# Some Microfoundations of Collective Wisdom

Lu Hong and Scott E Page

July 22, 2011

## Abstract

Collective wisdom refers to the ability of a population or group of individuals to make an accurate forecast of a future outcome or an accurate characterization of a current outcome. Using basic statistical arguments, it can be shown that the collective will always be more accurate than it's average member, and that in some circumstances, the collective can be more accurate than any of its members. And yet, collective wisdom need not emerge in all situations. Crowds can be unwise as well as prescient. In this paper, we unpack what underpins and what undermines collective wisdom using two models. The first is the standard statistical model of aggregated signals. The second is a model of agents who possess predictive models. This second approach builds upon and extends traditional statistical approaches to characterizing collective wisdom by demonstrating how collective accuracy requires either individual sophistication/expertise or collective diversity. A lack of both characteristics necessarily leads to breakdowns in collective wisdom.

In describing the benefits of democracy, Aristotle observed that when individuals see distinct parts of the whole, the collective appraisal can surpass that of individuals. Centuries later, von Hayek in describing the role of information in decentralized markets made a related argument that suggested the market can accurately determine prices even if the average person in the market cannot (von Hayek 1945). To be sure, institutional structures such as democracies and markets rests substantially on the emergence of collective wisdom. Without a general tendency for groups of people to make reasonable appraisals and decisions, democracy would be doomed. The success of democracies, and for that matter markets, provides broad stroke support that at least some collective wisdom exists in the aggregate. Abundant anecdotal and small to large scale empirical examples also suggest the potential for a highly accurate collective forecasts, a so called ”wisdom of crowds” (Suroweicki 2004).

Collective wisdom, as we shall define it here, exists when the crowd outperforms the individuals that comprise it at a predictive task. This is a restrictive notion. Wisdom has a broader conception than mere forecast accuracy. A society deliberating on laws or common purpose must exercise wisdom in judgement. That task is much richer and nuanced then estimating the value of a stock or the weight of a steer. And yet, if we characterize wisdom in these contexts as anticipating multiple implications and interactions, then wisdom can be seen as the ability to forecast the existence and magnitude of multiple effects, and our conception of collective wisdom becomes less circumscribed.

The statistical foundations for collective wisdom are well established. A straightforward mathematical calculation demonstrates that the average prediction of a crowd *always* outperforms the crowd's average member (Page 2007). Second, this same calculation implies that with some regularity, crowds can outperform *any* member or all but a few of their members.

Mathematics and lofty prose not withstanding, the claim that the whole of a society or group somehow exceeds the sum of its parts occurs to many to be over idealized. Any mathematician or philosopher who took a moment to venture out of his or her office would find no end of committee decisions, jury verdicts, democratic choices, and market valuations that have proven far wide of the mark. Collective wisdom, therefore, should be seen as a potential outcome, as something that can occur when the right conditions hold, but one that is in no way guaranteed.

The gap between theory, which points to an avoidable wisdom of crowds, and the less than ideal reality can be explained by the starkness of existing theory. The core assumptions that drive the mathematical necessity of collective wisdom may be too convenient. In particular, the idea that people receive independent signals that correlate with the truth has come to be accepted without thought. And, as we shall show, this assumption produces the near inevitability of collective wisdom.

In this paper, we first describe the standard statistical model used to explain collective accuracy of forecasts. We then describe a richer theoretical structure that can explain the existence of collective wisdom as well as the lack thereof. In this second model, individuals possess predictive models. Hong and Page (2009) refer to these as *interpreted signals* to capture the fact that these predictions can be thought of as statistical signals but that their values depend on how people interpret the world.

The second model differs from the standard statistical model of aggregation. As mentioned above, in the statistical model, individuals receive signals that correlate with the value or outcome of interest. Each individual's signal may not be that accurate but in aggregate, owing to a law of large numbers logic, those errors tend to cancel. In the canonical statistical model, errors are assumed to be independent. More elaborate versions of the model include both negative and positive correlations, a modification we take up at some length in the paper owing to the fact that negative correlation proves to be crucial for collective wisdom.

In our formulation,the prediction of a crowd of people can be thought of as an average of the forecasts produced by models contained within the individual's heads. Thus, collective wisdom depends on characteristics of the models people carry around in their heads. For collective wisdom to emerge those models must be sophisticated, or they must be diverse. Ideally, crowds will possess both.

These two features refer to different units of analysis. Diversity refers to the collection seen as a whole. The people within it, or their models, must differ. Sophistication /expertise refers to the capabilities of individuals within the collection. The individuals must be smart. There need not be a tradeoff between these two aspects of a crowd. The members of a crowd can become both more individually sophisticated and more collectively diverse. They can also become less sophisticated and less diverse. In the former case, the crowd becomes more accurate, and in the latter case, they become less so.

A tradeoff does exist in the necessity of these characteristics for an accurate crowd. Homogeneous crowds can only be accurate if they contain extremely sophisticated individuals, and groups of unsophisticated individuals can only be collectively accurate if they possess great diversity.

