UNDERSTANDING INFORMATION DIVERSITY IN THE ERA OF REPURPOSABLE CROWDSOURCED DATA

by

© Oluwafemi Shawn Ogunseye

A Thesis

Submitted to the school of Graduate Studies

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Faculty of Business Administration

MEMORIAL UNIVERSITY OF NEWFOUNDLAND

MAY, 2020

ST. JOHN'S NEWFOUNDLAND

Abstract

Organizations successfully leverage information technology for the acquisition of knowledge for decision-making through information crowdsourcing, which is gathering information from a group of people about a phenomenon of interest to the crowdsourcer. Information crowdsourcing has been used to drive business insight and scientific research, providing crowdsourcers access to information outside their traditional reach. Crowdsourcers seek high-quality data for their information crowdsourcing projects and require contributors who can provide data that meet predetermined requirements. Crowdsourcers recruit contributors with high levels of relevant knowledge or train contributors to ensure the quality of data they collect. However, when crowdsourced data needs to fit more than a single usage scenario because the requirements of the project changed or the data needs to be repurposed for tasks other than the one(s) for which it was initially collected, the ability of contributors to provide diverse data that can meet multiple requirements is also desirable.

In this thesis, I investigate how the domain knowledge a contributor possesses affects the diversity and quality of data they report. Using an experiment in which 84 students randomly assigned to three knowledge conditions reported information about artificial stimuli, I found that explicitly trained contributors provided less diverse data than either implicitly trained or untrained contributors.

In addition, I looked at the longitudinal effect of knowledge on the diversity of data reported by contributors. Using review data from Amazon.com and organism sighting data from NLNature.com (a citizen science data crowdsourcing platform), I studied the impact of knowledge on the diversity and quality of crowdsourced data. The results show that experience reduced the diversity and usefulness of contributed data. The study provides insights for crowdsourcers in

industry and academia on how to manage and utilize their crowds effectively to collect high-quality reusable data.

Acknowledgments

Thank God for the grace to complete this Ph.D. program.

I will always be grateful for the support I have received and continue to receive from my Ph.D. supervisor, Dr. Jeffrey Parsons. He is a remarkable mentor and friend.

I am also grateful to the members of my thesis committee: Dr. Roman Lukyanenko and Dr. Yolanda Wiersma. Their unwavering support saw me through the thick and thin of the thesis writing and oral defence process.

Also, a big thank you to everyone at the Faculty of Business Administration who supported me in one way or the other: Dean Isabelle Dostaler, Dr. Rachelle Shanahan, Dr. Travor Brown, Dr. Kara Arnold, Dr. Sherrie Komiak, Dr. Katherine Gallagher, Dr. Wieslaw Kubiak, Ms. Donna Fitzgerald, Ms. Sherry Whelan, the Rosemarys, and Dr. Mekaela Gulliver. Dr. Gallagher's timely words of encouragement were always refreshing.

To all the great Ph.D. students I have had the honour of sharing the Ph.D. lab with, and who have become my dear friends, and to all my dear friends outside the Ph.D. program, I say thank you for your support. I would mention names, but I want to limit the number of pages.

I am incredibly grateful to my wife, Dolapo Ogunseye. She stood by me through the Ph.D. program. Thank you! I am also grateful to my daughter Tara. You came at the right time and brightened our lives with your light. I hope where I stop is where you'll start. And to my late mum, Dolapo Kehinde Ogunseye, who is my most ardent supporter, my dad and siblings, thank you all.

Finally, I am grateful to Professor Olusegun Folorunso, who introduced me to academic research and nudged me on into academia.

Dedication

To my mum, who didn't get to see me finish

Table of Contents

List of Tables

List of Figures and Illustrations

Chapter One: **Introduction**

1.1 Background and Motivation

In 2016, information was reported to have become the world's most valuable resource (The Economist, 2017). Of course, information has always been valuable to both public and private sector organizations, helping to guide the correct allocation of business resources. What has changed in recent times is the ability of businesses and individuals to collect and store vast amounts of data from internal and external sources and to analyze these data in creative ways to generate business insights. More importantly, advancements in our ability to analyze collected data have made it possible to use such data in contexts different from the ones they were originally collected, which can generate unanticipated insights (Günther, Mehrizi, Huysman, & Feldberg, 2017). The ability to generate insights through data analytics is, therefore, a major driver of competitive advantage for many businesses; for example, top-performing organizations use analytics "five times more" than lower-performing ones (LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011, p. 22).

