How Diversity Helps Jury Decisions

Patrick Grim,[1] Daniel J. Singer,[2] Aaron Bramson,[3]
Bennett Holman,[4] Jiin Jung,[5] and William J. Berger [6]

[1] Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan, United States of America and Department of Philosophy, Stony Brook University, Stony Brook, New York, United States of America
[2] Department of Philosophy, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America
[3] AI Strategy Center, GA Technologies, Tokyo, Japan
[4] Underwood International College, Yonsei University, Seoul, South Korea
[5] Department of Psychology, Lehigh University, Bethlehem, Pennsylvania, United States of America
[6] Department of Philosophy, University at Buffalo, Buffalo, New York, United States of America

Forthcoming as a chapter in Coen, C. A., & Larson, J. R. Jr. (Eds.), Agent-Based Modeling for Research on Groups, Networks, and Organizations (American Psychological Association).

Many factors contribute to whether juries reach correct verdicts. Here we focus on the role of diversity. Direct empirical studies of jury deliberation are severely limited for conceptual, practical, and ethical reasons. Using an agent-based model to avoid these difficulties, we argue that diversity, in the form of subgroups within a jury, can play at least four importantly different roles in affecting jury verdicts. We show that when different subgroups have access to different information, equal representation can strengthen epistemic jury success, and if one subgroup has access to particularly strong evidence, epistemic success may demand participation by that group. Diversity among subgroups can also reduce the redundancy of the information on which a jury focuses, which can have a positive impact. Finally, and most surprisingly, we show that limiting communication between diverse groups in juries can favor epistemic success as well.

Traditionally, juries are conceived of as finders of fact, and the question of what makes a good jury is a question of social epistemology. On the standard characterization, it is the jury's job to find a criminal defendant guilty if and only if they really are guilty. On a more sophisticated characterization, the jury's job is to find the defendant guilty if and only if the prosecution has successfully shown that the accused is guilty. On either characterization, the

jury's job is to make a judgment about what is true—an essentially epistemic task. Grim and Singer, et al. (2024) recently constructed a simple model to track several ways in which diversity (or the lack of it) can contribute to or hinder the epistemic success of juries. We discuss the model from that paper here.

Computational simulation of jury decision-making has a venerable history, with the work of Penrod and collaborators as a seminal beginning (Penrod & Hastie, 1980; Hastie, Penrod, & Penington, 1983; Tanford & Penrod, 1983; see also Stasser & Davis, 1981). These early models used a basic 'strength-in-numbers' mechanism for opinion updating, in which individuals' opinions are swayed by the number of other jurors in each 'camp,' although individuals could differ in 'persuadability.' Garold Stasser's (1988) model for small group decision-making is the clearest predecessor of our work here. Stasser's model used items of information with weights and valences, a parameter for advocacy in 'sharing' items with the group, probabilistic 'recall' of that information from memory, and different levels of consensus as decision rules. A major focus of the Stasser model was the failure of groups to find 'hidden profiles'—superior decision alternatives that are revealed only when members pool information initially held by just a few. In discussing these hidden profiles, Stasser hypothesizes that "in the presence of a hidden profile, a minority may act as a catalyst facilitating the discovery of the superior decision alternative," suggesting that diverse groups might perform better than more homogenous ones (1987, p. 416). Here, we focus on that idea by exploring models of jury information sharing where certain information may be accessible (or more accessible) to only certain members of a group, bringing the issue of epistemic diversity to discussions of the

dynamics and outcomes of jury decision-making (see also Hong and Page 2004; Thompson 2014; Page 2015, 2018; Kuehn 2017; Singer 2019; Grim et al. 2019).

Recent models and mathematical results support an important role for cognitive diversity in group epistemic success, largely because diversity either increases the amount of unique information and cognitive skills brought to the group or because it reduces information or skill redundancy (Hong & Page, 2004; Weisberg & Muldoon, 2009; Zollman, 2010a). However, the precise conditions under which diversity allows groups to outperform high-ability individuals remain contested (Thompson, 2014; Kuehn, 2017; Singer, 2019). To test these dynamics specifically within the parameters of a courtroom setting, we develop and apply an agent-based model of collective information aggregation to the study of diversity in jury decision-making.

Our modeling has a limited focus within a far broader literature across a range of disciplines. There are many factors besides diversity that affect jury verdict accuracy and that have previously been considered by scholars in the law, in empirical studies, and in some cases with formal modeling, including jury size (Saks and Marti 1997), standards of legal proof (Koch and Devine 1999, Laudan 2008), and deliberation styles (Devine et al. 2007). There is a particularly interesting divergence on whether unanimity or a mere majority decision should be required of a jury. A range of empirical results indicating that unanimity has an epistemic advantage (Devine 2012) conflict with previous analytic and computational results that recommend majority voting with a threshold as low as >50% (List, 2004; Angere, Olsson, & Genot, 2016). Below we use three decision rules (>50%, ≥80%, and unanimity) to test the role of diversity in jury decision-making.

Method

There are two kinds of entities in our model: jurors (the agents) and reasons, which support or oppose verdicts. The reasons exist independently of the jurors, each can be held by a juror (or not), and the same reason can be held by multiple jurors. Deliberation consists of jurors sharing reasons, and deliberation ends when the jurors all agree. Diversity is modelled by specifying the number of jurors in each of two different subgroups (**A** and **B**). The model as described here should be thought of as a "base version." In the Results section, we show the output for this base version, and we vary how diversity is operationalized in order to draw out different ways diversity can impact group decision making.

Agents and their Properties

In our model, jurors have reasons and exchange those reasons with other jurors. Here we discuss jurors and reasons in turn and then describe how jurors get their initial reasons.

Jurors

We assume 12 jurors per run, mimicking the most common US jury size, though the model allows users to modify this. Jurors have two properties. The first is their memory, which holds their collection of reasons. We assume that agents can 'remember' an unlimited number of reasons. Second, agents have a fixed subgroup membership, which specifies them as either in subgroup **A** or subgroup **B**. We use these subgroups to model diversity, and we describe in detail further below how subgroup membership affects agents.

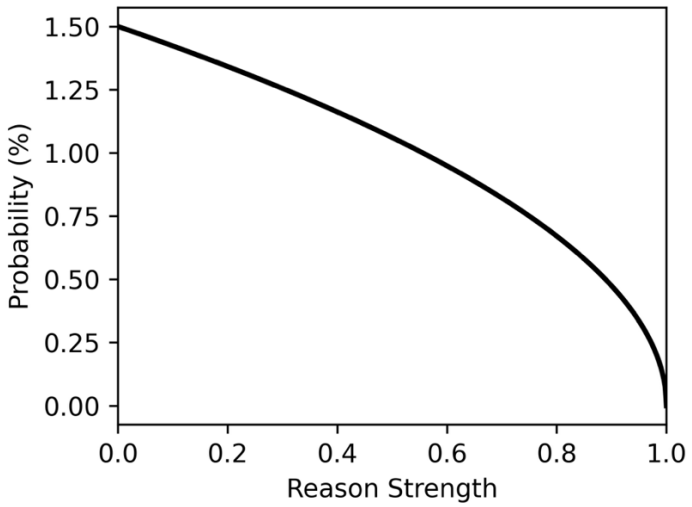
Reasons

Each reason has a valence—positive or negative, indicating whether it is a reason to think the defendant is guilty (positive) or not guilty (negative)—and a real-valued strength between 0

and 1.[1] At the start of each run, 100 reasons are created and remain fixed throughout. Following Singer et al. (2019, 2021), we think of the reasons as all the facts or pieces of evidence relevant to the verdict in any given case, and we regard the 'truth' in a run of the model to be guilty if the reasons with positive valence have a higher total strength than the reasons with negative valence (and not guilty otherwise).

Formally, we create each reason by drawing a strength value at random from a beta distribution with $\alpha = 1$ and $\beta = 1.5$ (Figure 1), then assigning it a positive or negative valence uniformly at random. With $\alpha = 1$ and $\beta = 1.5$, the beta probability density distribution has a mean of 0.40, so relatively weak reasons are more common than strong ones, and the resulting values (after a valence is assigned) are constrained on the [-1, 1] interval, following Singer, et al. (2019). Exploratory tests show that the qualitative results are invariant to changes in α and β that roughly preserve the prevalence of weak versus strong reasons.

