The Limits of Reasons: Re-Interpreting The Condorcet Jury Theorem

Lu Hong
Loyola University Chicago

Scott E Page
University of Michigan, Santa Fe Institute[1]

December 7, 2018

[1]We thank participants in the MacArthur Foundation working group on collective problem solving. We received valuable feedback on this work from audiences at the American Political Science Association 2018 Annual Meeting, INSEAD, NTU-Singapore, Northwestern University, the University of Illinois, the Santa Fe Institute, the University of Michigan and at the Deliberation, Belief Aggregation, and Epistemic Democracy Conference held in Paris, June 2018

Abstract

In this paper, we derive necessary and sufficient conditions for a collection of individuals with interpreted signals to classify with perfect accuracy. Interpreted signals assume classification models defined over a state space. We show that perfect accuracy requires that the outcome function be measurable with respect to intersections of sets in the individuals' classification models and that the outcome function can be expressed as a linear sum of those models. Given that result, it follows that inaccuracy can arise from lack of model diversity, insufficient model granularity, randomness of outcomes, or outcome function complexity. We then re-analyze the Condorcet Jury Theorem assuming interpreted signals and show that none of the standard results necessarily hold. While increasing individual accuracy has a monotonic (setwise) effect on collective accuracy, increasing group size does not. Moreover, collective accuracy is equally likely to increase or decrease in model diversity for low to moderate levels of diversity. However, reductions in collective diversity can be reversed through deliberation or strategic behavior.

The Condorcet Jury Theorem considers a collection of individuals choosing between two alternatives. It applies to two problem domains: prediction of future outcomes and classification (or adjudication) of a present or past state. Both are core tasks that arise within organizations and confront society writ large. Businesses, governments, and non profits try to identify the current state of the world and make forecasts when formulating actions and strategies. Society must decide the guilt or innocence of the accused, the legality or illegality of actions, and between candidates and on referenda (Kleinberg, Ludwig, Mullainathan, and Obermeyer 2015).

Not surprisingly, the Condorcet Jury Theorem occupies a prominent position within the social sciences. In political science, it provides theoretical justification for the use of juries to decide criminal cases, judicial panels to adjudicate laws, legislative committees to approve bills, the separation of powers to approve laws, and, to some extent, democracy itself. In economics, psychology, and management, the theorem explains the predominance of groups in decision making, such as the reliance on corporate boards to make key strategic decisions, on teams of stock analysts to make investment decisions, and on multiple interviewers to make hiring decisions (McLean and Hewitt 1994, Page 2018).

An enormous number of models in economics, political science, and psychology borrow Condorcet's formalization of classifications or votes as independent draws from a distribution. The core framework can be embedded in models of markets and hierarchies as well as democracies (Sah and Stiglitz 1986). There exist any number of extensions of the original model that allow for correlation (Borgers, et al 2013, Ladha 1992), bias (Stone 2015), dependency (Berg 1993), common signals (Grofman, et al 1983), multiple truths (List and Spiekermann 2016), and strategic voting (Austen-Smith and Banks 1996, Feddersen and Pesendorfer 1996).

The formal theorem comprises four distinct claims (i) that a collection of individuals using majority rule will make more accurate binary classifications or predictions than its individual members, that group accuracy increases in both (ii) the accuracy of the individuals

and (iii) the group size, and (iv) that large groups approach perfect accuracy (Austen-Smith and Banks 1996). Empirical evidence corroborates two of these claims. Groups generally outperform individuals, and groups comprised of more accurate individuals are more accurate. That evidence comes from a variety of domains—economic forecasts, interviews, jury decisions—in both observational and experimental settings (Miller 1996, Palfrey 2009, Kagel and Roth 2016).

Support for the third claim is much weaker. Google found that the gain in accuracy from additional interviewers effectively vanished after four interviewers (Sharper 2017). Prediction contests involving thousands and even millions of participants in the domains of international politics, economics and sports found no benefit beyond a dozen forecasters (Goldstein, McAfee, and Suri 2014, Mellers et al 2015). An analysis of tens of thousands of economic forecasts by experts as well as a meta analysis of hundreds of experiments also shows no support for group accuracy increasing once a group exceeds a dozen members (Mannes, et al 2014). It therefore also follows that the fourth claim lacks empirical support: very large groups do not, as a rule, approach perfect accuracy.

In this paper, we re-evaluate the claims of the Condorcet Jury Theorem using alternative micro foundations. We abandon Condorcet's *generated signal* assumption, a construction in which each individual receives a signal generated by some process, in favor of an *interpreted signal* framework (Hong and Page 2009). [1] Within that framework, we find that none of the four claims necessarily hold, but, in agreement with empirical findings, that in expectation we should find much stronger support for the first two claims, that is groups outperform individuals, and more accurate individuals result in more accurate groups.

The interpreted signal framework assumes that people make predictions and classifications by constructing categories and using models. Thus, collective accuracy arises through function approximation. Each person constructs a classification function, and through voting, these are averaged to form a collective function. The group's accuracy depends on how

[1] See also Green and Stokey (unpublished).

well the collective function approximates the actual outcome function. The contrast between this logic and the standard logic is rather stark. The Condorcet Jury Theorem assumes independent or commonly correlated signals. Each vote is akin to a flip of a biased coin, with head more likely than tails. The more times the coin is flipped, the more likely the majority will be heads. That reasoning, though sound, relies on an assumption of independence that has little empirical support and, as we show later in the paper, lacks micro foundations.

The formal interpreted signal framework assumes a set of states of the world together with an outcome function that maps each state into one of two outcomes. In a jury trial, the state would consist of all knowable, germane information. In an investment decision regarding an initial public offering, the state would consist of information about the firm's profits and losses, its market potential, its management team, and its intellectual and physical capital.

Individuals construct signals (or classifications) by first partitioning the state space into categories and then assigning one of the two possible outcomes to each category. That assignation of outcomes can be thought of as a person's classification model. From the individual's standpoint, any two states in the same category are treated identically (Hong and Page 2009). Given this construction, an individual's accuracy corresponds to the difference between her model's classification of the state space and the outcome function's classification.

A group's classification can be represented similarly to that of an individual. The sets in the group's partition consist of the meet (the intersection) of its members partitions. The outcome the group assigns to a set in this partition corresponds to the majority outcome of the intersecting sets, and the group's accuracy corresponds to the proportion of states it classifies correctly. The collective classification equals the intersection of the individual classifications. If individuals in a group rely on distinct classifications, then the group partitions the world more finely, that is, with more granularity, than the individuals that comprise. That finer partitioning makes possible but does not guarantee increased group accuracy. As we will show, the group can even be less accurate than its members.

The fineness of the collective partition depends on two features of the individuals' parti-

tions: their fineness, that is their accuracy, and their diversity, that is the different ways they categorize the states. Thus, as noted above, in the interpreted signal framework, collective accuracy depends on both individual accuracy and collective diversity (Page 2008a, 2017).

That same logic, and not the probabilistic logic of the generated signals, explains the success of random forest algorithms and other ensemble methods used to make classifications (Dietterich 2000, Brieman 2001). Each classification tree in a random forest creates a coarse partition of the state space. Each is also accurate more than half the time. A forest (group) of classification trees voting on a classification outperforms each tree because the diverse classifiers create a finer categorization of the space of possibilities.

We do not mean to imply say that the standard model is not useful. It formalizes the insight that if people make different types of mistakes then groups will outperform individuals. We take exception to the assumption of independence of the signal. It hard codes a particular, and implausible, degree of diversity of errors. It also treats as identical any situations in which individuals have the same accuracy. Thus, a group of professional stock analysts forecasting whether a company's revenues will increase would be indistinguishable from a group of school children predicting whether a cheetah can outrun a lion, if in each group, individuals have the same probability of a correct classification, say 60%.

The conflation of these two cases matters for two reasons. First, for people to make predictions that even approximate independence, they must use diverse models or have different information (or both). A relatively large set of diverse sophisticated predictors exist for complicated, i.e. high dimensional tasks. They likely do not for low dimensional classifications made by less sophisticated predictors. Second, sophisticated people have more opportunity to improve accuracy through dialogue and strategic behavior than do unsophisticated people by virtue of the fineness of their respective classifications.

The remainder of the paper consists of six parts. In the first part, we provide a brief overview of the main implications of the Condorcet Jury Theorem. In the second part, we define classification tasks, describe the interpreted signal framework, and derive necessary

and sufficient conditions for a group to classify with perfect accuracy. In the third part, we re-evaluate the Condorcet Jury Theorem within the interpreted signal framework. We show that group accuracy requires a diversity of classification models, but that diversity is not sufficient. A group of individuals using diverse classifications can be less accurate than each member. In fact, for low to moderate levels of diversity, the maximal possible decrease in group accuracy equals the maximal possible increase. We also show that increasing individual accuracy has a monotonic effect (set wise) on that interval but increasing the group size enlarges the interval in both directions for low to moderate levels of diversity. Thus, our findings suggest less regularity in large groups than in small groups, in marked contrast to the results based on probabilistic foundations.

In the fourth part, we evaluate collective accuracy assuming that individuals drawn from a distribution. We again find that large groups need not be more accurate and that asymptotic collective accuracy can fail for any of four reasons: lack of model granularity, lack of diversity, outcome function randomness, and outcome function complexity. In the fifth part, we discuss the potential for deliberation and strategic behavior to overcome errors due to model interactions. In the sixth part, we conclude and discuss extensions.

The Condorcet Jury Theorem

For over two centuries, the Condorcet Jury Theorem has provided a core logical justification for relying on groups of people and collections of models to make classifications, decisions, and adjudicate truth. Proposed in 1785 by the French enlightenment philosopher and mathematician, Nicolas de Condorcet, and made formal by Laplace (1815), the theorem applies probability theory to show that collective classifications will be more accurate than those of individuals. The theorem considers voters who each receives a signal that is independently correct with the same probability. If that probability exceeds one-half, four results follow: the majority decides correctly with a higher probability than each individual, the group's

accuracy increases in the individual accuracy of the signals and in the group's size, and large groups approach but never achieve perfect accuracy. There exist multiple extensions, generalizations, and variations of the model (Grofman et al 1983).

Condorcet Jury Theorem: Given an outcome which is either true (T) or false (F) and $N = 2\tau + 1$ individuals, indexed by i, who receive signals, s_i, that satisfy **equal accuracy**: s_i is correct with probability $p > \frac{1}{2}$ and **independence**: the probabilities of s_i and s_j being correct are independent for all i and j, the following hold:

Group Accuracy: For all $\tau \geq 1$, a majority vote classifies correctly strictly more often than each individual.

Monotonicity in Accuracy: Collective accuracy strictly increases in individual accuracy, p, if and only if $p \in (0.5, 1)$.

Monotonicity in Group Size: Collective accuracy strictly increases in the group size, N.

Asymptotic Perfection: Accuracy approaches one as $N \to \infty$.

pf: See Ladha (1992) and Grofman, et al (1983).

Modeling individual votes or classifications as independently correct random variables oversimplifies how people make classifications. When evaluating Condorcet's modeling assumption, we must keep in mind that, at the time, he was working along a mathematical frontier. Only recently had Pascal and Fermat formalized the concept of probability.[2] Those assumptions enabled Condorcet to established one intuition for why groups make more accurate classifications than individuals. However, those assumptions imply a regularity of group accuracy and a benefit to increasing group size that do not hold generally.

