https://doi.org/10.48009/1_iis_2010_724-732

EVALUATING THE WISDOM OF CROWDS

Christian Wagner, City University of Hong Kong, iscw@cityu.edu.hk
Tom Vinaimont, City University of Hong Kong, tom.vinaimont@cityu.edu.hk

ABSTRACT

Estimating and forecasting are difficult tasks. This is true whether the activity requires the determination of uncertain future event outcomes, or whether the estimation effort is complex in itself and based on insufficient information. Consequently, such tasks are frequently assigned to experts. Surprisingly, recent research suggests that collectives of non-experts can outperform individual experts, as long as certain conditions are met. The resulting capability has been described as collective intelligence or "wisdom of crowds". Our research explores this collective intelligence, for real and simulated crowds. An empirical test demonstrates that even a relatively small crowd of 30 subjects can demonstrate expert-like performance. A further investigation through simulation shows that the performance of a collective predictably compares to that of an expert, with the expert outperforming small crowds but being outperformed by large collectives. The relationship between performance and log of collective size follows a linear function.

Keywords: Collective Intelligence, Wisdom of Crowds, Simulation, Diversity, Expertise

INTRODUCTION

Making risky decisions and predicting [1] unknown events are two activities which people cannot avoid in their lives. Yet estimating is difficult and error-prone, as illustrated by events such as unforeseen oil spills, economic crisis, or outcomes of sports events. Difficulties exist whether the task requires forecasting uncertain future events, or whether the estimation task is complex in itself and based on insufficient information. Consequently people are not good at performing estimation tasks [1], frequently exhibiting biases and making errors [2, 3]. For example, physicians faced with the difficult question "Doc how long do I have left to live" systematically demonstrate an optimistic bias [4]. Research has uncovered that systematic, non-pathological biases in cognitive processes, which are neither dependent on intelligence, nor on education, are distributed equally in the population [5]. Thus experts are as prone to biases and errors which may influence their prediction and decision making behavior as are non-experts.

Ability of Crowds to Guess Well

While individual non-experts and possibly even experts are not good in estimating, collectives apparently are. This interesting insight can be drawn from scenarios such as the TV game show "Who Wants to Be a Millionaire" (with its ask-the-audience feature) or prediction games such as Yahoo's "College Pickem", where the crowd year after year equals or beats expert punters in predicting football game outcomes. Collective intelligence has come to widespread attention through Surowiecki's influential book *The Wisdom of Crowds* [6], which describes both the condition under which collective intelligence manifests itself and which illustrates, through scenarios, the power of crowd wisdom. Validating this suggested ability of crowds to predict was part of the purpose of the research described in this article.

BACKGROUND AND FOUNDATIONS

Collectives

Collectives differ from traditional groups, not only in term of their size, but also with regard to their characteristics. Groups are frequently defined as "social aggregates that involve mutual awareness and potential mutual interaction. … are relatively small and relatively structured or organized" [7, p. 7] While their structure and cohesion gives groups an advantage in a number of tasks, these characteristics can also lead to reasoning biases such as group polarization [8] or representativeness fallacy [9]. Collectives, which are not 'normed' [10] and don't necessarily share the same attitudes may perform worse on tasks requiring integrated action, but in turn benefit from members' relative independence. Surowiecki [6] identified three requirements for collective wisdom to emerge, diversity, decentralization of opinion, and independence. Diversity of opinion refers to the availability of multiple viewpoints (ideally many), within the group, as each further viewpoint may help to explain the phenomenon better. Independence means that

[1] We use the terms predicting, forecasting, estimating, and guessing interchangeably in this article.

peoples' opinions are not determined by the opinions of those around them, which is typically the case in groups or teams. Decentralization requires that people can draw on their local and specific knowledge and make independent decision.

Requirements for Performance

For collectives to be able to predict, several performance criteria have to be fulfilled. First, the variable to be determined must not be completely random. Asking anyone to predict the outcome of a coin flip is futile. Second, individuals have to have some reasoning capability and information. If individuals guess at random because they lack either information of reasoning ability, the collective outcome will carry no information. Individuals, however, do not have to know the exact right value. Most importantly is the ability to eliminate some values. To illustrate, if three individuals (I1-I3) seek to determine the right answer among choices A-D, and each one is able to eliminate two choices, then a process of approval voting might yield: I1: A and B, I2: B and C, I3: B and D. Correspondingly, A, C, and D each would receive one vote, while B would receive three votes, thus making it the favorite guess. Similarly, in guessing the quantity of candies in a container, ascertaining the right number will not be as important as being able to exclude impossible ranges. To achieve this outcome, we need the crowd to eliminate as many impossible or improbable values as possible. If one person is able to eliminate at least one such value, then multiple people can eliminate many, as long as they approach the problem with different estimation mechanisms so as not to replicate the same elimination outcome time and again. In other words, the crowd needs diversity.

