How Wise Is the Crowd: Can We Infer People Are Accurate and Competent Merely Because They Agree With Each Other?

Jan Pfänder (janlukas.pfaender@gmail.com)
Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, France

Benoît De Courson (benoit2c@gmail.com)
Max Planck Institute for the Study of Crime, Security and Law, Freiburg im Breisgau, Germany

Hugo Mercier (hugo.mercier@gmail.com)
Institut Jean Nicod, Département d'études cognitives, ENS, EHESS, PSL University, CNRS, France

Abstract

Are people who agree on something more likely to be right and competent? Evidence suggests that people tend to make this inference. However, standard wisdom of crowds approaches only provide limited normative grounds. Using simulations, we argue that when individuals make independent and unbiased estimates, under a wide range of parameters, individuals whose answers converge with each other tend to have more accurate answers and to be more competent. In 2 experiments (UK participants, total N = 399), we show that participants infer that informants who agree have more accurate answers and are more competent, even when they have no priors, and that these inferences are weakened when the informants are systematically biased. In conclusion, we speculate that inferences from convergence to accuracy and competence might help explain why people deem scientists competent, even if they have little understanding of science.

Keywords: consensus; convergence; majority rules; wisdom of crowds; trust; trust in science

Introduction

Imagine that you live in ancient Greece, and a fellow called Eratostenes claims the circumference of the earth is 252000 stades (approximately 40000 kilometers). You know nothing about this man, the circumference of the Earth, or how one could measure such a thing. As a result, you discard Eratostenes' opinion and (mis)take him for a pretentious loon. But what if other scholars had arrived at very similar measurements, independently of Eratosthenes? Or even if they had carefully checked his measurement, with a critical eye? Wouldn't that give you enough ground to believe not only that the estimates might be correct, but also that Eratosthenes and his fellow scholars must be quite bright, to be able to achieve such a feat as measuring the Earth?

In this article, we explore how, under some circumstances, we should, and we do infer that a group of individuals whose answers converge are likely to be correct, and to be competent in the relevant area, even if we had no prior belief about either what the correct answer was, or about these individuals' competence.

The literature on the wisdom of crowds has treated separately situations with continuous answers, such as the weight of an ox in Galton's famous observation (Galton, 1907), and with categorical answers, as when voters have to choose between two options, in the standard Condorcet Jury Theorem (De Condorcet, 2014). The continuous and the categorical case are typically modeled with different tools, and they have usually been studied in different empirical literatures (see below). Given that they both represent common ways for answers to converge more or less (e.g. when people give numerical estimates vs. vote on one of a limited number of options), we treat them both here, with different simulations and experiments.

In the continuous case, the most relevant evidence comes from the literature on 'advice-taking'. In these experiments, participants are called 'judges', and they need to make numerical estimates–sometimes on factual knowledge, e.g. 'What year was the Suez Canal opened first?' (Yaniv, 2004), sometimes on knowledge controlled by the experimenters, e.g. 'How many animals were on the screen you saw briefly?' (Molleman et al., 2020). To help answer these questions, participants are given estimates from others, the advisors.

One subset of these studies manipulate the degree of convergence between groups of advisors, through the variance of estimates (Molleman et al., 2020; Yaniv, Choshen-Hillel, & Milyavsky, 2009), or their range (Budescu & Rantilla, 2000; Budescu, Rantilla, Yu, & Karelitz, 2003; Budescu & Yu, 2007). These studies find that participants are more confident about, or rely more on, estimates from groups of advisors who converge more.

In categorical choice contexts, there is ample and long-standing (Crutchfield, 1955) evidence from social psychology that participants are more likely to be influenced by majority opinions, and that this influence is stronger when the majority is larger, both in absolute and in relative terms (Morgan, Rendell, Ehn, Hoppitt, & Laland, 2012; Asch, 1956). This is true even if normative conformity (when people follow the majority because of social pressure rather than a belief that the majority is correct) is unlikely to play an important role, e.g. because the answers are private (Asch, 1956). Similar results have been obtained with young children (Fusaro & Harris, 2008; Corriveau, Fusaro, & Harris, 2009; Bernard, Harris, Terrier, & Clément, 2015; Bernard, Proust, & Clément, 2015; Chen, Corriveau, & Harris, 2013; Herrmann, Legare, Harris, & Whitehouse, 2013; Morgan, Laland, & Harris, 2015).

