An Agent-Based Model of Collective Decision-making in Correlated Environments

Louisa Jane Di Felice[1] and Payam Zahadat[1]

[1]IT University of Copenhagen, DK, 2300
louisajane.df@gmail.com

Abstract

In many complex systems, from robot and insect swarms to human social systems, agents take decisions collectively, using information retrieved from the environment and from each other. This information is usually correlated to some extent – e.g., voters reading the same media outlets, animals receiving the same cues from their environment, or people listening to the same opinion leaders. Taking inspiration from human social systems, we consider the case of collective decision-making between two choices, one being the correct one. We break down the problem of collective decision-making in correlated environments into two components: (i) how likely different configurations of information environments are to show the correct option and (ii) how likely different configurations of collectives are to detect the majority shown by the environment. An agent-based model is presented, where agents scan an information environment, composed of correlated and uncorrelated sources, and form individual opinions based on the information perceived. Once individual opinions are formed through a majority function, agents take a majority vote to determine their chosen option. Preliminary results show how different population parameters lead to different decision accuracy in similar information environments, and how the two steps of opinion formation and collective vote can skew the collective's perception of the environment positively or negatively. Future work will expand these results by allowing agents to form local groups before taking decisions collectively.

Introduction

Across different domains of sociality, agents in collectives interact to take decisions together, be it eusocial insects finding new nest sites, or people voting on high-stake referendums. From an information perspective, collective decision-making can be seen as a process of collective computation, where there is a slow phase of information accumulation followed by a faster phase of information aggregation (Flack (2017); Daniels et al. (2017)). In the first step of information accumulation, agents can perceive information cues from spatial environments or from semantic ones. In the former, information sources are organized by spatial proximity; in the latter, by semantic similarity. We refer to such semantic environments as "information environments". When agents in collectives form individual opinions by scanning these environments, information cues tend to present some degree of correlation because of their semantic closeness (e.g., people in an organization reading the same news sources, or members of a collective listening to charismatic opinion leaders). Due to this correlation, the phenomenon of the "wisdom of the crowds", where larger collectives converge to more accurate decisions, does not always hold – this has been studied and proven across different domains (see references in Galesic et al. (2018)). Considering an environment composed of several (correlated and uncorrelated) information cues, collective decision-making problems can be seen as problems of (i) how likely an information environment is to show the correct cue as the majority and (ii) how "good" a collective is at detecting the environment's majority. We present an agent-based model (ABM) that allows exploring the relationship between different degrees of correlation in an information environment, different population parameters (the size of the population, and the information capacity of individual agents), and decision accuracy. A group of agents scans an information environment composed of information sources that can show one of two options. Each agent forms an opinion on which information source to vote for, based on the one they have seen the most through their exploration. Then, agents vote collectively with their opinions, through a majority vote. We refer to *information abstraction* as any process where information at a given level is lost or compressed onto a higher level of analysis, through an abstraction function that could be sampling, averaging, taking the maximum, etc. In our case, opinion formation is the first abstraction, since each agent compresses the information sources they have perceived into a single opinion (through an internal majority vote); and the vote is the second abstraction, since agents' opinions are compressed onto a final vote (through a collective majority vote). Following Ladha (1995), we refer to the majority option shown by a given environment as the Full Information Majority Rule (*FIMR*). In addition, we consider the environment's Full Information Majority Strength (*FIMS*) – i.e., how strong the environment's majority is. Preliminary results show how different degrees of correlation and accuracy of correlated and uncorrelated information sources lead to different types of information environments, described by *FIMR* and *FIMS*. Population parameters, then, determine how collectives distort information from those environments, either positively or negatively, through the two abstraction steps of opinion formation and voting.