The intuition for why collective wisdom requires sophisticated individuals when those individuals are homogeneous should be straightforward. We cannot expect an intelligent whole to emerge from incompetent parts. The intuition for why diversity matters, and matters as much as it does, proves more subtle, so much so that several accounts misinterpret the mechanism through which diversity operates and others resort to hand waving.

The logic for why diversity matters requires two steps. First, diverse models tend to produce negatively correlated predictions.[1] Second, negatively correlated predictions produce better aggregate outcomes. If two predictions are negatively correlated when one tends to be high, the other tends to be low, making the average more accurate.

In what follows, we first describe the statistical model of collective wisdom. This approach dominates the social science literature on voting and markets as well as the early computational literature on ensemble learning. That said, computational scientists do a much more complete job of characterizing the contributions of diversity.

[1] In the case of a yes or no choice, Hong and Page (2009) show that when people use maximally diverse models (we formalize this in the paper), their predictions are necessarily negatively correlated.

Social science models tend to sweep diversity under the rug – calling it noise. In fact, we might even go so far as to say that social scientists consider diversity to be more of an inconvenience than a benefit.

We then formally define *interpreted signals* (Hong and Page 2007). These form the basis for what we will call the *cognitive model* of collective wisdom. This approach dominates computational learning models models (see Kuncheva 2005 for an overview). The cognitive model does not in any way contradict the statistical models. In fact, we rely on the statistical model as a lens through which to interpret the cognitive model.

In characterizing both types of models, we describe a general environment that includes both binary choice environments, i.e. simple yes or no choices, and cardinal estimation, such as when a collection of people must predict the value of a stock or the rate of inflation. When necessary for clarity, we refer to the former as *classification problems* and to the latter as *estimation problems.* The analysis differs only slightly across the two domains, and the core intuitions prove to be the same. We conclude our analysis with a lengthy discussion of what the theoretical results imply for the the existence or lack thereof of collective wisdom in markets and democracies and we discuss what we call *the paradox of weighting.* Our analysis is by no means exhaustive, but is meant to highlight the value of constructing deeper micro foundations.

Before beginning, we address three issues. First, a large literature in political science and in economics considers the implications of and incentives for strategic voting and revelation of information. For the most part, we steer clear of strategic considerations. When they do come into play, we point out what their effect might be. We want to make clear from the outset that regardless of what motivates the votes cast, the possibility of collective wisdom ultimately hinges on a combination of collective diversity and individual sophistication /expertise.

Second, we would be remiss if we did not note the irony of our model's main result: that collective wisdom requires diverse or sophisticated models. Yet, in this paper, we have constructed just two models - a statistical model and a model based on individuals who themselves have models. If our theory is correct, these two cannot be enough. Far better that we have what Page (2007) calls "a crowd of models." Complementing these models with historical, empirical, sociological, psychological, experimental, and computational models should provide a deeper, more accurate picture of what conditions must hold for collective wisdom to emerge. Clearly, cultural, social, and psychological distortions can also bias aggregation. We leave to other papers in this volume the task of fleshing out those other perspectives on collective intelligence. We note in passing that historical accounts, such as that of Ober in this volume, also identify diversity and sophistication as crucial to the production of collective wisdom, but rely on slightly different conceptions of those terms.

Finally, we provide a heads up that this paper does involve some mathematical formalism. We present the two main theorems of the statistical framework and one main theorem from the cognitive framework. The mathematics is necessary to clarify distinctions and to identify magnitudes of effects. It is one thing to say that diversity and sophistication contribute to collective accuracy. It is quite another to say that they contribute equally, which we show to be the case.

## The Statistical Model of Collective Wisdom

We begin with the statistical model of collective wisdom. The model considers the predictions or votes of individuals to be *random variables* drawn from a distribution. That distribution can be thought of as generating random variables conditional on some true outcome. In the most basic of models, the accuracy of the signals is captured by an *error term.* In more elaborate models, signals can also include a *bias.* In the canonical model, the signals are assumed to be independent. This independence assumption can be thought of as capturing diversity. How diversity translates into signal independence is left implicit. More nuanced theoretical results that allow for degrees of correlation also leave implicit how that correlation arises.

In all of the models that follow, we assume a collection of individuals of fixed, finite size.

*The* ***set of individuals****:* $N = \{1, ..., n\}$.

The voters attempt to predict an *outcome* that will arise in the future. As mentioned above that outcome can either be a simple yes /no or it could be a numerical value.

*The* ***outcome*** $\theta \in \Theta$. *In classification problems* $\Theta = \{0, 1\}$, *and in estimation problems* $\Theta = [0, \infty]$

In the statistical model, individuals receive signals. To distinguish these signals from those produced by cognitive models, we refer to this first type as *generated signals* and to the latter as *interpreted signals.* This nomenclature serves as a reminder that in the statistical model the signals are generated by some process that produces signals according to some distribution whereas in the cognitive model the signal an individual obtains depends on how she interprets the world.

The value of a signal depends on the value of the outcome. Therefore, we write the distribution of generated signals as being *conditional* on the true value of the outcome, $\theta$.

