The Downside of Expertise: Does Domain Knowledge have a Negative Effect on the Quality of Crowdsourced Data?

Shawn Ogunseye
Memorial University of Newfoundland
osogunseye@mun.ca

Jeffrey Parsons
Memorial University of Newfoundland
jeffreyp@mun.ca

ABSTRACT

Subject matter expertise is widely believed to have a positive effect on information quality in crowdsourcing. Many crowdsourcing systems are therefore designed to seek out contributions from experts in the crowd. We argue that expert contributors of data in crowdsourcing projects are proficient rule-based classifiers, and are efficient because they attend only to attributes of instances that are relevant to a classification task while ignoring attributes irrelevant to the task at hand. We posit that this selective attention will negatively affect the tendency of experts to contribute data outside of categories anticipated in the design of a class-based data crowdsourcing platform. We propose hypotheses derived from this view, and outline two experiments to test them. We conclude by discussing the potential implications of this work for the design of crowdsourcing platforms and the recruitment of expert versus novice data contributors in studies of data quality in crowdsourcing settings.

Keywords

Crowdsourcing, data quality, classification, expertise

INTRODUCTION

Crowdsourcing systems are an increasingly popular tool to collect data outside of traditional organizational settings (Castriotta & Guardo, 2011; Hosseini, Phalp, Taylor, & Ali, 2014; Tarrell et al., 2013; Tripathi & Tahmasbi, 2014). A claimed advantage of crowdsourcing is the possibility to broaden the base of potential data contributors. At the same time, some crowdsourcing applications (e.g., citizen science) demonstrate a preference for expert contributors, to ensure that information collected is of high quality (Budescu & Chen, 2014; Roman Lukyanenko, Parsons, & Wiersma, 2016). This preference reflects an underlying assumption that there is a positive correlation between a contributor's level of expertise in a given domain and the quality of information the contributor provides. In other words, there is at least a tacit belief in many cases that *the expert crowd is the better crowd*.

A popular strategy to constrain data collection (in both traditional and crowdsourcing applications) is the use of class-based abstractions to represent, collect and store data. These require contributors to report data according to a predetermined and fixed set of abstractions. For example, a common strategy in citizen-science crowdsourcing in a natural history context is to require participants to identify the *species* of observed organisms when reporting data. This strategy limits the participation of some crowd members, being appropriate only for those who share the schema imposed in the data collection interface. It excludes users who do not understand the schema or, worse, forces them to guess and contribute possibly erroneous data (Lukyanenko et al., 2014). Another strategy is to explicitly or implicitly identify experts in the crowd and only process contributions from identified experts. In other cases, contributors may be trained to the required level of expertise to perform the classification task (e.g., www.galaxyzoo.org). These restrictive strategies, which are intended to ensure that data collected are of sufficient quality for intended use(s), may be detrimental to the acquisition of unanticipated information – a desirable feature in distributed knowledge crowdsourcing projects (Parsons & Wand, 2014).

Parsons, Lukyanenko, & Wiersma (2011) present a more flexible alternative to data collection, which entails limiting the use of classes (abstractions) and focusing instead on instances and their attributes in order to improve the IQ of the crowd, especially among those lacking expertise in the domain of the task. In the citizen science context, they argue that compared to class-based designs, instance-based designs will foster usability for contributors and engender novel discoveries and uses of data for different consumers. The potential trade-off however, is an increase in data that may not align with predetermined goals of a crowdsourcing application (such as collecting data about the prevalence and distribution of known species in a geographic location).

We argue that, to fully appreciate the ramifications of alternative approaches to the design of distributed knowledge crowdsourcing systems, consideration must be given to not just the intended, positive consequences of strategies implemented to elicit information from experts, but also the potential negative, unintended ones. This follows from the "surprising" findings of a number of studies involving distributed knowledge projects where there was no substantial difference between the performance of experts and non-experts, when the design of the information system (IS) was not restrictive. For instance, examining three groups of participants (assigned into groups by their levels of expertise) in the detection of malicious data amongst a measure of "good" data, researchers found that, contrary to their expectations and the projections from prior research, participants' level of expertise had no significant influence on their performance (Biros, George, & Zmud, 2002). Also, by reconceptualising and redesigning a crowdsourcing system to focus more on enabling "lay" information contributors provide content in ways familiar to them – via attributes of instances rather than classes based on inclusion rules – the knowledge requirement for making contributions was reduced and the quality of information contributed by non-experts improved (Lukyanenko, Parsons, & Wiersma, 2014; Parsons, Lukyanenko, & Wiersma, 2011). Similarly, there was no statistically significant difference between the performance (accuracy) of experts and non-experts in the reported experiment (Lukyanenko, Parsons, & Wiersma, 2014).