If many studies have demonstrated that participants tend to infer that more convergent answers are more likely to be correct, few have examined whether participants thought that this convergence was indicative of the individuals' competence. One potential exception is a study with preschoolers (Corriveau et al., 2009) in which the children were more

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

likely to believe the opinion of an informant who had previously been the member of a majority over that of an informant who had dissented from the majority. However, it is not clear whether the children thought the members of the majority were particularly competent, since their task–naming an object–was one in which children should already expect a high degree of competence from (adult) informants. This result might thus indicate simply that children infer that someone who disagrees with several others on how to call something is likely wrong, and thus likely less competent at least in that domain.

Is Inferring Competence From Convergence Justified?

To provide a normative answer, we conducted simulations, both for a continuous and a categorical scenario.

Continuous Case

In these simulations, groups of agents, with each agent holding varying degrees of competence, provide numerical answers. We measure how accurate these answers are, and how much they converge (i.e. how low their variance is). We then observe how convergence is associated with accuracy and competence of the agents.

More specifically, agents provide an estimate on a scale from 1000 to 2000 (chosen to facilitate the experiments below). Each agent is characterized by a normal distribution of possible answers. All of the agents' distributions are centered around the correct answer, but their standard deviation varies, denoting varying degrees of competence. The agents' standard deviation varied from 1 (highest competence) to 1000 (lowest competence). Each agent's competence level is drawn from a population competence distribution, expressed by a beta distribution, which can take different shapes. We conducted simulations with a variety of beta distributions which cover a large range of possible competence populations (see Fig. 1 A).

In the simulation, a population of around 990000 agents with different competence levels is generated. An answer is drawn for each agent, based on their respective competence distribution. The accuracy of this answer is defined as the squared distance to the true answer. Having a competence and an accuracy value for each agent, we randomly assign agents to groups of, e.g., three. For each group, we calculate the average of the agents' competence and accuracy. We measure the convergence of a group by calculating the standard deviation of the agents' answers. We repeat this process for different sample sizes for the groups, and different competence distributions. Fig. 1 C displays the relation of convergence with accuracy (left), and competence (right) across different underlying competence distributions and group sizes.

Categorical Case

We simulate agents who have to decide between a number of options, one of which is correct, and whose competence varies. Competence is defined as a value p, which can differ for each agent, and which corresponds to the probability of selecting the right answer (the agents then have a probability (1-p)/(m-1), with m being the number of options, of selecting any other option). Competence values range from chance level (p = 1/m) to always selecting the correct option (p = 1). Individual competence levels are drawn from the same population competence beta distributions as in the numerical case (see Fig. 1 A). Based on their competence level, we draw an answer for each agent. We measure an agent's accuracy as a binary outcome, namely whether they selected the correct option or not. In each simulation 999000 agents are generated, and then randomly assigned to groups (of varying size in different simulations). Within these groups, we first calculate the share of individuals voting for each answer, allowing us to measure convergence. For each level of convergence occurring within a group, we then compute (i) the average accuracy and (ii) the average competence of agents. Across all groups, we then compute the averages of these within-group averages, for each level of convergence.

We repeat this procedure varying population competence distributions and the size of informant groups, holding the number of choice options constant at n = 3. Fig. 1 B shows the average accuracy (left), and the average competence (right) of informants, as a function of the share of votes for their answer within their groups, across different underlying competence distributions and group sizes.

The simulations for the numerical and categorical case demonstrate a similar pattern, which can be summarized as follows: (i) Irrespective of group size and of the competence distribution, there is a very strong relation between convergence and accuracy: more convergent answers tend to be more accurate. (ii) For any group size and any competence distribution, there is a relation between convergence and the competence of the agents: more convergent answers tend to stem from more competent agents. The strength of this relation is not much affected by the number of agents whose answers are converging, but, although it is always positive, it ranges from very weak to very strong depending on the initial competence distribution. (iii) The relation between convergence and accuracy is always much stronger than the relation between convergence and average competence of the agents.