However, even if an organization collects large amounts of data, insights do not naturally follow if they are absent from the data. Therefore, researchers and practitioners are looking beyond the amount of data available to organizations and are instead focusing on the capacity for available data to produce insights when viewed from different perspectives through analytics. When considering the value of data, " 'big' is no longer the defining parameter, but, rather, how 'smart' [data] is—that is, the insights that the volume of data can reasonably provide" (George, Haas, & Pentland, 2014, p. 321). Yet, "[r]egrettably…[m]anagement tends to think that the larger the Big Data project is (e.g., the largest amount of data involved in the project), the larger benefits (e.g., the soundest knowledge) can be obtained" (Merino, Caballero, Rivas, Serrano, & Piattini, 2016, p.124).

Even though a very creative and innovative analytics team can elicit insights from limited data, perspectives available in data will limit the amount of insight that analytics can provide (Ghasemaghaei & Calic, 2019). The diversity of perspectives in collected data takes precedence over the depth or breadth of analytics skills available to an organization. Organizations seeking to gain competitive advantage through analytics, therefore, can benefit from collecting diverse data in the first place. Determining how to collect diverse data begins with understanding the data sources, i.e., humans and human-programmed machines that observe and report data about phenomena of interest to businesses. This thesis considers explicitly human data contributors who provide data to address specific information needs of organizations or individuals through crowdsourcing. Crowdsourcing is "outsourcing a task to a 'crowd,' rather than to a designated 'agent' …" (Afuah & Tucci, 2012, p 355). Industry and research have successfully outsourced the task of information gathering from large groups of people[1] using purpose-built integrative crowdsourcing systems, i.e., systems that "pool complementary inputs from the crowd" (Schenk & Guittard, 2011, p 98). One example is Statistics Canada, which uses crowdsourcing to map buildings across Canada to acquire "national-level statistics on buildings—and their attributes—that can be used to compare specific local areas"

[1] Crowdsourcing systems that gather distributed information for decision making are referred to as integrative crowdsourcing in Schenk & Guittard (2011) and observational crowdsourcing in Lukyanenko & Parsons (2018).

(www.statcan.gc.ca/eng/crowdsourcing). Another example is the Great Sunflower Project, which recruits people to count the number of pollinators that visit sunflowers in their environments and uses these counts to investigate how the decline of the bee population is affecting the pollination of plants. (www.greatsunflower.org).

When designing crowdsourcing systems, it is essential for crowdsourcers – organizations and individuals that use crowdsourcing to collect information – to determine the composition of an appropriate crowd from which to collect data (Malone, Laubacher, & Dellarocas, 2009). Crowdsourcers usually require potential contributors to possess relevant knowledge of the crowdsourcing task and implement recruitment strategies that favour knowledgeable contributors in order to mitigate the risk of collecting low-quality information. Training volunteers before they are allowed to participate and recruiting experienced contributors who have previously participated (or are presently participating) in a similar project (Gura, 2013; Wiggins, Newman, Stevenson, & Crowston, 2011) are strategies employed by crowdsourcers to ensure potential contributors to their projects have the requisite knowledge to provide quality data. Several studies in the literature support these strategies, based on the assumption that more knowledgeable contributors provide higher quality data than less knowledgeable contributors. This thesis aims to develop a better understanding of how knowledge affects contributors' ability to provided diverse yet high-quality data.

Erickson, Petrick, & Trauth (2012) identified several types of knowledge relevant in the crowdsourcing context. These are:

Situational Knowledge – refers to knowledge that comes from contributors' access to the setting in which the phenomena relevant to the crowdsourcing task occurs. It does not necessitate any knowledge of the domain of the crowdsourcing project or how to carry out the crowdsourcing task. For example, citizen journalism does not require the citizen to be knowledgeable about journalism or the subject matter being reported. It only requires equipment and access to the location of the newsworthy event.