Figure 1. *Beta distribution with* $\alpha = 1$ *and* $\beta = 1.5$



[1] Although we talk as though a reason with a positive valence is a reason to see the defendant as guilty and a negative one as a reason against guilt, it's also possible to adopt a proof-centric conception by viewing reasons as reasons to think that *a guilty verdict has been proven by the prosecution or not*. The model is also completely symmetric with regard to guilt and innocence, so the valences can be swapped without affecting the results.

How Jurors Get Initial Reasons

At the beginning of each run, each agent receives 15 reasons. A juror's starting reasons represent their background knowledge and what they find salient from the trial. To investigate the role of diversity in jury verdicts, we start by thinking of different subgroups of the population as having differential access or sensitivities to bodies of evidence on different sides of the question of guilt. The idea is that members of different social groups, because of their background information or experience, may be more sensitive to different bodies of evidence. So, rather than simply assigning jurors reasons at random, in the base version of the model we assign jurors reasons as a function of their membership in either subgroup **A** or subgroup **B**, in what we think of as a model of 'sided access.'

Like the jurors, we divide the reasons into two types, designated with letters **A** and **B**. We start with reasons with positive valence designated as **A** reasons and those with negative valence designated as **B** reasons. We then create a pool that contains both **A** and **B** reasons ("the **AB** reasons" or "the reasons in common"). We form the overlap pool of **AB** reasons by drawing a particular percentage of reasons (controlled by the **probability-overlap** slider on the model interface) equally from the **A** and **B** sets. So now we have **A** reasons (all positive), **B** reasons (all negative), and **AB** reasons (a mix of positive and negative reasons).

After we divide up the reasons in this way, jurors of each subgroup (**A** and **B**) are assigned 15 reasons chosen uniformly at random from reasons that either match their subgroup or are in the reasons in common. That is, each **A** juror starts with 15 distinct reasons randomly drawn from the **A** ∪ **AB** reasons; similarly for **B** jurors. The model thus includes two subgroups of jurors who, in expectation, have partially distinct and partially overlapping access to the reasons, with **A** jurors more likely to have positive reasons (because of their exclusive access to

the **A** reasons), and **B** jurors more likely to have negative reasons (because of their exclusive access to the **B** reasons). Note that the same reason may be held by multiple jurors, and some reasons may be held by no one. In different subsections below, we modify this assignment method; the above is the base version.

Once jurors have reasons, we think of them as having a belief or opinion about the case. At any given moment, the opinion or belief of a juror is modelled as the sum of the reasons they hold weighted by their valences. So, for example, if a juror has three reasons of strength 0.2 and one reason of strength 0.05 to believe the defendant is guilty (positive valence), and one reason of strength 0.35 for not guilty (negative valence), the juror will, all-things-considered, believe the defendant is guilty and will do so with strength 0.3 (i.e., 0.2 + 0.2 + 0.2 + 0.05 – 0.35). On the model's interface, jurors are displayed as circles, with shades of blue representing belief in guilt, shades of red representing belief in innocence, and darker colors indicating stronger beliefs.

Agent Interaction

The dynamics of our model are very simple: The model proceeds in time steps, and in each step, a random juror is chosen to share a random reason they have. The shared reason might be a reason they started with or a reason they have subsequently heard. In the simplest versions of the model, anytime a juror shares a reason, all of the other jurors hear it and add it to their collection of reasons. In the last Results subsection, we consider a modification of the model, where jurors of different subgroups only hear each other probabilistically. The base model is simple in that each time step consists of a random juror sharing a random reason. If a juror already has that reason, they do nothing, but if it's new to them, they add it to their memory.[2]

---

[2] In other work, we consider more sophisticated ways jurors might share information and manage a limited memory (e.g., Singer et al. 2019, 2021). Here, however, to focus on the effects of jury diversity, we put aside those complications.

Agent Decision Rules

At the end of each time step, we assess the status of the jury by examining the valence of the beliefs held by each juror. If all 12 jurors have a belief with a positive valence, the jury unanimously agrees that the defendant is guilty; if only 7 do, then the jury is split with >50% believing the defendant is guilty but not the full group. Because jurors continually accumulate reasons, hung juries are vanishingly rare.

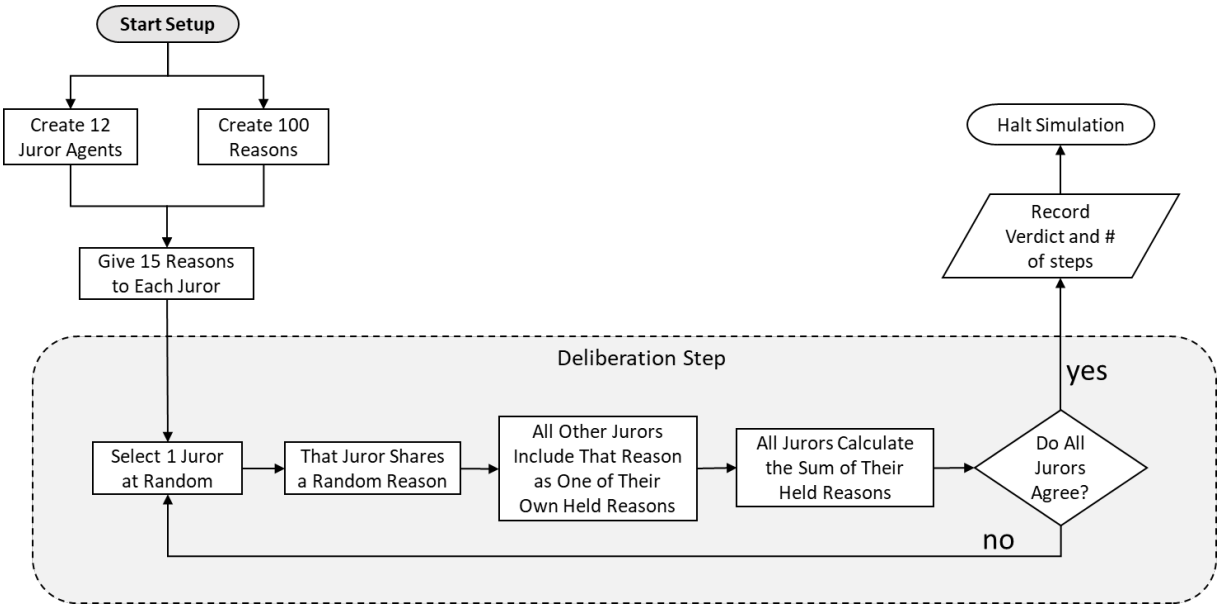
As mentioned above, we think of reasons as pieces of information that are relevant to the verdict, and we model the 'truth' as what is supported by all of the reasons, that is, what one would believe if one had access to all the facts relevant to the verdict. Using that notion of truth, we can assess how the jury performs. Perfect jury success happens when the jury unanimously agrees that the defendant is guilty when the defendant is in fact guilty and unanimously agrees that the defendant is not guilty when the defendant is in fact not guilty. We also measure performance at less-than-full unanimity using simple (>50%) and super (≥80%) majorities, though these variations do not substantively affect our main claims about diversity's impact on accuracy.

Importantly, this framework allows that a jury may agree on an incorrect result through no fault of its own. This can happen because truth is defined by all the reasons, not just those the jury has access to. We think this accurately reflects real jury trials, where relevant information may never reach the jury. That said, none of our results depend on this modeling choice. To verify this, the model has a variable, *initial-best-evidence*, the valence of which tracks the valence of the sum of the reasons actually held by jurors. More advanced readers will find that they can recreate the results using that variable instead of *truth*.

Process Scheduling

Figure 2 diagrams the model's operation. It begins by creating jurors and reasons, and then giving jurors their initial reasons. It then cycles through repeated rounds of deliberation until all agree. We also check verdict proportions after each step and record accuracy for simple (>50%) and super (≥80%) majorities the first time they occur, in addition to unanimity.

Figure 2. *Workflow diagram of the model, including setup, each step of the model, and the halting condition for ending a run.*



Note: The verdict and the number of steps is also recorded the first time >50% and ≥80% of the jurors agree, but the model continues running until all jurors agree. We repeat the process from start to consensus 10,000 times.

Output Variables

Fixing the percentage of 'reasons in common' (by setting the **probability-overlap** slider on the model interface to 50%, for example) and a decision rule (simple majority, super majority, or unanimity), we track diversity's effect by varying the **A**/**B** jury composition (0/12, 1/11, 2/10,

etc.) and noting epistemic success—the proportion of correct verdicts. Below we report this measure as the proportion of correct verdicts across 10,000 runs.