[2]Pascal and Fermat were not thinking of how to design democratic institutions but instead engaged in the less lofty pursuit of formulating betting strategies for a dice game. Jacob Bernoulli later extended that framework to apply to an even broader class of games. The axiomatic basis of probability used today was put forth by Kolmogorov and did not appear until the mid twentieth century (Apostol 1969).

interpreted Signals and Perfect Accuracy

In this section we describe the *interpreted signal* framework of Hong and Page (2009) which assumes a classification task, Ψ, that consists of a set of states of the world and an outcome function G, that classifies each state as either true (T) or false (F).

*A **classification task**, Ψ, consists of an **outcome function**, G, defined over a set of **states of the world** Ω, where $G : \Omega \to \{T, F\}$*

It will be useful to characterize the outcome function by its **level sets**, $\Omega_\theta = \{x \in \Omega \mid G(x) = \theta\}$ for $\theta \in \{T, F\}$. A binary **interpreted signal** consists of an **interpretation** which consists of a partitioning of the states of the world, along with a **classification model** mapping each set in that partition into an outcome, either T or F.[3]

*Individual i's **interpretation** $\Phi_i = \{\phi_{i1}, \phi_{i2}, ..., \phi_{im}\}$ partitions Ω, i.e., $\phi_{ij} \cap \phi_{ik} = \varnothing$ and $\bigcup_{j=1}^{m} \phi_{ij} = \Omega$.*

For $x \in \Omega$, we let $\phi_i(x)$ denote the set in the partition Φ_i that contains x.

Individual i**'s classification model** $M_i : \Omega \to \{T, F\}$

We say that classification model, M_i, is *based on interpretation* $\Phi_i = \{\phi_{i1}, \phi_{i2}, ..., \phi_{im}\}$ if and only if $x, y \in \phi_{ij}$ implies $M_i(x) = M_i(y)$. In other words, the classification model must assign the same outcome to all states that belong to the same set in the interpretation. In the analysis that follows, we need not consider the full interpretation. It will suffice to consider **individual i's level sets**, $\Pi_i = \{\Pi_{iT}, \Pi_{iF}\}$ where $\Pi_{i\theta} = \{x \in \Omega \mid M_i(x) = \theta\}$ for $\theta \in \{T, F\}$

Given this construction, the *accuracy* of an individual's classification model corresponds to the proportion of states classified correctly. Classification errors could arise for a variety

[3]Notice that this construction resembles an information partition (Aumann 1976), where the the interpretation plays the role of an *information set*. The key difference is that in the interpreted signal framework, the state of the world does not include an outcome value in the partitions. Instead, two individuals could assign different values to the same state.

(a) An Outcome Function (b) Level Sets of a Classification Model

Figure 1: Classification Error: Interpretation Misalignment

of reasons. If a majority of the elements in a set in an individual's interpretation map to one outcome, but the classification model maps the set to the other outcome, we say that the classification model makes a *mistake*. We restrict attention here to *mistake free* classification models.

Classification errors can also arise from *misalignment* between the outcome function and the classification model. In these cases, the level sets of the interpretation fail to align with the level sets of the outcome function as shown in Figure 1. These misalignments could be due to having an incorrect model of how states get mapped into outcomes or relying on an interpretation that creates regions that overlap the level sets of the outcome function.

The next three causes of of errors—*lack of interpretive granularity*, *outcome function complexity*, and *randomness*—all stem from individual level sets that contain both types of outcomes as shown in Figure 2. The left diagram shows the outcome function where grey regions correspond to true states (x, s.t. $G(x) = T$) and white regions to false states, (x, s.t. $G(x) = F$). The right diagram shows a classification model based on an interpretation in which each level set Π_i consists of the two sets for $i = T, F$. As evident from the figure, the classification model misclassifies, or makes errors in each of the circular regions which are subsets of the level sets of the individual.

Figure 2 shows how a lack of interpretive granularity produces errors. The interpretation creates only four sets. Classifying the outcome function correctly would require eight sets. The individual would have to separate out a circular region from each of the four rectangular

sets in her partition. This lack of granularity could result from missing information or a lack of precision. In Figure 2 each state of the world can be represented as a two dimensional vector (x, y). An individual's interpretation could be based on two imprecise signals. One might tell if x is above or below its median value, and another might tell her whether y is above its median value. Given this signaling structure, her optimal classification model would be that showed on the right of the figure. As noted it is insufficiently granular to classify outcomes perfectly. This example is analogous to those found in Barelli, Bhattacharya, and Siga (2018), who derive conditions on the signal space that are sufficient for a large group to make an accurate classification. These conditions are analogous to those required for an individual to classify with perfect accuracy.

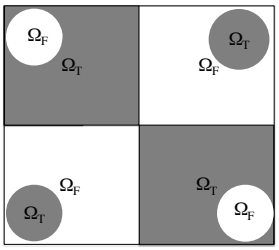
Insufficient granularity can also result from bounded awareness or inattention (Gabaix forthcoming). Bounded awareness occurs when people do not consider relevant information when making a decision. (Chugh and Bazerman 2007). People may ignore relevant information. They may fail to seek it out. Or, even if they have information, they may lack the cognitive repertoire to put the information to use when making a classification.

Rational inattention models derive lack of granularity by assuming individuals possess limited information capacity, expressed as a bound on information entropy (Sims 2003).[4] That constraint limits the number and size of the sets that can belong to an interpretation. If the level sets of the outcome function have an information content of C, and if an individual is constrained to having an interpretation with an information content $c < C$, then the individual's interpretation must lack sufficient granularity.

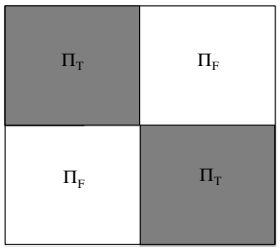
Figure 2 can also capture a complex function, where complexity refers to level sets of small measure such as Such fractal mappings, mappings that exhibit extreme sensitivity to initial conditions, and highly nonlinear functions.(Al-Najjar, Casadesus-Massanell, and Ozdenoren 2003).[5] In those cases, the outcome function's level sets consist of many small

[4]Tools from information theory can be used to measure differences in classifications as well as levels of sophistication. Given a probability distribution over the space of outcomes, the entropy of an interpretation, would loosely correspond to the sophistication of a model.

[5]This conception of complexity is nonstandard. See Prokopenko et al (2009), and Page (2008) for surveys



(a) An Outcome Function



(b) Level Sets of a Classification Model

Figure 2: Classification Error: Lack of Granularity, Model Complexity, or Randomness

sets or sets with complicated boundaries.

Finally, the outcome could depend on an unknown past event or on future random events. As an example, consider a Polya Process based on an urn containing red and blue balls. In each period, a ball is drawn from the urn and placed back in the urn along with a new ball of the same color. The state of the world would correspond to the initial number of red and blue balls. The outcome function might correspond to whether the urn contains a majority of red balls or blue balls ten periods into the future. Given a state of the world, the outcome can be written as a probability distribution. If the current state of the world consists of five red balls and two blue balls, a majority of red balls will be more likely in period ten, but with some probability a majority of blue balls could occur.

To make this case fit the figure, we can expand the state space to include the future draws. An interpretation cannot include outcomes of future draws, so, a set in the interpretation must contain outcomes of both types. Therefore, the situation, appears "as if" a lack of granularity holds, that the individual cannot distinguish between two states of the world that have different outcomes.

Outcomes in markets, elections, and sporting contests all include random components. By construction, an interpretation cannot distinguish between states of the world that differ with respect to future events, and, therefore, must classify some of those outcomes incorrectly.

To summarize, a classification model may make errors because of mistakes, misalignment,

of complexity measures.

lack of granularity, outcome function complexity, or randomness. The first of these corresponds to the individual selecting the wrong outcome. The second can be represented by the individual sets being of proper size and number but not quite matching reality. The final three cases can all be represented by sets in her interpretation containing sets within both level sets of the outcome function. The next claim states necessary and sufficient conditions for a mistake free classification model to be perfectly accurate.

Claim 1. *A mistake free classification model, M_i based on interpretation Φ_i, has perfect accuracy if and only if Φ_i refines the partition created by the level sets, $\{\Omega_T, \Omega_F\}$.*

The proof of the claim is straightforward. If each set in the partition lies entirely within a level set, and the model makes no classification errors, then the classification model is perfectly accurate. If the classification model makes a classification error, then by definition, the model makes errors. And, if the interpretation Φ_i does not refine the level set partition, then there must exist two states in some set $\phi_{ij} \in \Phi_i$ that have distinct outcomes.[6]

The previous claim implies that, all else equal, mistake free classification models based on finer interpretations cannot be less accurate. To make this intuition formal, we say that individual $\hat{i}$ is *more sophisticated* than individual i if the interpretation on which $\hat{i}$'s classification model is based refines the interpretation on which i's model is based.

Observation 1. *If individual $\hat{i}$ is more sophisticated than individual i, and both use mistake free classification models, then individual $\hat{i}$'s classification model cannot be less accurate than individual i's.*

Collective Accuracy

We now derive necessary and sufficient conditions for a group using majority voting to classify with perfect accuracy. Given $N = 2\tau + 1$ classification models, we define the *majority*

[6] Given that there exist only two outcomes, an individual also classifies a state of the world correctly if it makes both types of errors. If a state of the world belongs to a set in an individual's partition in which a majority of states produce the other outcome, and if the individual misclassifies that set, then the state will be classified correctly.

classification model $M_{\text{MAJ}}(x)$ for an $x \in \Omega$ as the classification of the majority of the models and denote its level sets by, $\Pi_{\text{MAJ},\theta}$ for $\theta \in \{T, F\}$. First, we need that the collection of all subsets created by finite intersections of the individuals' classification models' level sets refine the level set partition of the outcome function. Second, we need that majority voting is a mistake free classification model on the sets created by finite intersections. To state the formal claim, we first construct a σ-algebra based on the level sets.

The ***level set*** σ ***- algebra****,* $\sigma(\{\Pi_i\}_{i=1}^N)$ *consists of all finite intersections and unions of sets in* $\{\Pi_i\}_{i=1}^N$

The ***collective partition*** $\sigma(\{\Pi_i\}_{i\in N})$ *equals the partition comprised of the minimal nonempty sets produced by taking finite intersections of the sets in* $\{\Pi_i\}_{i\in N}$*.*

Given their classification models, the collective does not distinguish among states in any minimal nonempty set in the level set σ- algebra. Therefore, for an outcome function to be classifiable with perfect accuracy it must be *measurable* with respect to the level set σ-algebra, $\sigma(\{\Pi_i\}_{i\in N})$, i.e. any minimal nonempty set of $\sigma(\{\Pi_i\}_{i\in N})$ must lie within a level set of G. We introduce the following definition.

An outcome function G *is* ***collectively measurable*** *given a set of* N *level sets* $\{\Pi_i\}_{i=1}^N$ *if every minimal nonempty set of* $\sigma(\{\Pi_i\}_{i=1}^N)$ *lies entirely within one of the level sets of* G*.*

We can now state a necessary condition for a collective to classify with perfect accuracy. The proof is contained in the proof for Claim 3 in the Appendix.

Claim 2. *If a collection of individuals using majority voting classifies with perfect accuracy, then the outcome function* G *is collectively measurable*

The following corollary states that if the majority voting classifies with perfect accuracy, then the outcome function can be written as a function of binary strings of length N, where the value of the ith bit in the string corresponds to the level set of individual i. This construction will prove useful in future claims.