Overview Experiments

We test whether people infer that individuals–we will call them informants henceforth–are more likely to give accurate answers, and to be competent, when their answers converge, both in a continuous tasks (Experiment 1), and in a categorical tasks (Experiment 2). By contrast with previous experiments, participants were not given any information about the tasks–how difficult they were–and the informants–how competent they might be. We also manipulated whether the convergence of the answers could be explained by something else than accuracy, which should reduce participants' reliance on the convergence of the answer as a cue to accuracy and to competence.

All experiments and analyses were pre-registered, unless

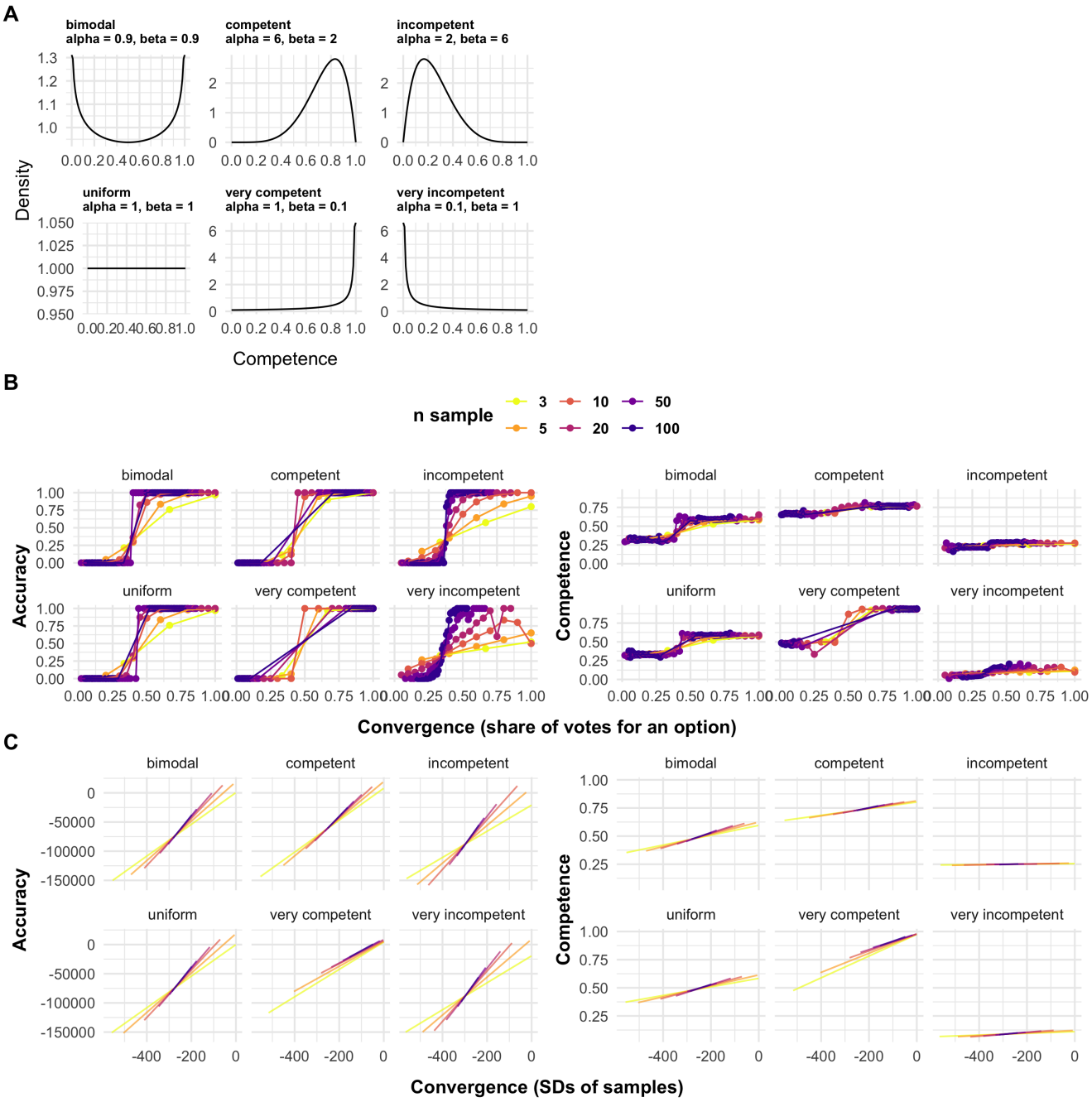

Figure 1: **A** Shows the different population competence distributions we considered in our simulations. In the continuous simulations, competence values of 0 correspond to a very large standard deviation (1000, with a mean of 1500, on a scale from 1000 to 2000), thereby practically taking the form of a uniform distribution, while competence of 1 corresponds to a standard deviation of 1. In the categorical simulations, a competence value of 0 corresponds to chance (e.g. in a 3-choice-options scenario, an individual picking the correct answer with a probability of $1/3$), while competence of 1 corresponds to definitely picking the correct answer. **B** Shows the results of simulations in a categorical setting with three choice options. Points represent average accuracy (left)/competence (right) values by degree of convergences (measured by the share of votes for an options), for different population competence distributions and sample sizes. **C** Shows the results in a continuous setting. Regression lines represent the correlation between accuracy (left; measured by squared distance to true mean and reversed such that greater accuracy corresponds to top of the figure) or competence (right), respectively, and convergence (reversed such that greater convergence corresponds to being more right on the x-axis)