*Individual i's* ***generated signal*** $s_i \in \Theta$ *is drawn from the conditional distribution* $f_i(\cdot \mid \theta)$*.*

The notation $f_i$ allows for each individual's signals to be drawn from a different distribution function. We refer to the distribution of all of the individual's signals as the *collective distribution function.*

The *squared-error* of an individual's signal equals the square of the difference between the signal and the true outcome.

*The* ***sq-error*** *of the ith individual's signal* $SqE(s_i) = (s_i - \theta)^2$

*The* ***average sq-error*** $SqE(\vec{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - \theta)^2$

For the moment, we assume that the *collective prediction* equals the average of the individuals' signals. In the last section of the paper, we take up differential weighting of signals.

*The* ***collective prediction*** $c = \frac{1}{n}\sum_{i=1}^{n} s_i$.

We denote the squared error of the collective prediction by $SqE(c)$.

Finally, we define the *predictive diversity* of the collective as the variance of the predictions.

*The* ***predictive diversity*** *of a vector of signals* $\vec{s} = (s_1, s_2, ..s_n)$ *equals the variance of the signals.*

$$PDiv(\vec{s}) = \frac{1}{n}\sum_{i=1}^{n}(s_i - c)^2$$

If the signals differ greatly, then predictive diversity will be high. If all individuals make the same prediction, then predictive diversity will be zero.

## Statistical Model Results

With this notation in hand, we can now state what Page (2007) calls the *Diversity Prediction Theorem* and the *Crowds Beat Averages Law.* These widely known results provide statistical logic for the wisdom of crowds. The first theorem states that the squared error of the collective prediction equals the average squared error minus the predictive diversity.

**Theorem 1.** *(Diversity Prediction Theorem) The squared error of the collective pre-*

*diction equals the average squared error minus the predictive diversity.*

$$SqE(c) \ = \ SqE(\vec{s}) \ - \ PDiv(\vec{s})$$

Here, we see hard evidence that collective accuracy depends on sophistication (low average error) and diversity. And, moreover, we see that the two components matter equally, as mentioned in the introduction.

Given how counter intuitive this result can appear – does diversity really matter as much as ability? – we feel that its proof merits working through. Fortunately, the proof requires only a few lines of algebraic manipulaton. Expanding the left hand side of the equation gives:

$$(c-\theta)^2 = c^2 - 2c\theta + \theta^2$$

Substituting the value for $c$, this can be written as follows

$$(c-\theta)^2 = \left[\sum_{i=1}^{n} \frac{s_i^2}{n}\right] - 2c\theta + \theta^2 - \left[\sum_{i=1}^{n} \frac{s_i^2}{n}\right] + 2c^2 - c^2$$

Pulling out the $\frac{1}{n}$ terms gives

$$(c-\theta)^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(s_i^2 - 2s_i\theta + \theta^2)\right] - \frac{1}{n}\left[\sum_{i=1}^{n}(s_i^2 - 2s_ic + c^2)\right]$$

Which can be rearranged to give the desired result

$$(c-\theta)^2 = \frac{1}{n}\left[\sum_{i=1}^{n}(s_i - \theta)^2\right] - \frac{1}{n}\left[\sum_{i=1}^{n}(s_i - c)^2\right]$$

An corollary of this theorem states that the collective squared error must always be less than or equal to the average of the individuals' squared errors. Thus, it is a mathematical fact that *the crowd is always at least as good at predicting than the average of the individuals who make up the crowd.*

**Corollary 1.** *(Crowd Beats Averages Law) The squared error of the collective's prediction is less than or equal to the averaged squared error of the individuals that comprise the crowd.*

The fact that predictive diversity cannot be negative implies that the corollary follows immediately from the theorem. Nevertheless, the corollary merits stating. It provides a clean description of collective wisdom. In aggregating signals, the whole cannot be less accurate than the average of its parts.

The previous two results beg the question: How do we ensure diverse predictions? We now describe more general results from the statistical model of collective wisdom to address this question. That characterization will get us part way there, but to fully understand the basis of diverse predictions, we need a cognitive model, a point we take up in the next section. First though, we focus on the statistical foundations of diverse predictions.

Note that the previous two results describe a particular instance of a prediction. Here, we derive results in expectation over all possible realizations of the generated signals given the outcome. Up to now we could think of each signal as having an error. Hereafter, we average over a distribution of signal values.

In this richer framework errors can take two forms. A person's signal could be systematically off the mark, or it could just be off in a particular realization. To differentiate between systematic error in an individual's generated signal and idiosyncratic noise, statisticians refer signal *bias* and signal *variance*.

Here, the notation becomes a bit more cumbersome. We now think of each individual's signal as a random variable. That random variable has a mean, a bias, and a variance.

*Let* $\mu_i(\theta)$ *denote the* ***mean*** *of individual i's signal conditional on* $\theta$*. Individual i's* ***bias****,* $b_i = (\mu_i - \theta)$

*The* ***variance*** *of individual i's signal* $v_i = E[(s_i - \mu_i)^2]$

We can also define the *average bias* and the *average variance* across the individuals.