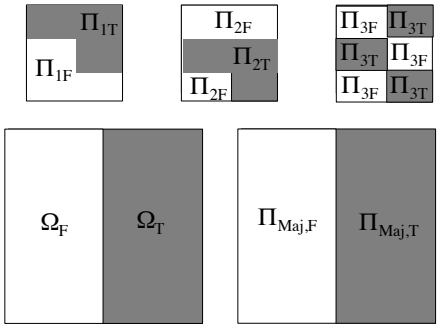


Figure 3: An Accurate Collective Classification

Corollary 1. *If a collective classifies an outcome function with perfect accuracy, then the outcome function G is* collectively measurable *with respect to the level set σ- algebra, $\sigma(\{\Pi_i\}_{i\in N})$, and there exists an $\hat{G} : \{T, F\}^N \to \{T, F\}$, such that*

$$\hat{G}(M_1(x), M_2(x), \ldots, M_N(x)) = G(x) \;\; \textit{for all} \;\; x \in \Omega$$

Figure 3 shows how a majority vote of individuals who each makes classification errors can be perfectly accurate. Notice that the intersections of the level sets of the individuals refine, that is, lie within the level sets of the outcome function. Formally, the outcome function is collectively measurable with respect to the level set σ-algebra.

To see the necessity of collective measurability consider the example shown in Figure 4. The outcome function is shown in the diagram in the bottom left. All states that lie further than distance $\sqrt{\frac{2}{\pi}}$ from the origin are true, i.e., take value T. All others are false, i.e., take value F. Given this construction, exactly half of the states have value T. The first individual's level sets, shown in the upper left, partition the unit square based on whether the state's x coordinate is less than one-half. The second individual's level sets partition the unit square based on the state's y coordinate is less than one-half. The third individual's level first partitions the unit square based on whether the sum of the x and y coordinates is less than one and then selects out a circle from each region.

The majority vote level sets partition the space into points whose x and y values sum

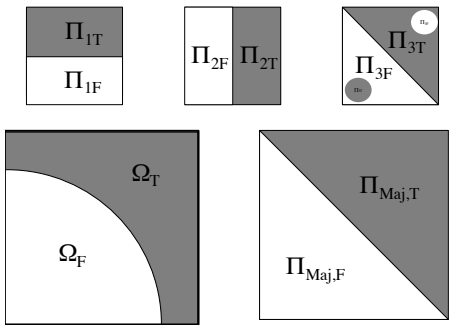


Figure 4: A Violation of Collective Measurability

to less than one, which it classifies as F, and those which sum to greater than one, which it classifies as T. Majority voting makes mistakes because it cannot classify states in the curved regions above and below the diagonal.

Our next example shows that measurability, though necessary, is not sufficient for the collective to classify with perfect accuracy. It is straightforward to verify that the outcome function satisfies collective measurability. However, majority rule classifies every state of the world as F. Therefore, on average, the majority of better than random classifiers is less accurate than its members. This is the opposite of the Condorcet Jury Theorem in which the majority must be more accurate than its members.

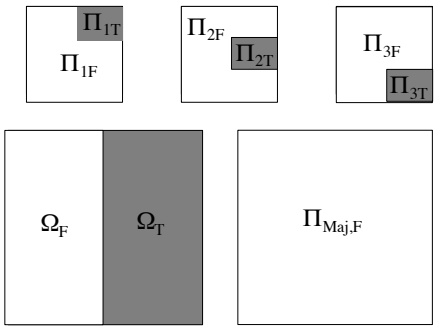


Figure 5: Majority Rule Classification Error With Collective Measurability

For sufficient conditions, we also need that the outcome function be representable as a *separable majority threshold function.*

The function $\hat{G}$: $\{T, F\}^N \rightarrow \{T, F\}$ *is a* ***separable majority threshold function*** *if and only if*

$$\hat{G}(s_1, s_2, \ldots, s_N) = T : \Longleftrightarrow : |\{i : s_i = T\}| > \tau$$

We can now state a necessary and sufficient condition for majority rule to produce perfectly accurate classifications.

Claim 3. Conditions for Perfect Accuracy: *A collection of* N *classification models collectively classifies the outcome function* G *with perfect accuracy if and only if* G *is collectively measurable with respect to* $\{\Pi_i\}_{i=1}^N$ *and there exists a separable majority threshold function,* $\hat{G} : \{T, F\}^N \rightarrow \{T, F\}$ *such that* $\hat{G}(M_1(x), M_2(x), \ldots, M_N(x)) = G(x)$ *for all* $x \in \Omega$

This formulation provides the key intuition for when a majority vote of classification models makes perfectly accurate classifications. We can think of first intersecting all of the level sets of the individual models to create a partition of the set of possible states. On each of these sets, we then assign an outcome corresponding to majority rule. That process creates a classification model, the *majority rule classification model.* That model must be identical to the outcome function.

To summarize, for a majority of non perfectly accurate classification models to be perfectly accurate, they must create different level sets, that is they must be based on *diverse interpretations.* In addition, the classification models based on those diverse interpretations, when intersected, must refine the level sets of the outcome function. Each individual may see a crude version of the outcome function and make mistakes but provided those crude versions intersect fortuitously, the majority could be perfectly accurate. Whereas, in the standard Condorcet Jury Theorem accuracy improves because of biased independent draws, in the reinterpreted version, accuracy depends on the intersections of the level sets of the individuals creating a partition that refines the outcome function's level sets.

These two logics differ markedly. The first relies on statistical reasoning. The second

relies on the logic of function approximation. As a result, the implications of the logics will also differ. In fact, as the next claim states, none of the main results from standard Condorcet Jury Theorem necessarily hold.

Classification Model Jury Theorem: *Given $N = 2\tau + 1$ individuals who apply classification models that are each accurate with probability $p > \frac{1}{2}$, none of the implications of the Condorcet Jury Theorem: group accuracy, monotonicity in accuracy, monotonicity in group size, and asymptotic accuracy need hold.*

Violations of group accuracy and asymptotic accuracy have already been shown. In Figure 5, the majority is less accurate than its members, and, in Figure 3, a group of three is perfectly accurate, violating asymptotic accuracy. We show violations of monotonicity in accuracy and group size in the next section. The next corollary states that it is even the case that an individual can be made more sophisticated, that is have a finer interpretation, and be more accurate, yet make the group less accurate.

Corollary 2. *It is possible to make an individual more sophisticated and strictly more accurate yet make the collective less accurate even if the individual applies a mistake free classification model.*

The Condorcet Jury Theorem with interpreted Signals

In this section, we reconstruct the Condorcet Jury Theorem assuming interpreted signals. First, we show what assumptions must be made on classifications in order for independence to hold. Second, we characterize the possible collective accuracy levels as a function of group size and diversity. That approach reveals a wide range of possible accuracies. In order to make the classification model framework commensurate with the Condorcet Jury Theorem, we make the following assumptions:

Equal State Probability: *uniformly distributed state space*

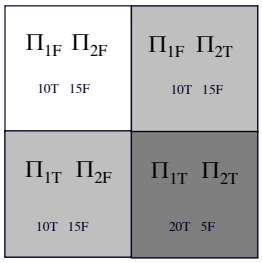


Figure 6: Independent Level Sets and Negatively Correlated Correct Classifications

Equal Outcome Probability: $\mid \Omega_0 \mid = \mid \Omega_1 \mid$

Equal Model Accuracy: $\frac{|\Pi_{i\theta} \cap \Omega_\theta|}{|\Omega_\theta|} = p > \frac{1}{2}$ *for all* i *and* θ

We first draw a distinction between *independent level sets* and *independently correct classifications*. Independent level sets exist if and only if the probability that a state of the world exists in level set T (resp F) for individual i is independent of the probability that it exists in level set T (resp F) for individual j. Figure 6 shows an example of independent level sets. Each of the four sets created by intersections of level sets contains one-fourth of the states.

In the example, each model classifies correctly with probability 0.6. Both models classify correctly with probability 0.35, which implies negatively correlated correct classifications. The next claim states that this holds generally. Thus, independent interpretations of the states of the world produces correlated, not independent correct classifications.

Claim 4. *Any two classification models with independent level sets produce negatively correlated correct classifications.*

pf. see Hong and Page (2009).

Recall that the Condorcet Jury Theorem assumes independently correct classifications: the probability that any one individual classifies correctly is independent of any other individual classifying correctly. It follows from the previous result that independently correct

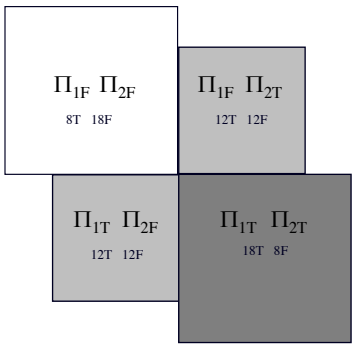


Figure 7: Independently Correct Classifications Imply Positively Correlated Level Sets

classifications must imply positively correlated level sets. Figure 7 shows an example. Each model classifies correctly with probability 0.6, and the two models make independently correct classifications. However, the level sets are not independent. The two classification models assign the same level sets with probability 0.52. In general, if we construct any two classification models satisfying our assumptions above, then their level sets must overlap with probability $p^2 + (1-p)^2$. That same condition must hold for each pair of classifications as we increase the group size. Independently correct classifications, therefore, require that the level sets of each new classification model intersects with the level sets of every other classification model with the exact same probability.

In fact, up to relabeling (homomorphism), Figure 7 shows the only possible example. All sets of N interpretations that satisfy the independently correct classification assumption (up to relabeling) take a common form: the states of the world must be expressible as an N dimensional space and each of the models must take into account a distinct N-1 dimensions. In other words, each person must neglect to see a distinct dimension. Moreover, if we add another two more individuals to the group, we have to assume that the state space can be reformulated as if it consists of $N + 2$ dimensions and each person considers a unique set of $N + 1$ of those dimensions (Hong and Page 2009). The assumption strains credibility, and it is not surprising at all that independence is not supported by data (Satopää, et al 2015).

Varying Diversity, Individual Accuracy, and Group Size

We now derive collective accuracy as a function of diversity, individual accuracy, and group size. We let *accuracy gain* denote any positive difference between the accuracy of the group and that of each individual and *accuracy loss* denote a negative difference. We first state a lemma that provides the central intuition for subsequent claims.

Lemma 1. *Given equal state probabilities, equal outcome probabilities and equal model accuracy* $p > \frac{1}{2}$*, given a group of* $N = 2\tau + 1$ *individuals. (a1) When individual accuracy is not too high,* $p < \frac{\tau+1}{N}$*, maximal group accuracy is reached when on some states, individual classifications are unaminously wrong and on all other states, margin of correct classifications is one. (a2) When individual accuracy is high,* $p \geq \frac{\tau+1}{N}$*, maximal group accuracy is reached when on all states, margin of correct classifications is at least one. (b) Minimum group accuracy is reached when on some states, classifications are unaminously correct and on all other states, margin of incorrect classifications is one.*

This lemma suggests a complicated relationship between unanimity and collective accuracy. Groups consisting of individuals with similar interpretations and predictive models often vote the same way. That means the group cannot be very accurate (unless the problem is simple). Conditional on unanimity, one must infer an increased likelihood of similar interpreted signals. That lowers confidence in the group. On the other hand, a two to one vote means that two people have disagreed. That means the third person is pivotal on a state of the world for which the outcome was relatively hard to classify. If it were not, the other two people would not have disagreed.

The inference of unanimity becomes much less complicated if the group makes multiple classifications. Consistent unanimity implies either a lack of diversity or simple problems. In many cases, the outcome of a vote rendering any determination of accuracy is impossible. We do not know if a defendant really was guilty or innocent or whether a policy would have been effective. In those contexts, unanimity should be viewed with suspicion.

