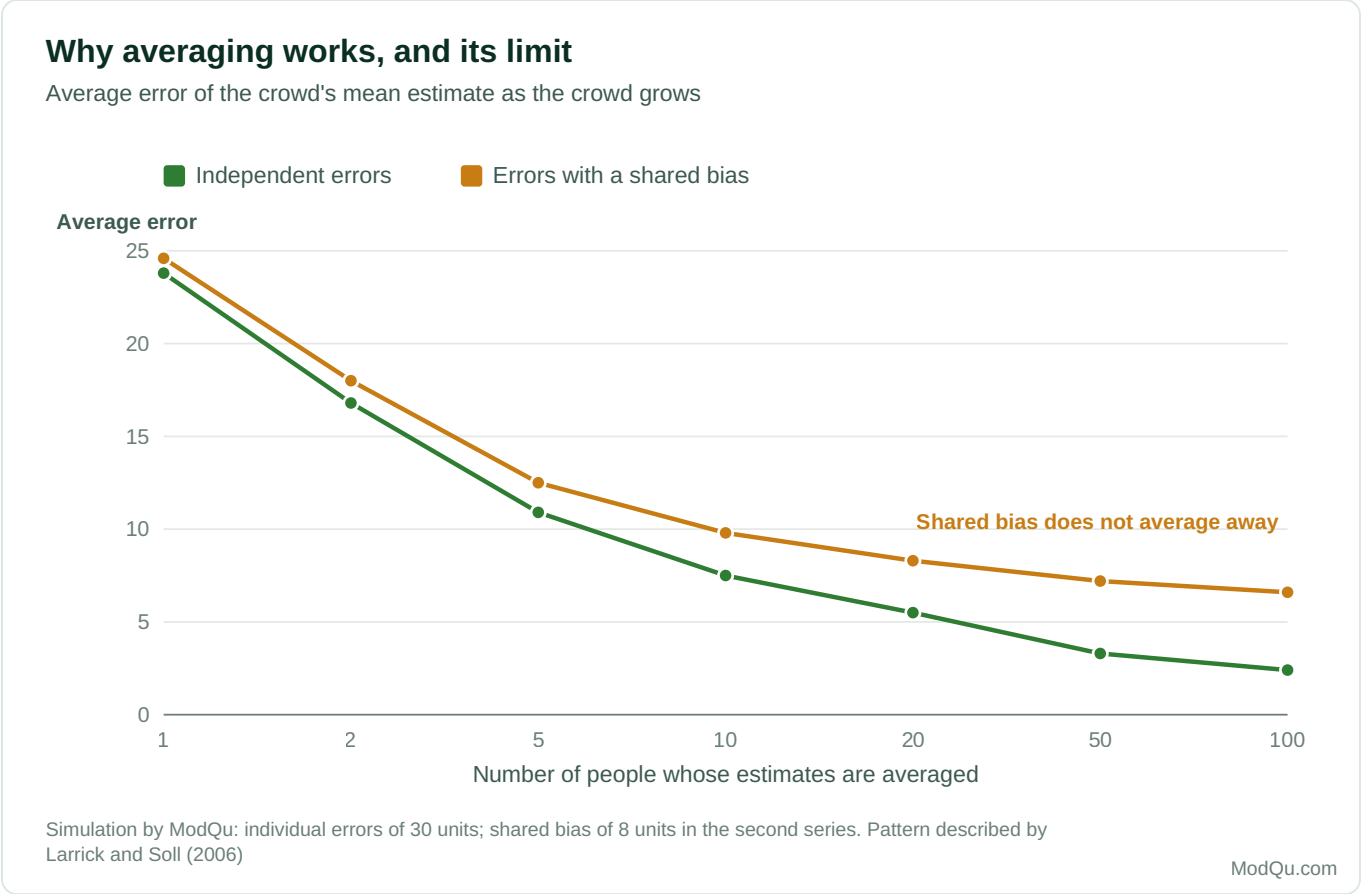


Figure 1. Why averaging works, and its limit. When errors are independent, the average becomes steadily more accurate as more people are added. When people share a common bias, adding more people helps much less. Simulation by ModQu; pattern described by Larrick and Soll (2006).

The limit matters as much as the effect. Averaging removes independent error but not shared error. If everyone has been misled by the same news report or shares the same blind spot, the average inherits that mistake no matter how many people are asked.

An older idea

The mathematical case for group judgement is older than Galton. In 1785, the Marquis de Condorcet proved what is now called the **jury theorem**: if each member of a group is more likely than not to be right on a yes-or-no question, and members judge independently, then the probability that the majority is right rises towards certainty as the group grows (Condorcet, 1785). The same conditions, competence and independence, recur throughout later research.

Hayek (1945) made a related argument about economies. Hayek wrote that knowledge of local circumstances is dispersed among many people and can never be gathered in one place. Prices work because they combine that dispersed knowledge into a single signal. Collective intelligence methods in organisations try to do something similar for questions that markets do not price.