Product/Service knowledge – refers to knowledge that is specific to the crowdsourcing project. This can include familiarity with the use of a crowdsourcing platform, the procedure required to complete a crowdsourcing task, and other details limited to a particular crowdsourcing project. In this thesis, this type of knowledge will be referred to as **task knowledge.** This knowledge is usually acquired in crowdsourcing by training potential participants on the task to be performed in the project and assesses their knowledge of the training. An example is the GalaxyZoo project (www.galaxyzoo.org), in which volunteers receive training on how to identify features of galaxies that help in their classification. Volunteers practice and are tested to determine if they have gained sufficient knowledge to be allowed to participate. Participants do not need prior knowledge of the domain of study.

Domain Knowledge – refers to *a priori* knowledge of the topic and focus of the crowdsourcing project. Participants with domain knowledge have prior knowledge about the phenomenon under study. This knowledge may have been acquired through some training and is usually broad, covering more than just the particular entity or phenomenon to be studied in a crowdsourcing project. For example, in eBird domain knowledge consists of knowledge about birds, including the ability to identify species.

Over time, as participants continue to contribute to a crowdsourcing project, regardless of their level or type of knowledge before participation, they gain experience. Our focus is on task knowledge. Task knowledge is gained by participation or training and can lead to task expertise, while domain knowledge refers to knowledge of the area of the crowdsourcing project and can lead to domain expertise (see Mukhopadhyay, Singh, & Kim, 2011). Understanding the impact of crowd selection strategies that prioritize some task knowledge based on a desire for high quality crowdsourced data will affect design decisions (i.e. decisions about the recruitment, task, system and motivational strategy) made by crowdsourcers, especially concerning crowd recruitment (Wang & Strong, 1996; Wiggins et al., 2011).

1.2 Thesis Objectives

Our focus in this thesis is on integrative crowdsourcing systems rather than selective crowdsourcing systems. Unlike integrative crowdsourcing systems that pool all inputs from the crowd, selective crowdsourcing systems seek inputs from a crowd, rank these inputs, and choose the best ones (Schenk & Guittard, 2011). Integrative crowdsourcing systems typically have the following characteristics: (a) the goals, level of expertise, and motivation of members of the contributing crowd are typically unknown; (b) the types of data and ways in which crowd members will contribute the data are unpredictable; (c) the uses for the contributed data can be predetermined or emergent; (d) there is potential for high contributor turnover, and perhaps most importantly, (e) the events being reported may be transient; in many cases, crowdsourcers may only have one chance to collect high-quality information from contributors. Therefore, it is essential to be sure that contributors can deliver high-quality multidimensional data over time.

Significantly, the goal of selective crowdsourcing is to select the best input(s) from a number of competing inputs by a crowd of people, whereas integrative crowdsourcing uses all inputs for decision-making. In other words, the eventual outcome of the selective crowdsourcing process is a reflection of the "best" contributors, while integrative crowdsourcing is a reflection of the entire crowd. An example of a selective crowdsourcing project is the General Mills Worldwide Innovation Network (G-WIN) which accepts ideas from the public that can help the company in its areas of business, reviews them and depending on the outcome of their review, selects the ones to pursue and the ones to reject (gwin.secure.force.com). In contrast, integrative crowdsourcing considers all crowd inputs for decision-making. For instance, the Great Sunflower Project recruits people to count the number of pollinators that visit sunflowers in their environments and uses these counts to investigate how the decline of bee populations is "affecting the pollination of gardens, crops and wild lands" (www.greatsunflower.org). Success in integrative crowdsourcing projects is achieved when a sufficient number of people report data about the target entity that is usable for decision-making.

The tasks in the integrative crowdsourcing systems we address in this thesis would be ill-defined, usually open-ended tasks, and require volunteers to report observations about their "broader environment" continuously. This types of crowdsourcing tasks are classified as observational crowdsourcing (Lukyanenko & Parsons, 2019, p. 4). The crowdsourcing tasks we address in this thesis do not include small scale, well-defined tasks that may take place primarily online and where the workers are typically paid.

Integrative crowdsourcing systems align with Aksulu and Wade's (2010) framework depicting the properties of classes of information systems based on systems theory. The subclass, integrative crowdsourcing, represents systems that collect data from the crowd about phenomena of interest. Similar to open source software projects, integrative crowdsourcing systems can include loosely defined data collection projects that mature over time, with a lifespan that is organically defined, and are flexible to internal and external changes. This includes changes to data requirements and changes to contributors. Integrative crowdsourcing systems, therefore, represent open-source data collection platforms. Whether citizen science, social media platforms or online review systems, these shared properties include them as members of the integrative crowdsourcing systems class. However, the extent to which each member implement these properties vary.

