Issues in Information Systems
Volume 16, Issue IV, pp. 209-214, 2015

OPENING THE "BLACK BOX" OF COLLECTIVE INTELLIGENCE: A CRITICAL REFLECTION

Christian Wagner, City University of Hong Kong, c.wagner@cityu.edu.hk
Ayoung Suh, City University of Hong Kong, ayoung.suh@cityu.edu.hk

ABSTRACT

Challenging and complex problems may be solved better by a collective with knowledge or information drawn from a broad range of perspectives. While the ability of collectives to perform better than individual experts is well recognized (e.g., [18]), the results are not unequivocal. Collectives sometimes perform better ("wisdom of crowds") and sometimes worse ("madness of the masses"). Thus the phenomenon requires a further understanding of (1) why collectives outperform individuals, (2) which problems this strength is best applied to, and (3) how this strength can be leveraged further. The article seeks to answer these questions, by "opening the black box" of collective problem solving and explaining mechanisms for solution finding, the impact of problem difficulty (perceived and actual) and of collective size. We argue that by applying these insights, a cognitively complex problem solving task, if suitably decomposed, may be relatively easily solved by non-expert collectives, just as a suitably decomposed mechanical task can be easily completed by non-experts working in industrial environments.

Keywords: Information Technology (IT), Collective Intelligence, Wisdom of Crowds, Problem Decomposition, Law of Large Numbers, Virtual Work

INTRODUCTION

Collective intelligence describes a phenomenon, whereby non-experts' aggregate judgments provide higher quality outcomes than those generated by individuals, even when these individuals are experts. This was illustrated for instance in the "jellybean jar" experiment (e.g., Scoles [14]) where the average estimate for the bean count, drawn from a number of people, exceeded the best individual estimate. In other words, not even the best performer was able to beat an average that included wild guesses and outliers. This phenomenon can be observed widely. For example, prediction markets were repeatedly able to predict delays in Boeing 787 delivery, months before the CEO knew [20]. Collectives in such settings differ from traditional groups on a number of characteristics. Most importantly, whereas groups are often described as "social aggregates that involve mutual awareness and potential mutual interaction" [12], and are consequently relatively small and organized, collectives lack this awareness and interaction. By contrast, collectives, as defined for instance by Jeppesen and Lakhani [9], must meet three requirements for collective intelligence to emerge [8], the possession of multiple viewpoints (diversity), opinions that are based on the opinions of others in the collective (independency), and separate knowledge sources to draw in in order to form judgments (decentralization). Collective intelligence, as a tool for management to improve judgment, decision making, and problem solving, has become possible through the combination of the Internet, providing the reach to a far-flung collective, and computing power to quickly aggregate individual responses into collective insights. These aggregate judgments then produce of higher knowledge validity than most or all individual contributors. But why so? Opening the black box of collective intelligence, by explaining the reasoning processes of individuals in the collective, the aggregation of individual heuristics, and the role of task difficulty in altering reasoning processes, is the purpose of this article.

COLLECTIVE INTELLIGENCE PRINCIPLES

Law of Large Numbers vs. Aggregation of World Views

Collective intelligence is not simply an outcome of the Law of Large Numbers [6]. The Law of Large Numbers suggests that results improve when the number of observation grows, because error terms average each other out, thus reducing the error term ε in a function such as $y = ax + b + \varepsilon$ through aggregation. Positive outliers cancel out

negative ones. People's imprecise or somewhat incorrect guesses can clearly be improved through error reduction, so the Law of Large Numbers should have some impact. Yet this does not explain the main quality of "crowd wisdom". Instead, collective intelligence appears to emerge from an aggregation of partial, imperfect estimation heuristics of many. Let's take for instance an estimation function that predicts the price of a condominium. To illustrate, we assume a two-factor prediction model for values, conisting of a component for the size of the house and the number of bedrooms, resulting in a value computation such as $V = a_s \times size + a_b \times bedrooms$. Members of the collective who are not experts in property valuation may instead use single factor prediction models, considering either the size or the number of bedrooms only. By aggregating the individual models, however, the combined model would incorporate both factors. Figure 1 demonstrates this with a numeric example, where the "true" value function is relatively closely approximated through the aggregation (V_{12}) of two single-factor estimation functions.