The conditions for a wise crowd

Surowiecki (2004) popularised the wisdom of crowds and set out four conditions under which crowds tend to judge well. Figure 2 summarises them.

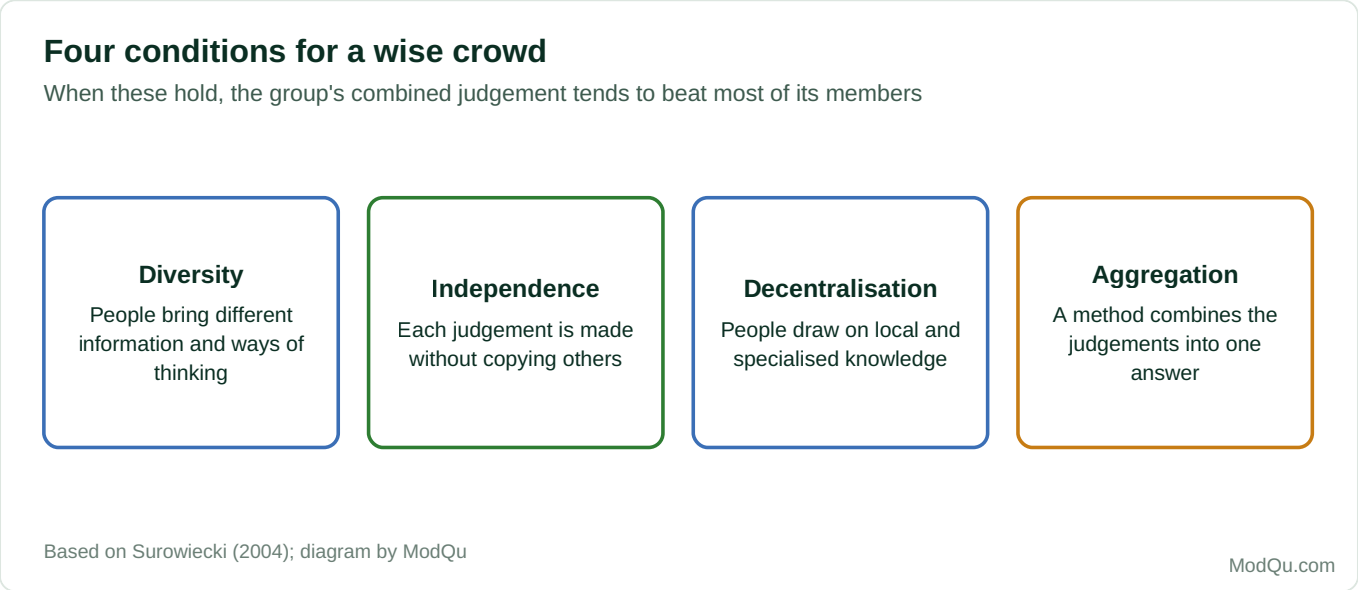


Figure 2. Four conditions for a wise crowd. Based on Surowiecki (2004); diagram by ModQu.

Diversity matters because people with different information and ways of thinking make different errors, which cancel. Hong and Page (2004) showed in a formal model that a group of diverse problem-solvers can outperform a group of the individually best problem-solvers, because the best performers tend to think alike. Page (2007) developed the argument into a broader account of why cognitive diversity improves group prediction and problem-solving.

Independence matters because it keeps errors from becoming correlated. Lorenz et al. (2011) showed experimentally that when people saw others' estimates and revised their own, the range of opinion narrowed and confidence rose, but accuracy did not improve. Social influence made the crowd less wise while making it feel wiser.

Decentralisation lets people use specialised or local knowledge. **Aggregation** provides a method, such as an average, a vote or a market price, that turns many judgements into one answer.

Smarter groups

Groups that work together, rather than simply having their answers averaged, can also be more or less intelligent. Woolley et al. (2010) gave small groups a range of tasks and found that a group's performance on one task predicted its performance on others, suggesting a general **collective intelligence factor**. Surprisingly, this factor was not strongly linked to the average or maximum intelligence of members. It was linked to how evenly members took turns in conversation, to members' skill at reading others' emotions, and to the proportion of women in the group, which itself was associated with social sensitivity. How a group interacts can matter more than who is in it.

Malone (2018) argues that the most powerful forms of collective intelligence now combine people and computers, forming what Malone calls "superminds". Online platforms let thousands of people contribute ideas, judgements or work. Howe (2006) coined the term **crowdsourcing** for the practice of opening tasks traditionally done by employees to a large, undefined group of people online.

Collective intelligence in forecasting

The strongest evidence for collective intelligence in intelligence work comes from forecasting tournaments. In a large research programme sponsored by the American intelligence community, teams competed to forecast geopolitical events. Mellers et al. (2014) found that the winning approach combined selection of skilled forecasters, training in probabilistic reasoning, teamwork, and statistical aggregation that gave more weight to better forecasters and adjusted the combined forecast to be more decisive. Atanasov et al. (2017) compared prediction markets with aggregated individual forecasts and found that well-designed statistical aggregation performed at least as well as markets.