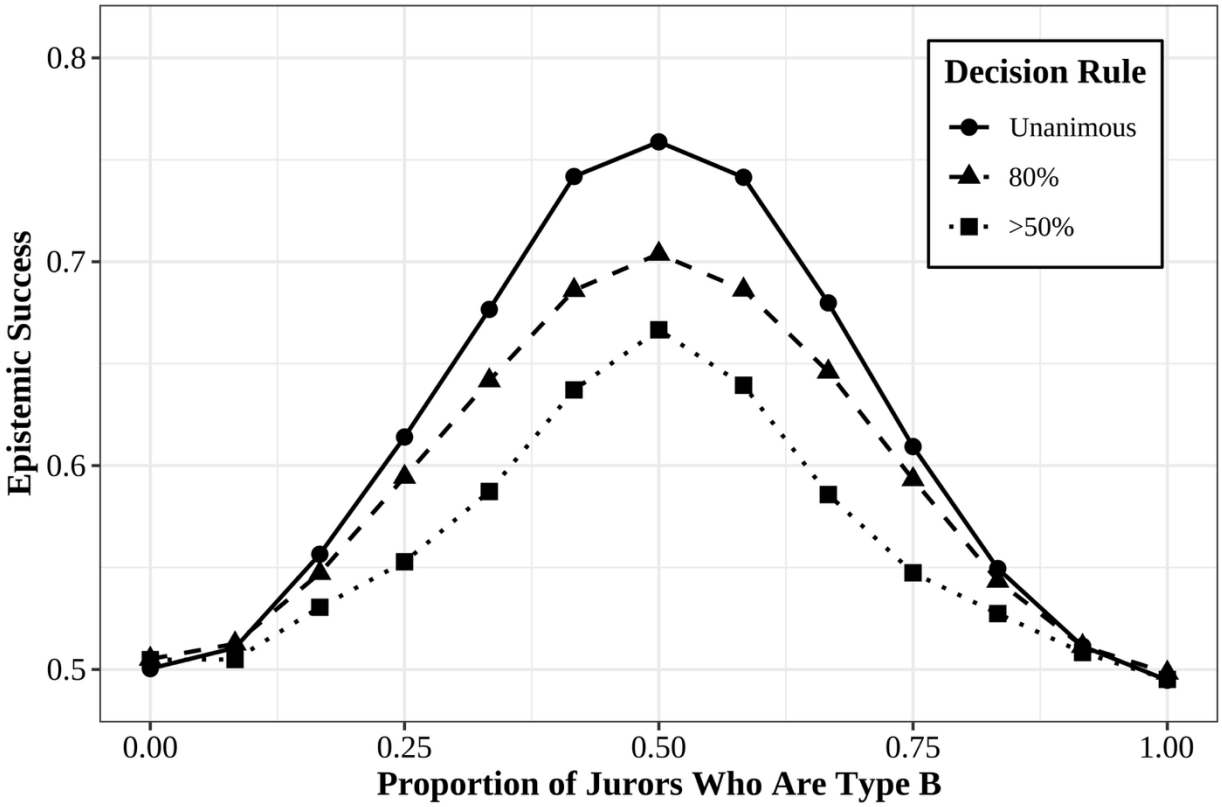
Results

The following four subsections explore four contributions that subgroup diversity can make to epistemic jury success, using the base model and three variations. Using the base model, the first subsection shows that 'sided access'—different subgroups accessing different information—makes equal representation favor jury success. The second subsection highlights the importance of diversity in cases of 'sided access to *better* information,' where one subgroup has access to particularly strong evidence. The third subsection explores diversity in cases of 'random differential access,' where differences in access to information reduce redundancy, which benefits the jury. The fourth subsection presents the surprising result that 'enclave communication'—limiting communication between diverse subgroups—can also favor jury success.

Diversity and Sided Information Access

The base model has jurors in different subgroups accessing different reasons (plus reasons in common), thus modeling diversity as differential sensitivity to trial information or background knowledge. Figure 3 shows jury success for different **A**/**B** juror compositions (with **AB** reasons always set at 50%). This is plotted for each of the three decision rules. As throughout, the results summarize 10,000 runs using juries of 12. They show that unanimous juries do best across all **A**/**B** proportions, and that the gap between the unanimous and the two lower-majority decision rules (i.e., >50% and ≥80%) is most pronounced with equal representation.

Figure 3. *Epistemic jury success (proportion of correct verdicts) for juries of different A/B compositions when Type A jurors and Type B jurors have differential access to positively valent (incriminating) and negatively valent (exculpatory) reasons.*



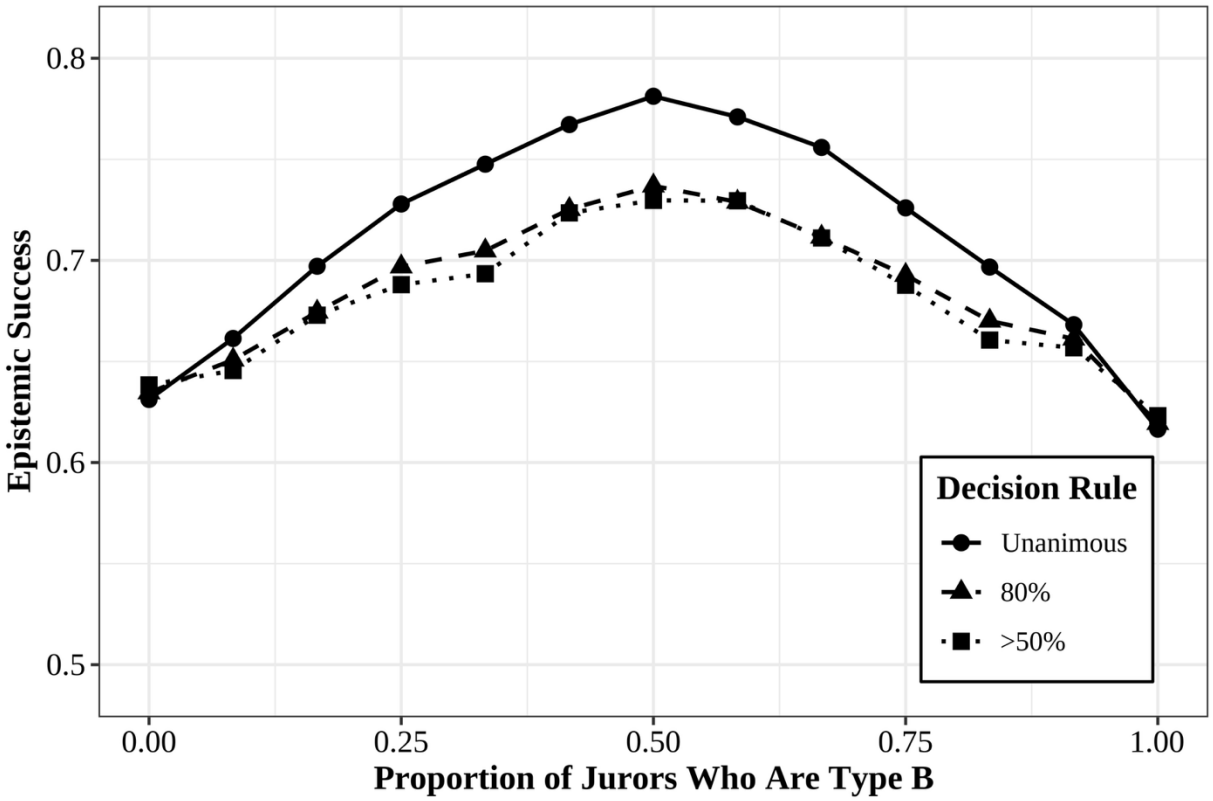
Note: Type A jurors have exclusive access to half of the positively valent reasons (the Type A reasons), Type B jurors have exclusive access to half of the negatively valent reasons (the Type B reasons), and both types of jurors have access to all remaining reasons (the Type AB reasons). Results are aggregated over 10,000 runs using juries of size 12 and are plotted for three verdict decision rules: unanimity, 80%, and simple majority.

In this first simulation, we have assumed that all **A** reasons are positive in valence, supporting a guilty verdict, and all **B** reasons are negative, supporting a not guilty verdict. In this model, it is still possible for **A** jurors to draw both negative and positive reasons among their initial 15, since they draw from a pool of reasons in common as well as from what we might call a pool of 'exclusively **A** reasons' (i.e., **A** reasons outside of 'reasons in common'). The same holds for **B** jurors, who draw from a pool of reasons in common as well as from a pool of 'exclusively **B** reasons' (i.e., **B** reasons outside of 'reasons in common').