Whereas it may seem unlikely that individuals could consider simple one-factor models only, Shanteau [15] demonstrates that individuals single- or few-factor models. Aggregating simple, but diverse models would invariably lead to a more precise multi-factor factor model, and higher precision.

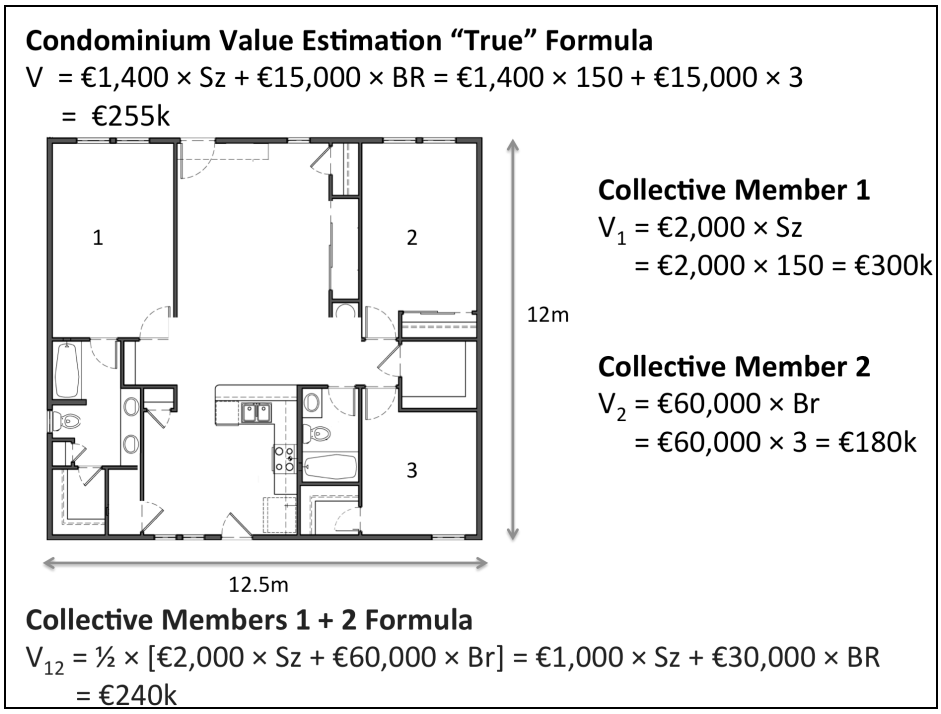

Figure 1. House Value Estimation Model

The use of such simple linear models in judging and decision making has been recognized repeatedly (e.g., [1]), but not explicitly in the context of collective intelligence. Page [13], however, has argued for the aggregation of models, and Wagner and Suh [22] have offered evidence for the use of simple models by collectives.

Underlying Assumptions

For collective intelligence to emerge, several task related criteria have to be met. First, the outcome which the collective is judging cannot be random. Asking a collective for the expected result of a (random) coin flip, for instance, will not predict the actual outcome, but at best a prediction reflecting the probabilistic outcome (e.g., 50/50 percent heads/tails prediction). Next, individuals need to have some reasoning capability and information concerning the task. The aggregation of random guesses generates no additional information. Needed is the ability to eliminate at least some outcomes, so that multiple processes of elimination of unlikely outcomes, combined, let the best result(s) emerge as most likely, or shift the aggregate estimate towards the true value. Referring back to our earlier example in Figure 1, let us assume that two individuals jointly are asked to estimate a property price for a 3-bedroom house of 150m^2, whose actual value is €255k. Neither of them knows the right answer. One, however is

quite certain that the value is not less than €180k, the other that it will not exceed 300k. Together they can bracket the value between €180k and €300k (with their average being 240k).

Finally, members of the collective must address the problem using different reasoning approaches. In other words, the collective needs diversity, with members drawing on different information sources, and using different heuristics, independent from other members of the collective. In our example, the first person might have observed that properties in the area rarely ever sell for more than €2,000 per m^2. The second person might "know" (believe) that houses sell for at least €60k per bedroom. With different individuals applying different judgment heuristics to the task, the aggregate estimation heuristic becomes more robust, in our example leading to a combined formula of V_{est2} = ½ × [€2,000 × size + €60k × bedrooms] for just two people. While each partial heuristic is able to predict only part of the variance in the observed outcome, the aggregate explains increasingly larger amounts of variance, thus improving precision. Thus, collective intelligence is not the cancellation of errors predicted by the Law of Large Numbers, but instead an aggregation of function terms [7]. Multiple collective members with similar partial models could be considered as more "votes" for one or the other partial algorithm, thus shifting the overall result to a differently weighted average of these partial algorithms. For example, if three individuals believed in value calculation based on size, and only one in calculation based on the number of bedrooms, the resulting algorithm would become V_{est4} = ¼ × [3 × €2,000 × size + €60k × bedrooms] for a final aggregate V_{est4} = €1,500 × size + €15k × bedrooms. Thus, a larger collective would help to rebalance the weight of the functional terms, in addition to creating a more complex model and neutralizing random errors.