*The* ***average bias****,* $\bar{b} = \frac{1}{n}\sum_{i=1}^{n}(\mu_i - \theta)$

*The* ***average variance*** $\bar{V} = \frac{1}{n}\sum_{i=1}^{n} E[s_i - \mu_i]^2$

To state the next result, we need to introduce the idea of covariance. The covariance of two random variables characterizes whether they tend to move in the same direction (positive covariance) or in opposite directions (negative covariance). If covariance is positive, when one signal is above its mean, the other is likely to be above its mean as well. Negative covariance implies the opposite. Thus, negatively correlated signals tend to cancel out idiosyncratic errors.

*The* ***average covariance*** $\bar{C} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} E[s_i - \mu_i][s_j - \mu_j]$

Keep in mind the core assumptions of this framework. Each individual has an associated distribution function that generates signals. It's as though each person stands in front of a slot machine, pulls a lever, and gets a signal. A collective prediction is just the aggregate of all of those lever pulls. To evaluate a prediction's accuracy, we must therefore use the measure of expected squared error, $E[SqE(s_i)]$. The expected squared error can be decomposed into the systematic error and the idiosyncratic noise:

$$E(s_i - \theta)^2 = (\mu_i - \theta)^2 + E(s_i - \mu_i)^2.$$

**The expected squared error for the collective has the same decomposition:**

$$E[SqE(c)] = (\frac{1}{n}\sum_{i=1}^{n} \mu_i - \theta)^2 + E[\frac{1}{n}\sum_{i=1}^{n}(s_i - \mu_i)]^2.$$

Note that $c = \frac{1}{n}\sum_{i=1}^{n} s_i$ **and the mean of the collective prediction is equal to the average of the individual means,** $\frac{1}{n}\sum_{i=1}^{n} \mu_i$

The decomposition above reveals the key to collective wisdom in the statistical model. First, the collective's systematic error (the first term on the right side) is smaller if individuals' signals are biased in different ways, i.e. if some bias upward and others bias downward. This is because the bias for the collective is the average of the individuals biases. In the collective, biases get averaged out.

Second, for any realization of the signals, the collective's deviation from its mean is the average of the individual deviations from their respective means. So if individuals deviate in different ways, some deviate by being too high and others deviate by being too low, then the deviation of the collective is reduced, leading to a smaller idiosyncratic error. In other words, in the collective, idiosyncratic errors get averaged out. Finally, if enough such realization happens, taking expectation over the distributions, idosyncratic error for the collective is small. We characterize these situations with variance and covariance in the following result commonly known as the *bias-variance-covaraince decomposition.*

**Theorem 2.** *Bias-Variance-Covariance Decomposition (BVCD): Given $n$ generated signals with average bias $\bar{b}$, average variance $\overline{V}$, and average covaraince $\overline{C}$, the following identity holds:*

$$E[SqE(c)] = \bar{b}^2 + \frac{1}{n}\overline{V} + \frac{n-1}{n}\overline{C}$$

[2]

According to the BVCD, increasing the variance of signals increases expected error. This would appear to contradict the Diversity Prediction Theorem which states that variation in predictions reduces error. There is in fact no contradiction. In the first result, variance refers to realized differences in predictions. In the second result, variance refers to noisy or inaccurate signals. Realized differences in predictions would come about if the signals are negatively correlated.

An example makes the logic clearer. Consider a collective consisting of two people with unbiased signals. The signals of these two people have the same variance, denoted by $v$.So $\overline{V} = v$. However, assume that their signals are perfectly negatively correlated

[2] The proof of this result requires a little more algebra than the Diversity Prediction Theorem but relies on a similar approach. We include it in an appendix for the interested reader.

which means that any time one person's prediction deviates by being too high, the other deviates by being too low. Since the correlation is perfect, the magnitude of their covariance equals the variance, expect that it is negative. Formally, $\overline{C} = -v$. Plugging these values into the BVCD, the error for the collective equal to zero. This makes sense because, with each realization of the signals, the average deviation is also zero - one person's deviation cancels out the other's.

To be even more precise, in the Diversity Prediction Theorem, variance (or what we called predictive diversity) measures the difference between the individual signal realization and the resulting collective prediction. For the expected predictive diversity to be high, it has to be that for most of any given realizations, high deviaitons to the right by some are balanced out by high deviations to the left by others simply because the collective prediction is the average of individual signals. When bias is taken out of the picture, this implies negative correlations in signals which lowers the expected squared error of the collective according to the BVC decomposition.

Taking the biases as given, if we consider generated signals that are negatively correlated to be diverse, then the theorems provide two alternative ways of seeing the benefits of accuracy and diversity for collective prediction.

We conclude our analysis of the statistical model with a corollary that states that as the collective grows large, if generated signals have no bias and bounded variance and covariance, then the expected squared error goes to zero.

**Corollary 2.** *(Large Population Accuracy) Assume that for each individual average* *bias $\overline{b} = 0$, average variance $\overline{V}$ is bounded from above, and that average covaraince $\overline{C}$ is weakly less than zero. As the number of individuals goes to infinity, the expected collective squared error goes to zero.*

This proof is relatively straightforward. From the BVCD, we have that

$$E[SqE(c)] = \overline{b}^2 + \frac{1}{n}\overline{V} + \frac{n-1}{n}\overline{C}$$

By assumption $\overline{b} = 0$ and there exists a $T$ such that $\overline{V} < T$. Furthermore, $\overline{C} \leq 0$. Therefore $E[SqE(c)] < \frac{T}{n}$ which goes to zero as $n$ approaches infinity.