Related Work

Kao and Couzin (2014) show how smaller groups can outperform larger ones when taking collective decisions in correlated environments. In their model, animals in a collective take a decision based on a low-correlation information cue, characterized by a reliability R_L between 0.5 and 1; a high-correlation information cue, characterized by a reliability R_H, also between 0.5 and 1; and p, the probability of choosing the low-correlation cue (with $1 - p$ being the probability of choosing the high-correlation one). Varying the population size N, they find that smaller groups can outperform larger ones when correlation is high, since the noise in their information aggregation allows them to explore higher decision accuracy areas that are not available to larger groups. We take this set-up as our starting point, translating it into an ABM where information sources are distributed across a semantic environment. Differently from Kao and Couzin (2014), in our model agents do not see all the information sources around them, but partially scan their environment. In doing so, we add the variables of information capacity (how much information each agent can detect) and an algorithm of exploration (how the agents move to scan their environment). Translating notions of correlation to media consumption, Pescetelli et al. (2020) show how Kao and Couzin (2014)'s ideas can be applied to human social systems, making a statistical argument for the inclusion of independent (uncorrelated) news sources in the information landscape used to take decisions where the collective payoff needs to be maximized. In different fields, existing ABMs have explored the theme of collective decision-making. For a review grounded in ecology, see DeAngelis and Diaz (2019). In human social systems specifically, social choice theory addresses the fundamental question of how people combine their individual preferences into a collective decision (Laughlin, 2011). For a recent overview and conceptual framework for studying collective adaptation, including references to models of human social systems, see Galesic et al. (2022). The framework through which we look at these problems in our ABM is that of information abstraction in multi-scale systems. A multi-scale system is a system where there is at least one layer of macro-scale information abstracted from a micro-scale, through an abstraction function leading to less syntactic information at the macro-scale (Diaconescu et al., 2019). In our case, both individual opinions and the final vote are abstractions from the information environment.

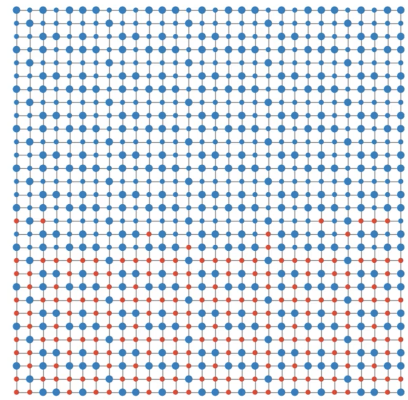


Figure 1: An example information environment. Larger dots are global information sources; smaller ones are local information sources; blue dots show the correct option (1); red dots show the incorrect option (0). The global information in this example is presenting the correct option.

Model Description

Information Environment

The model was built using the *AgentPy* Python package (Foramitti, 2021). In its current iteration it remains abstract, loosely simulating people forming opinions on binary choices (e.g., a referendum) from online news consumption. The information environment (Fig. 1) has a fixed number of 900 information sources, distributed across a 30x30 grid. We refer to uncorrelated information sources as local sources, and to correlated information sources as global ones. Local sources are distributed in a polarized way, with the correct sources (those showing option 1, blue in Fig. 1) in the top part of the environment, and the incorrect sources (those showing option 0, red in Fig. 1) in the bottom part of the environment. The transition area between the two parts is composed of three rows (10% of the information space). This polarized landscape reflects information environments used by people who take decisions through online news consumption, since media consumption tends to be highly polarized (Garimella et al. (2021)). When agents are placed on the grid, they can detect nine information sources at once (the one they are on, plus the eight neighbors around it). Global sources (the bigger dots in Fig. 1) are distributed uniformly throughout the grid, so that agents always perceive the same number of global sources in their direct environment.

The environment is described by three parameters:

1. The reliability of local information cues, R_L, between 0.5 and 1, meaning that we assume that local information sources have a higher probability of being correct than incorrect. At each run, R_L determines how many of the local information sources show the correct option (in Fig. 1, how many of the smaller dots are blue vs. red).

2. The reliability of global information cues, R_G, between 0.5 and 1. Global sources all show the same option within each run (in Fig. 1, the larger dots, in this run showing the correct option). R_G determines in how many runs across each sample the global sources show the correct option.

3. The ratio of global sources to total sources, G – since the direct environment perceived by each agent is of 9 sources, this ratio can range from 1/9 to 8/9.

Initially, agents are distributed randomly on the grid. Thus, R_L reflects the probability that an agent will see local sources showing the correct option.

Population parameters

The population is characterized by two parameters: (i) a population size N; (ii) an information capacity C_I. Once N agents are placed randomly on the grid, they explore the space until their C_I are full. The exploration algorithm is a simple random walk, meaning that larger C_I lead to a larger exploration of the space, and to a higher probability of crossing the border between local sources showing different options. This reflects the fact that people who consume more information to take a decision have a higher probability of reading information sources with opposing views. The random walk, on the other hand, reflects the fact that agents tend to move within semantic spaces with similar information sources. When collecting information through their random walk, agents have a 1-step memory for redundancy. When moving one step in one of the possible eight directions, they only scan the new sources they see, without re-scanning the sources they have already stored. At subsequent steps, agents can re-scan a source from which they had already collected information.

