Some Microfoundations of Collective Wisdom

Lu Hong and Scott E Page

September 15, 2008

Abstract

Collective wisdom refers to the ability of a population or group of individuals to make an accurate prediction of a future outcome or an accurate characterization of a current outcome. Without feedbacks, the collective will always be more accurate than it's average member, and in some circumstances, it can be more accurate than any of its members. 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 a model of agents with predictive models. Our model extends traditional statistical approaches to characterizing collective wisdom. Within our model, we demonstrate 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 collective wisdom often does exist. Abundant anecdotal and small to large scale empirical examples also suggest at least the potential for a "wisdom of crowds" (Suroweicki 2004).

Collective wisdom, as we shall define it here, exists when the crowd outperforms the people in it at a predictive task. This is a restrictive notion. Wisdom often has a broader conception than mere accuracy. A society deliberating on laws or common purpose must exercise wisdom in judgement. The task is much richer and nuanced then estimating the value of a stock or the weight of a steer. And yet, if we see wisdom in these contexts and anticipating multiple implications and interactions, then we might also wisdom as the ability to recognize the multiplicity of effects and to accurately predict the magnitude of each. If so, we might see our conception of collective wisdom as less circumscribed.

The logical statistical foundations for collective wisdom are well known. First, 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. We describe how that can be the case in detail.

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 it is in no way guaranteed.

The gap between theory and 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 argue, it is this assumption that creates the near inevitability of collective wisdom.

In this paper, we describe a richer theoretical structure that can explain the existence of collective wisdom as well as the lack thereof. In this model, individuals possess predictive models. Hong and Page (2007) 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 prediction of a crowd of people can be thought of as some type of average of the models contained within those peoples heads. Thus, collective wisdom depends on characteristics of the models people carry around in their heads. We show that 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. Crowd members can become both more sophisticated and more 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 naive individuals can only be collectively accurate if they possess great diversity. [1]

The intuition for why collective wisdom requires sophisticated individuals when those individuals are homogeneous should be straightforward. We cannot expect an

[1]Our approach borrows ideas from ensemble learning theory. In ensemble learning, collections of models are trained to make a prediction or a classification. The predictions of the individual models are then aggregated to produce a collective prediction.

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 that others resort to hand waving. The logic for why diversity matters requires two steps. First, diverse models tend to produce negatively correlated predictions.[2] 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.

The model we describe differs from the standard approach in political science and economics, or what we call the *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 correlation, a modification we take up at some length as negative correlation proves to be crucial for collective wisdom.

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, the computational

[2]In the case of a yes or no choice, Hong and Page (2007) show that when people use maximally diverse models (we formalize this in the paper), their predictions are necessarily negatively correlated.

scientists do a much more complete job of characterizing the contributions of diversity. 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 the current computational science models. This cognitive model does not in any way contradict the statistical model. 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 consider 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*. That discussion is by no means exhaustive, but is meant to highlight the value of constructing deeper micro foundations.

Before beginning, we must address two issues. First, a growing literature in political science and in economics considers the implications of and incentives for strategic voting. 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.

The Statistical Model of Collective Wisdom

The statistical model of collective wisdom 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 but how that diversity translates into the signals is left implicit. More nuanced theoretical results will allow for degrees of correlation as we shall show.

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

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

The voters attempt to predict the *outcome*. 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* $\Theta = [0, \infty]$

As mentioned, individuals receive signals. A signal can be thought of as the prediction or opinion of the individual. 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.

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

The notation allows for each individual's signals to be drawn from a different distribution function. The collection of all signals can be characterized by a *collective distribution function.*

The ***collective distribution function*** $g(s_1, s_2, ...s_n \mid \theta)$ *describes the joint distribution of all of the signals conditional on* θ, $g_i(s_1, s_2, ...s_n \mid \theta) = f_i(s_i \mid \theta)$

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

In what follows, 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 $SqE(c)$ The squared error gives a measure of the accuracy of the collective. We can measure the *predictive diversity* of the collective by taking the variance of the predictions.

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

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

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 the basic 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. Here, we see the first evidence that collective accuracy depends both on expertise (low average error) and diversity.

Theorem 1. *(Diversity Prediction Theorem) The squared error of the collective prediction equals the average squared error minus the predictive diversity.*

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

pf. Expanding each term in the expression, it suffices to show that

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

Note that in the proof that the third equality follows from the second by the definition of c. A 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.

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. 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.

We first note that the previous two results describe a particular instance. Here, we derive results in expectation over all possible realizations of the generated signals given the outcome. Before, we considered a single instance, so we could think of each signal as having an error. Now, we average over a distribution and 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 the systematic error in an individual's generated signal and the idiosyncratic noise, statisticians refer to these as the *bias* and the *variance* of the signal.

Let $\mu_i(\theta)$ *denote the* ***mean*** *of individual* i*'s signal conditional on* θ*. 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 or in opposite directions. 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 one another's idiosyncratic errors.

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

Note the implicit mode of thinking that underpins this analysis. Each person has an associated distribution function that generates signals. Any one prediction can be thought of as a random draw from that distribution.

To evaluate a prediction's accuracy prior to a random draw, we use the measure of expected squared error, $E[SqE(s_i)]$. Mathematically, it is straightforward that 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 bias in different ways, 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 (BVC) Decomposition) Given n generated signals with average bias* $\overline{b}$*, average variance* $\overline{V}$*, and average covaraince* $\overline{C}$*, the following identity holds:*

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

pf. From the discussion of the previous decomposition, we only need to show that $E[\frac{1}{n}\sum_{i=1}^{n}(s_i - \mu_i)]^2 = \frac{1}{n}\overline{V} + \frac{n-1}{n}\overline{C}$.