explicitly said so. Pre-registration documents, data and code can be found on Open Science Framework project page[1] (`http://tinyurl.com/3229xu5y`). All analyses were conducted in R (version 4.2.2) using R Studio. For statistical models, we relied on the `lme4` package and its `lmer()` function. Unless mentioned otherwise, we report unstandardized model coefficients that can be interpreted in units of the scales we use for our dependent variables. Interaction effect coefficients mark the difference in convergence effects between independent and conflict of interest conditions.

Experiment 1

In Experiment 1, participants were told that they would be looking at (fictional) predictions of experts for stock values. They were provided with a set of numerical estimates which were more or less convergent, and asked whether they thought the estimates were accurate, and whether the experts making the estimates were competent. In a conflict of interest condition, the experts had an incentive to value the stock in a given way, while they had no such conflict of interest in the independence condition. We formulated four hypotheses:

H1a: Participants perceive predictions of independent informants as more accurate when they converge compared to when they diverge.

H1b: Participants perceive independent informants as more competent when their predictions converge compared to when they diverge.

H2a: The effect of convergence on accuracy (H1a) is more positive in a context where informants are independent compared to when they are in a conflict of interest.

H2b: The effect of convergence on competence (H1b) is more positive in a context where informants are independent compared to when they are in a conflict of interest.

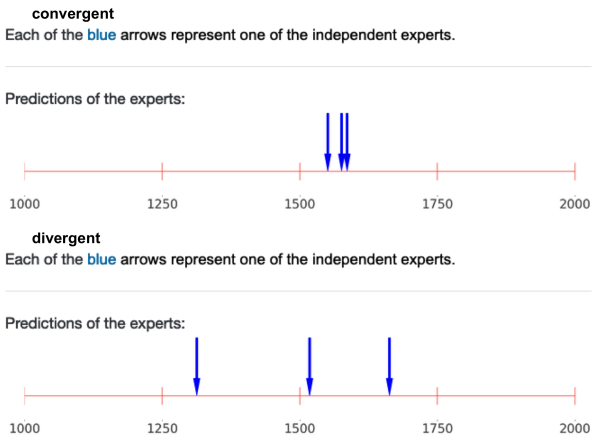

Figure 2: Example of stimuli as used in Experiment 1, independence condition.

[1]Note that on the OSF, the enumeration of experiments differs, since we report only a subset of experiments here (pattern: article at hand ˜ OSF; Exp.1 ˜ Exp.3; Exp. 2 ˜ Exp.5).

Methods

Participants We ran a power analysis by simulation. The simulation code is available on OSF, and the procedure is described in the pre-registration document. The simulation suggested that 100 participants provide a significant interaction term between 95% and 97% of the time, given an alpha threshold for significance of 0.05. Due to uncertainty about our effect size assumptions and because we had resources for a larger sample, we recruited 199 participants for this study, from the UK and via Prolific (99 female, 100 male; age_{mean}: 40.30, age_{sd}: 12.72, age_{median}: 38).