We can also relax these assumptions to model cases where differential access does not perfectly align with valence, creating a new parameter—the correlation between reason type (exclusively **A** or exclusively **B**) and valence—controlled by the **reason-type-valence-correlation** slider on the model interface. With a perfect 1.00 correlation (as above), exclusively **A** reasons all support guilt and exclusively **B** reasons all support innocence. With lower correlations, exclusively **A** reasons are less likely to be positive and exclusively **B** reasons less likely to be negative. Figure 4 repeats the simulation with a 0.50 correlation—75% of positive reasons are **A** and 75% of negative reasons are **B**, with 50% of each type converted to **AB**—achieved by setting the **reason-type-valence-correlation** slider to 0.50.

In this case, equal representation again shows epistemic superiority, though the disadvantage at the unequal ends is dampened by the reduced correlation (compare Figures 3 and 4). Unanimous juries remain clearly superior, and the difference between using the 80% and the simple majority rules has nearly vanished.

Figure 4. *Epistemic jury success (proportion of correct verdicts) for juries of different A/B compositions when there is only a 0.50 correlation between reason Type A vs. B and reason valence.*



Note: As before, Type **A** jurors have exclusive access to the Type **A** reasons, and Type **B** jurors have exclusive access to the Type **B** reasons, and both types of jurors have access to all remaining reasons (the Type **AB** reasons). However, now only 75% of the Type **A** reasons are positive, and only 75% of the Type **B** reasons are negative. Results are aggregated over 10,000 runs using juries of size 12 and are plotted for three verdict decision rules: unanimity, 80%, and simple majority.

When Evidence is Stronger on One Side

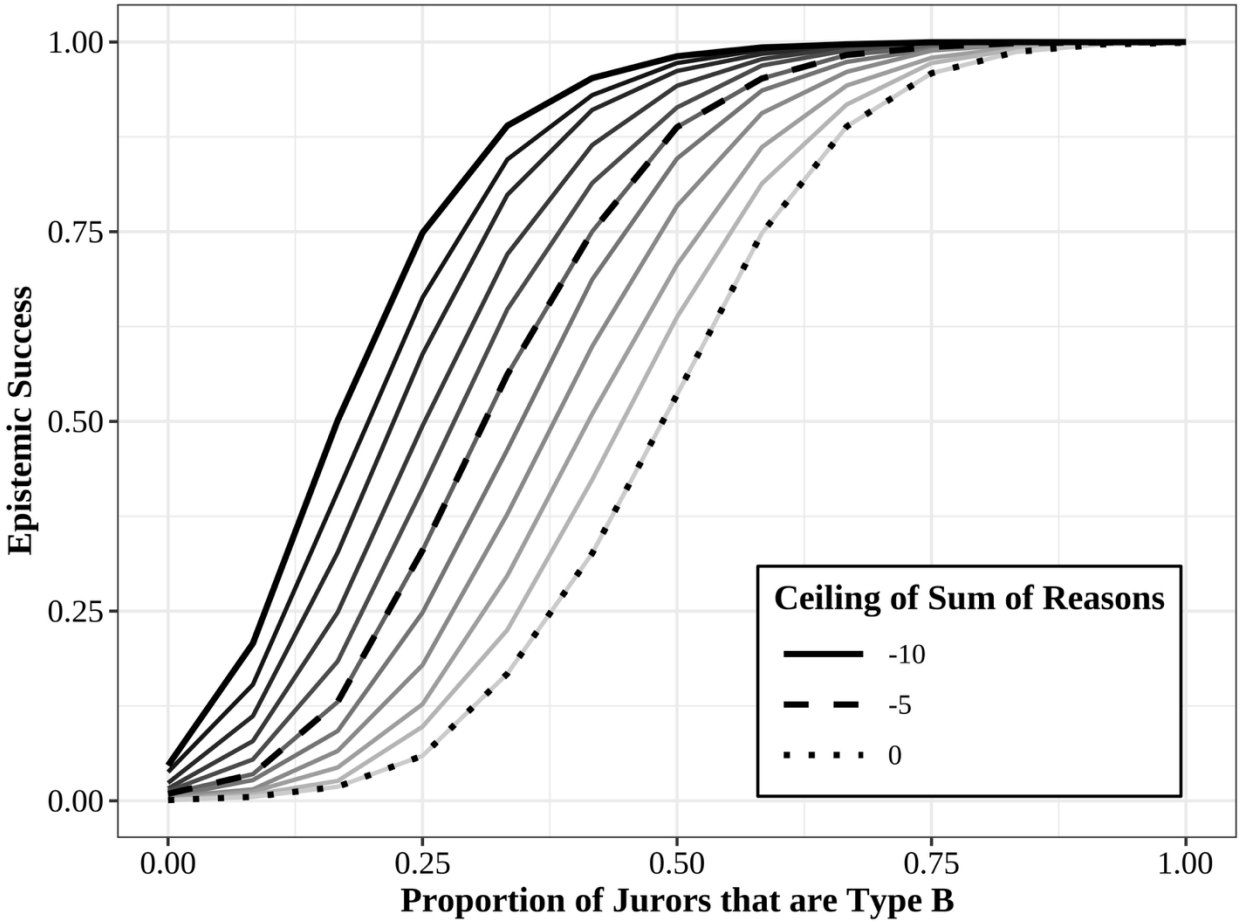
In the base model, sided access was symmetric: with valences assigned uniformly at random, we could expect that the average strength of reasons accessible to **A** and **B** jurors was roughly equal. Here we ask what happens when evidence is stronger on one side—'sided access to better information.' What if **B** jurors are not only more sensitive to exculpatory evidence, but that evidence is also stronger? In such a case, initial 'reasons in common' as well as those

accessible only to **A** jurors ('exclusively **A** reasons') might indicate a close call between guilty and not guilty, yet there might still be significant exculpatory evidence that is initially available only to **B** jurors ('exclusively **B** reasons'). If so, at what strength and at what proportion of **B** jurors can we expect correct verdicts in these cases?

To model this, we again assume that 50% of our reasons are reasons in common (**AB** reasons), that all exclusively **B** reasons are negatively valent, and that all exclusively **A** reasons are positively valent. We restricted our sample data to cases in which (by chance) the sum of the entire pool of reasons was negative (so, the truth is "not guilty"). We collected 10,000 randomly-produced runs where the sum of the entire pool of reasons fell between -11 and -10 (including -10 but excluding -11), between -10 and -9 (including -9, excluding -10), etc. until the interval -1 to 0 for every possible combination of **A/B** jurors (0/12, 1/12, etc.). Figure 5 shows the results.

Not surprisingly here, jury success is driven by two factors: evidence strength and the representation of the better-informed subgroup. With a strong case (sum between −11 and −10), a jury composed of as few as 25% **B** jurors produces success rates comparable to the highest in earlier simulations. These results confirm two expectations: (a) representation by the better-informed subgroup is essential, and (b) when their evidence is particularly strong, even minority representation can suffice. As the case grows more subtle (the sum of reasons approaches 0), more members of subgroup **B** are required for high success.

Figure 5. Epistemic jury success (proportion of correct verdicts) for juries of different A/B compositions under a unanimity decision rule when the sum of the full reason pool is negative at different strengths.



Note: Type A jurors have exclusive access to half of the positively valent reasons (the Type A reasons), Type B jurors have exclusive access to half of the negatively valent reasons (the Type B reasons; so, a perfect 1.00 correlation between A vs. B reason type and reason valence), and both types of jurors have access to all remaining reasons (the Type AB reasons). Results are aggregated over 10,000 runs using juries of size 12.