Collective Wisdom vs. Law of Large Numbers

Frequently it is assumed that the wisdom of crowds is merely a manifestation of the Law of Large Numbers, such that when a crowd estimates, the error terms of the extreme cases will cancel each other out and thus the mean estimate approximates a good guess. While one aspect of having a large crowd is the reduction of errors and noise, this is not the main effect the crowd has. In fact, as mentioned already, diversity is sought. The logic of collective intelligence is that different individuals will apply different "theories", or r heuristics, to the guessing task, the aggregate of which results in a highly precise estimate of the variable in question. While each "theory" would only be able to predict part of the variance in the observed outcome, the collection of theories brought together, can explain much of the variance and lead to a highly precise result. Asked "will it rain tomorrow", one individual might observe the clouds to make a prediction, someone else may view the barometer, someone else may refer to the Farmer's Almanac e.g., [11], and so on. Either method by itself may be imprecise. Taken together their precision is expectedly high. Consequently also, crowd members must use different heuristics, or else their value is diminished to the law of large numbers. This is the reason why diversity in the crowd is fundamentally important.

Rationale of Crowd Wisdom

Page [12] categorized cognitive diversity into four dimensions: diversity of perspectives (ways of representing situations and problems), diversity of interpretations (ways of categorizing or portioning perspectives), diversity of heuristics (ways of representing situations and problems) and diversity of predictive models (ways of inferring cause and effect). He formalizes the range of diversity in a prediction diversity (PD) variable, and further conceptualizes two other parameters of collective intelligence, collective error (CE), and (average) individual error (IE).

Prediction diversity is defined as the aggregate squared difference between individual guesses and the average guess. It reflects how far individuals, on average, veer from the group.

$$\text{PD} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \ ;$$

Individual error aggregates the squared errors of all the participants. It thus captures the average accuracy of the individual guesses.

$$\text{IE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{true})^2 \ ;$$

Collective error, the squared error of the collective prediction, represents the difference between the correct answer x_{true} and the average guess aggregated $\bar{x}$ from all individuals.

$$\text{CE} = (\bar{x} - x_{true})^2$$

Page then formulates the *diversity prediction theorem*, which states:

Collective Error (CE) = Individual Error (IE) – Prediction Diversity (PD)

A low collective error signifies high overall collective intelligence and accuracy. The collective error, being composed of individual error and

prediction diversity, pits these two parameters against each other. In other words, we can lower collective error by reducing individual error (raising individual expertise) or by raising prediction (group) diversity.

Research Evidence of Collective Intelligence

The capability and values of crowds has been demonstrated in research in a number of scenarios and through several aggregation mechanisms. Dalkey's empirical investigation of the Delphi method [13] may be considered an early exploration of collective intelligence. In the Delphi method, multiple decision makers first provide independent guesses, followed by feedback on the collective performance, and another round of individual guessing. Collaborative filtering methods such as GroupLens [14] emerged as another early form of collective intelligence, whereby the opinion of crowd members was used to effectively filter large data amounts for other users. Kittur and others [15] demonstrated the value of the crowd in increasing the accuracy of collectively created content. The ability of so called "good samaritans" within the crowd, who rarely participated, but if so, corrected critical mistakes was the key contributor to raising collective performance. Wolfers and Zitzewitz [16], among others, demonstrated the power of prediction markets. Prediction markets are markets for events, where participants "bet" on the outcome of uncertain events. The power of prediction markets, such as the Hollywood Stock Exchange or Newsfutures, has been ascribed to their ability to elicit and propagate otherwise private information.

RESEARCH METHODS

Operationalization

To formally analyze the impact of collective intelligence on estimation, we carried out two types of tests. The first was an empirical test of the ability of 30-person crowd to estimate, the second a simulation study comparing the performance of a simulated expert against simulated crowds of various sizes and abilities.