Procedure After providing their consent to participate in the study and passing an attention check, participants read the following introduction: "You will see four scenarios in which several experts predict the future value of a stock. You have no idea how competent the experts are. It's also possible that some are really good while others are really bad. As it so happens, in all scenarios, the predictions for the value of the stock have to lie between 1000 and 2000. Other than that, the scenarios are completely unrelated: it is different experts predicting the values of different stocks every time." Participants then saw four scenarios (Fig. 2), each introduced by a text according to the condition the participant was assigned to (see 'Design'). Participants were then asked to rate the experts' accuracy ("On average, how accurate do you think these three predictions are?" on a 7-point Likert scale - (1) "not accurate at all" to (7) "extremely accurate"), and their competence ("On average, how good do you think these three experts are at predicting the value of stocks?" - (1) "not good at all" to (7) "extremely good").

Design We manipulated two factors: informational dependency (two levels, independence and conflict of interest; between participants) and convergence (two levels, convergence and divergence; within participants). In the independence condition, the participants read "Experts are independent of each other, and have no conflict of interest in predicting the stock value - they do not personally profit in any way from any future valuation of the stock." In the conflict of interest condition, the participants read "All three experts have invested in the specific stock whose value they are predicting, and they benefit if other people believe that the stock will be valued at [mean of respective distribution] in the future."

Results

To account for dependencies of observations due to our within-participant design regarding convergence, we ran mixed models, with participants as random factor for the intercept and the convergence slope. We find evidence supporting all four hypotheses. For the first set of hypotheses, we reduced the sample to half of the participants, namely those who were assigned to the independence condition. As for accuracy, participants rated informants in convergent scenarios (mean = 5.28, sd = 1.05) as more accurate than in divergent ones (mean = 3.40, sd = 1.08; $\hat{b}_{Accuracy}$ = 1.88 [1.658, 2.102],

p = < .001). As for competence, participants rated informants in convergent scenarios (mean = 5.24, sd = 0.99) as more competent than in divergent ones (mean = 3.61, sd = 1.11; $\hat{b}_{Competence}$ = 1.62 [1.411, 1.839], p = < .001).

The second set of hypotheses targeted the interaction of informational dependency and convergence (Fig. 4 A). In the independence condition, the effect of convergence on accuracy was more positive ($\hat{b}_{interaction,Accuracy}$ = 0.99 [0.634, 1.348], p = < .001) than in the conflict of interest condition ($\hat{b}_{baseline}$ = 0.89 [0.636, 1.142], p = < .001). Likewise the effect of convergence on competence was more positive ($\hat{b}_{interaction,Competence}$ = 0.80 [0.474, 1.13], p = < .001) than in the conflict of interest condition ($\hat{b}_{baseline}$ = 0.82 [0.591, 1.056], p = < .001).

Experiment 2

Experiment 2 is a conceptual replication of Experiment 1 in a categorical instead of a numerical case:

H1a: Participants perceive an estimate of an independent informant as more accurate the more it converges with the estimates of other informants.

H1b: Participants perceive an independent informant as more competent the more their estimate converges with the estimates of other informants.

H2a: The effect of convergence on accuracy (H1a) is more positive in a context where informants are independent compared to when they are biased (i.e. share a conflict of interest to pick a given answer).

H2b: The effect of convergence on competence (H1b) is more positive in a context where informants are independent compared to when they are biased (i.e. share a conflict of interest to pick a given answer).

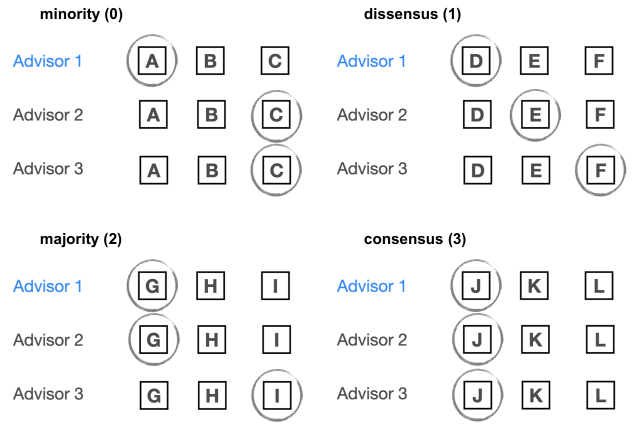

Figure 3: Example of stimuli as used in Experiment 2.