Parsons and Wand (2014) refer to these types of information systems as operating in open information environments where the sources of their data are unknown, and the uses of their collected data can be emergent and unanticipated. They identified that such systems would need flexible, quality, and semantically diverse information to meet the needs of different information users and contributors. In order to address this need for flexible, semantically diverse, and high-quality data, there is a need for more understanding of the limitations of the traditional information quality dimensions and how diverse data may impact these dimensions. Consequently, we first address the following research question:

Research Question 1: What are the factors that enable or inhibit information diversity in integrative crowdsourcing projects?

Researchers and practitioners would benefit from a theoretical and conceptual grounding of the underlying factors that drive information diversity and how they can avoid pitfalls that limit the ability of their crowds to provide diverse data.

Since crowdsourcers assume that knowledgeable contributors provide better quality data, they resort to training potential contributors (e.g., galaxyzoo.com) to mitigate the scarcity of expert contributors. To test this assumption about knowledge and information quality, we examine how training affects the diversity of information that contributors provide and the relationship between diversity and traditional information quality dimensions. Also, recruiting knowledgeable contributors either directly or by training and testing them first limits the available participant pool and increases the costs of acquiring contributors for crowdsourcing projects. Moreover, so does restricting participation in crowdsourcing tasks to contributors with a predetermined level of knowledge of the task, such as experts. Based on the literature, we posit that crowdsourcer-provided training leads to the acquisition of knowledge by contributors and different types of training lead to different types of knowledge. Therefore, besides investigating the effect of training on the diversity of information contributed to crowdsourcing tasks, we also study the impact of contributors' level of task knowledge on the diversity of crowdsourced data collected in integrative crowdsourcing projects. Correspondingly, the second research question addressed in this thesis is as follows:

Research Question 2: How does knowledge affect the diversity and quality of crowdsourced data?

Answering this research question will help crowdsourcers to address the crucial design decision: "who should be recruited to a crowdsourcing project?", which is a topical issue and a necessary research focus (Austen, Bindemann, Griffiths, & Roberts, 2016; Crall et al., 2011; Lukyanenko, Wiggins, & Rosser, Forthcoming; Ogunseye & Parsons, 2018). Crowdsourcers, including organizations, researchers, and crowdsourcing platforms like Amazon Mechanical Turk (MTurk), would, therefore, benefit from a better understanding of the effect of knowledge-based recruitment on the quality and diversity of crowdsourced data.

We claim that contributor knowledge increases as a result of participating in crowdsourced projects, and this negatively affects information diversity. After volunteering, crowd members interact with the crowdsourcing system, contribute to the project, communicate (directly or indirectly) with other participants, and sometimes get more training, therefore gaining experience in the crowdsourcing task. Crowdsourcers concerned about the quality of crowdsourced data in their projects may recruit these experienced contributors outright, and exempt (or not actively pursue the recruitment of) inexperienced or novice contributors from their projects. This capacity to limit members of the crowd to experienced contributors is a central part of the business model of some crowd hiring and online review platforms. For example, Amazon Mechanical Turk (www.mturk.com) continuously ranks crowd workers (contributors) based on their capacities to complete tasks, and they charge a premium for their most experienced crowd workers, referred to as "master workers." At the same time, due to the cost of acquiring crowd members, crowdsourcers genuinely aim to keep their crowds for as long as possible, implying that the experience of members of crowds generally increases with participation on a crowdsourcing project. Crowdsourcers would benefit from a better understanding of the longitudinal effect of increasing knowledge of the task or task experience on the quality and diversity of data reported by crowds. Therefore, we ask the following research question:

Research Question 3: What is the longitudinal impact of task experience on the diversity and quality of crowdsourced data?

Answering this research question will shed more light on how contributor experience affects information quality and provide clear, empirical guidance on how crowdsourcers should organize their crowds.