Diversity and the Reduction in Informational Redundancy

One purported benefit of diversity is an increase in the total information and skills in the group (Singer, 2019). Lu Hong and Scott Page show a 'diversity trumps ability' effect: groups with randomly diverse heuristics can outperform groups of individuals with the highest individual performance (Hong & Page, 2004; Page, 2007; Page, 2011;

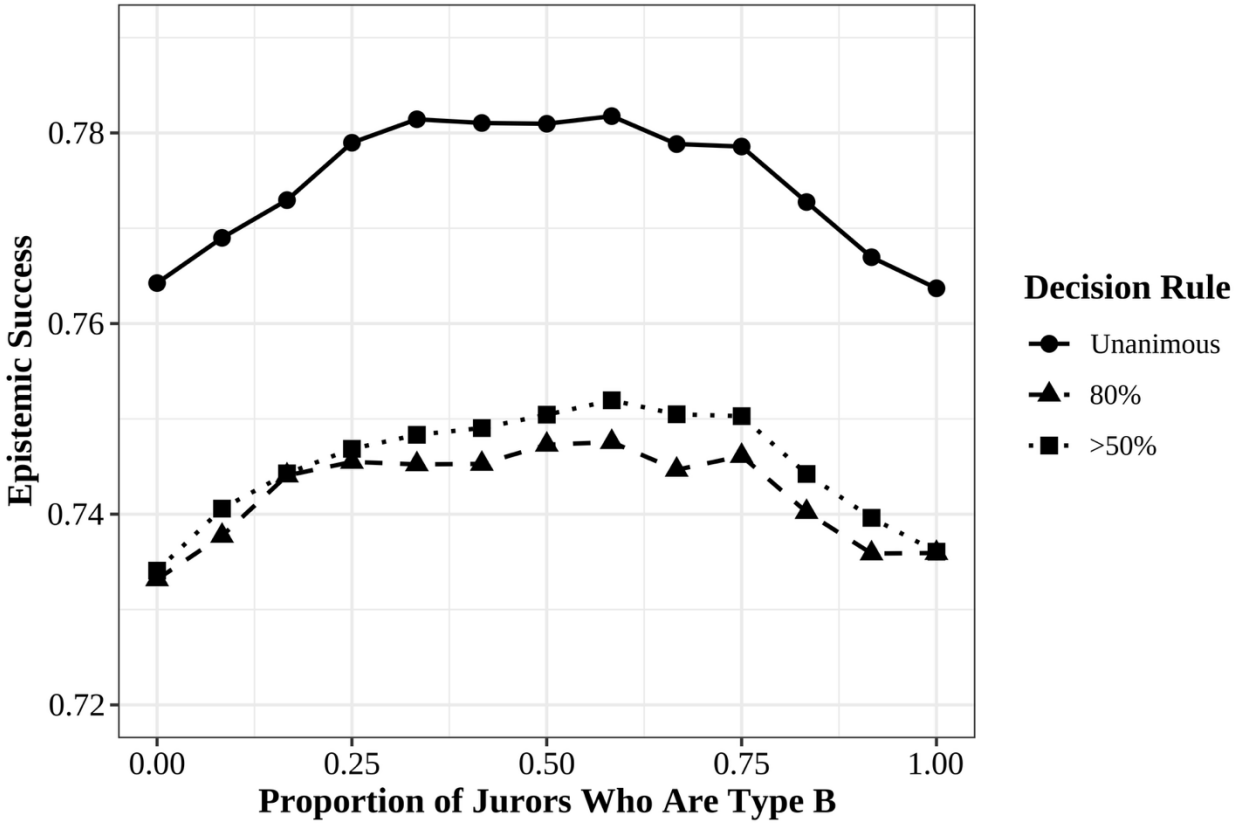
Page, 2017). Although the landscape that Hong and Page (2004) used in their model appears to have been crucial to this result (undercutting its generality), diverse heuristics have nevertheless been shown to play an important role across a range of problems (Grim et al., 2019).

The informal explanation given for the Hong and Page 'diversity trumps ability' effect, where it does hold, is that a group of highest-performing individuals will share the same heuristics. A random group will likely not. It is the redundancy of heuristics in the group of highest-performing individuals, and the non-redundancy in the random group, that explains the greater success of the latter.

In this subsection, we consider cases where different subgroups have different initial access to different pieces of information, but there is nothing special or distinctive about the kind of information they have access to. This stands in contrast to the base version of the model, where we assumed that the sided access was (at least likely to be) information on one side of the question. In cases of 'random differential access,' we see a benefit of diversity that mirrors the Hong and Page result, where the mere fact that different subgroups have access to different pieces of information provides a benefit (albeit a smaller one) to the group.

In this variation of the model, juries again have two subgroups with exclusive access to different reasons, but with no correlation between reason type and valence (**reason-type-valence-correlation** = 0) and no difference in average evidence strength on either side. The only difference between subgroups is which individual reasons they initially have access to. The results are shown in Figure 6.

Figure 6. *Epistemic jury success (proportion of correct verdicts) for juries of different A/B compositions when there is no correlation between reason Type A vs. B and reason valence.*



Note: As before, Type A jurors have exclusive access to the Type A reasons, and Type B jurors have exclusive access to the Type B reasons, and both types of jurors have access to all remaining reasons (the Type AB reasons). However, now reasons are assigned randomly to the A and B types, without regard to either the sign or strength of their valence. Results are aggregated over 10,000 runs using juries of size 12, and are plotted for three verdict decision rules: unanimity, 80%, and simple majority.

Success is again highest when juries use the unanimous decision rule, with the 80% and >50% rules performing worse. The difference is small (note the y-axis range) but clearly evident. Success is fairly uniform with substantial representation of both subgroups, falling off only at the less-diverse ends.

Why does this happen? Since the true verdict is fixed by all the reasons, the more reasons that are considered, the greater the chance of a correct verdict. Since different subgroups have

access to different pieces of information in these runs, in a more diverse jury, jurors are less likely to bring the same reasons to the discussion, increasing the number of different pieces of information considered. So here, non-redundancy explains diversity's positive epistemic effect, and redundancy explains the weaker results of homogeneous juries.

The Benefits of Limited Communication with Diversity

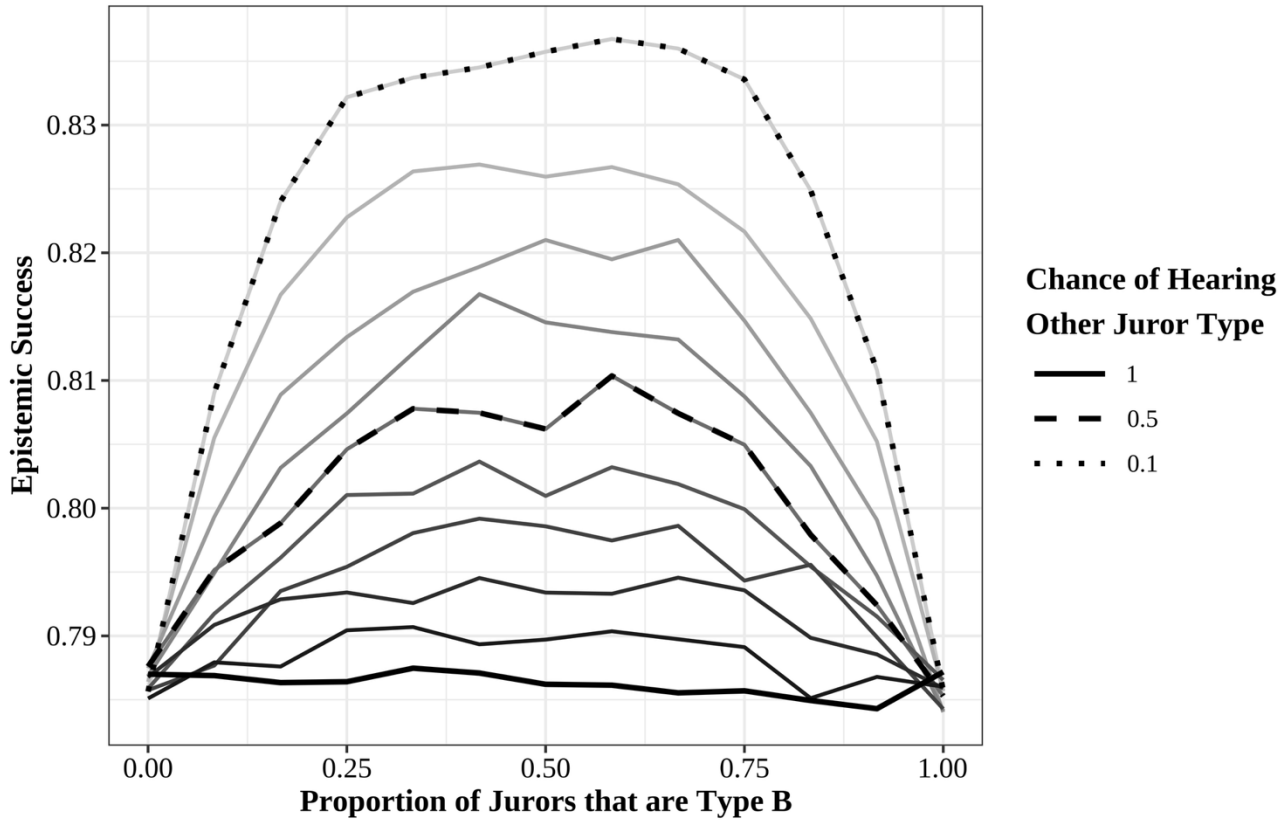
It is a standard part of liberal belief (at least as old as John Stuart Mill) that diversity has epistemic benefits. For example, this is a central tenet of the recent literature on the epistemic benefits of democracy (Anderson, 2006; Estlund, 1993, 2008; Goodin & List, 2001; Landemore, 2012; Berger & Sales, 2020). All of our results so far support this idea.