Task Structure

The tasks we used, both in the test with people and in the simulation, were simple. Subjects had to predict the value of a variable. In the real-life test, the crowd of 30 guessed the quantity of candies in a container (based on observation of the transparent container), and the temperature in the city one week into the future. One task was essentially deterministic, as the true result was determinable at the time, given enough observation and analysis effort. The other task was heuristic and forward looking into the future. Hence, through the combination of these tasks, we examined both the ability to use analytic methods and possible heuristics for the determination of a certain number, and the ability to make meaningful educated guesses. Financial incentives were used to motivate subjects to put forward their best estimates.

The simulation task similarly required the "guessing" of a value. The target number in the simulation was always 1000, and simulated experts and non-experts were defined by different model-based random guessing algorithms bounded at the upper and lower end based on pre-defined precision ranges. Thus, if a simulated non-expert's guesses were bounded at [500; 1500], then a 10x more precise expert would be defined by a narrower precision range of [950; 1050]. Simulated estimates were uniformly distributed within the boundaries.

Determining Intelligence

A significant challenge for the evaluation of crowd intelligence is the definition of expertise. For our first two tests involving human subjects, we did not even use a human expert. Yet even if a human expert is used, the expertise level of that individual can be challenged, especially if the expert were outperformed by the crowd. Hence, for the tests with human subjects, we had to seek different measures of expertise that were independent of identifying a true human expert. For instance, a crowd might be argued to be intelligent if its collective error were lower than the average individual error. However, as per diversity prediction theorem, this is true for any crowd which demonstrates at least some prediction diversity. Thus, to confirm intelligence, we devised a stricter measure. We created two metrics for intelligence, one based on expert-novice performance differences, the other based on statistical reasoning.

Expert-novice differences

The literature on expertise has repeatedly, although not entirely consistently, found evidence of experts outperforming non-experts (often called "novices") by considerable margins, generally measured in multiples. [17] for instance found in a study of computer programmers that, on average, experts completed a software debugging task requiring diagnosis and planning almost twice as fast as non-experts (taking 44.2% less time than novices). Lee et al. [18] similarly found an almost doubling of performance, indicated by reduction in task

completion time to 57.8% for a complex problem solving task. Lee et al. further observed effectiveness increases of between 64.0% and 212.% for a multi-stage competitive bidding exercise. Thus, with expertise raising productivity by a factor of 1.79 (time) and up to 3.13 (effectiveness), we chose to define an expert performance threshold accordingly, namely exceeding 3.13. For convenience, we actually chose a slightly higher value of 3.16 ($\sqrt{10}$) and thus required IE / CE > 10, as both IE and CE represented squared terms. Consequently, we considered the error reduction expressed by the IE/CE quotient to be the "collective intelligence quotient" (CIQ).

Proximity of collective guess to the true value

We also considered a statistical definition of collective intelligence which captured the distance between true value and collective guess. Based on the t-statistics, we would consider a collective guess to be expert-like, if its proximity to the true value (confidence interval) would be so precise that it allowed for no more than p = 0.05 likelihood for this to happen based on pure chance. This formulation does not exactly represent the nature of collective intelligence, as it considers absolute quality of the guess, not the comparison of a collective versus an individual guess. After all, a collective could beat any individual expert, but still be relatively far off the true value for any particular estimate. For example, the crowd in Yahoo's 2008 College Pickem was correct only 77.1% (216 out of 280 guesses), but this was the best performance (http://rivals.yahoo. com/ ncaa/ foot-ball/pickem?w=16), a tie.

Simulation Design

The simulation was used to address the question of prediction quality from a different direction. Namely, we sought to explore the probability of a crowd being able to beat the estimate of an expert in repeated trials. Hence, the simulation system allowed for an adjustment of relative precision of expert and non-expert (e.g., 10-to-1, 20-to-1), as discussed before, as well as the size of the crowd (10 to 999), among other things. In varying these parameters, the objective was to determine significant relationships between crowd size and performance, as well as expert precision and performance. Simulations were run as 100-repetition experiments. For each repetition, the simulated expert would produce an estimate, which would be compared to the average of the n non-experts. Results would be tallied across the 100 repetitions to determine the expert's and crowd win ratios. Furthermore, IE/CE values were aggregated for the computation of an average collective intelligence quotient (CIQ). For each experiment, we replicated the 100 repetitions five times (500 repetitions in total), providing five data points each.