1.3 Organization of the Thesis

The thesis uses a manuscript format. To address the first research question, we review the literature in Chapter 2 to understand information diversity in crowdsourced data more thoroughly. We also address how information diversity can be measured and theoretically link information diversity with information usefulness – a consequence of information quality, accuracy, and completeness. We examine from literature the effects of knowledge on information diversity in two types of directed integrative crowdsourcing systems: (i) Online reviews are a type of crowdsourcing, where members of the crowd post their opinions on products and/or services (Kleemann, Voß & Rieder, 2008), and in doing so, they help guide future shoppers in their decision-making processes (Edelman, 2010). (ii) Citizen science is the "partnership between volunteers and scientists to address research questions" (Crall et al., 2011 p. 433), usually culminating in citizens assisting with data collection and analysis while gaining scientific knowledge through their involvement in the research. Using these exemplars of integrative crowdsourcing aids the generalizability of the findings of this study. For the second research question, we explore the effect of training on how contributors report data (Chapter 3). Using selective attention and classification theories from cognitive psychology, we develop and test hypotheses about how training or not training contributors affects the information they report in crowdsourcing projects. Hypotheses about training and information diversity are tested using an experiment with 84 participants conducted over one year. Furthermore, we developed and tested a hypothesis about the effect of contributors' levels of knowledge on the quality of information they contribute in Chapter 3. Chapter 4 reports our investigation of the effect of experience on information diversity, and the relationship between information diversity and information usefulness. The hypotheses developed in this chapter are evaluated using review data from Amazon.com and comments from NLNature – a citizen science project. We employ natural language processing and machine learning algorithms to test these hypotheses and answer Research Question 3. In addition, we made recommendations about how to prevent the negative effects of experience in crowdsourcing projects. In Chapter 5, we discuss the general contributions of the thesis. A Glossary of terms used throughout the thesis is provided on page 188.

References

Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review*, *37*(3), 355–375.

Aksulu, A., & Wade, M. (2010). A comprehensive review and synthesis of open source research. Journal of the Association for Information Systems, 11(11), 576.

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. *Scientific Reports*, *6*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5028888/

Crall, A. W., Newman, G. J., Stohlgren, T. J., Holfelder, K. A., Graham, J., & Waller, D. M. (2011). Assessing citizen science data quality: An invasive species case study. *Conservation Letters*, *4*(6), 433–442. https://doi.org/10.1111/j.1755-263X.2011.00196.x

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

Ghasemaghaei, M., & Calic, G. (2019). Can big data improve firm decision quality? The role of data quality and data diagnosticity. *Decision Support Systems*, *120*, 38–49.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*.

Gura, T. (2013). Citizen science: Amateur experts. *Nature*, *496*(7444), 259–261. https://doi.org/10.1038/nj7444-259a

Kleemann, F., Vo\s s, G. G., & Rieder, K. (2008). Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, Technology & Innovation Studies*, *4*(1), PP–5.

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2), 21.

Lukyanenko, R., & Parsons, J. (2019). Beyond micro-tasks: Research opportunities in observational crowdsourcing. In *Crowdsourcing: Concepts, Methodologies, Tools, and Applications* (pp. 1510–1535). IGI Global.

Lukyanenko, R., Wiggins, A., & Rosser, H. K. (n.d.). Citizen Science: An Information Quality Research Frontier.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2009). *Harnessing crowds: Mapping the genome of collective intelligence*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1381502

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2016). A data quality in use model for big data. *Future Generation Computer Systems*, *63*, 123–130.

Ogunseye, S., & Parsons, J. (2018). Designing for Information Quality in the Era of Repurposable Crowdsourced User-Generated Content. *International Conference on Advanced Information Systems Engineering*, 180–185. Springer.

Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*, (1), 93–107.

The Economist. (2017). The world's most valuable resource is no longer oil, but data—Regulating the internet giants. Retrieved May 31, 2019, from https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. *2011 IEEE Seventh International Conference on E-Science Workshops*, 14–19. https://doi.org/10.1109/eScienceW.2011.27

Chapter Two: **What Information Quality Should Mean in this Era of Repurposable Data: A Case for the Information Diversity Dimension**

Abstract

The ability to repurpose crowdsourced data through analytics leads to the generation of valuable insights. For large organizations, reusing data through repurposability also saves organizations the cost of reacquiring and storing data. To ensure crowdsourcers – individuals and organizations who use crowdsourcing for data collection – can collect insightful data, we must be able to measure insightfulness. While traditional information quality dimensions measure factors like accuracy and completeness, there is also a need for more knowledge about how to measure the quality of data based on its repurposability.