At the same time, it is often noted that diversity brings communication difficulties (Beyer, Edwards, & Fuller, 2015; Rickford & King, 2016; Jones, et al., 2019; Putnam, 2007). In this subsection, we model the effect of limiting communication between diverse enclaves. Our results challenge the liberal idea that increased communication between subgroups is always epistemically advantageous (Berry, 1999; Mutz, 2002; Grönlund, et al. 2015; Pettigrew & Tropp, 2006, 2013; Iyengar, et al. 2019). Creating epistemic enclaves, by *restricting* communication between subgroups, can in some cases result in greater epistemic success of the full group.

Our previous results assumed perfect communication: all shared reasons were heard and incorporated by all jurors. To examine the effect of imperfect communication, we start with the base version of the model, with 12 jurors divided into two subgroups, **A** and **B**. We add a variable probability that a reason shared by a member of one's own subgroup will be heard by a member of the other subgroup (the **chance-of-hear-offtype-speaker** slider), while within-subgroup sharing remains perfect. For instance, when an **A** juror shares a reason, all other **A** jurors hear it, but each **B** juror hears it only with some probability. To isolate the effect of

imperfect communication, we assume 100% reason overlap (all reasons are **AB** reasons) and 0 valence correlation. Figure 7 shows jury success under unanimity when cross-subgroup communication is limited.

Figure 7. Epistemic jury success (proportion of correct verdicts) for juries of different A/B compositions as a function of the probability that a reason shared by a juror of one type will be heard by a juror of the other type.



Note: Both Type A and Type B jurors have access to all reasons (i.e., all reasons are AB reasons), unanimity is the verdict decision rule, and results are aggregated over 10,000 runs using juries of size 12. Results are shaded according to the probability [0.1 to 1] of an A juror incorporating a reason announced by a B juror, and vice versa.

As Figure 7 shows, overall jury success *increases* with each *decrease* in the probability of successful communication between members of the two juror subgroups. Here the effect is again small but noticeable, especially when there is substantial representation of both juror subgroups. When we look at groups that are perfectly split between **A** and **B** jurors, with perfect

communication (**chance-of-hear-offtype-speaker** = 1), jury success rate is about 78.6% (see Figure 6). But with a probability of 0.1 of hearing off-type jurors, the success rate of the whole group reaches about 83.5% (Figure 7). In these runs, there is no difference, on average, between the subgroups in terms of what reasons they can access or what the valence of those reasons is likely to be. So, unlike the sections above, there is no way in which the benefits of diversity we see here could be made sense of in terms of diversity increasing the number of reasons held by the jury as a whole. The effect we see here is purely due to members of different groups communicating in an enclaved way. Further, although we do not show the specific results here, this effect is one that remains no matter how much we decrease overlap and increase the correlation between reason type and valence.

In the previous section we argued that the epistemically positive non-redundancy effects of diversity echoed aspects of Hong and Page's (2004) 'diversity trumps ability' effect. The explanation for the epistemically positive effects of enclave communication that we see here, we propose, echoes another familiar effect from a model of the social structure of science. In a series of papers, Kevin Zollman (2007, 2010a, 2010b) shows that there is an epistemic benefit from 'transient diversity.' In the simplest case, Zollman considers networks in which agents independently explore payoffs from two one-armed bandits (essentially two simple slot-machines that pay either +1 or 0). The agents also communicate results with linked neighbors in different network structures, including fully connected networks, hub networks, and ring networks. The communication network that performs best is not the fully connected network in which all agents have access to results from all others. Rather, the best network is the least connected one, the maximally distributed ring network. That is also the network that converges

to a universally agreed result most slowly. Related results regarding the benefits of restricted communication appear in Grim et al. (2013).

As in our model, information exchange in Zollman's models ends when the group converges on a shared answer. Convergence happens because evidence sharing eventually gets everyone on the same side. In networks that are less connected, it takes longer for a piece of evidence to be shared across the network, meaning that some agents who might have changed their mind upon hearing the evidence (potentially resulting in convergence for the group) have more time to explore. What Zollman saw was that agents in less connected networks generally have more time to explore on their own before the group converges. That extra exploration results in the whole group collectively producing more insight into which bandit is best, thereby increasing the chances that the whole group will converge on the right answer. The increased success of less connected networks in Zollman's models is explained by the agents in the less connected networks collectively exploring more than agents do in the more connected networks.

Our jury model exhibits a similar effect. Limiting communication makes it less likely that members of one subgroup will hear reasons put forward by members of the other subgroup. As in Zollman's models, this means that agents who would have been convinced by such reasons (potentially resulting in unanimity for the group) are sometimes able to maintain their view when those reasons go unheard. This results in longer times to convergence in models with more limited communication. The longer times to convergence mean that more reasons are shared and a higher average number of reasons end up being held by individual jurors. We see this in the data: when communication is unlimited in an evenly-split jury (with 100% overlap and 0 correlation), the number of reasons held by an agent when a unanimous verdict is reached is 36.4 (averaged over 100,000 runs). That number shoots up to 45.1 (a 24% increase) when

communication between the groups is limited to a 0.1 chance of success. So even though limiting communication reduces the chance that a shared reason will be taken up by a juror of the other subgroup on any particular occasion, overall, limiting communication increases the total number of reasons heard by increasing the time to unanimity. This explains their higher success rates. So, like in Zollman's models, the increased success of juries with limited communication can be explained by their exploring more reasons before agreeing.

Our model thus challenges the liberal idea that more cross-group communication is always better: insulating subgroups can be epistemically beneficial. Of course, taking this to the extreme would be detrimental—without any cross-subgroup communication, consensus would rarely occur and success would plummet.

Discussion

We have used a simple agent-based model to explore an overlooked aspect of jury information aggregation: the role of diversity in epistemic success. Diversity proves important in four ways. When different groups are sensitive to different kinds of reasons, jury success demands the maximal diversity of equal representation between the two groups. When reasons are particularly strong on one side and the differential sensitivity remains, diversity is of even more importance. Even without assumptions regarding correlation and strength of reasons, lack of diversity threatens jury success in ways analogous to Hong and Page's 'diversity trumps ability' result. Finally, and somewhat surprisingly, diversity is of value to jury success in limiting communication between the groups—a result evocative of Zollman's 'transient diversity' effect.

It is worth emphasizing the limits and artificialities of the model. Our 'truth' is reflected simply in the sum of the strengths of the initial reasons weighted by their valences. Further, reasons are communicated randomly during decision-making, and jurors have unlimited

memory. Although we refer to jury opinion-formation as 'deliberation,' our emphasis is on collective information aggregation and consensus based on diverse individual information introduction, exchange, processing, and retrieval. Other aspects of full deliberation—for example, challenging the structure of other jurors' arguments, questioning the credibility of witnesses or the interpretation and relevance of circumstantial evidence—are clearly beyond the simple model offered here. We do not claim that practical or policy implications follow directly. These results are best understood as theoretical insights into mechanisms that could operate, at a high level, in real-world phenomena. Despite those limitations, our results suggest four important ways in which jury diversity can influence epistemic success.

As noted, empirical results indicate that unanimity has an epistemic advantage (Devine, 2012), even though some previous analytic and computational results have not shown such benefit (List, 2004; Angere, Olsson, & Genot, 2016). Our results, which incorporate diversity in ways that those models do not, suggest that unanimous juries *do* likely have an importantly higher success rate across all the conditions considered. The suggestion, worthy of further investigation, is that the epistemic case for unanimity is ultimately tied to fully exploiting the benefits of diversity.

