Diversity promotes collective intelligence in large groups but harms small ones

NICCOLO PESCETELLI, ALEX RUTHERFORD, IYAD RAHWAN, MIT Media Lab

1. INTRODUCTION

In modern days, online interactions among millions of individual contribute to shape the public discourse on social issues and world events, and influence political agendas and decisions about our future. Understanding how large collectives navigate such complex environments is a crucial and timely research question. Intelligent collective decision-making in human and non-human animals has fascinated intellectuals for centuries and more recently popularized under the name of ”collective intelligence”. Local information aggregated across multiple individuals can produce better judgments, with examples ranging from juries to bee hives [Condorcet 1785]. Social interaction can have positive outcomes [Bahrami et al. 2010] but can also lead the group astray due to unwanted social biases (*e.g.*, herding, group polarization, etc.) that reduce the independence of individual judgments [Kerr and Tindale 2004]. The learning strategies used to explore and exploit the solution space $\mathbb{S}$ interact with task complexity, or the ”ruggedness” of $\mathbb{S}$, to produce the wisdom or madness of interactive crowds [Surowiecki 2004]. Solitary learning performs well in simple tasks (*e.g.*, characterized by single peaked $\mathbb{S}$), but becomes increasingly more costly and unreliable in more complex solution spaces. In such environments, social learning achieves better performance and increased use of social information has been observed in both human and non-human animals [Toyokawa et al. 2019; Barkoczi and Galesic 2016; Wisdom et al. 2013].

Characterizing the dynamics in which information (and thus errors) is allowed to flow between individuals can improve our understanding of the exploration and exploitation patterns observed and the resilience of groups to sub-optimal solutions or shocks in the environment $\mathbb{S}$ [Mason and Watts 2012; Becker et al. 2017; Lazer and Friedman 2007; Barkoczi and Galesic 2016; Goldstone et al. 2013; Toyokawa et al. 2019]. These studies have shown that sometimes, pruning ties among individuals makes collectives more resilient. Group size or the number of groups interrogated is here an important factor [Navajas et al. 2018; Kao and Couzin 2019]. Although greater gains in accuracy are expected as more independent judgments are pooled together, larger group sizes in interacting crowds produce greater rates of herding and thus decrease independence [Condorcet 1785; Wisdom et al. 2013; Toyokawa et al. 2019]. Small groups tend to maximize accuracy in environments characterized by inter-judgment correlations thanks to their inherent noise and greater exploratory behavior [Kao and Couzin 2014; Toyokawa et al. 2019]. It has recently been suggested that in order to avoid the downsides of group interaction while maintaining its benefits, multiple smaller groups should be interrogated, so that groups will err independently from one another [Navajas et al. 2018]. We call this feature ”modularity”. The benefits of modularity stem from an optimal balance between the benefits of social exposure and the benefits of maintaining independence across judgments. Modularity maintains information diversity (across groups) in spite of herding (within groups). Whether the same results generalize to more complex realistic problems is unknown.

Another factor that supposedly breaks the correlation between judgments is group diversity. Although diversity is a highly heterogeneous construct touching several disciplines [van Knippenberg and Schippers 2007], an extensive literature in psychology has suggested that group diversity is as-

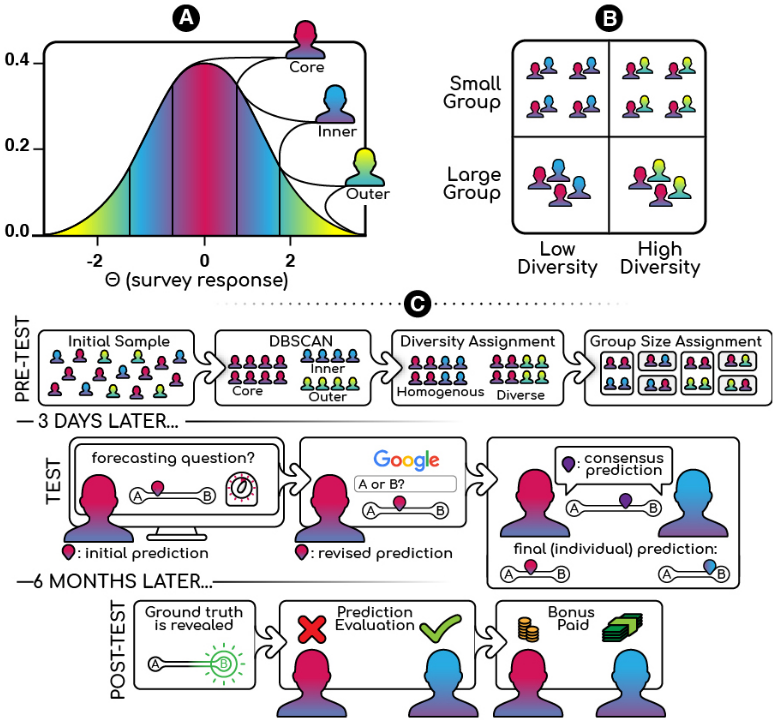

Fig. 1. Experimental design. At pre-test, participants were administered a battery of tests to sort them into conditions. Three days later, they answered real geopolitical forecasting problems whose ground truth was revealed only 6 months after the experiment.