Note that *independent unbiased generated signals* are a special case of this corollary. If each individual's generated signal equals the truth plus an idiosyncratic error term, then as the collective grows large, it necessarily becomes wise. We see this as a weakness in the model. It's not the case that all large groups necessarily make correct predictions. To get rid of this result in the statistical model, one has to introduce some common bias. If so, then the crowd's error just equals this bias – an equally unsatisfying result.

# The Cognitive Model of Collective Wisdom

We now describe a cognitive model of collective wisdom. This cognitive model allows us to generate deeper insights than the statistical model. In the cognitive model, for collective wisdom to emerge the individuals must have relatively sophisticated models of the world. Otherwise, then cannot collectively come to the correct answer.

Furthermore, the models that people have in their heads must differ. If they don't, if everyone in the collective thinks the same way, the collective cannot be any better than the people in it. Thus, collective wisdom must depend on moderately sophisticated and diverse models. Finally, sophistication and diversity must be measured relative to the context. What is it that these individuals are trying to predict?

When we think of collective wisdom from a cognitive viewpoint, we begin to see shortcomings with the statistical model. The statistical model uses accuracy as a proxy for sophistication or expertise as well as for problem difficulty and uses covariance as a proxy for diversity. The cognitive model that we describe considers expertise, diversity, and sophistication explicitly. To do so, the model relies on a different type of signals called *interpreted signals.* These signals come from predictive models.

## Interpreted Signals

Interpreted signals can be thought of as model based predictions that individuals make about the outcome.[3] Those models, in turn, can be thought of as approximations of an underlying *outcome function.* Therefore, before we can define an interpreted signal, we must first define the outcome function that the models approximate. To do this, we first denote the set of all possible states of the world.

[3] For related models, see Barwise and Seligman, (1997), Al-Najjar, N., R. Casadesus-Masanell and E. Ozdenoren (2003), Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler (2005), and Fryer, R. and M. Jackson (2008).

*The set of states of the world* $X = \{x_1, x_2, ..x_n\}$

These states of the world correspond to a complete description of all relevant parts of reality. In the case of predicting the outcome of a policy, they would be a full description of the policy.

We next assume an *outcome function*, $F$ that maps the state of the world into an outcome. Note that this outcome is denoted by the same variable $\theta$ that denoted the value of interest in the statistical model.

*The* ***outcome function*** $F : X \rightarrow \Theta$

In the interpreted signal framework, each individual has an *interpretation* (Page 2007, Fryer and Jackson 2008) which is a partition of the set of states of the world into distinct categories. These categories form the basis for the individual's predictive model. For example, an individual might partition politicians into two categories: liberals and conservatives. Another voter might partition politicians into categories based on identity characteristics such as age, race, and gender.[4]

*Individual i's* ***interpretation*** $\Phi_i = \{\phi_{i1}, \phi_{i2}, ...\phi_{im}\}$ *equals a set of* ***categories*** *that partition* $X$*.*

We let $\Phi_i(x)$ denote the category in the interpretation to which the state of the

[4] An interpretation is similar to an information partition (Aumann 1976). What Aumann calls a information set, we call a category. The difference between our approach and Aumann's is that he assumes that once a state of the world is identified individuals know the value of the outcome function. We do not. In our model, given a category, they make a prediction.

world $x$ belongs. If an individual categorizes politician $x$ as a liberal, then $\Phi(x)$ is just the mapping of $x$ into the category named 'liberal.'

Most often two individuals' interpretations won't be identical. People will differ in how they assign states to categories. Note also that given this framework, it should be clear that individuals with finer interpretations can be thought of as more sophisticated. If a person mapped all states into a single category, then he would always make the same prediction. Loosely speaking, the number of categories is a proxy for subtlety of mind. We will say that one individual is more sophisticated than another if every category in its interpretation is contained in a category of the other's.[5]

*Individual i's interpretation is* ***more sophisticated*** *than individual j's interpretation if for any x,* $\Phi_i(x) \subseteq \Phi_j(x)$*, with strict inclusion for at least one x. A collection of individuals* ***becomes more sophisticated*** *if every individual's interpretation becomes more sophisticated.*

Individuals have *predictive models* which map their categories into outcomes. Predictive models are coarser than the outcome function. Whereas the objective function maps states of the world into outcomes, predictive models maps sets of states of the world, namely categories, into outcomes. Thus, if an individual places two two states of the world in the same category, the individual's predictive model must assign the

[5] Admittedly, this strong restriction may often fail to hold across individuals, but it is the natural definition given our construction.

same outcome to those two states.

*Individual i's* ***predictive model*** $M_i : X \to \Theta$ *s.t. if* $\Phi_i(x) = \Phi_i(y)$ *then* $M_i(x) = M_i(y)$.