Engaging with the Model

To deepen understanding of the dynamics between diversity, information sharing, and jury verdicts, we encourage readers to download and explore the model. It can be found on the book's companion website. Below we suggest four experiments, ranging from simple parameter adjustments to code modifications, all bearing on diversity, unanimity, and jury success.

First, try exploring a "Separate Worlds" scenario. In the results discussed above, we maintained a moderate overlap in the information available to different subgroups. But what

happens when that common ground vanishes? Readers can test this by setting the **probability-overlap** slider to 0 while keeping the **reason-type-valence-correlation** slider set at 1.0. In this scenario, **A** and **B** jurors begin with completely distinct facts pointing in opposite directions, as if from separate epistemic worlds. By running the model a few times and watching the three monitors below the heading **Time (tick) when jury reached…**, as well as the **Percent of Jurors with True Belief** plot, readers can observe how these "echo chambers" struggle to reach even a simple majority, let alone full consensus. This is most evident if you compare these runs to a few runs with the **probability-overlap** slider set to .5 instead of 0. In the latter case, jurors share a portion of their reasons in common (i.e., their worlds are only partially separate), and they are generally able to come to a decision more quickly. Does the lack of common ground in the first case simply delay the verdict, or does it destabilize the jury entirely? If a unanimous verdict does still occur, consider why that happens in the model. Comparing these two types of runs will highlight the role of shared information in bridging diverse groups.

Second, simulate a 'Lone Dissenter' scenario, reminiscent of the classic movie *12 Angry Men*. Set **num-people** to 12 and **proportion-type-B-jurors** to 0.08 (type this in the box above the slider). This will generate a jury of eleven members of one subgroup and one of the other. If the reader also sets **reason-type-valence-correlation** to 1.0 and the **probability-overlap** slider to 0, that lone juror becomes the sole possessor of a specific set of exculpatory or incriminating evidence. How often can this lone agent persuade the majority? Run the scenario 20 times and tally the number of correct verdicts under each decision rule (watch the three monitors below the heading **Jury converged on truth?**). Compare what you see to what happened when you ran the completely Separate Worlds scenario (for a fair comparison, you may want to re-run that Separate Worlds scenario 20 times). How would you characterize the performance of the juries

with a "lone dissenter" in terms of both speed and accuracy? How does this differ from the "Separate Worlds" juries? Based on the results you generated, how realistic does the plot depicted in the movie *12 Angry Men* seem to you?

The preceding activities are easy, but their conclusions rest on small samples—20 runs per condition—which can mislead. More experienced readers can use a BehaviorSpace experiment to run many trials quickly, with results stored automatically in a spreadsheet. We have provided three such experiments: one for the 'Separate Worlds' scenario, one for the 'Lone Dissenter' scenario, and a third that is explained below. These pre-set experiments can be found by going to **Tools** >> **BehaviorSpace**, and then click open the one you wish to run.[3]

The Separate Worlds experiment runs the model 1,000 times with **proportion-type-B-jurors** set at 0.5, **probability-overlap** set at 0, and **reason-type-valence-correlation** set at 1.0 (the completely separate worlds condition), and 1,000 times with **proportion-type-B-jurors** set at 0.5, **probability-overlap** set at .5, and **reason-type-valence-correlation** set at 1.0 (the partially separate worlds comparison condition). We recommend selecting 'table output' (click **run**, then **browse** next to table output). This produces a data file with one output line per run, which can be easily tallied in a spreadsheet.

The Lone Dissenter experiment works similarly, but in this case the model runs 1,000 times with **proportion-type-B-jurors** set at 0.08, **probability-overlap** set at 0, and **reason-type-valence-correlation** at 1.0 (the lone dissenter condition), and 1,000 times with **proportion-type-B-jurors** set at 0.5, **probability-overlap** set at 0, and **reason-type-valence-correlation** at 1.0 (the same completely separate worlds condition as above). Again, we recommend choosing

---

[3] Additional information about creating and running BehaviorSpace experiments can be found in the BehaviorSpace section of the NetLogo User Manual: https://docs.netlogo.org/7.0.3/behaviorspace.

table output. With 1,000 runs per condition, you should now be able to answer more confidently the questions raised above.

The third BehaviorSpace experiment concerns our finding that limiting communication between jury members (creating enclaves) can improve jury verdict accuracy. As mentioned, however, if communication drops too low, the jury will fail to communicate at all. So, up to what point does limiting communication help? Readers can explore this question with the Limited Communication BehaviorSpace experiment. It systematically lowers the value of the **chance-of-hear-offtype-speaker** slider from 0.1 down to 0.0 in increments of 0.01, running the model 50 times at each step. This is done for both a small jury (**num-people** = 6) and a large one (**num-people** = 50)—1,100 runs total (50 × 11 × 2). Again, select table output. Despite fewer runs than the other experiments, this one takes longer—be patient. When done, open the output file and tabulate the percentage (out of 50) reaching a correct verdict under each decision rule in each of the 22 conditions.[4] Note in particular the percentage of 6-person and 50-person juries that reached a correct decision when **chance-of-hear-offtype-speaker** dropped to 0. Are the larger juries more or less successful than the smaller juries in this minimum communication situation? Does the decision rule (>50% vs. ≥80 vs. unanimity) make a difference? And how is it possible for a jury to reach unanimity in such a situation? The answer to this last question lies in the trade-off between the redundancy provided by a large group and the insulation provided by restricted communication.

---

[4] If a jury deliberates for more than 100,000 ticks without coming to a unanimous decision, the model run for that jury is terminated, the jury is declared to be hung with respect to the unanimity decision rule, and no result is output. The same jury may or may not also be hung according to the simple majority and 80% decision rules, depending on whether or not those less stringent criteria were exceeded before reaching the 100,000 tick limit.

Finally, we have several suggestions for more sophisticated modelers who feel comfortable modifying the model's underlying code. One is to introduce a third juror subgroup. Real-world juries are rarely perfectly binary; they often contain factions with distinct theories of the case or varying focus on procedural versus factual evidence. Adding a **C** subgroup of jurors can change the social dynamic from a polarized dyad to a complex triad. This requires adding a **C** subgroup and **C**-exclusive reasons, then creating **AC**, **BC**, and **ABC** reasons in common. There are many new possibilities to test in this kind of model, since moving from a dyadic dynamic (**A** vs. **B** jurors) to a triadic one (**A** vs. **B** vs. **C** jurors) may reveal non-linear effects in how diversity impacts verdict accuracy, particularly if the third group acts as a bridge or a buffer between the other two.

Another suggestion is to modify how jurors' memory works. As noted, jurors retrieve reasons randomly from a perfect, unlimited memory—not an accurate description of real jurors. Several modifications could help. One is to make retrieval (and thus sharing) a function of reason strength, with stronger reasons more likely to be recalled. Another is to make retrieval also a function of the juror's current belief about guilt—an idea from Stasser (1988). Finally, a forgetting feature could be added: reasons heard during deliberation are forgotten with a probability that varies with their strength and recency (stronger and more recently heard reasons are less likely to be forgotten). Determining whether these features—individually or in combination—preserve the epistemic benefits of diversity described here would bring the model meaningfully closer to the realities of jury deliberation.

References

Anderson, E. (2006). The epistemology of democracy. *Episteme, 3*(1-2), 8–22. https://doi.org/10.3366/epi.2006.3.1-2.8

Angere, S., Olsson, E. J., & Genot, E. J. (2016). Inquiry and deliberation in judicial systems: the problem of jury size. In C. Baskent (Ed.), *Perspectives on Interrogative Models of Inquiry: Developments in Inquiry and Questions* (pp. 35-56). Springer. https://doi.org/10.3366/epi.2006.3.1-2.8

Anwar, S., Bayer, P., & Hjalmarsson, R. (2022). Unequal jury representation and its consequences. *American Economic Review: Insights*, *4*(2), 159-174. https://doi.org/10.1257/aeri.20210149

Batterman, R. W., & Rice, C. C. (2014). Minimal model explanations. *Philosophy of Science*, *81*(3), 349-376. https://doi.org/10.1086/676677

Berger, W. J., & Sales, A. (2020). Testing epistemic democracy's claims for majority rule. *Politics, Philosophy & Economics*, *19*(1), 22-35. https://doi.org/10.1177/1470594x19870260

Berry, J. W. (1999). Intercultural relations in plural societies. *Canadian Psychology/Psychologie Canadienne*, *40*(1), 12-21. https://doi.org/10.1037/h0086823

Beyer, T., Edwards, K. A., & Fuller, C. C. (2015). Misinterpretation of African American English by adult speakers of standard American English. *Language & Communication*, *45*, 59-69. https://doi.org/10.1016/j.langcom.2015.09.001

Chopra, S. (2014, Summer). Preserving jury diversity by preventing illegal peremptory challenges: how to make a Batson/Wheeler motion at trial (and why you should). *The Trial Lawyer*, 12-23, 29. http://www.njp.com/wp-content/uploads/article/article30.pdf

Devine, D. J. (2012). *Jury decision making: The state of the science*. New York University Press.