An even more recent study compared the accuracy of reviews of many expert movie critics and ordinary viewers at predicting movie acceptance and their earnings. In a system that had predominantly relied on experts, this study showed that not only do non-expert reviewers better predict movie market success, even the review of a small crowd of 40 people was as accurate as the reviews of 30 of the best critics in the industry (Escoffier & McKelvey, 2015).

The reason(s) for and consequences of such findings are of increasing interest as open information environments (OIEs) – informal communities in which "data are not necessarily created or used within, by, or for, a formal organization" (Parsons & Wand, 2014) – have become more pervasive. Researchers and practitioners need to understand the extent to which expertise improves (or decreases) the IQ of crowdsourced data. To this end, we first examine the methodologies used in estimating the level of expertise of contributors and therefore crowds. We then present cognitive principles based on classification theory and their consequences for understanding the impact of expertise on performance in crowdsourcing tasks. Subsequently, we describe a proposed laboratory experiment to investigate the impact of expertise on IQ in crowdsourced data. Finally, we discuss the potential implications of our experiment on research in the design of OIEs.

MEASURING EXPERTISE

Expertise has been used to operationalize the presence of knowledge (Stein, 1992), and is related to competence, familiarity, and job experience (Schultze & Leidner, 2002). A contributor's level of expertise, compared to others in the crowd – termed "relative expertise" in Budescu & Chen (2014) – can be measured objectively through the assignment of weights to contributors' level of education, seniority, professional status and historical track record (e.g., Biros et al. 2002). Expertise can also be measured subjectively using ratings provided by the experts themselves or by others like their peers or superiors (e.g. Lukyanenko, Parsons, & Wiersma, 2014). However, assigning weights and judging a contributor's expertise based on their performance on tests similar to the intended task is claimed to be a more efficient approach (Clemen, 2008; Davis-Stober, Budescu, Dana, & Broomell, 2014; Lin & Cheng, 2009). Accordingly, we adopt the idea of performance-based measurement and definition of expertise. In the next section, we explore classification theory as our overarching guide.

THEORETICAL FOUNDATION AND HYPOTHESES

Classification Theory

Classification (or categorization) is a fundamental human capability. We classify to make efficient use of our cognitive resources by organizing our existing knowledge about phenomena mainly through their similarities, which allows us to make predictions about new instances and events (Best, Yim, & Sloutsky, 2013; Parsons & Wand, 2008, 2013). Classes are therefore useful abstractions of the similarities of the classified phenomena. Classification theory and its import to information system (IS) analysis and design have been extensively discussed (e.g., Parsons, 1996; Wand, Monarchi, Parsons, & Woo, 1995). Classes are considered from the perspective of either the inclusion rules that form them (rule-based classification) or their instances (exemplars and prototypes) (Kloos & Sloutsky, 2008; Murphy, 2002).

In rule-based classification, it is necessary to pay selective attention to only relevant (or diagnostic) features for identifying instances of the class, while irrelevant features (not useful for predicting class membership) can be safely ignored. Although selective attention leads to efficient learning, especially when making connections between instances with very sparse similarities and dense dissimilarities, it has costs. The primary cost of selective attention is a learned inattention to features that are not diagnostic in the present context (Hoffman & Rehder, 2010). If these features, however, become diagnostic in another context, then the ability to make predictions and transfer knowledge is lost. Adapting Hoffman & Rehder's example, an observer tasked with distinguishing rose bushes from raspberry bushes who considers the presence of berries as the most diagnostic feature may ignore other features of both plants (e.g., thorns and leaves). However, if the observer must later distinguish raspberry from cranberry bushes, thorns are suddenly diagnostic as both have red berries and only the raspberry has thorns. The observer may, therefore, have difficulty distinguishing both bushes due to an inattention to additional features of the raspberry bushes earlier observed.