We can use the previous lemma to derive explicit bounds on group accuracy. The first claim states that maximal accuracy equals one.

Claim 5. *Given equal state probabilities, equal outcome probabilities and equal model accuracy, if each of $N = 2\tau + 1$ individual is correct with probability p, then maximal group accuracy equals $p_N^{\max} = \min(\frac{pN}{\tau+1}, 1)$. In particular, if $p > \frac{\tau+1}{N}$, maximal group accuracy equals one.*

The second claim states that minimal accuracy decreases in group size and that groups comprised of less accurate individuals have lower minimal accuracy. The first of these results runs counter to the logic of the Condorcet Jury Theorem. The second result aligns with the logic that more accurate individuals comprise more accurate groups.

Claim 6. *Given equal state probabilities, equal outcome probabilities and equal model accuracy, if each of $N = 2\tau + 1$ individual is correct with probability p, the minimal group accuracy equals*

$$p_N^{min} = \frac{pN - \tau}{\tau + 1}$$

which decreases in N and converges to $p^{min*} = (2p - 1)$ as N becomes large.

This implies that a large group consisting of individuals each of whom is 51% accurate need only be 2% accurate, and not nearly 100% accurate as implied by the Condorcet Jury Theorem. To achieve 2% accuracy, the majority must be incorrect in 98% of cases and unanimously correct in the other 2%. Each individual will then be correct in approximately 50% of the incorrect votes (or 49% of all votes) and always correct in the remaining 2% of votes for an average of 51%.

Figure 8 gives an example of minimal group accuracy for three individuals who each classifies correctly with probability three-fifths. The group accuracy equals two-fifths. In this case, the majority rule classification model satisfies collective measurability. That must hold generally.

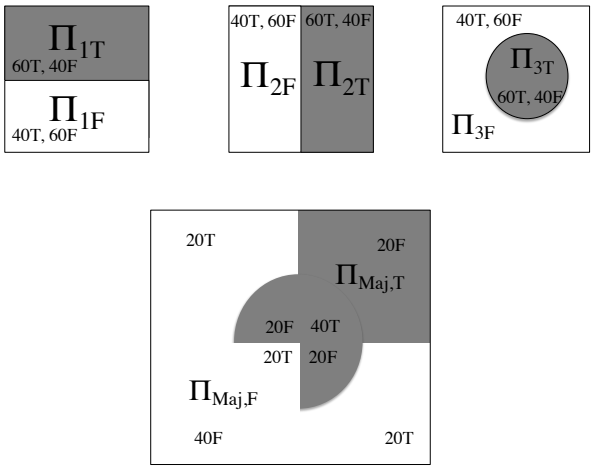


Figure 8: Worst Case Scenario: Individual Accuracy = 0.6; Group Accuracy = 0.4

Corollary 3. *Maximal accuracy loss or minimal group accuracy occurs when majority rule makes classification errors and the collective partition refines the outcome partition.*

Our next set of results characterize minimal and maximal accuracy as a function of the diversity of classification models. If each member of a group uses the same classification model, then the group is no more accurate than each person. If the classification models are maximally diverse, then, as in the standard results, the group will be more accurate than its members. But, if the classification models differ by small to moderate amounts, the group can either be more accurate or less accurate.

Our next claim characterizes maximal and minimal accuracy as a function of average disagreement. To state the claim, we first formally define the **disagreement** of two classifications i and j, $d(i,j)$ as the proportion of states of the world classified differently by individuals i and j. We let d denote the average pairwise disagreement for the group. The value of d can range from 0, when all classifications agree, to $D(p)$, which denotes maximal disagreement given individual accuracy p. With three individuals, $D(p) = 2(1-p)$ for $p \geq \frac{2}{3}$.

Claim 7. *Let* $N = 2\tau + 1$ *and* $\frac{\tau+1}{2\tau+1} \leq p \leq \frac{2\tau}{2\tau+1}$.[7] *Given equal state probabilities, equal*

[7]We focus on this region of individual accuracy p with the idea that the classification problem is difficult for a variety of reasons discussed earlier, but each classification model is strictly better than random.

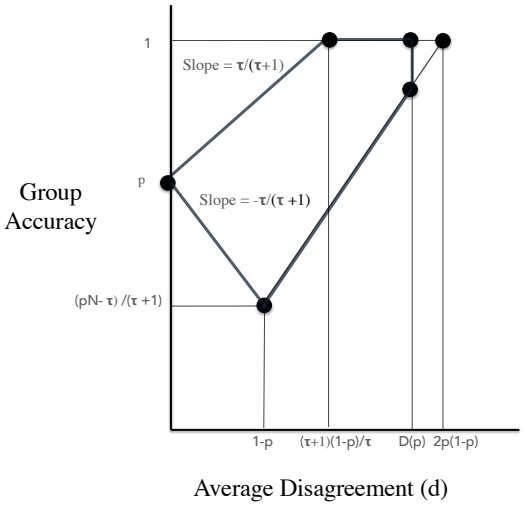


Figure 9: Collective Accuracy and Diversity

outcome probabilities and equal model accuracy, the set of possible average disagreement accuracy pairs (d, P)*, the* ***diversity-accuracy set****,* $DA(p, N)$*, equals the convex set defined by the following five points:*

$$\left\{(0,p), \left(1-p, \frac{pN-\tau}{\tau+1}\right), \left(\frac{(\tau+1)(1-p)}{\tau}, 1\right), (D(p), 1-\frac{(2(1-p)-D(p))N}{\tau+1}), (D(p),1)\right\}$$

where $D(p) = \frac{2(\tau-1)(1-p)}{\tau} + \frac{2}{\tau(2\tau+1)}$.

In the proof, we show that $\frac{(\tau+1)(1-p)}{\tau} < D(p) < 2(1-p)$ for any $\tau > 1$ (or group containing more than three individuals) and $\frac{\tau+1}{2\tau+1} < p < \frac{2\tau}{2\tau+1}$ so that the last three points are distinct points.[8] Figure 9 illustrates the claim.[9]

Figure 9 shows that, initially, increasing levels of disagreement results in less predictability in group classifications. Maximal accuracy gain and accuracy loss both increase. In addition,

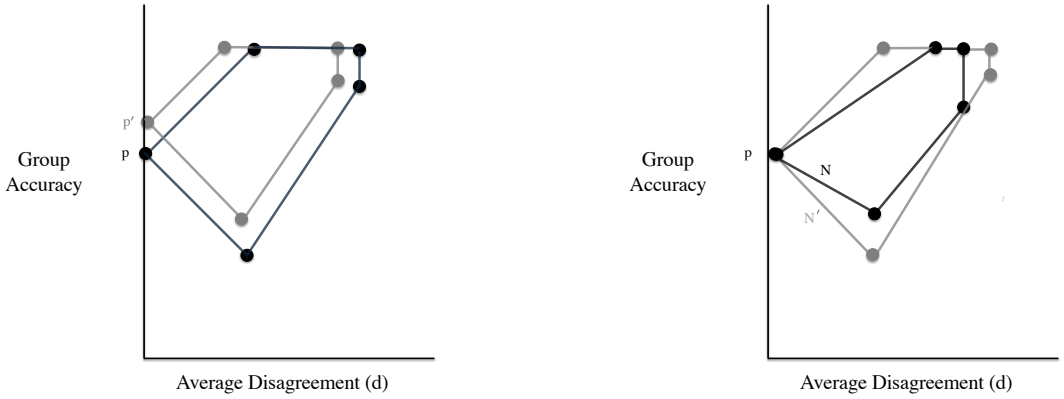
[8]When $\tau = 1$ (or group with three individuals), these values are equal so that the last three points collapse to one point $(2(1-p), 1)$.

[9]In the traditional Condorcet Jury Theorem (CJT), average vote correlation is given. This correlation implies more than a corresponding level of average disagreement d here. The group accuracy of the CJT is a value in the interval $(P_d^{\min}, P_d^{\max})$ which are the group accuracy bounds valued at d. In fact, the CJT states that the group accuracy lies within the interval $(p, P_d^{\max})$.

it shows that as disagreement increases, accuracy in group classifications becomes more predictable. However, at maximal diversity, perfect accuracy is not insured.

The above characterization of the diversity-accuracy set given an individual accuracy p and a group size N allows us to gauge how group accuracy is impacted by the individual accuracy and the group size which we state as formal corollaries below. Given a diversity-accuracy set $DA(p, N)$, we define the **possible accuracy set given** d, $DA_d(p, N)$, to consist of all collective accuracies P, such that the pair (d, P) belongs to $DA(p, N)$. Possible accuracy sets can be written as closed intervals $[P_d^{\min}(p, N), P_d^{\max}(p, N)]$.

The first corollary states that increasing individual accuracy results in monotonic improvements in the possible accuracy sets. This result aligns in spirit with the Condorcet Jury Theorem in which collective accuracy increases in individual accuracy. The result is shown graphically on the left hand side of Figure 10.



(a) Increasing Individual Accuracy ($p' > p$) (b) Increasing Group Size ($N' > N$)

Figure 10: Effect of Individual Accuracy and Group Size on Collective Accuracy

Corollary 4. *For any feasible level of disagreement, i.e.* $d > 0$ *such that* $DA_d(p, N) \neq \emptyset$*, the possible accuracy set monotonically improves in* p*, i.e.* $p' > p$*, then* $P_d^{\min}(p, N) < P_d^{\min}(p', N)$ *and* $P_d^{\max}(p, N) \leq P_d^{\max}(p', N)$*.*

The second corollary states that increasing group size results in a set wise increase in the possible accuracy set as shown on the right hand side of Figure 10.

Corollary 5. *For any feasible level of disagreement, i.e.* $d > 0$ *such that* $DA_d(p, N) \neq \emptyset$*, the possible accuracy set increases setwise in group size,* N*, i.e.* $N' > N$*, implies* $DA_d(p, N) \subset DA_d(p, N')$*.*

With independent generated signals, accuracy increases in group size. We obtain a different result with interpreted signals because increasing group size allows for more possible overlaps in the level sets creating the possibility of both greater or less accuracy.

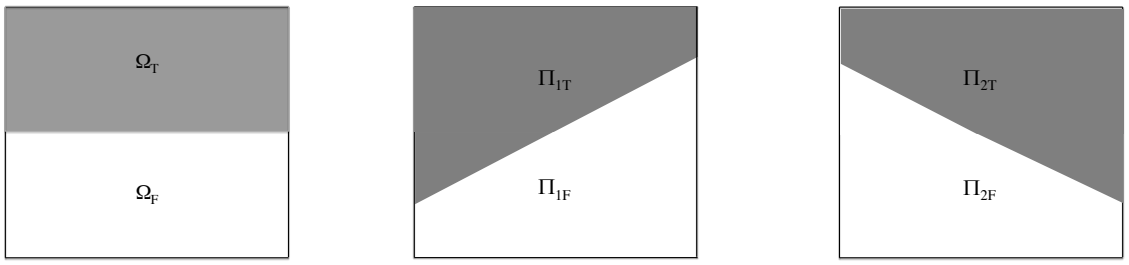
Collective Accuracy of Randomly Drawn Groups

The analysis so far considers majority voting outcomes for a given set of individual models. Alternatively, we can consider a group or jury as a draw from a population of types. Formally, we can define a distribution over a finite set of classification models. A group of size N can be modeled as a random draw of N classification models from that distribution. We can then calculate the likelihood that a group makes the correct classification.