sociated with better group performance [Hong and Page 2004; Surowiecki 2004; Kerr and Tindale 2004]. From an informational stand point, diverse groups should be more resilient to group biases, engage in complex thinking, be more creative and able to explore a greater region of the solution space [Woolley et al. 2010; Mannix and Neale 2005; Hong and Page 2004]. According to this view, the recent emphasis on the dangers of filter bubbles and online echo-chambers is justified as homophily increases the possibility of herding and *groupthink*. According to social categorization theories however, diversity—here more often defined based on demographics indicators—can create socio-cognitive barriers to interaction [Mannix and Neale 2005], which online filter bubbles avoid by design. These two views and definitions of diversity are often invoked to explain the wide range of positive and negative results found in the diversity literature [van Dijk et al. 2012; Mannix and Neale 2005; de Oliveira and Nisbett 2018]. We notice however, also in light of recent press scandals, that demographics as well as cognitive and personality features are often used to segment users on online platforms, and consequently to customize their searches, content recommendations and overall experience. One should then expect that, although some of these features (like demographics) tend not to affect information diversity *per se* [de Oliveira and Nisbett 2018], they might do so in an online environment that maps inter-individual differences along a heterogeneous and high dimensional profiling space into different information access. Thus, rather than focusing on one specific type of diversity as previous literature has done, we are interested in whether a heterogeneous definition of diversity as trait—namely the objective distance on a profiling space including demographic, cognitive, psychological and relational factors—would result in variability in information retrieved online, and thus variability in opinions or 'state' diversity. Given the difficulty of disentangling the causal contributions of diversity on performance, we here employ an experimental design, which allows us to better understand realistic online group decision-making.

In this preregistered study (available via OSF), we investigate the causal contributions of group diversity and group size (modularity) on individual and group performance, in complex problems with high external validity. We look at diversity through the lens of social learning and network science, which have been consistently applied in other fields (*e.g.*, ecology, cultural evolution, social psychology etc.) and generated a bounty of experimental evidence, new hypotheses and novel predictions [Kendal et al. 2018]. Based on this theoretical background and the evidence reviewed above, we expect group diversity to interact with variables such as task complexity and group size. Small groups ($\sim$5 people) were approximately the square root of large groups ($\sim$25 people), which recent theoretical work in animal behavior suggests being the optimal ratio in complex environments [Kao and Couzin 2019]. We use a clustering method to experimentally manipulate group diversity, here defined as average Euclidean distance of team members along demographic, cognitive, personality and relational dimensions. This allows us to randomize a subset of participants into teams of more or less similar individuals and to groups of varying sizes (Figure 1a-b). Carefully controlled experiments can sometimes neglect external validity. For this reason, we use real-world questions and online deliberation rather than simpler estimation tasks [van Dijk et al. 2012; Mannix and Neale 2005; Hong and Page 2004]. Performance was evaluated as forecasting accuracy on real geo-political forecasting problems, characterized by a complex problem space, high degrees of uncertainty and correlated information between judges. Forecasting problems—whose ground truths were unknown at the time when the forecasts were made—were randomly chosen from a larger pool of binary questions independently provided by the IARPA Hybrid Forecasting Competition (HFC), a national program aiming to improve analysts' forecasting accuracy. We find that although group diversity benefits large groups, this effect is reversed in smaller groups. This is true both when looking at individual and aggregated forecasts. Linguistic analysis of chat data shows that this pattern of results seems to be explained by greater conflicting evidence (defined here as the standard deviation of the group's opinions [Lorenz et al. 2011]) affecting small diverse groups but not large diverse groups. We also find that forcing individuals to reach a consensus as opposed to simply being exposed to social information benefits their ability to forecast future events. These findings inform how social interaction online can affect real-life problem solving in complex information environments, and can potentially inform other domain-general tasks. We discuss these results in light of the recent literature on collective behavior in ecology and social science.

REFERENCES

Bahador Bahrami, Karsten Olsen, Peter E Latham, Andreas Roepstorff, Geraint Rees, and Chris D Frith. 2010. Optimally interacting minds. *Science (New York, N.Y.)* 329, 5995 (8 2010), 1081–5. DOI:http://dx.doi.org/10.1126/science.1185718

Daniel Barkoczi and Mirta Galesic. 2016. Social learning strategies modify the effect of network structure on group performance. *Nature Communications* 7 (10 2016), 13109. DOI:http://dx.doi.org/10.1038/ncomms13109

Joshua Becker, Devon Brackbill, and Damon Centola. 2017. Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences* (6 2017), 201615978. DOI:http://dx.doi.org/10.1073/pnas.1615978114

Marquis de Condorcet. 1785. *Essai sur l'application de l'analyse à la probabilité des decisions rendues à la pluralité des vois*. de l'Imprimerie Royale, Paris.