Even though rule-based classification has the cost of learned inattention to the non-diagnostic features of instances, it is dominant in IS and informs the design of interfaces, databases and programs (Lukyanenko, Parsons, & Samuel, 2015). In OIEs (e.g., citizen science projects), where unanticipated applications, uses and users of data may be paramount, instance-based designs are argued to be better because they encourage contributions that include non-diagnostic features (Parsons & Wand, 2014). Additionally, Lukyanenko et al. (2014) argued that designs based on instances rather than classes are more accommodating of less knowledgeable contributors.

We address the implications of instance-based and rule-based classification for the IQ of contributors from two perspectives.

The expert-adult and novice-infant dichotomy

Interestingly, instance-based perception and reasoning about entities in our world is "a developmental default" (Gelman & Coley, 1990; Gelman & Markman, 1986, 1987; Kloos & Sloutsky, 2008). Rule-based classification is learned with development and distinguishes adults from children. Experiments conducted by Best, Yim, & Sloutsky (2013), comparing the ability of infants and adults to form inclusion rules and selectively attend to attributes of instances based on such rules, show that infants do not have the capacity for selective attention.

Infants reason about classes by observing all the features of individual instances without any *a priori* class inclusion rules (Gelman, Collman, & Maccoby, 1986). We contend they are naturally comparable to non-experts in the domain of a distributed knowledge crowdsourcing project. Like infants, non-experts in a domain lack *a priori* class-forming rules for phenomena in that domain. Infants can, therefore, help us understand how non-experts and "expert amateurs" (people with incomplete knowledge) perceive instances (Keil, 2011; Kloos & Sloutsky, 2008). Gopnik – a researcher in the psychology and philosophy of children's learning and development – explained (in an interview available at bigthink.com) from her research findings, that adults can "functionally" (and perhaps "phenomenologically") "tune in into the mental advantages infants have" when they are exposed to something new to them, for which they do not have a previous schema. She states:

> "… going to a new place is an example of a situation in which you put yourself in the position of a baby. So if I go to Beijing for the first time, everything around me is brand new, everything is different. I'm soaking up lots of information at once, about everything going on. The doors and the tables and the way people look and everything about the place is new".

We contend that a non-expert's exposure to an instance in a citizen science project is also an example of a situation that activates the default ("infant") kind of reasoning about classes. Conversely, the tendency of adults to employ rule-based classification can help us understand expert contributors. We, therefore, hypothesize:

H1a: Non-experts will report more attributes of instances than will experts when asked to report these instances in a crowdsourcing context.

This proposition is consistent with the findings of Lukyanenko et al.'s (2014) research, and with an example cited in Lukyanenko et al. (2016). In this example, a non-expert contributor to Galaxy Zoo – a citizen science project aimed at classifying galaxies – went beyond the scope of the defined task to report atypical information about an observed instance, which lead to the discovery of an important astronomical phenomenon (Hanny's Voorwerp).

In contrast, rule-based classification allows experts to focus on relevant features for identifying instances of classes, resulting in *cognitive economy* (efficiency of classification). Thus, they are less likely to attend to non-diagnostic attributes than will novices.

H1b: Non-experts will report more attributes that are irrelevant to an expected classification than will experts when asked to report instances in a crowdsourcing context.

These non-diagnostic data (irrelevant to the classification task) reported by non-experts lead to a richer dataset, emphasizing the "data completeness" component of IQ. It is valuable to OIEs especially as interest in big data continues to grow because it can lead to novel discoveries and unanticipated uses.