Abstraction steps

Agents scan their environment, form an opinion based on what they have seen (through an internal majority vote), and then vote collectively, each with their opinion (through a second majority vote). We refer to these two steps (opinion formation and final vote) as two abstraction steps, since each time multiple information sources are being abstracted through a majority function onto a single opinion or vote.

Experiments

We set the following discrete parameter ranges:

- N: 3, 5, 7, 11, 21, 31, 41, 51, 61, 75, 101
- C_I: 9, 18, 27, 36, 45, 54, 63, 72, 81, 90, 99
- R_L: 0.5, 0.6, 0.7, 0.8
- R_G: 0.5, 0.6, 0.7, 0.8
- G: 1/9, 2/9, 3/9, 4/9, 5/9

Each parameter combination is run 400 times and results are averaged out per sample. In terms of stochasticity, within each run R_L gives the probability that a local source will show option 1. Across runs, R_G gives the probability that global sources will (all) show option 1 or 0.

Decision quality in correlated environments

Decision quality as a function of *FIMR* & *FIMS*

Each information environment (given by a combination of R_L, R_G and G) is described by two variables: the Full Information Majority Rule (*FIMR*) and Full Information Majority Strength (*FIMS*). At each run of a given sample, *FIMR* is given by the source shown as the majority option (either 1 or 0). *FIMS* is given by the total number of correct sources divided by the total number of sources (a number between 0 and 1). For each sample, *FIMR* and *FIMS* are averaged over the 400 runs, leading to a number between 0 and 1. These two variables reflect how likely a given information environment is to show the correct option as the majority option (*FIMR*), and how strong that majority is (*FIMS*). For each sample, decision accuracy is calculated as the average of the final vote across runs (either 0 or 1, corresponding to the final vote for the incorrect or correct option), leading to a value between 0.5 and 1. Fig. 2 shows how decision accuracy varies as a function of these two environmental variables, for all N and C_I. As expected, *FIMS* is a better predictor of decision accuracy than *FIMR*, since it describes in more detail what the information environment looks like. Population parameters, then, can help explain the noise in the curves – i.e., why a similar *FIMS* can lead to a range of decision accuracy, as shown by the confidence intervals in Fig. 2.

The role of population size and information capacity

Fig. 3 shows how decision accuracy varies as a function of N and C_I, for different environmental parameters R_L, R_G and G. In Fig. 3a, $R_L = 0.7$, $R_G = 0.5$, $G = 2/9$. Under these conditions, decision accuracy increases as a function of N, converging to 1 for $N > 21$, while different C_I do not affect the outcome of the collective vote. As the global ratio increases to 4/9 (Fig. 3b), decision accuracy decreases overall, except for $C_I = 9$, where agents are placed randomly on the grid without exploring (they form an opinion based on the nine sources they see upon landing on the grid). Once the agents start exploring their environment, higher N and C_I lead to more accurate votes. In the second row (Fig. 3c, 3d), the situation is inverted: $R_L = 0.5$ and $R_G = 0.7$. As expected, decision accuracy improves as G increases from 2/9 to 4/9. However, this does not hold for C_I = 9.

These results show how, in low-correlation environments where the accuracy of local information sources is high (Fig. 3a), higher population sizes outperform lower ones, irrespective of how much agents explore the information en-