RESULTS

Crowd Guesses

The results of our analysis are captured in Table 1. The absolute values suggest that the crowd of 30 performed quite well. In one of the two crowd guesses, the quotient IE/CE ("collective intelligence quotient" QIC), at 89.6 was considerably higher than 10, our threshold for collective intelligence. In the other scenario, the QIC, at 8.3 approached almost our definition of expert performance.

Table 1. Group Results

Task	x_{true}	$\bar{X}$	s.d.	CE	IE	PD	IE/ CE
Candy Qty	46	44.37	15.62	2.66	238.4	235.7	89.6
Future Temp	29	29.89	2.46	0.8	6.66	5.86	*8.3*

Our more stringent quality assessment, comparing the collective estimates against true values, showed some interesting results. In terms of absolute values, the collectives did in fact quite well, guessing temperatures correctly within one degree (Celsius) and jelly bean quantities within 1.67 and 2.53 beans. A more stringent test of accuracy, however, is to assess whether the true value would fall into a +/-5% confidence interval around the sample mean. In these statistical terms, there performance was not conclusive, as Table 2 shows.

Table 2. Absolute Accuracy of Collectives

Group	x_{true}	$\bar{X}$	s.d.	*t*	*p*
Candy Quantity	46	44.37	15.62	0.573	0.571
Future Temperature	29	29.89	2.46	1.987	0.056

Analyzing the data this way, the 30-person collective performed nearly at p = 0.05 for future temperature, but estimated in a much wider (> 50%) confidence interval for candy quantity guesses.

Simulation Results

Results of the simulation show clear relationships between crowd size and the possibility of the collective to outperform an expert.

Expert precision 10-to-1
When expertise was set at 10-to-1, we found a highly significant linear relationship between the crowd's win ratio (CWR) and the log of crowd size (CS), indicated by an R^2 = .946, with F=490.24 (p=.0000).

CWR_{10} = -.124 + .338 * log(CS).

This equated to a win ratio of 25% with a crowd size of 13, and a 75% win ratio for a crowd of 385.

We also found a less strong but still highly significant relationship between collective intelligence quotient (CIQ) and log of crowds size, indicated by an R^2 = .580, with F=38.81 (p=.0000).

CIQ_{10} = -51.26 + 58.83 * log (CS).

Expert precision 20-to-1
When expertise level was raised to 20-to-1, the likelihood for the collective to prevail expectedly decreased. As a result, the slope of the crowd win ratio function remained about the same, but with an intercept lower by approximately .14. The relationship remained highly significant with an R^2 = .941, and F=446.67 (p=.0000).

CWR_{20} = -.267 + .311 * log(CS).

A win ratio of 25% thus would require a crowd size of 46, and a 75% ratio a crowd size of 1,863.

The collective intelligence quotient CIQ_{20} relationship was much weaker than CIQ10's relationship, with an R^2 of only .121, with F = 3.85 (p = .0599), thus only significant at the 6% level.

CIQ_{20} = -100.54 + 98.89 * log (CS)

Figure 1 depicts crowd win ratios for relative precisions of 10-to-1 and 20-to-1.

Crowd size 999
Our third simulation scenario compared CWR for four different levels of expert precision (10-to-1, 20-to-1, 50-to-1, 100-to-1) with a constant crowd size of 999. Once again, we observed significant results, demonstrating a strong linear relationship between CWR and, this time, the log of precision (P) for the range of simulated precision levels.

CWR_{999} = 2.342 – 1.539 * log(P).

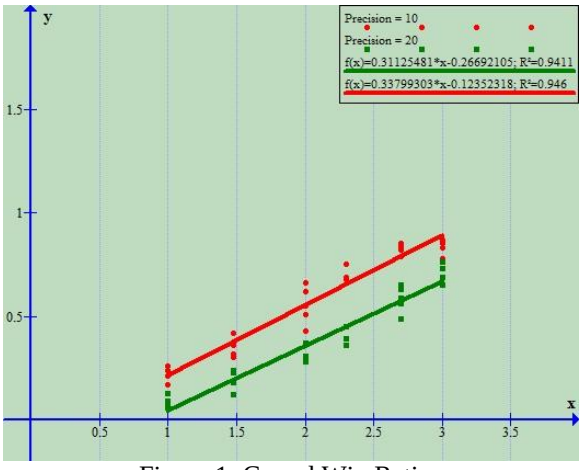

Figure 1: Crowd Win Ratios

The slope of the crowd win ratio function was expectedly negative. The relationship of crowd win ratio CWR_{999} to log of precision was strong and highly significant, with an R^2 of .992, and F = 240.64 (p = .0000). At the same time, we also observed a strong but less significant relationship between CIQ and log of precision, indicated by an R^2 = .828, and an F=9.63 (p=.08).