Methods

Participants We ran two different power analyses, one for each outcome variable. inform our choice of sample size. All assumptions and details on the procedure can be found on the OSF. The power simulation for accuracy suggested that for 80 participants, we would have a power of at least 90% for the interaction effect. The simulation for competence suggested that with already 40 participants, we would detect an interaction, but only with 60 participants we would also detect a main effect of convergence. Due to uncertainty about our assumptions and because resources were available for a larger sample, we recruited 200 participants, again in the UK and via Prolific (99 female, 100, 1 non-identified; age_{mean}: 41.88, age_{sd}: 13.94, age_{median}: 39).

Procedure After providing their consent to participate in the study and passing an attention check, participants read the following introduction: "We will show you three financial advisors who are giving recommendations on investment decisions. They can choose between three investment options. Their task is to recommend one. You will see several such situations. They are completely unrelated: it is different advisors evaluating different investments every time. At first you have no idea how competent the advisors are: they might be completely at chance, or be very good at the task. It's also possible that some are really good while others are really bad. Some tasks might be difficult while others are easy. Your task will be to evaluate the performance of one of the advisors based on what everyone's answers are." They were then presented with eight recommendation scenarios (Fig. 3). To assess perceptions of accuracy, we asked: "What do you think is the probability of advisor 1 making the best investment recommendation?". Participants answered with a slider on a scale from 0 to 100. To assess perceptions of competence, we asked: "How competent do you think advisor 1 is regarding such investment recommendations?" Participants answered on a 7-point Likert scale (from (1)"not competent at all" to (2)"extremely competent").

Design We manipulated convergence within participants, by varying the ratio of advisors choosing the same recommendation option as a focal advisor (i.e. the one that participants evaluate). The levels of convergence are: (i) consensus, where all three advisors pick the same option [`coded value = 3`]; (ii) majority, where either the third or second advisor picks the same option as the first advisor [`coded value = 2`]; (iii) dissensus, where all three advisors pick different options [`coded value = 1`]; (iv) minority, where the second and third advisor pick the same option, but one that is different from the first advisor's choice [`coded value = 0`]. In our analysis, we treat convergence as a continuous variable, assigning the coded values in squared parenthesis here. We manipulated conflict of interest between participants. In the conflict of interest condition, experts were introduced this way: "The three advisors have already invested in one of the three options, the same option for all three. As a result, they have an incentive to push that option in their recommendations." For the independence condition: "The three advisors are independent of each other, and have no conflict of interest in making investment recommendations."

Materials All participants saw all four conditions of convergence (Fig. 3), with two stimuli per condition, i.e. eight stimuli in total (4 convergence levels x 2 stimuli).

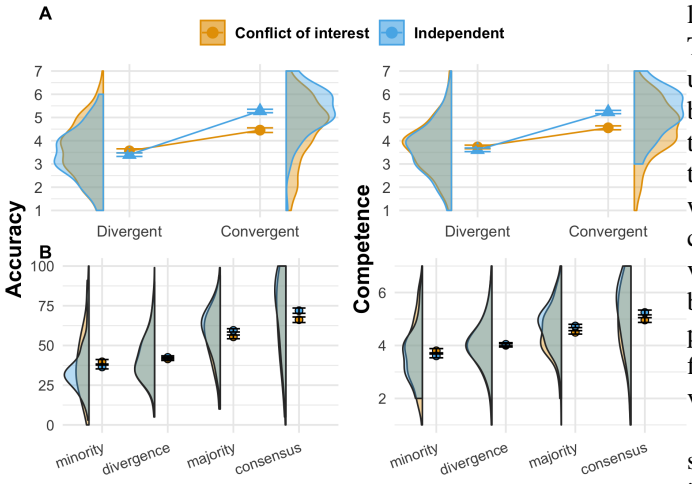

Figure 4: Interaction of convergence and informational dependency in **A** Experiment 1 (continuous) and **B** Experiment 2 (categorical).

Results

To account for dependencies of observations due to our within-participant design regarding convergence, we ran mixed models, with participants as random factor for the intercept and the convergence slope.