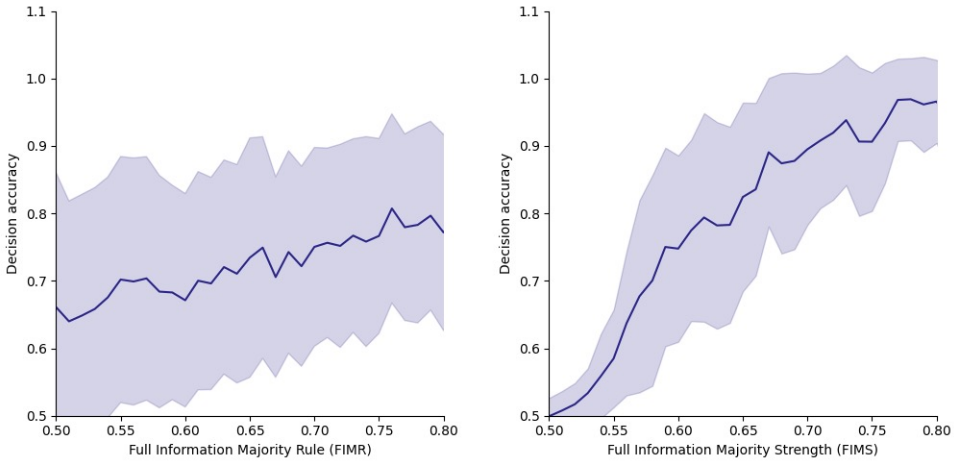
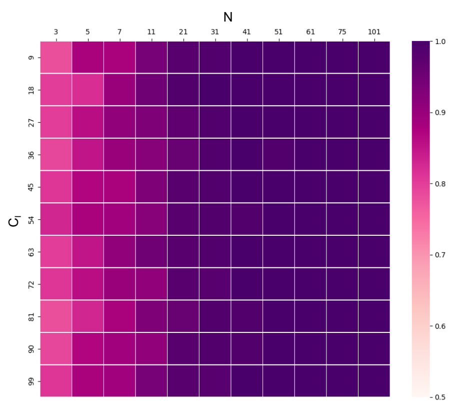
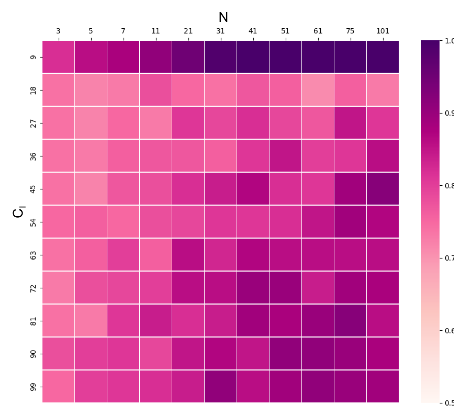


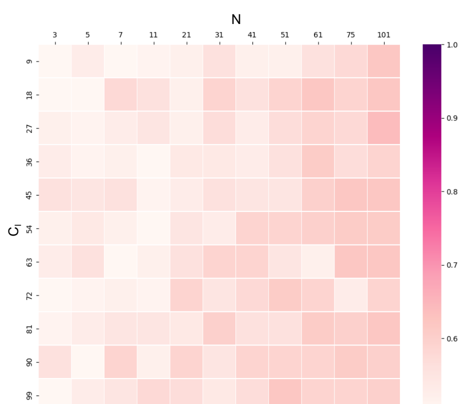
Figure 2: Decision accuracy as a function of *FIMR* and *FIMS*