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

At first glance, according to the BVC decomposition, increasing the variance of signals increases expected error. This seems to contradict the Diversity Prediction

Theorem which implies that variation in predictions reduces error. There is in fact no contradiction. First, the impact of the variance of signals on the collective error can not be analysed alone. A change in the variance of signals often changes covariance and therefore, the impact of the variance of signals should be analysed along with the impact of the covariance.

Consider a collective consisting of two people. The signals of these two people have the same variance, denoted by v.So $\overline{V} = v$. However, their signals are perfectly negatively correlated 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. So $\overline{C} = -v$.

According to the BVC decomposition, the idiosyncradic error for the collective is 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. Now let the variance increase but keep everything else the same. Then the idiosyncratic error for the collective remains zero. That is, even though each individual signal is less accurate due to the increased variance, the collective does equally well.

Second, in the Diversity Prediction Theorem, variance is defined differently - it 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.*

pf. From the BVC Decomposition, 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 signals equal the truth plus an idiosyncratic error term, then as the collective grows large, it necessarily becomes wise.

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 this section we show that for collective wisdom to emerge, the individuals must have relatively sophisticated models of the world otherwise, we cannot expect them collectively to 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.

The set of states of the world X

The *outcome function*, F maps each possible state into an outcome.

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

Each individual has an *interpretation* (Page 2007) which is a partition of the set of states. An interpretation partitions the states of the world into distinct categories. These categories form the basis for the individual's predictive model. For example, one 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]

[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).

[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 assume that once a state of the world is identified individuals know the value of the outcome function.

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 world x belongs. Individuals with finer interpretations can be thought of as more sophisticated. Formally, we 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 what we call *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 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* $M_i(x) = M_i(y)$.

[5] Admittedly, this strong restriction may often fail to hold across individuals, but it is the natural definition if we think of an individual as becoming more sophisticated.

An individual's prediction equals the output of his or her predictive model. 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*. The *collective prediction* of a population of individuals we take to be the average of the predictions of the individuals

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

The ability of a collection of individuals to make an accurate prediction depends upon their predictive models. Intuitively, 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 can occur.

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 using 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	**F(x)**	**M_1(x)**	**Error(M_1)**	**SqE(M_1)**	$\overline{M}$**(x)**	**Error(**$\overline{M}$**)**	**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. This is a result of negative correlation in each pair of interpreted signals. Covariance of interpred signals 1 and 2 (or 2 and 3 or 1 and 3) $= \frac{1}{3}[(1-\frac{4}{3})(1-\frac{4}{3})+(1-\frac{4}{3})(2-\frac{4}{3})+(2-\frac{4}{3})(1-\frac{4}{3})] = -\frac{1}{9}$.In the collective, idiosyncratic errors of individuals cancel each other out. The remaining error comes from the squared average bias.

In the example, each individual considered a distinct attribute. Hong and Page (2007) refer to these as *independent interpreted signals.*

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

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

If two individuals use independent interpretations, then they look at different dimensions given the same representation. Hong and Page (2007) 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 $F : X \to \{0, 1\}$, 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.*

pf. See Hong and Page (2007)

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.

The statistical approach focuses on the size of the expected error as a function of bias, error, and correlation of generated signals. The cognitive model approach does not assume any randomness in the prediction, though uncertainty about the state of the world does exist. Therefore, a natural question to ask within the cognitive model

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

approach is whether a collection of individuals can, through voting, produce the correct outcome. In other words, we can 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 or express opinion about the likely value of an outcome, then what they are doing is aggregating these functions.

If the outcome function can be defined over the categories of individuals, then it would seem possible that the individuals can combine their models and approximate the outcome function. However, suppose that the outcome function assigns an extremely high value to states of the world in the set S but that no individual can identify S, i.e. for each individual i, S is strictly contained within a category in Φ_i. 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. In addition, 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 should be seen as especially strong. It limits what outcomes a crowd can always predict correctly.

Sophistication and Diversity in Cognitive Models

In the statistical model, bias and error are meant to be proxies for sophistication and correlation is thought to capture diversity (Ladha 1992) In the cognitive model framework, sophistication refers to the numbers and sizes of the categories. 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 come from diversity in predictive models.

Even with a large number of individuals, we might expect some limits on the amount of diversity present. In the statistical model, as the number of individuals tends to infinity, then in the absence of bias, the collective becomes perfectly accu-

rate. That will not happen in the cognitive model unless we assume that each new individual brings a distinct predictive model.

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 might be expected to correlate with sophistication and diversity, while the cognitive model approach includes sophistication and diversity directly.

The cognitive micro foundations that we have presented also help to explain the potential for the madness of crowds. A collection of people becomes more 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. Note though that in other venues where collections of individuals do not make mistakes, they are not necessarily more accurate individually, they may just be more diverse collectively.

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. It's important though to recall that ramping up individual level sophistication can have costs: de-

creases in diversity can more than offset increases in individual accuracy.

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, the weight assigned to each generated signal would be proportional to its accuracy. In any particular group setting we have no guarantee that such weighting will emerge. And, in fact, improper weightings may lead to even worse choices. In our cognitive model, persuasion can have a similar effect. However, instead of changing weights people may abandon models because they find another person's model more convincing. Often such behavior proves to make the collective worse off. 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.

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

[3] Aumann, Robert. 1976, "Agreeing to Disagree", Annals of Statistics 4, 1236-9

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

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

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

[7] Hong L. and S. Page (2007) "Interpreted and Generated Signals "working paper

[8] Hong L. and S. Page (2008) "On the Possibility of Collective Wisdom"working paper

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

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

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

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

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

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