Degredation of Collective Intelligence: Madness of the Masses

Collective intelligence will not always produce superior results than individuals or groups would. It is already well understood that for tasks that are purely random (e.g., coin flips), or where information is virtually unavailable (e.g., Papal elections), collectives will not provide any performance improvements. For example, if people estimated house values based on the house color (e.g., green = €200k, red = €400k), because they had previously seen a red house selling for €400k and a green one for €200k, the outcomes would, individually or in aggregate, bear little resemblance to true values.

With increased randomness in the prediction algorithms or lack of information, collective predictions will degrade, lose their model-based character, and worst-case, introduce a systematic bias which results in systematic under- or over-estimation, leading to a "madness of the masses" effect. Evidence for this phenomenon is provided by Wagner and Suh [22]. Wagner and Suh found individual guesses for very uncertain events to be apparently random and highly inaccurate. When systematic errors are introduced into such situations, e.g., through collective mis-information, or through anchoring on a single, bad estimate, the overall result is not only inaccurate, but also biased and thus "mad".

In essence, we posit that the bias revealed in systematic mistakes of collective judgments is the most problematic aspect of collective intelligence, so much so, that avoidance of such bias, rather than the improvement of individual precision, is the most important aspect of seeking insights from collectives.

Effect of Task Difficulty

Task difficulty is an important factor for collective intelligence. After all, if tasks become difficult to the point where nobody can even provide a partially reasoned answer or single-variable model, we expect performance degredation similar to that for random phenomena. Individuals will simply guess an outcome without reason, thus leading to noise or systematic error, but no performance gains. At the opposite end, for simple tasks, most or all individuals will provide good guesses. With individual guesses already being good, aggregation will lead to only small performance improvements, and not yield significant advantages for the collective. Hence, we posit the existence of a "sweet spot" of collective performance in the area of medium difficulty. More precisely, collectes are expected to demonstrate the largest performaance advantage for medium difficulty tasks, since at medium difficulty, individuals perform with some non-random ability, but can still benefit from aggregation. This finding stands in direct oppostion to past research, which conjectured that collectives would perform either particularly well on unfamiliar, difficult tasks, or on familiar and thus easy tasks [3, 10]. Our research, e.g. [21], places the sweetspot between the very easy and the very difficult.

Defining difficulty, however, is in itself a challenge. Difficulty may be determined by the number of relationships between problem elements, the complexity of these relationships, and the amount of information available (cf. Simon [17]). At the same time, *perceived* task difficulty can be significantly affected by the format in which it is presented. A complex problem can be presented as several decomposed sub-problems (problem shaping), each structured enough to be within the grasp of the members of the collective [16]. Simon illustrated his arguments concerning complexity with the unequal ability of two watch makers. Both build watches of overall identical complexity (number of parts), but one uses sub-assemblies and thus becomes more successful in task completion. Based on this logic, knowing how to decompose can effectively reduce task complexity to those working on the subtasks and enable completion, which might otherwise have been impossible. For example, estimating the monthly cost of living in a large, metropolitan city such as Hong Kong, may be considered difficult by many who don't know the city. Such a problem will become easier, though, when asked for partial costs, such as the cost of housing, cost for food, cost for clothing, transportation, and so on. Decomposition allows more specialization, better contextualization, and thus better use of information. Each individual task is therefore better defined and thus easier, yet in aggregate, their result will provide the desired answer. Judgments may likewise be simplified by restricting the solution set, such as by setting upper and lower bounds (solution shaping). Problem and solution shaping are well known techniques in task completion and even problem solving (e.g., Dalkey and Helmer [2]), but there exists little understanding on how to shape and what the positive or negative impacts of task shaping might be.