Consider the outcome function shown at the left of Figure 11. All states of the world in the top half of the set are true, and all states in the bottom half are false. Assume that the population consists of two types of classification models and that each is equally likely. The first type, shown in the middle of Figure 11 classifies all states above an upward sloping line as true and all states below that line as false. The second type, shown on the right, classifies based on downward sloping line.

If we draw individuals from these two types with equal probability, then all individuals classify as true (false) any state in the region above (below) a bow tie region shown in Figure 12. Thus, any state outside of the bow tie region will be classified correctly by unanimous vote. In the bow tie region, a randomly selected individual is equally likely to be correct or incorrect depending upon where in the region it occurs and the individual's type.

In this example, a randomly drawn group with an odd number of members will classify outcomes outside the bow tie regions correctly but will classify outcomes inside the bow tie



(a) Outcome Function (b) Type 1 Classification Model (c) Type 2 Classification Model

Figure 11: An Outcome Function and Two Types of Classification Models

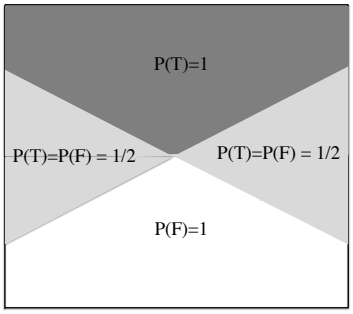


Figure 12: Probability a Random Individual Classifies Correctly Given Two Types

region correctly only half of the time. If we assume that the bow tie region has an area equal to one-fourth of the state space, then the group will classify correctly with probability seven-eighths. Increasing the jury size has no effect on the likelihood of a correct classification.

This example reveals three key features of the model when considering random draws of jurors. First, the relevant partitioning is the intersection of the interpretations of all possible types, what we defined as the collective partition. Second, within each set in the collective partition, there exists a distribution of outcomes. If within a set, all outcomes take the same value, that is, if a set lies within a level set of the outcome function, as in the regions above and below the bow tie region, then the group could be perfectly accurate in that set. If not, the group must make mistakes regardless of its size. Third, the classification models can differ on intersection of the interpretations. If some proportion, $p > 0.5$, classifies correctly, by that we mean that they assign the more likely outcome, then we can apply the standard

logic of the Condorcet Jury Theorem. Larger groups will be more likely to classify correctly on that particular set in the intersection.

The figure shows that errors can occur for three reasons. First, the collective partition may suffer from mismatch or lack of granularity. Second, even if a majority of the models classify a set in that partition correctly, a particular random draw may result in the wrong choice. Third, a majority of models could classify incorrectly on a set. In which case, making the group larger decreases accuracy on that set.

We now make these insights formal. Given an outcome function, we assume a population distributed across T types of classification models.[10] A group consists of a random draw of size $N = 2\tau + 1$ from the population of T models according to the population distribution. The group classification is determined by majority rule.

We introduce the following notation to clarify the presentation. We define the $T-$partition of Ω as $\Pi_{T,\Omega} = \{\Omega_1, \Omega_2, ..\Omega_S\}$ where each Ω_s corresponds to a unique intersection of the level sets of the T classification models.

Given the T-partition of Ω, we can define an *alignment profile*, $\vec{r} = (r_1, r_2, ..., r_S)$ where r_s equals the proportion of the more likely outcome in set Ω_s. We refer to one minus those probabilities, $((1 - r_1), (1 - r_2), ...(1 - r_S))$, as the *misalignment profile.* Perfect alignment, $r_s = 1$, means that every state in the set Ω_s has the same value. By construction, $r_s \geq \frac{1}{2}$.[11]

We first state a straightforward lemma.

Lemma 2. *Given a T-partition of Ω with a probability measure μ, and an alignment profile $\vec{r} = (r_1, r_2, ...r_S)$, the expected accuracy of a randomly drawn group of N voters cannot exceed the T-alignment, $A(\vec{r})$, which we define as follows:*

[10] We assume $T = 2M + 1$ is odd. When T is even, intuition and qualitative results here are the same but the formulation requires a slight modification.

[11] We can use the T-partition to show the severity of Condorcet's assumptions. If we let T go to infinity and assume that each classification model partitions the states into random sets each of which contains a proportion p of one type of outcome and a proportion $(1 - p)$ of the other, then we again fit Condorcet's assumption. The limiting T-partition will consist of singletons, so that the group approaches perfect accuracy.

$$A(\vec{r}) = \sum_{s=1}^{S} \mu(\Omega_s) r_s$$

The proof of the lemma goes as follows. By construction, the group cannot distinguish among the states in any Ω_s. Therefore, at best, the group classifies each set in the T-partition as the more likely outcome. This calculation establishes an upper bound because we have no guarantee that the group will classify correctly on each subset of the T-partition.

If we ascribe the non alignment to randomness, then the maximal possible alignment can be thought of as the predictability of the outcome function. The result of the lemma resonates with the *minimal causal state* model of Crutchfield and Shalizi (1999). The minimal causal states, a set of states that can produce the pattern in a time series minus the randomness, would correspond to the best possible T-partition. Relatedly, the maximal possible alignment would equal excess entropy which corresponds to non random component of total entropy in a time series (Prokopenko et al 2009).

To calculate the likelihood of a correct classification for a random jury requires more notation. We define the *accuracy profile* $\vec{q} = (q_1, q_2, ..., q_S)$, where q_s corresponds to the probability that a randomly drawn model agrees with the more likely outcome in set Ω_s. If the set Ω_s lies within a level set of the outcome function, then a correct classification will be perfectly accurate on that set.

If we assume each of the T models are equally likely to be drawn then each q_s can be written as $\frac{k}{T}$ where k equals the number of models that classify correctly. The accuracy profile probabilities operate like the probabilities in Condorcet Jury Theorem. If a particular q_s exceeds one-half, then as the group becomes larger it becomes more likely to make the correct classification of Ω_s with the important caveat that the set Ω_s may well contain both types of outcomes. We can now state the following claim.

Claim 8. *The expected accuracy of a group of N randomly drawn classification models using majority rule given an alignment profile $\vec{r}$ and an accuracy profile, $\vec{q}$, equals*

$$p_N^{MAJ} = \sum_{s=1}^{S} \mu(\Omega_s) \cdot \left\{ \left[\sum_{k=\tau+1}^{N} \binom{N}{k} q_s^k (1-q_s)^{N-k} \right] \cdot r_s + \left[\sum_{k=0}^{\tau} \binom{N}{k} q_s^k (1-q_s)^{N-k} \right] \cdot (1-r_s) \right\}$$

We can use the accuracy profile to classify sets in the T-partition as either *majority model correct* ($q_s > 1/2$) or *majority model incorrect* ($q_s < 1/2$). We can then state a straightforward corollary of the previous claim.

Corollary 6. *For large groups expected accuracy converges to the average alignment on majority model correct sets plus the average misalignment on majority model incorrect sets in the T-partition.*

$$lim_{N\to\infty} p_N^{MAJ} = \sum_{\{s\in S: q_s > \frac{1}{2}\}} \mu(\Omega_s) r_s + \sum_{\{s\in S: q_s < \frac{1}{2}\}} \mu(\Omega_s)(1-r_s)$$

The previous claim and corollary imply that larger groups need not be more accurate. Increasing the group size guarantees the correct classification on the majority model correct sets but it also guarantees incorrect classifications on the majority model incorrect sets (see Galesic et al 2018).

It remains to derive the worst and best possible cases. Here, we restrict attention to the cases where all of the classification models have identical accuracy denoted by p. By incorporating earlier results that maximal accuracy gain (loss) from a given set of models is achieved when all models are either unanimously incorrect (correct) or correct (incorrect) by a margin of one, we have the following results.

Claim 9. *Given $2M+1$ equally likely classification models with accuracy p, a lower bound on the accuracy of a randomly drawn group with $N = 2\tau + 1$ members is given by the following:*

$$\frac{p(2M+1) - M}{M+1} + \left[\sum_{k=\tau+1}^{N} \binom{N}{k} \left(\frac{M}{2M+1} \right)^k \left(\frac{M+1}{2M+1} \right)^{N-k} \right] \frac{(2M+1)(1-p)}{M+1}$$

The group becomes less accurate as its size increases and asymptotically converges to $\frac{p(2M+1)-M}{M+1}$, which is less than p. Minimal accuracy falls in the number of models and converges to $2p-1$ for M large.

The next claim describes an upper bound on collective accuracy.

Claim 10. *Given $2M+1$ equally likely classification models with accuracy p, an upper bound on the accuracy of a randomly drawn group with N members is given by the following:*

$$\left[\sum_{k=\tau+1}^{N} \binom{N}{k} \left(\frac{M+1}{2M+1}\right)^{k} \left(\frac{M}{2M+1}\right)^{N-k}\right] g^{*}$$

where $g^{*} = \min\{(\frac{2M+1}{M+1}p), 1\}$

The upper bound on group accuracy increases in group size and converges to g^{*}, which for a large number of models equals one. As expected, the upper bound on group accuracy increases with an increase in individual accuracy p for any group size.

Deliberation and Strategic Voting

Our analysis considers majority voting assuming a given set or distribution of classification models, i.e. interpreted signals. We have not considered the possibility of either deliberation or strategic voting. We take up deliberation first as that analysis will provide a foundation for considering strategic voting.

In the generated signal framework, deliberation involves sending signals about signals. (Feddersen and Austen-Smith 2006) In the interpreted signal framework, deliberation plays two very different functions: set identification and outcome re-classification. By the former, we mean that if each person in a group constructs an interpreted signal, then through deliberation the group could, with sufficient time and effort, differentiate among the sets in the finest collective partition. In more formal language, the group could construct the σ-algebra of the interpretations, as opposed to the coarser level σ-algebra of the sets used in

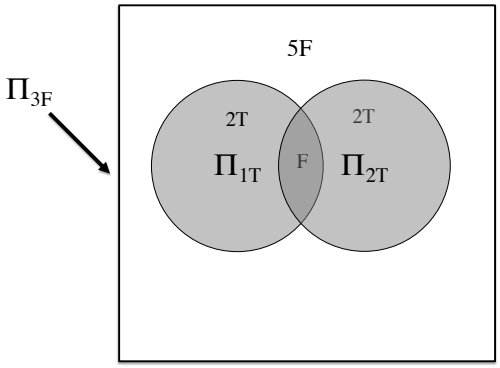


Figure 13: Potential for Refinement of Classification Models Through Deliberation

the proof above. We call this finer σ-algebra, the *collective interpretation*, $\mathcal{P}^*$

Deliberation improves outcome only if sets in the collective interpretation or the level set σ-algebra are re-classified. Figure 13 shows three classification models defined over ten states of the world, denoted by T's and F's. Person 3, whose level set we denote by Π_{3F} does not distinguish between any of the states. She classifies all as false and is correct with probability $\frac{6}{10}$. Person 1 has two level sets. She classifies as true anything in the circle on the left and classifies everything else as false. Person 2 classifies everything in the circle on the right as true and everything else as false. Both are correct with probability $\frac{7}{10}$.

Using majority rule, the three classify all states as false except for the single state denoted by F in the intersection of the two circles, which they classify as true. It follows that the group classifies correctly with probability one-half. Here, the group is less accurate than any of its members. Through deliberation (or repetition) the group could realize that the outcome is true if and only if either person 1 or person 2 (but not both) classifies the state as true. To arrive at that classification classification, the group would have to construct a new classification model, one different than the one created by majority rule (Landemore and Page 2015).