An individual's prediction equals the output of his or her predictive model. Note that the predictive model of an individual can be thought of as a signal. However, unlike a *generated signal*, this signal is not a random variable drawn from a distribution. It is produced by the individual's interpretation and predictive model To distinguish this type of signal, we refer to it as an *interpreted signal.*

As before, the *collective prediction* of a population of individuals we take to be the average of the predictions of the individuals. The only difference is that now, instead of averaging signals, we're averaging predictions from models.

*The* ***collective prediction*** $\bar{M}(x) = \sum_{i=1}^{n} M_i(x)$

Written in this way, the intuition that the ability of a collection of individuals to make an accurate prediction depends upon their predictive models becomes clear. If those models are individually sophisticated, i.e. partition the set of states of the world into many categories, and collectively diverse, i.e. they create different partitions, then we should expect the collective prediction to be accurate. The next example shows how this happens.

*Example* Let the set $X$ consist of three binary variables. Each state can therefore be written as a sequence of 0's and 1's of length three. Formally, $X = (x_1, x_2, x_3)$,

$x_i \in \{0, 1\}$. Assume that each state is equally likely and that the outcome function is just the sum of the variables, i.e. $F(x) = x_1 + x_2 + x_3$. Assume that individual $i$ partitions $X$ into two sets according to the value of $x_i$ which he can identify. $M_i(x) = 1$ if $x_i = 0$ and $M_i(x) = 2$ if $x_i = 1$. The table below gives the interpreted signals (the predictions) for each realization of $x$ as well as the collective prediction and the value of the outcome function.

| *State* | $M_1(x)$ | $M_2(x)$ | $M_3(x)$ | $\bar{M}(x)$ | $F(x)$ | $SqE(\bar{M})$ |
|---|---|---|---|---|---|---|
| 000 | 1 | 1 | 1 | 1 | 0 | 1 |
| 001 | 1 | 1 | 2 | 4/3 | 1 | 1/9 |
| 010 | 1 | 2 | 1 | 4/3 | 1 | 1/9 |
| 100 | 2 | 1 | 1 | 4/3 | 1 | 1/9 |
| 011 | 1 | 2 | 2 | 5/3 | 2 | 1/9 |
| 101 | 2 | 1 | 2 | 5/3 | 2 | 1/9 |
| 110 | 2 | 2 | 1 | 5/3 | 2 | 1/9 |
| 111 | 2 | 2 | 2 | 2 | 3 | 1 |

We can view these interpreted signals within the statistical framework. Though each prediction results from the application of a cognitive model, we can think of them as random variables. Since the statistic model presented before is with regard to any given outcome, we compute interpreted signal's bias and squared error conditional on a given value of $F(x)$. In what follows, we do our computation and comparison conditional on $F(x) = 1$. The case for $F(x) = 2$ is similar and the cases for $F(x) = 0$

$F(x) = 3$ are trivial since there is no randomness in the signals. By symmetry, it suffices to consider a single interpreted signal to compute bias and squared error.

| **State** | $\mathbf{F(x)}$ | $\mathbf{M_1(x)}$ | $\mathbf{Error(M_1)}$ | $\mathbf{SqE(M_1)}$ | $\overline{M}(\mathbf{x})$ | $\mathbf{Error}(\overline{M})$ | $\mathbf{SqE}(\overline{M})$ |
|---|---|---|---|---|---|---|---|
| **001** | **1** | **1** | **0** | **0** | **4/3** | **1/3** | **1/9** |
| **010** | **1** | **1** | **0** | **0** | **4/3** | **1/3** | **1/9** |
| **100** | **1** | **2** | **1** | **1** | **4/3** | **1/3** | **1/9** |
| **Expectation** | **1** | **4/3** | **1/3** | **1/3** | **4/3** | **1/3** | **1/9** |

As can be seen from the table, the bias of the interpreted signal equals $\frac{1}{3}$. A straightforward calculation shows that the variance of the interpreted signal equals $\frac{2}{9}$. Notice that each individual has an expected squared error equal to 1/3, but the collection has an expected squared error equal to just 1/9. So in this case, the collective is more accurate, in expectation, than any of the individuals.

We know from the earlier statistical models that this must result from negative correlation in each pair of interpreted signals. And, in fact, the covariance of interpreted signals 1 and 2 (or 2 and 3 or 1 and 3) equals $\frac{1}{3}[(1-\frac{4}{3})(1-\frac{4}{3}) + (1-\frac{4}{3})(2-\frac{4}{3}) + (2-\frac{4}{3})(1-\frac{4}{3})] = -\frac{1}{9}$.

Thus, in the collective, idiosyncratic errors of individuals tend to cancel each other out. The remaining error comes from the squared average bias.