Devine, D. J., Buddenbaum, J., Houp, S., Stolle, D.P., & Studebaker, N. (2007). Deliberation quality: a preliminary examination in criminal juries. *Journal of Empirical Legal Studies*, *4*(2), 273-303. https://doi.org/10.1111/j.1740-1461.2007.00089.x

Eason, R., Rosenberger, R., Kokalis, T., Selinger, E., & Grim, P. (2007). What kind of science is simulation? *Journal of Experimental & Theoretical Artificial Intelligence*, *19*(1), 19-28.

Epstein, J. (2006). *Generative social science: Studies in agent-based computational modeling*. Princeton University Press.

Estlund, D. (1993). Making truth safe for democracy. In D. Copp, J. Hampton & J. Roemer (Eds.), *The Idea of Democracy*, (pp. 71–100). Cambridge University Press.

Estlund, D. (2008). *Democratic authority: A philosophical framework.* Princeton University Press.

Grim, P., Singer, D. J., Bramson, A., Holman, B., Jung, J., & Berger, W. J. (2024). The epistemic role of diversity in juries: An agent-based model. *Journal of Artificial Societies and Social Simulation***,** 27(1), 12. https://doi.org/10.18564/jasss.5304

Grim, P., Singer, D. J., Fisher, S., Bramson, A., Berger, W. J., Reade, C., Flocken, F., & Sales, A. (2013). Scientific networks on data landscapes: Question difficulty, epistemic success, and convergence. *Episteme, 10*(4), 441–464. https://doi.org/10.1017/epi.2013.36

Grim, P., Singer, D. J., Fisher, S., Bramson, A., Holman, B., McGeehan, S., & Berger, W. J. (2019). Diversity, ability, and expertise in epistemic communities. *Philosophy of Science, 86*(1), 98-123. https://doi.org/10.1086/701070

Grönlund, K., Herne, K., & Setälä, M. (2015). Does enclave deliberation polarize opinions? *Political Behavior*, *37*, 995-1020. https://doi.org/10.1007/s11109-015-9304-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

Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: the effects of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behavior, 20*, 655-670. https://doi.org/10.1007/BF01499100

Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual Review of Political Science*, *22*, 129-146. https://doi.org/10.1146/annurev-polisci-051117-073034

Jones, T., Kalbfeld, J. R., Hancock, R., & Clark, R. (2019). Testifying while black: An experimental study of court reporter accuracy in transcription of African American English. *Language*, *95*(2), e216-e252. https://doi.org/10.1353/lan.2019.0042

Joshi, A. S., & Kline, C. T. (2015). *Lack of jury diversity: A national problem with individual consequences*. American Bar Association. https://www.americanbar.org/groups/litigation/committees/diversity-inclusion/articles/2015/lack-of-jury-diversity-national-problem-individual-consequences/

Koch, C. M., & Devine, D. J. (1999). Effects of reasonable doubt definition and inclusion of a lesser charge on jury verdicts. *Law and Human Behavior*, *23*(6), 653-674. https://doi.org/10.1023/A:1022389305876

Kuehn, D. (2017). Diversity, ability, and democracy: A note on Thompson's challenge to Hong and Page. *Critical Review*, *29*(1), 72-87. https://doi.org/10.1080/08913811.2017.1288455

Landemore, H. (2012). *Democratic reason: Politics, collective intelligence, and the rule of the many.* Princeton University Press.

Laudan, L. (2008). *Truth, error, and criminal law: An essay in legal epistemology*. Cambridge University Press.

List, C. (2004). On the significance of the absolute margin. *British Journal for the Philosophy of Science*, *55*(3), 521-544. https://doi.org/10.1093/bjps/55.3.521

Mutz, D. (2002). Cross-cutting social networks: Testing democratic theory in practice. *American Political Science Review, 96*(1), 111-126. https://doi.org/10.1017/S0003055402004264

Page, S. E. (2007). *The difference*. Princeton University Press.

Page, S. E. (2011). *Diversity and complexity*. Princeton University Press.

Page, S. E. (2015). Diversity trumps ability and the proper use of mathematics. *Notices of the American Mathematical Society, 62*(1), 9-10.

Page, S. E. (2018). *The model thinker*. Basic Books.

Penrod, S. D., & Hastie, R. (1980). A computer simulation of jury decision making. *Psychological Review*, *87*(2), 133-159. https://doi.org/10.1037/0033-295X.87.2.133

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751–783. https://doi.org/10.1037/0022-3514.90.5.751

Pettigrew, T. F., & Tropp, L. R. (2013). *When groups meet: The dynamics of intergroup contact*. Psychology Press.

Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century – The 2006 Johan Skytte Prize lecture. *Scandinavian Political Studies*, *30*(2), 137-174. https://doi.org/10.1111/j.1467-9477.2007.00176.x

Rickford, J. R., & King, S. (2016). Language and linguistics on trial: Hearing Racheal Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language*, *92*(4), 948-988. https://doi.org/10.1353/lan.2016.0078

Saks, M. J., & Marti, M W. (1997). A meta-analysis of the effects of jury size. *Law and Human Behavior 21*(5)*,* 451-467. https://doi.org/10.1023/A:1024819605652

Schutte, J. W., & Horsch, H. M. (1997). Gender differences in sexual assault verdicts: A meta-analysis. *Journal of Social Behavior and Personality, 12*(3), 759-772.

Singer, D. J. (2019). Diversity, not randomness, trumps ability. *Philosophy of Science, 86*(1), 178-191. https://doi.org/10.1086/701074

Singer, D. J., Bramson, A., Grim, P., Holman, B., Jung, J., Kovaka, K., Ranginani, A., & Berger, W. J. (2019). Rational social and political polarization. *Philosophical Studies, 176*, 2243-2267. https://doi.org/10.1007/s11098-018-1124-5

Singer, D .J., Bramson, A., Grim, P., Holman, B., Kovaka, K., Jung, J., & Berger, W. J. (2021). Don't forget forgetting: The social epistemic importance of how we forget. *Synthese, 198*, 5373-5394. https://doi.org/10.1007/s11229-019-02409-0

Stasser, G. (1988). Computer simulation as a research tool: The DISCUSS model of group decision making. *Journal of Experimental Social Psychology, 24*(5), 393-422. https://doi.org/10.1016/0022-1031(88)90028-5

Stasser, G., & Davis, J. H. (1981). Group decision making and social influence: A social interaction sequence model. *Psychological Review, 88*(6), 523-551. https://doi.org/10.1037/0033-295X.88.6.523

Tanford, S., & Penrod, S. (1983). Computer modeling of influence in the jury: The role of the consistent juror. *Social Psychology Quarterly, 46*(3), 200-212. https://doi.org/10.2307/3033791

Thoma, J. (2015). The epistemic division of labor revisited. *Philosophy of Science, 82*(3), 454-472. https://doi.org/10.1086/681768

Thompson, A. (2014). Does diversity trump ability? An example of the misuse of mathematics in the social sciences. *Notices of the American Mathematical Society, 61*(9), 1024-1030. https://doi.org/10.1090/noti1163

U.S. Courts. (2019, May 19). *Courts seek to increase jury diversity*. https://www.uscourts.gov/news/2019/05/09/courts-seek-increase-jury-diversity.

Weisberg, M., & Muldoon, R. (2009). Epistemic landscapes and the division of cognitive labor. *Philosophy of Science, 76*(2), 225-252. https://doi.org/10.1086/644786

Zollman, K. (2007). The communication structure of epistemic communities. *Philosophy of Science, 74*(5), 574-587. https://doi.org/10.1086/525605

Zollman, K. (2010a). The epistemic benefit of transient diversity. *Erkenntnis, 72*(1), 17-35. https://doi.org/10.1007/s10670-009-9194-6

Zollman, K. (2010b). Social structure and the effects of conformity. *Synthese, 172*(3), 317-340. https://doi.org/10.1007/s11229-008-9393-8