Learned classification and instance-based classification dichotomy

Hoffman and Rehder (2010) showed the need to differentiate supervised classification – engendered by some form of explicit training (e.g., by a teacher) with sufficient feedback to improve the classifier's skill – from unsupervised classification – classification formed without explicit training (self-taught). They argued that the latter "may involve less rule-based processing" and consequently, more attentiveness to attributes. As a result, their definition of expert classifiers precludes "amateur experts" or people who are deeply interested in a domain. Such amateur classifiers may or may not have accurate knowledge, as they are usually self-taught hobbyists. They also explained that "expert classification involves the same sort of attention optimization that characterizes supervised classification learning", which is due to extensive training and the type of task. To underscore their point, they cited Chi, Feltovich, and Glaser (1981), who showed that expert, but not novice, physics problem solvers ignored the superficial features of questions and attended instead to underlying principles.

Through a series of experiments comparing adults who learned categories using either inclusion rules or instances, Hoffman and Rehder's study emphasized that the use of classification is accompanied by the cost of limited knowledge (resulting from focusing narrowly) and of learned inattention. Their experiments show that when an individual learns to ignore certain attributes of an instance, they continue to ignore these attributes when they are present in novel instances (associated or not to previous instances), even if the ignored attributes have become critical to good performance. Following this, we hypothesize that:

H2a: Expert contributors in citizen science projects will misclassify novel instances into an existing schema more than non-expert contributors will.

H2b: Non-expert contributors will report more differences in attributes amongst instances than experts will.

While these hypotheses were not explicitly tested in Lukyanenko, Parsons, & Wiersma, (2014), that study's findings are consistent with hypothesis H2b, as the use of an instance-based design increased the use of their system and also the number of novel sightings reported. Other studies have also reported that users of rule-based classification perform poorly at tasks that require identifying a novel instance of a class (Best et al., 2013; Hoffman & Rehder, 2010; Kloos & Sloutsky, 2008; Sloutsky, 2003).

In the next section, we outline the experiments we propose to test the hypotheses presented here.

PROPOSED EXPERIMENTAL DESIGN

Participants

Following Budescu & Chen's (2014) approach to measuring and defining expertise, undergraduate students in an information systems course will be divided into two groups: Group 1 (experts) will be taught to perform rule-based classification, and Group 2 (non-experts) will not be taught any rules and will be expected to observe and report instances unassisted. The students will contribute to prototype citizen science systems designed based on instance-based and rule-based paradigms.

Stimuli

We will create images of artificial insects as done in Kloos & Sloutsky (2008) (see Figure 1). Each instance will have the following attributes: wings, tail, ridges, buttons, antenna, fingers and rings. There will be at least two variations of each attribute e.g. number of buttons may differ, length of tail (short or long), length of wings (short or long).

Manipulation of Expertise

To choose members of the first group, we will teach them to identify one class of artificial insects, which we will call "tyrans," using an inclusion rule. For example, the inclusion rule could be: *An organism is a **tyran** if and only if it has not more than one less finger than rings and fewer buttons than ridges.*

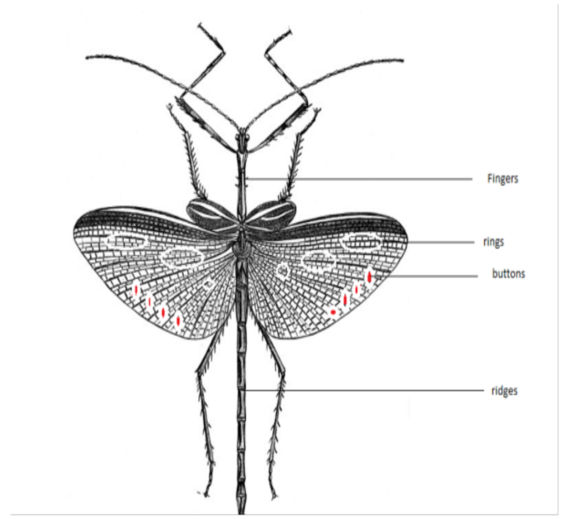

Figure 1. A Tyran

We will present 20 tyran instances varying other attributes except the inclusion rule. We will then present 20 new instances consisting 10 tyrans and 10 non-tyrans. Only, participants who can correctly distinguish tyrans from non-tyrans with at least 90% accuracy will be considered experts and assigned to Group 1. Others in the expert group will be excused from the experiment.