Stephanie de Oliveira and Richard E. Nisbett. 2018. Demographically diverse crowds are typically not much wiser than homogeneous crowds. *Proceedings of the National Academy of Sciences* 115, 9 (2 2018), 2066–2071. DOI:http://dx.doi.org/10.1073/pnas.1717632115

Robert L. Goldstone, Thomas N. Wisdom, Michael E. Roberts, and Seth Frey. 2013. Learning Along With Others. *Psychology of Learning and Motivation* 58 (2013), 1–45. DOI:http://dx.doi.org/10.1016/B978-0-12-407237-4.00001-3

Lu Hong and Scott E Page. 2004. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America* 101, 46 (11 2004), 16385–9. DOI:http://dx.doi.org/10.1073/pnas.0403723101

A. B. Kao and I. D. Couzin. 2014. Decision accuracy in complex environments is often maximized by small group sizes. *Proceedings of the Royal Society B: Biological Sciences* 281, 1784 (4 2014), 20133305–20133305. DOI:http://dx.doi.org/10.1098/rspb.2013.3305

Albert B. Kao and Iain D. Couzin. 2019. Modular structure within groups causes information loss but can improve decision accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences* 374, 1774 (6 2019), 20180378. DOI:http://dx.doi.org/10.1098/rstb.2018.0378

Rachel L. Kendal, Neeltje J. Boogert, Luke Rendell, Kevin N. Laland, Mike Webster, and Patricia L. Jones. 2018. Social Learning Strategies: Bridge-Building between Fields. *Trends in Cognitive Sciences* 22, 7 (7 2018), 651–665. DOI:http://dx.doi.org/10.1016/j.tics.2018.04.003

Norbert L. Kerr and R. Scott Tindale. 2004. Group Performance and Decision Making. *Annual Review of Psychology* 55, 1 (2 2004), 623–655. DOI:http://dx.doi.org/10.1146/annurev.psych.55.090902.142009

David Lazer and Allan Friedman. 2007. The Network Structure of Exploration and Exploitation. *Administrative Science Quarterly* 52, 4 (12 2007), 667–694. DOI:http://dx.doi.org/10.2189/asqu.52.4.667

Jan Lorenz, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America* 108, 22 (5 2011), 9020–5. DOI:http://dx.doi.org/10.1073/pnas.1008636108

Elizabeth Mannix and Margaret A. Neale. 2005. What Differences Make a Difference? *Psychological Science in the Public Interest* 6, 2 (10 2005), 31–55. DOI:http://dx.doi.org/10.1111/j.1529-1006.2005.00022.x

Winter Mason and Duncan J Watts. 2012. Collaborative learning in networks. *Proceedings of the National Academy of Sciences of the United States of America* 109, 3 (1 2012), 764–9. DOI:http://dx.doi.org/10.1073/pnas.1110069108

Joaquin Navajas, Tamara Niella, Gerry Garbulsky, Bahador Bahrami, and Mariano Sigman. 2018. Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. *Nature Human Behaviour* 2, 2 (2 2018), 126–132. DOI:http://dx.doi.org/10.1038/s41562-017-0273-4

James Surowiecki. 2004. *The Wisdom of Crowds. Why the Many are Smarter than the Few.* (abacus ed.). Little, Brown Book Group, London. 271 pages.

Wataru Toyokawa, Andrew Whalen, and Kevin N. Laland. 2019. Social learning strategies regulate the wisdom and madness of interactive crowds. *Nature Human Behaviour* 3, 2 (2 2019), 183–193. DOI:http://dx.doi.org/10.1038/s41562-018-0518-x

Hans van Dijk, Marloes L. van Engen, and Daan van Knippenberg. 2012. Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. *Organizational Behavior and Human Decision Processes* 119, 1 (9 2012), 38–53. DOI:http://dx.doi.org/10.1016/j.obhdp.2012.06.003

Daan van Knippenberg and Michaéla C. Schippers. 2007. Work Group Diversity. *Annual Review of Psychology* 58, 1 (1 2007), 515–541. DOI:http://dx.doi.org/10.1146/annurev.psych.58.110405.085546

Thomas N. Wisdom, Xianfeng Song, and Robert L. Goldstone. 2013. Social Learning Strategies in Networked Groups. *Cognitive science* 37 (2013), 1383–1425. DOI:http://dx.doi.org/10.1111/cogs.12052

Anita Williams Woolley, Christopher F Chabris, Alex Pentland, Nada Hashmi, and Thomas W Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science (New York, N.Y.)* 330, 6004 (10 2010), 686–8. DOI:http://dx.doi.org/10.1126/science.1193147