IMPLICATIONS

Collective Intelligence Rationalized

Our reflections into the nature of collective intelligence convey four important messages. First, collective intelligence is not a trivial outcome of the Law of Large Numbers, that is, not simply the reduction of error terms through aggregation. Second, collective intelligence is not a mysterious property that emerges when collectives come together, but is explained through the aggregation of partial explanatory models, which taken together explain increasing amounts of variance. Third, because collective intelligence requires the existence of partial models whose aggregation leads to improvement, collective tasks cannot be too easy (allowing everyone to formulate complete models) or too difficult (preventing anyone from creating a partial model). Fourth, because collective intelligence relies on the weighted aggregation of partial models, diversity and independence within the crowd are required to generate a rich, unbiased aggregate model. Earlier accounts (e.g., [18]) did explain conditions under which collective intelligence emerges, but not the reasons or mechanisms that bring this form of intelligence about. Without a deeper understanding, questions about the collective performance and collective failure (madness of the masses) cannot be answered. In this article we stressed the use of imperfect heuristics, which, when aggregated across individuals, produce meaningful results. Results become more meaningful as they allow the consideration of a larger number of factors and more information, than any individual, even an expert, could consider alone. Furthermore, the number of replications of individual heuristics within the collective leads to a proportional balancing of components in the aggregate heuristic (as illustrated in the V_{est4} property valuation formula), therefore enabling the aggregate to explain additional variance. Thus, by explaining the imperfect algorithmic nature of solution finding, we can also explain the need for large enough, but not infinite size collectives, the degredation of performance when tasks become too difficult, and mechanisms to enable collectives even on difficult tasks.

"Industrialization" of Problem Solving

Task performance and task difficulty are quite well understood in our mechanical, industrial world. There, under the logic of scientific management [19], production managers apply mechanisms that enable collectives of low-skill workers to jointly produce highly complex products, such as digital cameras or computers. These products are of such complexity, that most people, even highly educated Masters or PhD degree holders, would not be able to replicate their build, although it is carried out flawlessly by a low-skill factory work force.

We need to apply these industrial level techniques to the production of intellectual outcomes. In other words, by decomposing difficult intellectual tasks into smaller sub-problems, we may be able to source them to collectives of non-experts, each of which can provide a limited answer, which in aggregate lead to a very good solution, outperforming individual experts. The research we described above provides the parameters with which to undertake this effort. Tasks will need to be solvable by heuristic approaches. We need to decompose those tasks to reduce their

difficulty to manageable levels for the problem solvers. Problem solvers will need to be independent in their activity and able to use diverse sources of information. And collectives must be large enough to enable the explication of several different perspectives, while also repeating perspectives (and introducing redundancy), thus driving down the effects of random errors.

Wikification of Knowledge Creation

We see some form of "industrialized" problem solving already in the wikification of some knowledge creation tasks. For example, the largest English language encyclopedia, Wikipedia, has been co-authored as a virtual work project by a collective of over one million people, each of which possessing only a small fraction of the entire knowledge now contained in Wikipedia. In fact, editors need very little knowledge, sometimes only enough to correct a minor fact, such as a wrong date. Then, taken together, all contributions create a knowledge construct whose accuracy exceeds levels any individual could ever produce (cf. Giles [4]). Wiki collectives also use techniques for problem difficulty moderation, as outlined for instance by Majchrzak et al. [11]. Whereas many wiki participants add knowledge, some engage in wiki shaping, thereby decomposing and restructuring articles, so that it becomes less difficult for wiki adders to contribute their knowledge.

LIMITATIONS AND FUTURE WORK

This research article draws in its analysis on prior empirical work by the authors and other researchers. Therefore, questions arising from that work can be asked, but remain unanswered. Specifically, we can identify several limitations, which should become the impetus for future work. First, we need to explore the relative importance of the Law of Large Numbers versus the impact of partial model aggregation. For example, if a prediction model required ten independent variables, then a collective of hundreds would appear to largely serve as a mechanism to drive our random errors through large numbers. At the same time, the collective size may also be important to properly weigh the individual components in the prediction model. Thus, future work will need to explore the impact of collection size, through an experimental approach where subjects also explain their predictive reasoning, not simply make predicitions. Next, the work focused on decomposition and task complexity, but did not provide any advice on how these may be assessed ex-ante and then meaningfully moderated. Thus, future work must analyze the causes of problem difficulty, both from an informational and a structural perspective, and ideally find guidelines to ex-ante determine problem difficulty, as well as simplification strategies. Third, whereas this research relied on assumptions for collective intelligence that included diversity and independence, collective tasks frequently limit diversity and independence, yet still produce meaningful results. For example, prediction markets [5] allow "investors" to make decisions based on the current market price of a commodity, which reflects all prior beliefs. Similarly, wiki work shows any contributor the aggregate insights of all prior contributors. Future research thus must explore the effect of diversity and independence reduction on prediction outcomes.