In this chapter, we identify the limitations of traditional information quality dimensions for measuring insightfulness and repurposability of data and recommend the information diversity dimension as a solution to the identified limitations. We use ontology to show how information diversity can be measured, and we developed an information diversity framework based on three factors identified as essential for information diversity: the data model, the nature of the crowdsourced task, and the differences in contributors. Finally, we validate the information diversity dimension through requirements presented in Parsons and Wand (2014) and review two articles in ecology and agriculture to demonstrate the viability of the information diversity dimension. This study will inform research and practice on information diversity as a pertinent dimension for determining the quality of crowdsourced data today, providing a framework for gathering and measuring repurposable data.

Keywords: information diversity, repurposability, information quality, crowdsourcing, data analytics

2.1 Introduction

Information acquired from crowds can have unanticipated uses beyond the original purposes for which the information was collected, leading to valuable insights. For example, Yelp review data, which is intended to guide shoppers and merchants on the weaknesses and strengths of services provided by businesses, has been used to determine crime indexes of locations (Ballesteros, Carbunar, Rahman, Rishe, & Iyengar, 2014) and to identify restaurants with a high risk of health code violation and outbreaks of foodborne diseases (Harrison et al., 2014; Nsoesie, Kluberg, & Brownstein, 2014; Schomberg, Haimson, Hayes, & Anton-Culver, 2016). In conjunction with mobile check-in data from Foursquare, an app used to share location information with friends and family, and Yelp reviews have been used to accurately predict business failures (Wang, Gopal, Shankar, & Pancras, 2015). In the same fashion, public sentiments in Twitter data have been used to predict stock market price movements (Bollen, Mao, & Zeng, 2011; Mittal & Goel, 2012; Nisar & Yeung, 2018; Pagolu, Reddy, Panda, & Majhi, 2016). In this thesis, we refer to these uses of crowdsourced data—uses that deviate from the original purposes of data collection to meet previously unanticipated requirements—as data repurposing.

Repurposability, also called *portability*, is the ability to use data for purposes other than those for which it was collected. The ability to repurpose data is a major factor in the value of collected data (Günther, Mehrizi, Huysman, & Feldberg, 2017). In discussing the importance of being able to use data in different ways, the literature on data modeling and data quality emphasize that data are more valuable and provide more insights to users when they are not bound to any schema and are repurposable (e.g., see Günther, Mehrizi, Huysman, & Feldberg, 2017; Hunter, Alabri, & van Ingen, 2013; Parsons, 1996). Repurposable data can answer various questions from the same or different users, allowing decision-makers across an organization to answer new questions using existing data (see Tamm, Seddon, & Shanks, 2013), and enabling data to be useful beyond a single organization (Günther et al., 2017; Zuboff, 2015). Repurposable data can, therefore, address emerging user requirements and unanticipated needs for different consumers, including data procuring organizations.

Repurposability is necessary for most data analytics: the use of data that were collected from the operation of a business or sourced externally by data scientists, to derive business insights for competitive advantage (Woodall & Wainman, 2014, 2015). Our focus is on the repurposability of data acquired through *integrative crowdsourcing* systems: that is, systems that collect input from an undefined group of people (rather than known subjects, such as employees), regarding a phenomenon of interest to information consumers (Schenk & Guittard, 2011). When organizations and individuals—crowdsourcers—expend resources in the form of time and money to acquire data from a crowd through integrative crowdsourcing, they naturally want to maximize its use and value. Because business needs may be emergent or evolving, it is inadequate to evaluate the quality of data only by its ability to meet anticipated requirements. In other words, repurposing information implies that information is used for secondary purposes with different information quality requirements and “original quality levels may not be suitable for the secondary purpose” (Woodall & Wainman, 2015, p. 1).

The ability to adapt data to changes in business requirements should be considered when assessing the quality of crowdsourced data. For example, a researcher crowdsourcing information about a given phenomenon might discover a need for more information than initially anticipated. In this case, crowdsourced information that was considered of high quality becomes insufficient to answer research questions. Therefore, individuals and businesses using, selling, and procuring data would benefit from the ability to evaluate the repurposability of their data. Collecting repurposable data could help increase the reuse of data, reducing the need to commission new crowdsourcing projects because of evolving business needs.