We find evidence supporting all four hypotheses. To test H1a and H1b, we restricted our data on the independent condition. We find a positive effect of convergence on both accuracy ($\hat{b}_{Accuracy}$ = 12.34 [10.362, 14.311], p = $<$ .001) and competence ($\hat{b}_{Competence}$ = 0.56 [0.459, 0.665], p = $<$ .001).

The second set of hypotheses targeted the interaction of informational dependency and convergence (Fig. 4 B). In the independence condition, the effect of convergence on accuracy was more positive ($\hat{b}_{interaction,Accuracy}$ = 3.01 [0.027, 5.988], p = 0.048) than in the conflict of interest condition ($\hat{b}_{baseline}$ = 9.33 [7.232, 11.426], p = $<$ .001). Likewise, the effect of convergence on competence was more positive ($\hat{b}_{interaction,Competence}$ = 0.16 [0.014, 0.316], p = 0.032) than in the conflict of interest condition ($\hat{b}_{baseline}$ = 0.40 [0.291, 0.503], p = $<$ .001).

General Discussion

If we see a group of individuals, whose competence is unknown, converge on an answer, is it rational to infer that this answer is more likely to be correct, and that the individuals are likely to be competent? In simulations, we showed that the answer is yes on both counts–assuming there is no systematic bias among the individuals answering. In two experiments, we showed that participants (UK) tend to draw these inferences, and less so when the informants were systematically biased by a shared conflict of interest.

Our results–both simulations and experiments–are a novel contribution to the wisdom of crowds literature. In this literature–in particular that relying on the Condorcet Jury Theorem–a degree of competence is assumed in the individuals providing some answers. From that competence, it can be inferred that the individuals will tend to agree, and that their answers will tend to be accurate. Here we show that the reverse inference–from agreement to competence–is also warranted, and that it is warranted under a wide range of circumstances: If one does not suspect any systematic bias, convergence alone can be sufficient in determining who tends to be an expert. This finding qualifies work suggesting that people need a certain degree of expertise themselves, in order to figure out who is an expert (Nguyen, 2020; Hahn, Merdes, & von Sydow, 2018).

The current study has a number of limitations. In our simulations, we assume agents to be independent. Following previous work generalizing the Condorcet Jury Theorem (Ladha, 1992), more robust simulations would show that–while still assuming no systematic bias–our results hold even when agents influence each others' answers. Regarding our experiments, if the very abstract materials allow us to remove most of the priors participants might have, they might also reduce the ecological validity of the experiments.

Still, we believe that people might draw the type of inference discussed here in a variety of contexts, perhaps the most prominent one being science. Science is, arguably, the institution in which individuals end up converging the most in their opinions. For instance, scientists within the relevant disciplines agree on things ranging from the distance between the solar system and the center of the galaxy to the atomic structure of DNA. This represents an incredible degree of convergence. When people hear that scientists have measured the distance between the solar system and the center of the galaxy, if they assume that there is a broad agreement within the relevant experts, this should lead them to infer that this measure is accurate, and that the scientists who made it are competent. Experiments have already shown that increasing the degree of perceived consensus among scientists tends to increase acceptance of the consensual belief (Van Stekelenburg, Schaap, Veling, Van 'T Riet, & Buijzen, 2022), but it hasn't been shown that the degree of consensus also affects the perceived competence of scientists.

In the case of science, the relationship between convergence and accuracy is broadly justified. However, at some points of history, there has been broad agreement on misbeliefs, such as when Christian theologians had calculated that the Earth was approximately six thousand years old. To the extent that people were aware of this broad agreement, and believed the theologians to have reached it independently of each other, this might have not only fostered acceptance of this estimate of the age of the Earth, but also a perception of the theologians as competent.

Acknowledgments

We thank the anonymous reviewers for their very valuable feedback. HM received funding from the ANR (SCALUP, ANR-21-CE28-0016-01), and from the John Templeton Foundation ("An Evolutionary and Cultural Perspective on Intellectual Humility via Intellectual Curiosity and Epistemic Deference"). The ANR also supported this project through FrontCog (ANR-17-EURE-0017), and PSL (ANR-10-IDEX-0001-02).