The group could, through deliberation, assign a new classification for each set in the collective interpretation. In the example, that results in perfect accuracy. That need not always hold. In general, the group's accuracy is bounded by the alignment of the finest collective

partition, $\mathcal{P}^*$. Deliberation cannot improve upon total alignment unless it could differentiate among states that every person in the group interprets as identical. Thus, alignment of the collective partition seems the appropriate upper bound. That upper bound might not be always be achieved. One can construct examples in which the correct classification of a set in $\mathcal{P}^*$ differs from the classifications of every set that intersects to form that set.

in the interpreted signal framework, strategic behavior can take two forms. People can vote strategically, and they can choose classification models strategically. One shot strategic voting calculations have the same general form as in the generated signal framework. An individual would need to have priors on the probability of being correct conditional on others being correct and many of the results from existing, generated signal models would hold. For example, a unanimity rule for conviction would create an incentive for jurors to vote guilty given the conditional probability of being pivotal (Feddersen and Pesendorfer 1996). The key difference would be that in the interpreted signal setting, the calculations would be more complicated given the lack of independence.

People could also learn to vote strategically based on past accuracy. Each person could base her votes on the sets in her interpretation as well as those of others. If voting strategically, given the state of the world is in a particular set and taking the votes of everyone else as fixed, a person classifies each outcome to maximize accuracy conditional on being pivotal. A Nash equilibrium would consist of each person making the classification in each set in her interpretation that is more likely to be correct given the person is pivotal. Thus, the analysis of strategic voting also relies on the finest collective interpretation, $\mathcal{P}^*$.

There exist examples in which no Nash equilibrium achieves an accuracy equal to the alignment of the finest collective interpretation. That is true for the example shown in Figure 13. Suppose that person 1 and person 2 vote sincerely. The optimal response for person 3 is to vote true rather than false. Those voting rules constitute a Nash Equilibrium and result in an accuracy of 90%. Given that Person 3 cannot condition her vote on the votes of the others, it follows that this is the most accurate possible Nash Equilibrium. Recall that from

above, deliberation could, conceivably, produce perfect accuracy.

In addition, if models differed in their popularity, then if an individual knew (or learned) that she had a more common model, she has an incentive to abstain or vote the opposite of her model so that models, rather than votes, receive more equal weight.

Strategic behavior could also lead to new interpretations. Here, we have taken the set or distribution of classification models as exogenous. Strategic choice of classifications could mimic the boosting algorithms used to improve random forests, a machine learning algorithm. Boosting algorithms search for decision trees (a type of classification model) that are correct when a majority classifies incorrectly (Brown, et al 2005, Liu and Yao 1999). Recall that to create the worst case scenarios, we maximized the proportion of states of the world misclassified by a single vote. By constructing two classification models that were correct on those states of the world would produce a dramatic improvement in accuracy.[12]

For individuals to construct the appropriately diverse classification models, i.e. models that classify correctly on close votes, they must have the right incentives. Incentives to classify accurately will not produce this type of classification diversity. In fact, several models show that incentives for accuracy lead to less diversity not more (Hong et al 2012, Economo et al 2016, Shpitzer and Ogburn 2016, Mann and Helbing 2017, Page 2018).

Discussion

We have shown that by reinterpreting the Condorcet Jury Theorem that none of its results hold in a strict sense. Groups can be less accurate than their members. Groups comprised of more accurate people need not be more accurate nor need large groups. And, finally, as groups become very large they do not tend toward perfect accuracy.

Instead, our results paint a more nuanced and empirically substantiated set of results.

[12]Recall the example in figure 8 in which 60% of the states are inaccurately classified by two to one votes, call this set S_I, and 40% are accurately classified with three zero votes, call this set S_A. Two new voters who correctly classify states in S_I and incorrectly classify in set S_A, will produce a group of size five that is perfectly accurate.

First, we find that diversity creates a double edged sword. Without interpretive diversity – different ways of slicing up the world or making classifications, groups cannot be more accurate. However, for low to moderate levels of diversity, the group could be worse as well as better. Only for high levels of diversity is increased group accuracy assured. The independence assumption of the generated signal framework just so happens to bake into the model enough diversity to imply increased accuracy. We would conclude that while democratic institutions need diversity (Allen et al 2017), diversity is not sufficient unless there exists a lot of it.

Our analysis also reveals a sharp difference between the effects of increasing individual accuracy and group size. The Condorcet Jury Theorem implies both universally good. We find that increasing individual accuracy results in a set wise monotonic improvement in collective accuracy. Thus, we should expect groups of more accurate people to be more accurate but we have not guarantees. Increasing group size does not produce a monotonic set wise improvement. Instead, it increases the set of possible accuracies for the better and the worse. Thus, contrary to the standard theorem, we find that increasing jury size increases uncertainty over outcomes without increasing expected accuracy.

Perhaps most important, we have presented an alternative explanation for collective accuracy. The generated signal framework explains group success through a probabilistic logic based on the law of large numbers: as we draw larger or more accurate sample, the majority is more likely to be correct. In the interpreted signal framework, collective accuracy depends on a logic of function approximation. sets. Each person lumps states of the world into sets or categories that she sees as similar. In creating those sets, she makes mistakes. If different people partition the world differently, they make different mistakes. The group's classification corresponds to the sum of those maps. The collective ,therefore, classifies a state correctly if a majority of people lump that state with states that have the correct outcome.

This more set theoretic explanation reveals how lack of diversity, lack of granularity,

outcome function complexity, and randomness can all cause collectives to misclassify. In doing so, the explanation obliges us to look at any given situation through a finer lens. In doing so, we see how group accuracy depends on the dimensionality of the classification task, the complexity of the outcome function, the randomness in outcomes, and the sophistication and diversity of the group members before jumping to any conclusions. We should therefore not expect the same assumptions to capture a jury rendering a verdict about the guilt or innocence of a defendant in a case with limited information, venture capitalists deciding whether to make an investment based on abundant market data, college professors voting on tenure given a thick file of papers and letters of recommendation, and a collection of randomly generated decision trees classifying photographs of cats.

To summarize, applying the interpreted signal framework to the Condorcet Jury Theorem offers an alternative explanation for why groups outperform individuals based on aggregations of functions and not biased generated signals. The analysis also demonstrates how collective accuracy requires classification diversity and that while increasing individual accuracy, should on average result in higher collective accuracy, increases in group size generally will not. The framework also provides new avenues for the analysis of deliberation and strategic behavior.

Proof of Claim 3: We prove necessity (sufficiency is obvious): if G is classified by a collection of individuals using majority voting classifications with perfect accuracy, then (a) G is collectively measurable with respect to $\{\Pi_i\}_{i=1}^N$ and (b) there exists a separable majority threshold function, $\hat{G} : \{T, F\}^N \to \{T, F\}$ such that $\hat{G}(M_1(x), M_2(x), \ldots, M_N(x)) = G(x)$ for all $x \in \Omega$. Suppose (a) does not hold. This means that there are at least two different states, x and y, such that even though each individual classifies x and y identically, i.e., $M_i(x) = M_i(y)$ for all i, and therefore x and y get the same classification from the majority, the outcome function G classifies x and y differently, i.e., $G(x) \neq G(y)$. Therefore, the collective makes a mistake on either state x or state y. A contradiction. We now prove (b) by constructing a binary function defined on binary strings of length N, $\hat{G} : \{T, F\}^N \to \{T, F\}$, with the necessary properties. For any binary string of length N, $(s_1, s_2, ..., s_N)$, that is a classification profile of individuals for a state, i.e., there exists a state x s.t., $s_i = M_i(x)$ for each $i \in \{1, 2, ..., N\}$, define $\hat{G}(s_1, s_2, ..., s_N) = G(x)$. For any other possible values of $(s_1, s_2, ..., s_N)$, define $\hat{G}(s_1, s_2, ..., s_N) = T$ if and only if $|\{i : s_i = T\}| > \tau$. By construction, for any $x \in \Omega$, $\hat{G}(M_1(x), M_2(x), \ldots, M_N(x)) = G(x)$. Further, since the majority classification is perfectly accurate, $\hat{G}$ here is clearly a separable majority threshold function.

Proof of Classification Model Jury Theorem: The potential failure of the first, third, and fourth results have already been shown in the examples in Figures 3 and 5. It is trivial to construct examples that violate monotonicity in accuracy. For example, if we construct a single classification model that is correct with probability $\frac{4}{5}$, then a collecton of three identical classification models will also classify correct with probability $\frac{4}{5}$. This is less accurate than the group of size three in the example in Figure 3.

Proof of Corollary 2: Proof is by example. Assume the level sets of the outcome function can be written as follows: $\Omega_F = \{S_{F1}, S_{F2}, S_{F3}\}$ and $\Omega_T = \{S_{T1}, S_{T2}, S_{T3}\}$ with each S_{ij} of equal size. Let $\Pi_{1F} = \{S_{F1}, S_{F2}, S_{F3}, S_{T1}, S_{T2}\}$ and $\Pi_{1T} = \{S_{T3}\}$. Let $\Pi_{2F} = \{S_{F1}, S_{F2}, S_{T3}\}$ and $\Pi_{2T} = \{S_{T1}, S_{T2}, S_{F3}\}$, and $\Pi_{3F} = \{S_{F1}, S_{F3}, S_{T3}\}$ and $\Pi_{3T} = \{S_{T1}, S_{T2}, S_{F2}\}$. Each individual classifies correctly with probability $\frac{4}{6}$. It follows that the majority rule classification model is correct with probabililty $\frac{5}{6}$. Its level sets can be written as follows: $\Pi_{\text{Maj},F} = \{S_{F1}, S_{F2}, S_{F3}, S_{T3}\}$ and $\Pi_{\text{Maj},T} = \{S_{T1}, S_{T2}\}$.

Suffices to show that we can make an individual more sophisticated and more accurate, yet make the majority less accurate. Suppose that individual 1's initial interpretation corresponds to her level sets. Define a new interpretation consisting of the sets $\phi_{11} = \{S_{F1}, S_{F2}\}$, $\phi_{12} = \{S_{F3}, S_{T1}, S_{T2}\}$ and $\phi_{13} = \{S_{T3}\}$. If individual 1's classification model is mistake free, then her new level sets will be $\hat{\Pi}_{1F} = \{S_{F1}, S_{F2}\}$ and $\hat{\Pi}_{1T} = \{S_{F3}, S_{T1}, S_{T2}, S_{T3}\}$. Individual 1 now classifies correctly with probability $\frac{5}{6}$. However, the majority rule classification model is now correct with probababilty $\frac{4}{6}$. Its level sets can be written as follows: $\Pi_{\text{Maj},F} = \{S_{F1}, S_{F2}, S_{T3}\}$ and $\Pi_{\text{Maj},T} = \{S_{F3}, S_{T1}, S_{T2}\}$.