Information quality assessment focuses on information that was acquired for a specific use. Information quality metrics are tied to the intended use(s) of the data (Wand & Wang, 1996) and cannot measure their repurposability. Traditional information quality assessment uses metrics such as accuracy and completeness, with a focus on the information consumer, typically guided by the views of known consumers about what types of information are needed for known tasks. However, the strategic value of information lies in the amount of insight that it can provide (George, Haas, & Pentland, 2014). Repurposability drives such insight: it increases when contributed information includes a variety of views, and is manifested as differences in attributes, in perspectives, and in the amount of information provided about the subject (Günther et al., 2017; Parsons & Wand, 2014; Woodall, 2017). Because repurposability is centered on accommodating unknown views of information users (Lukyanenko, Parsons, Wiersma, & Maddah, 2019), its pursuit may be antithetical to the strategies traditionally used to ensure information quality, such as enforcing uniformity in contributed data and requiring that potential contributors have prior knowledge of the crowdsourced task. Therefore, organizations seeking to derive the most value from crowdsourced data will need to look beyond traditional data quality dimensions for guidance.

In this chapter, we refer to the number of unique attributes of entities present in information as information diversity, and we take a step towards better understanding information diversity as a metric for measuring and designing information repurposability in integrative crowdsourcing systems. We develop theoretical explanations for why the widely used top-down information quality model is inadequate for integrative crowdsourcing systems. Furthermore, we describe information diversity, grounding it in ontology, and show how it addresses the inadequacies of traditional information quality dimensions in the measurement and advancement of the repurposability of crowdsourced data. To address the inadequacies identified for traditional information quality dimensions, we introduce information diversity as a necessary dimension for measuring the evolving meaning of information quality for businesses that leverage crowds as external data sources.

2.2 Limitations of Traditional Information Quality Metrics

Contributed data should represent the state of an observed world at a given time and should help information users to reproduce that state whenever necessary. In the literature and practice, data quality is judged by the extent to which the data fit their intended use (Sadiq & Indulska, 2017; Wang & Strong, 1996) and is measured on several dimensions: most significantly, accuracy and completeness (Wang & Strong 1996). Accuracy means the degree to which the data provided are “correct,” “meaningful” and “objective”; completeness is “the degree to which all possible states relevant to the user population are represented in the stored information” (Nelson et al., 2005, p. 203).

However, many metrics for measuring information quality are *ad hoc* (Pipino, Lee, & Wang, 2002), lack any theoretical basis, and only apply to specific contexts (Wand & Wang, 1996). As we identify the problems of traditional information quality dimensions, we focus on accuracy and completeness. Wand and Wang’s described accuracy (which they termed correctness) as when reported data about an entity maps to a true state of the entity. And completeness as when the data properly represents the entity and maps back to the entity’s state without missing states. We adopt Wand and Wang’s (1996) view of accuracy and completeness, primarily because it accommodates the possibility that operationalizing accuracy can be automated, making it relevant as the adoption of machine learning and artificial intelligence by organizations is on the increase (Ransbotham, Gerbert, Reeves, Kiron, & Spira, 2018). Therefore, we discuss how the problems of traditional information quality may affect the collection and measurement of repurposable crowdsourced data, in the context of integrative crowdsourcing.

2.2.1 The Problem of Generalizability

Before technology-enabled crowdsourcing, organizations controlled their information management processes to ensure high data quality. For example, knowledgeable employees were assigned predefined data entry tasks. The information systems used for data collection were also designed with controls to ensure that collected data are validated. More generally, the sources, users, and uses of information were generally known, which made it possible not only to determine the quality of information collected but also to return to these sources should more clarification be needed. Accordingly, it was appropriate to strive for consistency in the information management processes and protect against variations in data resulting from diversity in employees or other data sources. Consistency was achieved in various ways, including specifying required input types and formats through system design, ensuring the employment of people with the knowledge needed to perform the task, and training potential employees to accomplish the task.