Notice that in the example, each individual considered a distinct attribute. Hong and Page (2009) refer to these as *independent interpreted signals.* Interpreted signals are independent if the category that one individual uses is independent of the category that another uses. This is somewhat cumbersome to make formal but it can be done as follows:

*The interpreted signals of individual* 1 *and* 2 *are based on* ***independent interpretations*** *if and only if for all $i$ and $j$ in* $\{1, 2, ...m\}$

$$\text{Prob}(\phi_{1j} \cap \phi_{2i}) = \text{Prob}(\phi_{1j}) \times \text{Prob}(\phi_{2i})$$

In effect, if two individuals use independent interpretations, then they look at different dimensions given the same representation. Hong and Page (2009) show that for classification problems, i.e. problems with binary outcomes, independent interpreted signals must be negatively correlated. The theorem requires mild constraints on the individuals' predictive models – namely that they predict both outcomes with equal probability and that they are correct more than half the time.[6]

**Theorem 3.** *If $F : X \rightarrow \{0, 1\}$, if each outcome is predicted equally often and if each individual's prediction is correct with probability $p > 1/2$ then independent interpreted signals are negatively correlated.*

[6] Extending the theorem to apply to arbitrary outcome spaces requiresstronger conditions on the predictive models and on the outcome function.

pf. See Hong and Page (2009)

This theorem provides a linkage between the models that individuals use and statistical properties of their predictions. For classification problems, model diversity implies negatively correlated predictions conditional on outcomes, which we know from the statistical models implies more accurate collective predictions.

This result provides a key insight into collective wisdom. We know from the BVCD framework that negative correlation improves accuracy. This result shows that model diversity implies negative correlation. Thus, model diversity improves collective accuracy.

We now turn to the question of how much diversity and sophistication are required for a collection of individuals through voting to predict the correct outcome. In other words, we ask - what has to be true of the individuals and of the outcome function, for collective wisdom to emerge?

The answer to that question is surprisingly straightforward. Individuals think at the level of category. They do not distinguish among states of the world that belong to the same category. Therefore, we can think of each individual's interpretation and predictive model as producing a function that assigns the same value to any two states of the world in the same category. If a collection of people vote, then what they are doing is aggregating these functions.

Here then lies the key insight: *If the outcome function can be defined over the categories used by the individuals, then the individuals can combine their models and* *approximate the outcome function.* However, if the outcome function takes on a unique value for states of the world in some set $S$ that no individual or group of individuals can identify, then we should not expect the individuals to be able to approximate the outcome function. Thus, a necessary condition for collective wisdom to arise is that, collectively, the interpretations of the individuals must be fine enough to approximate the outcome function. That will be our central result.

In addition, we will need one technical assumption. Namely that the outcome function must be an additive combination of the predictive models of the individuals (See Hong and Page (2008) for a full characterization). This additivity assumption must be included because voting merely adds up predictions. If we were to allow for deliberation, then the models could be combined in more nuanced ways.

## Sophistication and Diversity in Cognitive Models

Note the importance of sophistication. Interpretations that create more categories produce more accurate predictions. And, as just described, the ability of a collection of people to make accurate appraisals in all states of the world depends on their ability to identify all sets that are relevant to the outcome function. Therefore, as the individuals become more sophisticated, the collective becomes more intelligent.

As for diversity, we have seen in the case of classification problems that independent interpretations produce negatively correlated interpreted signals. That mathematical finding extends to a more general insight: *more diverse interpretations tend to produce more negatively correlated predictions.* Consider first the extreme case. If two individual's use identical interpretations and make the best possible prediction for each category, then their predictive models will be identical. They will have no diversity. Their two heads will be no better than one. If, on the other hand, two people categorize states of the world differently, they likely make different predictions at a given state. Thus, diversity in predictions arises from diversity in predictive models.

This is not to say that the statistical model ignores sophistication and diversity, only that they only enter implicitly. Bias and error are proxies for sophistication and correlation captures diversity (Ladha 1992). A problem with implicit assumptions is that they can blind us to their implications. In the generated signal framework, independence seems an obvious benchmark. Therefore, it doesn't seem problematic to assume a large number of independent signals. Recall also that n the statistical model, as the number of individuals tends to infinity, then in the absence of bias, the collective becomes perfectly accurate. Yet, if we think of independence as implicitly requiring some level of diversity in models, then we might expect some limits on the amount of diversity present.

# Discussion

In this paper, we have provided possible micro foundations for collective wisdom. We have contrasted this approach with the standard statistical model of collective wisdom that dominates the literature. While both approaches demonstrate the importance of sophistication and diversity, they do so in different ways. The statistical model makes assumptions that imply sophistication and diversity, while the cognitive model approach includes sophistication and diversity directly.

The cognitive micro foundations that we have presented provide an alternative lens on the wisdom of crowds and then also help to explain the potential for the madness of crowds. A collection of people becomes likely to make a bad choice if they rely on similar models. This idea aligns with the argument made by Caplan (2007) that people make systematic mistakes. If everyone leaves out some relevant feature of the world in constructing their models, then the crowd cannot be accurate as we have shown.

Note though that the cause of wise or mad crowds is not just the intelligence of the people that make up that crowd. Enough collective diversity can make up for lack of individual sophistication. This presents an interesting challenge for democracy. Not only should democratic institutions encourage sophisticated thinking, they should also support diversity. In the aggregate, collective diversity matters as much as individual sophistication. And, we might add, diversity might be easier to attain.

Empirically, the causes of diversity and sophistication are manifold and diverse. Diversity can be produced by differences in identity (see Nisbett, R. 2003). It can also result from different sources of experience and information (Stinchcombe 1990). Sophistication derives from experience, attention, motivation, and information.