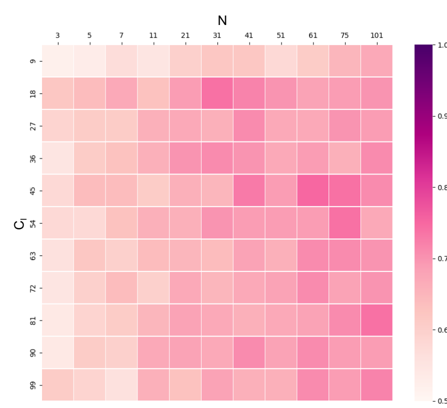
(a) R_L = 0.7, R_G = 0.5, G = 2/9



(b) R_L = 0.7, R_G = 0.5, G = 4/9

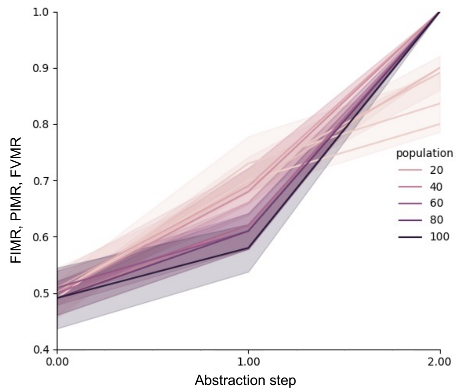


(c) R_L = 0.5, R_G = 0.7, G = 2/9

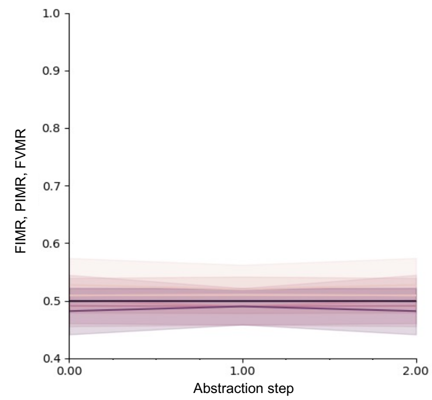


(d) R_L = 0.5, R_G = 0.7, G = 4/9

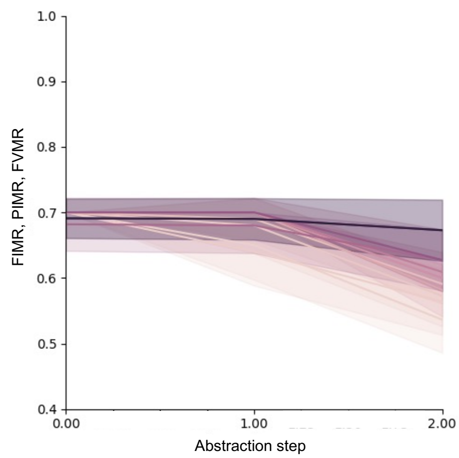
Figure 3: Decision accuracy as a function of population size N and information capacity



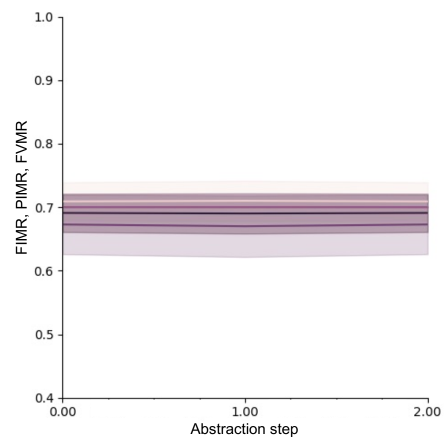
(a) R_L = 0.7, R_G = 0.5, G = 3/9



(b) R_L = 0.7, R_G = 0.5, G = 5/9



(c) R_L = 0.5, R_G = 0.7, G = 3/9



(d) R_L = 0.5, R_G = 0.7, G = 5/9

Figure 4: *FIMR* (abstraction step 0); *PIMR* (abstraction step 1) and *FVMR* (abstraction step 2)

vironment. If $R_G < R_L$ and the degree of correlation in the environment is increased (Fig. 3b), the best performance is given by agents who do not explore their environment, taking a vote with limited information. As agents start exploring, however, higher populations and information capacities lead to more accurate decisions. When $R_G > R_L$, and correlation is high (Fig. 3c), the situation is inverted and the lowest information capacity (with no exploration) leads to worse performance.

The role of information abstraction

While *FIMR* and *FIMS* characterize the information environment (they are calculated regardless of the agents), similar variables can be calculated to reflect the perception that agents have of their information environment before each abstraction step. When agents have explored their environment, before forming individual opinions, we refer to Partial Information Majority Rule (*PIMR*) as the majority option given by the pooled information collected by all agents, and to Partial Information Majority Strength (*PIMS*) as the associated strength (how many of the pooled information sources show the correct option, divided by the total number of pooled sources). They reflect the likelihood that the collective perceived environment shows the correct option as the majority (and the strength of that likelihood). They also show what the collective vote would be if agents pooled information sources before taking a decision, rather than forming an individual opinion and then voting based on that opinion. In this sense, they allow determining whether the abstraction of opinion formation is useful or not. The second abstraction step occurs when agents take a majority vote out of their aggregated opinion. In this case, the majority rule is the final vote (Final Vote Majority Rule – *FVMR*). This rule is also associated with a given strength (how many agents voted for the correct option), the Final Vote Majority Strength (*FVMS*). Looking at the evolution of the majority rules and strengths from full information, to partial information, to the final vote shows whether the abstraction steps distort the information environment in a positive way (increasing decision accuracy), in a negative way (decreasing it), or whether the abstraction steps have no effect. Fig. 4 gives an example of this, showing the evolution of *FIMR* to

PIMR and *FVMR* for different types of environments. The two top figures (Fig. 4a and 4b) show environments where $R_G < R_L$ ($R_L = 0.7$ and $R_G = 0.5$). In this case, when $G = 3/9$, the abstraction is positive, increasing the likelihood that agents will vote correctly. When more than half of the information sources are global (Fig. 4b), abstraction is irrelevant, as decision accuracy is dominated by those global sources. In the opposite case, where $R_G > R_L$ (Fig. 4c and 4d), abstraction has a negative effect, especially for lower population sizes. This suggest that abstraction plays a positive role when uncorrelated sources dominate and are more reliable than correlated ones.