Researchers have also developed cleverer ways to extract crowd knowledge. Prelec, Seung and McCoy (2017) proposed the "surprisingly popular" method: ask people both for their own answer and for their prediction of how others will answer. The answer that is more popular than people predicted is often correct, even when the majority is wrong, because it identifies the minority who know something the majority does not.

The crowd within

Even a single person can benefit from averaging. Vul and Pashler (2008) found that when people made a second guess at the same question weeks later, the average of their two guesses was more accurate than either guess alone. Herzog and Hertwig (2009) showed that asking people to deliberately assume their first estimate was wrong, and to think about why, produced a second estimate whose errors were more independent, making the average still better. The technique, which they call dialectical bootstrapping, is a cheap tool for any analyst making an estimate alone.

A business illustration

A food manufacturer must decide how many units of a new seasonal product to make. In past years, the head of sales set the figure, and the company was often left with large unsold stocks or empty shelves. This year, the intelligence team asks twenty-five people to estimate first-month sales independently: sales representatives in each region, two buyers from retail customers, marketing staff, production planners and a sample of call centre staff who hear customer comments. Each person also states a low and a high figure they would be surprised to see exceeded.

The team takes the median estimate, which is less affected by a few extreme guesses than the mean, and compares it with the head of sales' figure. The two differ by almost a third. In a short meeting, the people with the highest and lowest estimates explain their reasoning. A regional representative reports that a large retailer has cut its shelf space for seasonal lines, information the head of sales had not heard. After the discussion, everyone submits a second estimate independently, and the new median becomes the production plan. When actual sales come in, the crowd's figure is within 8% of the result; the original figure would have been 30% too high. The team records every estimate so that, over several seasons, it can see whose judgements prove most reliable.

Common mistakes

- **Discussing first and estimating second.** Open discussion before independent estimates lets the first or most senior speaker anchor the group.
- **Asking only people who think alike.** A crowd of specialists from one department shares the same information and the same blind spots.
- **Picking the "best" individual instead of averaging.** People are poor at identifying who is right in advance; the average usually beats the chosen expert.
- **Ignoring dissent.** A minority view that differs strongly from the average may carry information the majority lacks. Ask why before discarding it.
- **Never checking results.** Without a record of estimates and outcomes, there is no way to learn which methods and which people work.

Applying collective intelligence

For organisations, the practical lessons are clear:

1. **Ask many people, separately.** Collect judgements before any discussion, so that early speakers do not anchor the rest.
2. **Seek diversity.** Include people from different functions, regions and backgrounds, not only the most senior experts.
3. **Aggregate with a method.** Use averages, medians or structured scoring rather than letting the loudest voice decide.
4. **Track accuracy.** Record judgements and outcomes, and give more weight over time to those who prove reliable.
5. **Combine with discussion carefully.** After independent estimates are recorded, discussion can share information, followed by a second round of independent estimates.

Conclusion

Collective intelligence turns the knowledge scattered across many people into a single, often more accurate judgement. Its power depends on diversity, independence and sound methods of aggregation, and it weakens when people copy each other or share the same blind spots. Organisations that design their processes to protect these conditions can draw on a resource far larger than any individual expert.

References

Atanasov, P., Rescober, P., Stone, E., Swift, S. A., Servan-Schreiber, E., Tetlock, P., Ungar, L., & Mellers, B. (2017). Distilling the wisdom of crowds: Prediction markets vs. prediction polls. *Management Science, 63*(3), 691–706. https://doi.org/10.1287/mnsc.2015.2374

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Condorcet, M. de. (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale.

Galton, F. (1907). Vox populi. *Nature, 75*(1949), 450–451. https://doi.org/10.1038/075450a0

Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review, 35*(4), 519–530.

Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science, 20*(2), 231–237. https://doi.org/10.1111/j.1467-9280.2009.02271.x

Hong, L., & Page, S. E. (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. https://doi.org/10.1073/pnas.0403723101

Howe, J. (2006). The rise of crowdsourcing. *Wired, 14*(6).

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science, 52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

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. https://doi.org/10.1073/pnas.1008636108

Malone, T. W., & Bernstein, M. S. (Eds.). (2015). *Handbook of collective intelligence*. MIT Press.

Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together*. Little, Brown.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science, 25*(5), 1106–1115. https://doi.org/10.1177/0956797614524255

Page, S. E. (2007). *The difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton University Press.

Prelec, D., Seung, H. S., & McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature, 541*(7638), 532–535. https://doi.org/10.1038/nature21054

Surowiecki, J. (2004). *The wisdom of crowds*. Doubleday.

Vul, E., & Pashler, H. (2008). Measuring the crowd within: Probabilistic representations within individuals. *Psychological Science, 19*(7), 645–647. https://doi.org/10.1111/j.1467-9280.2008.02136.x

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science, 330*(6004), 686–688. https://doi.org/10.1126/science.1193147