CONCLUSIONS

Drawing on recent research findings, the article has offered explanations that "open the black box" towards a better understanding of collective intelligence, its effects on knowledge validity, and its underlying mechanisms. By recognizing that members of the collective engage in imperfect heuristic reasoning, it becomes clear that aggregation leads to better heuristics, but also that when task difficulty is too high, the heuristics used will be so far from perfect (i.e., random guessing") that aggregation yields little improvement. Improving collective intelligence thus needs to strive towards improving individual performance, to avoid random guessing, and thus also the possible introduction of biases. We propose the moderation of task difficulty as a very promising direction. Our research also leads to the conclusion that the biggest threat for collective intelligence is not random error by some individuals, but systematic bias within the collective. Any random error can be compensated for by increased collective size, yet judgement biases will be reinforced as the size of the collective grows.

ACKNOWLEDGMENT

This research was supported in part through GRF Project No. 194613 awarded to Christian Wagner and through GRF Project No. 21500714 awarded to Ayoung Suh by the Research Grants Council of the Hong Kong SAR.

REFERENCES

1. Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American Psychologist*, 26, 2, 180-188.
2. Dalkey, N. and Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, 9, 3, 458-467.
3. Farnsworth, P.R. and Williams, M.F. (1936). The Accuracy of the Median and Mean of a Group of Judgments. *Journal of Psychology*, 7, 237-239.
4. Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, 438*,* 7070, 900-901.
5. Hall, C. (2010). Prediction markets: Issues and applications. *Journal of Prediction Markets*, 4, 1, 27-58.
6. Hazewinkel, M., ed. (2001), "Law of large numbers", *Encyclopedia of Mathematics*, Springer.
7. Heath, C. and Tversky, A. (1991). Preference and Belief: Ambiguity and Competence in Choice under Uncertainty. *Journal of Risk and Uncertainty*, 4, 5-28.
8. Hueffer, K., Fonseca, M., Leiserowitz, A., and Taylor, K. (2013), The Wisdom of Crowds: Predicting a Weather and Climate-related Event. *Judgement and Decision Making*, 8, 2, 91-105.
9. Jeppesen, L. B., Lakhani, K. R. (2010). Marginality and Problem-solving Effectiveness in Broadcast Research. *Organization Science*, 21, 5, 1016-1033.
10. Klugman, S.F. (1945). Group Judgments for Familiar and Unfamiliar Materials. *The Journal of General Psychology*, 32, 103-110.
11. Majchrzak, A. Wagner, C., and Yates, D. (2013). The Impact of Shaping on Knowledge Reuse for Organizational Improvement with Wikis. *MIS Quarterly*, 37, 2, 455-469.
12. McGrath, J.E. (1984). *Groups: Interaction and performance*. Prentice Hall.
13. Page, S. E. (2007). *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Princeton University Press.
14. Scoles, S. (2007). The Wisdom of Crowds—A Better Way to Think About the Markets. *Risk and Rewards*, August, 10-13.
15. Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, 81, 1, 75-86.
16. Simon, H.A. (1962). The Architecture of Complexity. *Proceedings of the American Philosophical Society,* 106, 6, 467-482.
17. Simon, H. A. (1977). Causal ordering and identifiability. In *Models of Discovery,* 53-80. Springer Netherlands.
18. Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.
19. Taylor, Frederick Winslow. *The principles of scientific management*. Harper, 1914.
20. Thompson, D. N. (2012). *Oracles: How Prediction Markets Turn Employees into Visionaries*. Harvard Business Press.
21. Wagner, C. and Suh, A. (2013). The Role of Task Difficulty in the Effectiveness of Collective Intelligence. *Proceedings DEST Conference*, Stanford, USA.
22. Wagner, C. and Suh, A. (2014). The Wisdom of Crowds: Impact of Collective Size and Expertise Transfer on Collective Performance, *Proceedings HICSS-47*.