Finally, we have yet to discuss the potential for persuasion within a group. In the statistical model, persuasion places more weight on some individuals than on others. Ideally, Incidentally, optimal weighting depends not only on the accuracy of the various models but also on the diversity of the models. Models that tend to negatively correlate earn more weight (Lamberson and Page 2010). In any particular group setting we have no guarantee that deliberation will produce such a weighting or anything close to it. In fact, improper weightings resulting from deliberation often result in less accurate forecasts. Empirical evidence suggest that equal weighting of models tends to work best unless strong evidence supports placing more weight on some models than others (Armstrong 2001)

A similar phenomenon occurs in the cognitive model though the mechanisms through which deliberation hinders performance differs. In the cognitive model, persuasion and deliberation would involve abandoning and combining models. These attempts to improve individual accuracy could make the collective worse off by sacrificing collective model diversity on the alter of individual sophistication. It is better for the collective to contain a different and less accurate model than to add one more copy of any existing model, even if that existing model is more accurate (Lamberson and Page 2010).

To summarize, we have presented two formal models of collective wisdom, one based on statistics and one based on cognitive models. We've shown how both demonstrate that collective wisdom depends at its core on either crowds made of sophisticated accurate individuals or on crowds that are collectively very diverse in the models they employ. We also showed that Panglossian results that very large crowds will also be wise implicitly assume a level of diversity that probably does not and cannot exist. Taken together, the two models provide solid analytic grounding for how and why collective wisdom can exist. An additional benefit of these formalisms is that they define terms and clarify logic that bind and enrich the much richer historical, psychological, and empirical accounts of collective wisdom contained in this volume.

# Appendix

*Proof of the Bias-Variance-Covariance Decomposition*

From the discussion in the text it suffices to show that

$$E[\frac{1}{n}\sum_{i=1}^{n}(s_i-\mu_i)]^2=\frac{1}{n}\overline{V}+\frac{n-1}{n}\overline{C}.$$

Expanding the term on the left hand side gives

$$E\left[\frac{1}{n}\sum_{i=1}^{n}(s_i-\mu_i)\right]^2=\frac{1}{n^2}E\left[\sum_{i=1}^{n}(s_i-\mu_i)^2+\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}(s_i-\mu_i)(s_j-\mu_j)\right]$$

Which by rearranging yields

$$\begin{aligned}&=\frac{1}{n}\left[\frac{1}{n}\sum_{i=1}^{n}E(s_i-\theta)^2+(n-1)\cdot\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}E(s_i-\mu_i)(s_j-\mu_j)\right]\\&=\frac{1}{n}[\overline{V}+(n-1)\overline{C}]\\&=\frac{1}{n}\overline{V}+\frac{n-1}{n}\overline{C}\end{aligned}$$

# References

[1] Al-Najjar, N., R. Casadesus-Masanell and E. Ozdenoren (2003) "Probabilistic Representation of Complexity", *Journal of Economic Theory* 111 (1), 49 - 87.

[2] Armstrong, J. S., (2001) Combining Forecasts, in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Norwell, MA: Kluwer Academic Publishers.

[3] Aragones, E., I. Gilboa, A. Postlewaite, and D. Schmeidler (2005) "Fact-Free Learning", *The American Economic Review* 95 (5), 1355 - 1368.

[4] Aumann, R. (1976), "Agreeing to Disagree", *Annals of Statistics* 4, 1236-9

[5] Barwise and Seligman, (1997) *Information Flow: The Logic of Distributed Systems* Cambridge Tracts In Theoretical Computer Science, Cambridge University Press, New York.

[6] Caplan, Bryan (2007) *The Myth of the Rational Voter: Why Democracies Choose Bad Policies* Princeton University Press.

[7] Fryer, R. and M. Jackson (2008), "A Categorical Model of Cognition and Biased Decision-Making", *Contributions in Theoretical Economics, B.E. Press*

[8] Hong L. and S. Page (2009) ““Interpreted and Generated Signals” ”*Journal of Economic Theory* 144: 2174-2196.

[9] Hong L. and S. Page (2008) “On the Possibility of Collective Wisdom”working paper

[10] Kuncheva, L.I. (2005) *Combining Pattern Classifiers, Methods and Algorithms.* Wiley, New York, NY.

[11] Ladha, K. (1992) “The Condorcet Jury Theorem, Free Speech, and Correlated Votes”, *American Journal of Political Science* 36 (3), 617 - 634.

[12] Lamberson, P.J. and S. E. Page, “Optimal Forecasting Groups” working paper.

[13] Nisbett, R. (2003) *The Geography of Thought: How Asians and Westerners Think Differently...and Why* Free Press, New York.

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

[15] Stinchecombe, A. (1990) *Information and Organizations* California Series on Social Choice and Political Economy I University of California Press.

[16] Suroweicki, James (2004) *The Wisdom of Crowds* Doubleday Press, New York.

[17] Von Hayek, F. (1945) ”The Use of Knowledge in Society,” *American Economic Review*, 4 pp 519-530.