Proof of Lemma 1: Given equal model accuracy assumption, we can consider only the states where the outcome is true. First, we prove (a1). Assume $p < \frac{\tau+1}{N}$. Assume that the group achieves maximal accuracy. Let j denote the number of correct votes by the group on true events given $N = 2\tau + 1$. The group is correct on these states if and only if $j \geq \tau + 1$. Let q_j denote the proportion of states in which there exist j correct votes. It suffices to show that $q_j = 0$ for all $j \notin \{0, \tau + 1\}$. Suppose that there exists a $0 < j \leq \tau$

such that $q_j > 0$. If the j individuals who vote correctly in those cases all vote incorrectly in those cases, the group accuracy would not change. Thus, we can reallocate those votes. Similarly, suppose that there exists a $\tau + 1 < j' \leq 2\tau + 1$, such that $q_{j'} > 0$. By the same logic, we can change $j' - (\tau + 1)$ correct votes and make them incorrect. We now have $\sum_{j=1}^{\tau} q_j \cdot j + \sum_{j'=\tau+2}^{2\tau+1} q_{j'} \cdot [j' - (\tau + 1)]$ that can be used to create a proportion $q*_{\tau+1}$ groups with $(\tau + 1)$ correct votes in $\sum_{j=0}^{\tau} q_j$, where $q*_{\tau+1}$ is given by the following equation.

$$q*_{\tau+1} = \frac{\sum_{j=1}^{\tau} q_j \cdot j + \sum_{j'=\tau+2}^{2\tau+1} q_{j'} \cdot [j' - (\tau + 1)]}{\tau + 1}$$

This can be done as long as we do not run out of space. When $p < \frac{\tau+1}{N}$, once we smooth over votes, q_0 is still positive, that is we do not run out of space. When votes are either unanimously incorrect or margin of winning votes is one, $p = \frac{\tau+1}{N}(1 - q_0)$. This process improves improve group accuracy. A contradiction. Now we prove (a2). When $p \geq \frac{\tau+1}{N}$, when all states support votes with winning margin of one, there are extra correct votes to be allocated because individual accuracy is high. We simply spread these extra votes around. The group is perfectly accurate in this case. The proof for (b), minimal accuracy, follows by a symmetric logic.

Proof of Claim 5: By (a2) of Lemma 1, when $p \geq \frac{\tau+1}{N}$, maximal group accuracy equals 1. We consider $p < \frac{\tau+1}{N}$ case. By Lemma 1, let $q_{\tau+1}$ be the proportion of events with $\tau + 1$ correct votes and q_0 equal the proportion of events with 0 correct votes. Given p equals the proportion of correct votes for each individual, the following equation holds:

$$p = \frac{q_{\tau+1}(\tau + 1)}{N}$$

The result follows.

Proof of Claim 6: Let q_τ be the proportion of events with τ correct votes and q_N (for unanimous) equal the proportion of events with N correct votes. Given p equals the proportion of correct votes for each individual, the following equation holds:

$$p = \frac{q_\tau \tau}{N} + q_N$$

By Lemma 1, $q_\tau = (1 - q_N)$, therefore

$$p = \frac{(1 - q_N)\tau + q_N N}{N}$$

Rearranging terms gives

$$q_N = \frac{pN - \tau}{\tau + 1}$$

Proof of Claim 7: We first describe the proof for the case of three voter. We characterize maximal accuracy as a function of d, the average disagreement as follows: At $d = 0$, $q_3 = p$

and $q_0 = (1-p)$. For $d \leq p$, by Lemma 1, to maximize the probability of a correct classification, we increase q_2 and decrease q_3 and q_0 as follows. Given an average disagreement of d states, we must create $\frac{3d}{2}$ states within such that each pair of voters agrees on $\frac{d}{2}$ states and disagree on d states. All $\frac{3d}{2}$ states are now classified correctly by 2 to 1 votes. So that $q_2 = \frac{3d}{2}$. To maintain the probability of a correct classification at p, $q_3 + \frac{2}{3}q_2 = p$. Combining equations gives $q_3 + d = p$. Therefore, $q_3 = p - d$ and $q_0 = (1 - p - \frac{d}{2})$. Therefore, accuracy equals $q_3 + q_2 = p - d + \frac{3d}{2} = p + \frac{d}{2}$. Thus, the distribution $(q_3, q_2, q_1, q_0) = (p-d, \frac{3d}{2}, 0, (1-p-\frac{d}{2}))$ maximizes the probability of a correct classification for $d \leq p$. For $p \geq \frac{2}{3}$, $D(p) = 2(1-p) < p$. So maximal collective accuracy equals $p + \frac{d}{2}$ for any $d \in [0, D(p)]$. At the maximal disagreement $D(p)$, collective accuracy equals 1.

To characterize minimal accuracy as a function of d, we follow a similar approach. By Lemma 1, to minimize the probability of a correct classification, as a function of d for $d \leq (1-p)$, we increase q_1 and decrease q_3. Given disagreement d, $q_1 = \frac{3d}{2}$. To maintain the probability of a correct classification at p, $q_3 + \frac{q_1}{3} = p$. Combining equations gives $q_3 + \frac{d}{2} = p$. Thus, accuracy equals $p - \frac{d}{2}$ for any distribution of the form $(q_3, q_2, q_1, q_0) = (p - \frac{d}{2}, 0, \frac{3d}{2}, 1 - p - d)$. For $d > (1-p)$, to minimize accuracy, we consider distributions of the following form $(q_3, q_2, q_1, 0)$, where $q_3 + \frac{2}{3}q_2 + \frac{1}{3}q_1 = p$ and $q_2 + q_1 = \frac{3d}{2}$. It follows that $q_3 = 1 - \frac{3d}{2}$ and that $1 - \frac{3d}{2} + \frac{2}{3}q_2 + \frac{1}{3}(\frac{3d}{2} - q_2) = p$. The later equation can be written $1 - d + \frac{1}{3}q_2 = p$. Thus $q_2 = 3(p + d - 1)$, so accuracy equals $1 - \frac{3d}{2} + 3p + 3d - 3 = 3p - 2 + \frac{3d}{2}$ for $d > (1-p)$. Note that at $D(p)$, both minimal and maximal accuracy equal to 1.

Generalization of Claim 7 for 2τ+1 Voters: We first prove that for a given $N = 2\tau + 1$ and a given $\frac{\tau+1}{2\tau+1} \leq p \leq \frac{2\tau}{2\tau+1}$, the maximal possible average disagreement $D(p) = \frac{2(\tau-1)(1-p)}{\tau} + \frac{2}{\tau(2\tau+1)}$. Assume symmetry across voters. Let y denote the proportion of states where a given pair of voters are incorrect. The maximal disagreement is reached when there are no states on which more than two voters are incorrect and y is as small as possible. Since $(2\tau+1)(1-p-2\tau y) + \binom{2\tau+1}{2} y \leq 1$,we get $y \geq \frac{1}{\tau}(\frac{2\tau}{2\tau+1} - p) \geq 0$ when $p \leq \frac{2\tau}{2\tau+1}$. Since here by definition $d = 2(1-p-y)$, $d \leq 2(1 - p - \frac{1}{\tau}(\frac{2\tau}{2\tau+1} - p)) = \frac{2(\tau-1)(1-p)}{\tau} + \frac{2}{\tau(2\tau+1)}$. Thus, for $\frac{\tau+1}{2\tau+1} \leq p \leq \frac{2\tau}{2\tau+1}$, $D(p) = \frac{2(\tau-1)(1-p)}{\tau} + \frac{2}{\tau(2\tau+1)}$.

We now show that $\frac{(\tau+1)(1-p)}{\tau} < D(p) < 2(1-p)$ for any $\tau > 1$ and $\frac{\tau+1}{2\tau+1} < p < \frac{2\tau}{2\tau+1}$. Note that $D(p)$ can be equivalently written as $D(p) = 2(1-p) - \frac{2}{\tau}(\frac{2\tau}{2\tau+1} - p)$. When $p < \frac{2\tau}{2\tau+1}$, the second term being subtracted is positive and therefore $D(p) < 2(1-p)$. Proving $\frac{(\tau+1)(1-p)}{\tau} < D(p)$ is equivalent to proving $\frac{\tau-3}{\tau}(1-p) + \frac{2}{\tau(2\tau+1)} > 0$. This is obviously true when $\tau \geq 3$. For $\tau = 2$, we need $\frac{1-p}{2} < \frac{1}{5}$.This is true because when $\tau = 2$, $p > \frac{\tau+1}{2\tau+1}$ means $p > \frac{3}{5}$.

Now we characterize maximal accuracy as a function of d, the average disagreement. At $d = 0$, $q_{2\tau+1} = p$ and $q_0 = (1-p)$. By Lemma 1, to maximize the probability of a correct classification, we increase $q_{\tau+1}$ and decrease $q_{2\tau+1}$ and q_0.

First, we note that given $2\tau + 1$ voters, there exist $\binom{2\tau+1}{\tau+1}$ distinct simple majorities. The construction relies on creating equal percentages of outcomes spread across all of these outcomes in $q_{\tau+1}$. The number of simple majorities in which any two voters disagree and the first voter is in the majority equals $\binom{2\tau-1}{\tau}$. This also equals the number of simple majorities in which any two voters disagree and the first voter is in the minority. Therefore, the probability of any two voters disagreeing in these simple majorities equals

$$\frac{2(2\tau-1)!}{\tau!(\tau-1)!}\frac{(\tau+1)!\tau!}{(2\tau+1)!} :=: \frac{(\tau+1)}{(2\tau+1)}$$

Thus, if we create $\frac{(2\tau+1)d}{\tau+1}$ states in this way, each pair of voters agrees on $\frac{\tau d}{\tau+1}$ states and disagrees on d states, making average disagreement equal to d. All $\frac{(2\tau+1)d}{\tau+1}$ states are now classified correctly. It follows that $q_{\tau+1} = \frac{(2\tau+1)d}{\tau+1}$. To maintain the probability of a correct classification at p, $q_{2\tau+1} + \frac{\tau+1}{2\tau+1}q_{N+1} = p$. Combining equations gives $q_{2\tau+1} + d = p$. Therefore, $q_{2\tau+1} = p - d$ and accuracy is given by the following expression: $q_{2\tau+1} + q_{\tau+1} = p - d + \frac{(2\tau+1)d}{\tau+1} = p + \frac{\tau d}{\tau+1}$. So for any $d \in [0, \frac{\tau+1}{\tau}]$, maximal accuracy increases linearly in d with a slope of $\frac{\tau}{\tau+1}$. For $d \in [\frac{\tau}{\tau+1}, D(p)]$, maximal accuracy equals 1.

To characterize minimal accuracy as a function of d, we first consider $d \in [0, 1-p]$. By Lemma 1, minimizing the probability of a correct classification, we increase q_τ and decrease $q_{2\tau+1}$ and q_0. Using a parallel argument as the above, we get $q_\tau = \frac{(2\tau+1)d}{\tau+1}$, $q_{2\tau+1} = p - \frac{\tau d}{\tau+1}$, and $q_0 = 1 - p - d$. Minimal accuracy is equal to $q_{2\tau+1} = p - \frac{\tau d}{\tau+1}$ which decreases linearly with a slope of $-\frac{\tau}{\tau+1}$. This works as long as $d \leq 1-p$. If $d > 1-p$, similar to three individual case, minimizing the probability of a correct classification involves making $q_j = 0$ for all $j \notin \{2\tau+1, 2\tau, \tau\}$. We then have $d = \frac{\tau+1}{2\tau+1}q_\tau + \frac{2}{2\tau+1}q_{2\tau}$, $p = q_{2\tau+1} + \frac{2\tau}{2\tau+1}q_{2\tau} + \frac{\tau}{2\tau+1}q_\tau$, and $q_{2\tau+1} + q_{2\tau} + q_\tau = 1$. They allow us to solve for $q_{2\tau+1}, q_{2\tau}$, and q_τ as a function of d. In particular, $q_\tau = \frac{2\tau+1}{\tau+1}[2(1-p)-d]$. minimal accuracy $q_{2\tau+1}+q_{2\tau} = 1-q_\tau = 1+\frac{2\tau+1}{\tau+1}[d-2(1-p)]$. Thus for $d \in [1-p, D(p)]$, minimum accuracy increases linearly in d with a slope of $\frac{2\tau+1}{\tau+1}$, but does not attain 1 for any $\tau > 1$ and $\frac{\tau+1}{2\tau+1} < p < \frac{2\tau}{2\tau+1}$ since $D(p) < 2(1-p)$ as we proved earlier.