CIQ_{999} = 40.69 + 90.17 * log(P).

Aggregate measure of CWR
Finally, aggregating all simulation scenarios, we determined a model for the prediction of CWR a follows:
CWR = 0.5649 + .3249 * log(CS) – 1.539 * log(P).

The model accounted for 95% of the variance in the data (R^2 = .952) and was highly significant (F = 587.13, p =.0000), with each model component also being highly significant.

DISCUSSION

Our simulation results corroborate and formalize the findings of the earlier experiment, demonstrating that crowds can achieve performance equivalent to that of experts. Several specific lessons can be drawn.

First, while (real) crowds can achieve surprisingly good performance, approaching the true value of unknown variables at a significant level, is a challenging hurdle. For results to lie in a 5% interval around the true value requires crowd precision that is difficult to achieve. Our crowd did not meet this criterion although coming close once (p = 0.056). A less high hurdle of approaching a CIQ of 10 or higher was easier to meet. Our crowd over-achieved once,

with a CIQ of 89.6 while remaining close the second time, with a CIQ of 8.3. Yet even with a CIQ of 8.3, the crowd performed surprisingly well, guessing the future temperature in their city within 3% accuracy (29.89 degrees guessed vs. 29 degrees actual).

Second, crowd size matters significantly. While we earlier pointed out that wisdom of crowds is not simply a law of large numbers effect, a larger crowd can formulate more alternate models for its forecasts, and at the same time can also benefit from replication of the same modeling approach. Taken together, the impact is considerable. Thus, based on our simulation results, a crowd of 30 competing against a 10-times more precise expert would statistically outperform the expert in about 38% of all cases, while a crowd of 999 under the same conditions should have an 89% win ratio.

Third, despite the considerable impact of collective intelligence, the value of experts cannot be denied. For example, our findings for a large crowd of 999 reveal that when the expert has a precision advantage of about 15.7, expert and crowd performance will be comparable (50% win ratio each). Further, at a precision advantage of about 22.9, the expert would outperform the crowd in 75% of all situations. In other words, crowds may perform well and approach the true value of unknown variables in their estimates, yet true experts may still perform better, albeit the difference may not be highly important. Furthermore, the crowd may have an advantage over the expert, in that the notion that results were drawn from the aggregation of the views of many is confidence building. A single expert, when pressed for an estimate in a high stakes situation (even in a game show such as the "Who Wants to be a Millionaire") may develop self-doubts leading to failure and under-performance.

Finally our simulations also indicate that QIC, the collective intelligence quotient is a useful abstract measure of crowd performance over the average individual, but not a good predictor of crowd superiority over an expert. While high QIC values correlated with high crowd performance, the relationship between QIC and precision or crowd size was only weakly significant, according to our simulation characteristics.

IMPLICATIONS

It may be argued that point estimates as we used in our empirical and simulation studies are of interest academically, but are of little relevance to the practice of management. In our opinion, this is not so. Applications of collective intelligence are widely present in the financial markets. Earnings forecasts and targets for stock prices for instance are typically released by citing the average mean estimates of individual analysts as well as their range. Markets in general can be seen as a form of (imperfect) collective intelligence at work. Price discovery in markets is driven by supply and demand of a large number of market participants. Hence, they do draw together the perceptions of many with respect to the value of the respective asset class. Accountants often use prices obtained from forward or futures markets to value contracts on future in- or outflows of commodities or currencies. Risk managers of many international corporations will use forward and options prices to assess the risk of the company's commodity and currency positions and decide on hedging needs.

Of particular importance however is that the design of markets influences both the participation level and the weight of the collective. Therefore traded prices may not always be representing the best estimate for the fundamental price. The design of the market varies according to what is being traded and on the identity and size of the buyers and traders. Assets with broader demand and with a greater degree of uncertainty in their value (e.g. stocks and commodities) are typically traded on a single or a few exchanges to capitalize on the resulting concentration of different opinions. In contrast, securities with lesser degrees of uncertainty (e.g. government bonds) or smaller participation (e.g. exotic options and currencies) tend to be traded in distributed networks or over the counter.