References

Asch, S. E. (1956). Studies of independence and conformity: I. a minority of one against a unanimous majority. *Psychological Monographs: General and Applied*, *70*(9), 1–70. doi: 10.1037/h0093718

Bernard, S., Harris, P., Terrier, N., & Clément, F. (2015, 08). Children weigh the number of informants and perceptual uncertainty when identifying objects. *Journal of Experimental Child Psychology*, *136*, 70–81. doi: 10.1016/j.jecp.2015.03.009

Bernard, S., Proust, J., & Clément, F. (2015). Four- to six-year-old children's sensitivity to reliability versus consensus in the endorsement of object labels. *Child Development*, *86*(4), 1112–1124. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cdev.12366) doi: 10.1111/cdev.12366

Budescu, D. V., & Rantilla, A. K. (2000). Confidence in aggregation of expert opinions. *Acta Psychologica*, *104*(3), 371–398. doi: 10.1016/S0001-6918(00)00037-8

Budescu, D. V., Rantilla, A. K., Yu, H.-T., & Karelitz, T. M. (2003). The effects of asymmetry among advisors on the aggregation of their opinions. *Organizational Behavior and Human Decision Processes*, *90*(1), 178–194. doi: 10.1016/S0749-5978(02)00516-2

Budescu, D. V., & Yu, H.-T. (2007). Aggregation of opinions based on correlated cues and advisors. *Journal of Behavioral Decision Making*, *20*(2), 153–177. doi: 10.1002/bdm.547

Chen, E. E., Corriveau, K. H., & Harris, P. L. (2013). Children trust a consensus composed of outgroup members-but do not retain that trust. *Child Development*, *84*(1), 269–282. doi: 10.1111/j.1467-8624.2012.01850.x

Corriveau, K. H., Fusaro, M., & Harris, P. L. (2009). Going with the flow: Preschoolers prefer nondissenters as informants. *Psychological Science*, *20*(3), 372–377. doi: 10.1111/j.1467-9280.2009.02291.x

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist*, *10*(5), 191–198. doi: 10.1037/h0040237

De Condorcet, N. (2014). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Cambridge University Press.

Fusaro, M., & Harris, P. L. (2008). Children assess informant reliability using bystanders' non-verbal cues. *Developmental Science*, *11*(5), 771–777. doi: 10.1111/j.1467-7687.2008.00728.x

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

Hahn, U., Merdes, C., & von Sydow, M. (2018). How good is your evidence and how would you know? *Topics in Cognitive Science*, *10*(4), 660–678. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12374) doi: 10.1111/tops.12374

Herrmann, P. A., Legare, C. H., Harris, P. L., & Whitehouse, H. (2013). Stick to the script: The effect of witnessing multiple actors on children's imitation. *Cognition*, *129*(3), 536–543. doi: 10.1016/j.cognition.2013.08.010

Ladha, K. K. (1992). The condorcet jury theorem, free speech, and correlated votes. *American Journal of Political Science*, *36*(3), 617. doi: 10.2307/2111584

Molleman, L., Tump, A. N., Gradassi, A., Herzog, S., Jayles, B., Kurvers, R. H. J. M., & van den Bos, W. (2020). Strategies for integrating disparate social information. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1939), 20202413. (Publisher: Royal Society) doi: 10.1098/rspb.2020.2413

Morgan, T. J., Laland, K. N., & Harris, P. L. (2015). The development of adaptive conformity in young children: effects of uncertainty and consensus. *Developmental Science*, *18*(4), 511–524. doi: 10.1111/desc.12231

Morgan, T. J., Rendell, L. E., Ehn, M., Hoppitt, W., & Laland, K. N. (2012). The evolutionary basis of human social learning. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1729), 653–662. doi: 10.1098/rspb.2011.1172

Nguyen, C. T. (2020). Cognitive islands and runaway echo chambers: problems for epistemic dependence on experts. *Synthese*, *197*(7), 2803–2821. doi: 10.1007/s11229-018-1692-0

Van Stekelenburg, A., Schaap, G., Veling, H., Van 'T Riet, J., & Buijzen, M. (2022). Scientific-consensus communication about contested science: A preregistered meta-analysis. *Psychological Science*, *33*(12), 1989–2008. doi: 10.1177/09567976221083219

Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*, 1–13.

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 558–563. doi: 10.1037/a0014589