Discussion & Conclusions

Collective decision-making in correlated environments can be seen as a problem of how likely information environments are to show the correct option, and whether information abstractions by collectives distort the initial environment positively or negatively. Exploring this through an ABM, our preliminary results showed how decision accuracy improves with *FIMS*, as expected. However, there is noise in this relationship, that can be explained by the population parameters: how many agents there are, how they explore the semantic space, how much information they collect, and how they form opinions and vote based on that information. Our preliminary results showed that, when agents are placed randomly on the information environment, with no exploration, lower information capacities can have positive benefits when global information cues have low reliability, and negative ones when they have high reliability. Once agents start exploring the space, different patterns emerge. It is important to understand how much these patterns are a consequence of the way agents explore the space and store information. In particular, adjusting the level of redundancy (to what extent agents re-scan information they have already seen) will be crucial to further understand the dynamics at play. Moreover, agents with a higher information capacity have a higher chance of crossing the polarized divide between local sources showing different options. It will be interesting to explore how the ratio of agents who hold space for opposing views is correlated with decision accuracy. Looking at the system at subsequent abstraction steps (of opinion formation and voting) allows seeing whether the majority vote distorts the information environment positively or negatively for the collective. Here, preliminary results showed how abstraction is positive when the global information cues have low reliability, with different behaviours for different population sizes. More work is needed to understand these patterns and relate them to the existing literature. In its current iteration, the model remains highly abstract – to test for these theoretical results in real-world scenarios, controlled experiments such as the ones performed by Pescetelli et al. (2020) are needed. However, this level of abstraction also holds benefits, as it allows exploring particular relationships between chosen variables in ways that are applicable to a variety of complex systems. Future work will also expand the existing model by allowing agents to form local groups, determined by semantic proximity and a communication capacity, before voting. This will add a third abstraction step before a collective decision is taken, and will be useful in understanding whether there are information environments where some abstraction is positive, but excessive abstraction reverts those positive outcomes.

Acknowledgements

LJDF is grateful to Joël Foramitti for his patience and support with the *AgentPy* package.

References

Daniels, B. C., Flack, J. C., and Krakauer, D. C. (2017). Dual coding theory explains biphasic collective computation in neural decision-making. *Frontiers in Neuroscience*, 11:313.

DeAngelis, D. L. and Diaz, S. G. (2019). Decision-making in agent-based modeling: A current review and future prospectus. *Frontiers in Ecology and Evolution*, page 237.

Diaconescu, A., Di Felice, L. J., and Mellodge, P. (2019). Multi-scale feedbacks for large-scale coordination in self-systems. In *2019 IEEE 13th International Conference on Self-Adaptive and Self-Organizing Systems (SASO)*, pages 137–142. IEEE.

Flack, J. C. (2017). Coarse-graining as a downward causation mechanism. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 375(2109):20160338.

Foramitti, J. (2021). Agentpy: A package for agent-based modeling in python. *Journal of Open Source Software*, 6(62):3065.

Galesic, M., Barkoczi, D., Berdahl, A., Biro, D., Carbone, G., Giannoccaro, I., Goldstone, R., Gonzalez, C., Kandler, A., Kao, A., et al. (2022). Beyond collective intelligence: Collective adaptation.

Galesic, M., Barkoczi, D., and Katsikopoulos, K. (2018). Smaller crowds outperform larger crowds and individuals in realistic task conditions. *Decision*, 5(1):1.

Garimella, K., Smith, T., Weiss, R., and West, R. (2021). Political polarization in online news consumption. *ICWSM*, pages 152–162.

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

Ladha, K. K. (1995). Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. *Journal of Economic Behavior & Organization*, 26(3):353–372.

Laughlin, P. R. (2011). Social choice theory, social decision scheme theory, and group decision-making. *Group Processes & Intergroup Relations*, 14(1):63–79.

Pescetelli, N., Rutherford, A., Kao, A., and Rahwan, I. (2020). Collective learning in news consumption.