To illustrate the importance and impact of the design structure of markets, we contrast two popular market types: an auction market and an order-driven market. An auction market, by design, concentrates all demand for a specific good or number of goods such that highest bidders win and the prevailing price is set at or near the top bid price (or the lowest seller). Several auction forms exist. In English or open ascending price auctions (e.g., eBay), participants bid openly against one another, with each subsequent bid higher than the previous bid. In Dutch or open descending price auctions (popular for perishable goods such as flowers or fresh fish), an auctioneer begins with a high asking price which is lowered until a bidder accepts the auctioneer's price. Dutch auction prices therefore tend to reveal the opinion of participants on one side of the spectrum, the highest bidders, only. Not only goods are sold in this way,

US Treasury bills and bonds also find buyers via a Dutch auction.

Standardized commodities as well as financial products such as stocks and futures are often traded in order driven markets. In order driven markets the prices are continuously driven by buy and sell orders surrounding current market prices. Here each participant's opinion is represented by a simple buy or sell order specifying quantity and price, which does not necessarily reflect the "true value" of the commodity or security but rather his willingness to pay (or the willingness to give up the asset). A participant who thinks the price is too low by 50% would have a similar incentive to buy as a participant who thinks the price is off by 10%. Both participants would buy at the same current price, although they would differ in opinion about the fundamental price. Therefore their difference in opinion is lost within the price discovery process. At the same time, order book information can be available to other market participants and help in the determination of a collective estimate of the "true value". Unfortunately, because of limited resources and the continuous nature of the market, buyers and sellers will not be heard of again in the price setting process until they revise their opinion, the market price reaches their preset targeted level or they need to trade out of necessity, thereby revealing their opinion about the true price level only sporadically when effectively placing an order. Therefore, in order-driven markets, while the order book could reveal the crowd's estimate of the true value, some buyer and seller price opinions remain tacit. Thus, supply and demand are typically only seen around the current price. The entire supply and demand curves including price opinions which are remote from current 'average' prices will not be visible within the order book and collective intelligence cannot be fully displayed. In addition, because of the continuous nature of this market, supply and demand are time-sensitive. Large demand or supply shocks originating from liquidity traders can temporarily shift the market price from the true value.

One striking example of how the design of markets and participation level is essential in price discovery is the price difference of a stock between its initial public offering (IPO) and its first day trading on an exchange. If the design of the market did not matter, then resulting prices should not differ by much. For the price discovery at the Initial Public Offering often a book building system is used, where a large weight is assigned to a few experts (typically large institutional investors) and a small weight assigned retail investors). For the price discovery in the first trading day on an order-driven stock market, in contrast, all market participants play a role, albeit with weights depending on their initial investment. This transition from one market design to the other often results in large price jumps, with initial first day returns more often than not in the double digit range. In essence, the transition reflects a price determination by "expert" market makers to price determination by the crowd. The results are often windfall profits for favored parties. This apparent misjudgement on demand and price has led to some companies to experiment with alternative pricing systems at the the IPO stage. For instance, a Dutch auction like market was used for Google's IPO, allowing a price determination process which more closely reflected the collective wisdom and therefore was expected to result in less volatility and price jumps on the first trading day [19]. In contrast to predictions, Google's first day return was still 18% and volatility remained high as the stock reached an intra-day high of USD140 from its initial offering of USD85.

While volatility can be seen as a result of unsettled differences of opinion between market participants, it arguably also has a crucial role in price discovery [20, 21] and soliciting of opinions. During periods of high volatility the price moves rapidly over a large range, therefore inviting the opinion of broader strata of participants who are attracted to specific price ranges. The increasing number of participants in turn can induce further price changes. This effect might be an additional explanation why volatility is clustered and sudden spikes in volatility are followed by periods with above average volatility.

CONCLUSIONS, LIMITATIONS, FURTHER RESEARCH

Our study confirms several notions of collective intelligence, and extends them through the use of simulation studies. Thus we find that the answer to "how good are collectives" is that they are clearly better than (non-expert) individuals, and that they can be reasonably good in approaching true values (although not supported by statistics), as long as important conditions are met.

Our study has numerous limitations which create ample opportunity for further work. Specifically, in our simulations, we assumed uniform distribution of guesses. Preliminary test did not indicate significantly different outcomes when both expert and non expert guesses were normally distributed. However, further exploration, especially with differing expert and non-expert distributions functions may be revealing.