In addition, the inclusion rule will exempt at least one universal attribute of all tyrans. In other words, there will be one attribute in the examples that is common to all tyrans, but not part of the inclusion rule.

Experiment 1

To test hypotheses H1a, immediately after the training of Group 1 members, we will present members of both the expert and non-expert groups with images of tyrans and non-tyrans. As in a typical citizen science project, we will ask them to treat each image as a sighting to be reported, with the option of either stating the name of the class or entering the attributes of the instance (we may provide a simple instruction like "Tell us what you see" – the actual wording of the instruction will be pretested to ensure that participants understand the task). We expect experts to use a class name (e.g., tyran or non-tyran) when reporting the instances, while non-experts will describe the instance using its attributes.

For Hypothesis H1b, we will present an image of an insect in its surrounding (for instance, feeding on a leaf). We expect non-experts to not only report attributes of the insect but also information its environment while experts will only report its class.

Experiment 2

To test hypothesis H2a, participants in the Expert group will be shown an image of an insect that looks like a tyran and has all the attributes in the inclusion rule. The insect however, will lack at least one attribute common to all the tyrans previously presented during training. The absence of this universal feature(s), should imply that the instance is unusual and perhaps novel. Participants in the Non-expert group will be asked to observe an instance of a tyran unsupervised, and without any training. We will then introduce distractors before presenting the same insect presented to Experts We expect that Non-experts will recognize the novelty of this instance and report the missing feature while experts will go with the inclusion rule and report the instance as a tyran, ignoring its novelty.

To test hypothesis H2b, we will place two images side by side, both tyrans. One will have at least one new attribute which was not a part of the inclusion rule and has not been shown in previously presented tyrans. We will ask participants in both groups to report what they see. Similar to other citizen science projects, Group 1 will be provided with an interface that allows them enter the number of tyrans (or insects that are not tyrans) sighted, and allows additional information to be provided. The interface for Group 2 will request that they choose "yes" if the insects presented are the same and "no" if not. They will also be allowed to provide additional information. We expect that a larger number of non-experts (versus experts) will report the target attribute(s).

CONCLUSIONS AND POTENTIAL IMPLICATIONS

The studies proposed in this paper are expected to shed light on the relevance of expertise to OIEs, especially as it relates to realizing high information quality. If these hypotheses are supported, the findings will help researchers and practitioners better understand the costs and benefits of relying on expert contributors, and may serve as a precursor to research on how to maximize these benefits while mitigating the costs. It may also further underscore the need for inclusive designs in citizen science systems and OIEs, where the potential uses of information may not be fully known when a system is developed, or may evolve over time. Additionally, the hypotheses presented here, grounded in classification theory, may help us understand and explain prior research findings where expertise has been shown to be unrelated to performance. We therefore expect that the research will extend the impact of classification theory by adapting the theory to the field of IS design and also contribute to our understanding of expertise and its impact on the performance of classification tasks.

At this stage, we are in early development of the experiments. Our preliminary ideas will benefit from feedback at the conference to help us further define and articulate the experiments and our potential research contributions.

ACKNOWLEDGMENTS

This research was supported by a research grant from the Social Sciences and Humanities Research Council of Canada.