Proof of Claim 8: For any Ω_s, the probability of a randomly drawn N member group containing k models agreeing with the more likely outcome in set Ω_s and $N-k$ models agreeing with the less likely outcome in set Ω_s is equal to $\binom{N}{k}(q_s)^k(1-q_s)^{N-k}$. If $k > \tau$, majority will classify all states in set Ω_s to have the more likely outcome in set Ω_s so that majority is correct on r_s proportion of the states in Ω_s. Similarly, if $k \leq \tau$, majority will classify all states in set Ω_s to have the less likely outcome in set Ω_s so that majority is correct on $1-r_s$ proportion of the states in Ω_s.The result follows.

Proof of Claim 9: From the proof for Claim 6, to achieve minimal accuracy, $q_{2M+1} = \frac{(2M+1)p-M}{M+1}$, $q_M = \frac{(2M+1)(1-p)}{M+1}$ and all others equal to 0. On each state in Π_δ where $M+1$ models are incorrect by a margin of one, any randomly selected individual (assuming all models are equally likely) will be correct with probability $\frac{M}{2M+1} < \frac{1}{2}$. It follows that for a given such state, a group of N members drawn randomly from the population, will be correct with probability

$$\left[\sum_{k=\tau+1}^{N}\binom{N}{k}\left(\frac{M}{2M+1}\right)^k\left(\frac{M+1}{2M+1}\right)^{N-k}\right]$$

So the accuracy of a randomly drawn group with $N = 2\tau+1$ members is given by the

following:

$$\frac{p(2M+1)-M}{M+1}+\left[\sum_{k=\tau+1}^{N}\binom{N}{k}\left(\frac{M}{2M+1}\right)^{k}\left(\frac{M+1}{2M+1}\right)^{N-k}\right]\frac{(2M+1)(1-p)}{M+1}$$

Proof of claim 10: We consider the case where $p<\frac{M+1}{2M+1}$. Parallel to the above case and from the proof of Claim 5, to achieve maximal accuracy, $q_{M+1}=\frac{(2M+1)p}{M+1}$, $q_0=1-\frac{(2M+1)p}{M+1}$ and all others equal to 0. On each state in Π_δ where $M+1$ models are correct by a margin of one, any randomly selected individual (assuming all models are equally likely) will be correct with probability $\frac{M+1}{2M+1}>\frac{1}{2}$. It follows that for a given such state, a group of N members drawn randomly from the population, will be correct with probability

$$\left[\sum_{k=\tau+1}^{N}\binom{N}{k}\left(\frac{M+1}{2M+1}\right)^{k}\left(\frac{M}{2M+1}\right)^{N-k}\right]$$

So the accuracy of a randomly drawn group with $N=2\tau+1$ members is given by the following:

$$\left[\sum_{k=\tau+1}^{N}\binom{N}{k}\left(\frac{M+1}{2M+1}\right)^{k}\left(\frac{M}{2M+1}\right)^{N-k}\right]\frac{(2M+1)p}{M+1}$$

References

Allen, Danielle, Henry Farrell, and Cosma Shalizi (2017) "Evolutionary Theory and Endogenous Institutional Change." working paper.

Al-Najjar N, R. Casadesus-Masanell R, and E. Ozdenoren (2003). "Probabilistic representation of complexity" *Journal of Economic Theory*Vol 111:1 p 49-8.

Apostol, Thomas M. (1969) *Calculus, Volume II (2nd edition).* John Wiley & Sons, New York, NY.

Aumann, Robert (1976) "Agreeing to Disagree." *Annals of Statistics.* 4(6): 1236-1239.

Austen-Smith, David and Jeffrey S. Banks (1996) "Information Aggregation, Rationality, and The Condorcet Jury Theorem" *The American Political Science Review* 90: 34-45.

Barelli, Paulo, Sourav Bhattacharya, and Lucas Siga (2018) "On the Possibility of Information Aggregation in Large Elections." *unpublished manuscript.*

Berg, Sven (1993) "Condorcet's jury theorem, dependency among jurors." *Social Choice and Welfare.* 10(1): 87-95.

Borgers, Tilman, Angel Hernando-Veciana and Daniel Krahmer (2013) "When are Signals Complements or Substitutes?" *Journal of Economic Theory* 148: 165-195

Breiman, Leo (2001). "Random Forests". *Machine Learning* 45 (1): 5-32.

Brown, G., J. Wyatt, R. Harris, R. and X. Yao (2005) "Diversity creation methods: a survey and categorisation." *Information Fusion*, 6(1), pp.5-20.

Chugh, D, and M. Bazerman (2007) "Bounded Awareness: What You Fail to See Can Hurt You." *Mind &* Society 6(1): 1?18.

Crutchfield, J.P. and C. R. Shalizi, (1999) "Thermodynamic Depth of Causal States: Objective Complexity via Minimal Representations", *Physical Review E.* 59:275-283.

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In J. Kittler and F. Roli (Ed.) *First International Workshop on Multiple Classifier Systems* pp. 1-15. New York: Springer Verlag.

Economo, E, L. Hong and S.E. Page (2016) " Social Structure, Endogenous Diversity, and Collective Accuracy." *Journal of Economic Behavior and Organization* 125: 212-236.

Feddersen T. and D. Austen-Smith (2006) "Deliberation, Preference Uncertainty, and Voting Rules." *American Political Science Review.* 100: 209-217.

Feddersen, T. and W. Pesendorfer (1996) "The Swing Voters' Curse", *American Economic Review.* 86: 408-24.

Gabaix, Xavier (forthcoming) "Behavioral Inattention", in *Handbook of Behavioral Economics*, edited by Douglas Bernheim, Stefano DellaVigna and David Laibson, Elsevier.

Galesic, M., Barkoczi, D., and Katsikopoulos, K. (2018). "Smaller crowds outperform larger crowds and individuals in realistic task conditions." *Decision.* 5(1):1-15.

Gilboa, I. and D. Schmeidler. (1995). "Case-based decision theory", *The Quarterly Journal of Economics*,110: 605-639.

Goldstein, D.G., R. Preston McAfee, and Siddharth Suri (2014) "The Wisdom of Smaller, Smarter Crowds." *Proceedings of the 15th ACM Conference on Economics and Computation.*

Green, J. and N. Stokey (unpublished)"Two Representations of Information Structures and Their Comparisons." `http://www.people.hbs.edu/jgreen/two_reps_info_structures_comparisons.pdf`

Grofman, Bernard, Guillermo Owen, and Scott L. Feld. (1983). "Thirteen Theorems in Search of the Truth." *Theory and Decision* Volume 15 (3): 261-278.

Hong, Lu, and Scott E Page. (2009) "Interpreted and Generated Signals" *Journal of Economic Theory* 144: 2174-2196.

Hong, Lu, Maria Riolo, and Scott E. Page (2012) "Incentives, Information, and Emergent Collective Accuracy" *Managerial and Decision Economics* 33:5 pp 323-334.

Kagel, John, and Alvin E. Roth, eds. (2016) *The Handbook of Experimental Economics, Volume 2*, Princeton University Press, Princeton, NJ.

Kleinberg, Jon, Jens Ludwig, Sendhil Mullainathan, and Ziad Obermeyer (2015). "Prediction Policy Problems." *American Economic Review: Papers & Proceedings*, 105(5): 491-495.

Ladha, Krishna (1992). "The Condorcet Jury Theorem, Free Speech, and Correlated Votes". *American Journal of Political Science*: 36 (3): 617-634.

Landemore, Helene and Scott E. Page (2015) "Deliberation and Disagreement: Problem solving, Prediction, and Positive Dissensus." *Politics, Philosophy, and Economics.*

Laplace,P.S. de (1815),*Theorie analytique des probabilities* Paris.

List, Christian, and Goodin, R.E. (2001). "Epistemic democracy: Generalizing the Condorcet Jury Theorm." *Journal of Political Philosophy* 9(3): 277-301.

List, Christian and Spiekermann, Kai (2016) "The Condorcet jury theorem and voter-specific truth."In: McLaughlin, Brian P. and Kornblith, Hilary, (eds.) *Goldman and His Critics. Philosophers and their Critics.* Wiley-Blackwell, Hoboken, USA. pp 219-234.

Liu, Y. and X. Yao, (1999). "Ensemble learning via negative correlation." *Neural Networks*12:(10):1399-1404.

Mannes, Albert E., Jack B. Soll, and Richard P. Larrick. (2014). "The Wisdom of Select Crowds." *Journal of Personality and Social Psychology* 107: 276-299.

Martin, Travis, Travis Martin, Jake Hofman, Amit Sharma, Ashton Anderson, Duncan Watts (2016) "Exploring limits to prediction in complex social systems." *International World Wide*

Web Conference.

McLennan, A. (1998) "Consequences of the Condorcet Jury Theorem for beneficial information aggregation by rational agents." *American Political Science Review* 92: 413-419.

McLean, lain, and Fiona Hewitt. (1994) *Condorcet: Foundations of Social Choice and Political Theory.* Elgar. Brookfield, VT.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L and Tetlock, P. (2015). "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions." *Perspectives on Psychological Science.* 10(3):267-281.

Miller, Nicholas (1996) "Information, individual errors, and collective performance: Empirical evidence on the Condorcet Jury Theorem." *Group Decision and Negotiation* 5(3): 211-228.

Page, Scott E. (2006), "Essay: Path Dependence," *Quarterly Journal of Political Science*, 1: 87-115.

Page, Scott E. (2007), *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton University Press.

Page, Scott E (2008) *Diversity and Complexity*, Princeton University Press, Princeton, NJ.

Page, Scott E (2008b) "Uncertainty, Difficulty, and Complexity" *Journal of Theoretical Politics*, Vol. 20: pp. 115-149.

Page, Scott E (2018) *The Model Thinker* Basic Books. New York, NY.

Palfrey, Thomas R. (2009) "Laboratory Experiments in Political Economy." *Annual Review of Political Science.*12:(1):379-388.

Prokopenko, Mikhail, Fabio Boshectti, and Alex Ryan (2009) "An information-Theoretic Primer on Complexity, Self-Organization, and Emergene," *Complexity* 15(1):11-28.

Sah, Raaj Kumar and Joseph E. Stiglitz (1986) "The Architecture of Economic Systems: Hierarchies and Polyarchies" *The American Economic Review* 76(4):716-727.

Satopää, Ville, Shane T. Jensen, Robin Pemantle, and Lyle H. Ungar (2015) "Partial Information Framework: Model-Based Aggregation of Estimates from Diverse Information Sources" arXiv:1505.06472.

Sharper, Shannon (2017) "How many interviews does it take to hire a Googler?" https://rework.withgoogle.com/blog/google-rule-of-four/

Shpitser, Ilya and Elizabeth L. Ogburn "Collective problem solving, causal inference, and chain graphs" working paper.

Sims, Christopher (2003) "Implications of Rational Inattention" *Journal of Monetary Economics.* 50(3):665-690.

Stone, Peter (2015) "Introducing difference into the Condorcet jury theore." *Theory and Decision*. 78(3): 399-409.