The boundaries to our simulation results need to be explored further. While we found linear relationships, win ratios are obviously bounded at both low and high end (0.0; 1.0), leading to non-linearities in the CWR function as values approach these boundaries.

An extension of the study into the performance of collectives of experts versus collectives of non-experts will be valuable. As pointed out in the prior section, for instance in financial analysis, the estimates of analysts are usually aggregated and made available to the general public. This pits a collective of experts against a much larger collective of non-experts that make up the investor community. Consequently we might ask whether a group of 10 or 20 analysts following a stock, or some expert book makers, are better in-aggregate in predicting future performance than a crowd of thousands that shares its estimates in stock market related communities. Evidence from IPO scenarios such as Google's suggest that this may not be the case.

Finally, our measures of collective intelligence, specifically the "collective intelligence quotient" IE / CE raise questions. Determining appropriate measures and thresholds for collective intelligence is important, as is their interpretation in light of the peculiarities of statistics. After all, with current measures, one guess can be "better" than another not because it is closer to the true value, but because its internal variance (diversity) is higher. We saw that average QIC values in the simulation scenarios were only loosely related to other performance measures. Further research will hopefully help us to understand the differences between the quality of the process and the quality of the guess itself.

ACKNOWLEDGEMENT

This research was supported in part by GRF project No. 9041464 and by Strategic Research Grant No. 7002346 from the City University of Hong Kong.

REFERENCES

1. Kahneman, D. & Tversky, A. (1973). "On the psychology of prediction," *Psychologyical Review*, 80, 237-251.
2. March, J.G. (1978). "Bounded Rationality, Ambiguity, and the Engineering of Choice " *The Bell Journal of Economics*, 9, 587-608.
3. Oliven, K. & Rietz, T.A. (2004). "Suckers are born but markets are made: individual rationality, arbitrage, and market efficiency on an electronic future market," *Management Science*, 50(3), 336-351, 2004.
4. Lamont, E.B. (2008). "Foreseeing: Formulating an accurate prognosis," *Prognosis in Advanced Cancer*, P. Glare and N.A. Christakis, eds., Oxford: Oxford University Press.
5. Ravahi, S.-M. (2002). "Systematic biases in human cognition," in *IEEE International Conference on Systems*.
6. Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*, New York: Anchor.
7. McGrath, J. (1984). *Groups: Interaction and Performance*, Englewood Cliffs, NJ: Prentice-Hall.
8. Janis, I.L. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*, Boston: Houghton Mifflin.
9. Argote L., Devadas, R. & Melone, N. (1999). "The base-rate fallacy: Contrasting processes and outcomes of group and individual judgment," *Organizational Behavior and Human Decision Processes*, 46(2), 296-310.
10. Tuckman, B. (1965). "Developmental sequence in small groups," *Psychological Bulletin*, 63, 384-399.
11. Thomas, R. (1809). *The (Old) Farmer's Almanack*. Boston: Boston: E.G.House.
12. Page, S.E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton: Princeton University Press, 2007.
13. Dalkey, H. (1969). "The Delphi Method: An Empirical Study of Group Opinion," *RAND Report RM-5888-PR*.
14. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. & Riedl, J. (1994). "GroupLens: An open architecture for collaborative filtering of netnews," *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 175-186.
15. Kittur, A., Chi, E., Pendleton, B.A., Suh, B. & Mytkowicz, T. (2006). "Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie," *Proceedings SIGCHI*, 1-9.
16. Wolfers J. & Zitzewitz, E. (2004). "Prediction Markets", *The Journal of Economic Perspectives*, 18(2), 107-126.
17. Vessey, I. (1986). "Expertise in Debugging Computer Programs: An Analysis of the Content of Verbal Protocols," *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-16(5), 621-637.

18. Lee, Z., Wagner, C. & Shin, H. K. (2008). "The effect of decision support system expertise on system use behavior and performance," *Information & Management*, 45, 349-358.
19. Sherman, A.E. (2005). "Global trends in IPO methods: book building versus auctions with endogenous entry", *Journal of Financial Economics*, 78(3), 615-649.
20. French, K.R. and Roll, R. (1986). "Stock return variances: The arrival of information and the reaction of traders," *Journal of Financial Economics*, 17, 5-26.
21. Ross, S. (1989). "Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy," *Journal of Finance*, 44, 1-17.