REFERENCES

1. Best, C., Yim, H., & Sloutsky, V. (2013). The cost of selective attention in category learning: developmental differences between adults and infants. *Journal of Experimental Child Psychology*.
2. Biros, D., George, J., & Zmud, R. (2002). Inducing sensitivity to deception in order to improve decision making performance: A field study. *MIS Quarterly*.
3. Budescu, D., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science*.
4. Castriotta, M., & Guardo, M. Di. (2011). Open Innovation and Crowdsourcing: The Case of Mulino Bianco. *Information Technology and Innovation …*.
5. Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices*. *Cognitive Science*.
6. Clemen, R. (2008). Improving and measuring the effectiveness of decision analysis: Linking decision analysis and behavioral decision research. *Decision Modeling and Behavior in Complex and …*.
7. Davis-Stober, C., Budescu, D., Dana, J., & Broomell, S. (2014). When is a crowd wise? *Decision*.
8. Escoffier, N., & McKelvey, B. (2015). The Wisdom of Crowds in the Movie Industry: Towards New Solutions to Reduce Uncertainties. *International Journal of Arts Management*.
9. Gelman, S., & Coley, J. (1990). The importance of knowing a dodo is a bird: Categories and inferences in 2-year-old children. *Developmental Psychology*.
10. Gelman, S., Collman, P., & Maccoby, E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*.
11. Gelman, S., & Markman, E. (1986). Categories and induction in young children. *Cognition*.
12. Gelman, S., & Markman, E. (1987). Young children's inductions from natural kinds: The role of categories and appearances. *Child Development*.
13. Hoffman, A. B., & Rehder, B. (2010). The costs of supervised classification: The effect of learning task on conceptual flexibility. *Journal of Experimental Psychology: General*, *139*(2), 319–340. http://doi.org/10.1037/a0019042
14. Hosseini, M., Phalp, K., Taylor, J., & Ali, R. (2014). Towards crowdsourcing for requirements engineering.
15. Keil, F. C. (2011). Science Starts Early. *Science*, *331*(6020), 1022–1023. http://doi.org/10.1126/science.1195221
16. Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: Effects of statistical density on learning and representation of categories. *Journal of Experimental Psychology: General*, *137*(1), 52–72.

http://doi.org/10.1037/0096-3445.137.1.52

17. Lin, S., & Cheng, C. (2009). The reliability of aggregated probability judgments obtained through Cooke's classical model. *Journal of Modelling in Management*.
18. Lukyanenko, R., Parsons, J., & Samuel, B. (2015). Do We Need an Instance-Based Conceptual Modeling Grammar ? Do We Need an Instance-Based Conceptual Modeling, (MAY). http://doi.org/10.13140/RG.2.1.4613.9365
19. Lukyanenko, R., Parsons, J., & Wiersma, Y. (2014). The impact of conceptual modeling on dataset completeness: a field experiment.
20. Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2014). The IQ of the Crowd: Understanding and Improving Information Quality in Structured User-Generated Content. *Information Systems Research*, *25*(4), 669–689. http://doi.org/10.1287/isre.2014.0537
21. Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2016). Emerging problems of data quality in citizen science. *Conservation Biology*, n/a–n/a. http://doi.org/10.1111/cobi.12706
22. Murphy, G. (2002). The big book of concepts. *MIT Press*.
23. Parsons, J. (1996). An Information Model Based on Classification Theory. *Management Science*, *42*(10), 1437–1453. http://doi.org/10.1287/mnsc.42.10.1437
24. Parsons, J., Lukyanenko, R., & Wiersma, Y. (2011). Easier citizen science is better. *Nature*, *471*(7336), 37. http://doi.org/10.1038/471037a
25. Parsons, J., & Wand, Y. (2008). Using cognitive principles to guide classification in information systems modeling. *MIS Quarterly*.
26. Parsons, J., & Wand, Y. (2013). Extending Classification Principles from Information Modeling to Other Disciplines. *Journal of the Association for Information ….*
27. Parsons, J., & Wand, Y. (2014). A Foundation for Open Information Environments.
28. Schultze, U., & Leidner, D. (2002). Studying knowledge management in information systems research: discourses and theoretical assumptions. *MIS Quarterly*.
29. Sloutsky, V. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*.
30. Stein, E. (1992). A method to identify candidates for knowledge acquisition. *Journal of Management Information Systems*.
31. Tarrell, A., Tahmasbi, N., Kocsis, D., Pedersen, J., Tripathi, A., Onook, O., … DeVreede, G. (2013). Crowdsourcing: A snapshot of published research. *Proceedings of the Nineteenth Americas Conference on Information Systems, Chicago, Illinois, August 15-17, 2013*.
32. Tripathi, A., & Tahmasbi, N. (2014). Crowdsourcing typology: a review of is research and organizations. *Proceedings of the Ninth Midwest Association for Information Systems Conference, Ames, Iowa May 15-16, 2014*.
33. Wand, Y., Monarchi, D., Parsons, J., & Woo, C. (1995). Theoretical foundations for conceptual modelling in information systems development. *Decision Support Systems*.