In the current era of crowdsourced information, the contributors of information are not known and may be temporary sources of data. However, the information quality concerns of some integrative crowdsourcing systems still center on data accuracy and completeness. For example, in many citizen science applications—a type of integrative crowdsourcing—scientists rely on citizens to gather accurate and complete data about a phenomenon of interest to them, defining information quality in terms of accuracy (McKenzie, Long, Coles, & Roder, 2000; Oldekop et al., 2011; Salk, Sturn, See, & Fritz, 2016) or completeness (Jacobs & Zipf, 2017). We know that ordinary citizens are better at reporting the attributes of entities they observe than at accurately classifying them, as they may not have adequate knowledge to inform their classification (Lukyanenko, Parsons, & Wiersma, 2014). When crowdsourcers can use machine learning to determine entities from reported attributes, restricting participation in crowds to knowledgeable contributors might become unnecessary. Furthermore, there are citizen science systems that seek to facilitate discoveries and novel instances of phenomena (Lukyanenko et al. 2016). For these types of citizen science systems, there is a need to accommodate diverse perspectives, allowing contributors to report novel findings even when they do not fit a predetermined classification schema.

Moreover, traditional information quality metrics are less relevant for integrative crowdsourcing systems, such as online review systems that collect reviews from shoppers to guide them in their decision-making. Online reviews generally involve reporting experiences about products or services, and usually require classifying these products and services into abstract classes, such as “good” or “bad.” For example, a shopper who reviews a shoe purchased on Amazon as an “excellent product” (and gives it a 5-star rating) or as “very poor quality” (and gives it a 1-star rating), based on the shopper’s experience with the product, classifies the product into abstract categories of “excellent products” and “terrible products” that do not necessarily have well-defined inclusion criteria, but instead often rely on subjective criteria. The quality of the information provided by contributors to support such classification cannot be measured using traditional dimensions, such as accuracy or completeness, because it is difficult to determine whether a review is accurate. Instead, dimensions like usefulness, diversity, and informativeness are relevant aspects of information quality.

Information quality, therefore, means different things to different information users (Ardagna, Cappiello, Samá, & Vitali, 2018; De la Calzada & Dekhtyar, 2010), and in different crowdsourcing contexts (Hunter et al., 2013); it cannot be generalized even within the same class of information systems. Table 2.1 shows how crowdsourcers measure information quality in citizen science and online review systems. Nonetheless, while traditional dimensions of information quality do not apply uniformly across these integrative crowdsourcing systems, the perceived usefulness of contributed information—an outcome of information quality dimensions such as accuracy and completeness —is measurable, as evident in the literature (DeLone & McLean, 1992; Wixom & Todd, 2005; Xu, Benbasat, & Cenfetelli, 2013). Specific examples include Yelp and Amazon.com, which allow different users to rate the usefulness (called helpfulness on Amazon.com) of reviews. It is not just traditional dimensions of data quality that are responsible for the differences in perceived usefulness of contributed information (Cheung, Sia, & Kuan, 2012; Gobinath & Gupta, 2016; Jensen, Averbeck, Zhang, & Wright, 2013; Mudambi & Schuff, 2010). Table 2.1 summarizes these dimensions.

Table 2.1 Dimensions of information quality in two types of integrative crowdsourcing systems.

Information Quality Dimensions	Context	References
Correct identification (accuracy)	Citizen science	(Cox, Philippoff, Baumgartner, & Smith, 2012; Crall, Renz, Panke, & Newman, 2011; Nerbonne & Nelson, 2008; Salk et al., 2016)
Fitness for use		(Cox et al., 2012; Crall et al., 2011; Nerbonne & Nelson, 2008; Salk et al., 2016)
Context-dependent		(Hunter et al., 2013)
Usefulness		(Ballard, Dixon, & Harris, 2017; Gao, Barbier, & Goolsby, 2011)
Essential information		(Aceves-bueno et al., 2015)
Informativeness	Online review	(Gobinath & Gupta, 2016; Li, Hitt, & Zhang, 2011)
Expressiveness		(Korfiatis, GarcíA-Bariocanal, & SáNchez-Alonso, 2012)
Subjectivity of reviews including Self-involvement, other involvement, message involvement, and product involvement		(Dellarocas, Gao, & Narayan, 2010)

Table 2.1 shows that the focal aspects of information quality differ in different crowdsourcing contexts. Traditional information quality is tied to a specific use context (Nelson, Todd, & Wixom, 2005; Wang, Reddy, & Kon, 1995; Wixom & Todd, 2005), encouraging a shared understanding of the task requirement between the contributors and the project owner. Data collected with attention to traditional information quality dimensions might not be repurposable. Therefore, systems designed to focus on traditional quality dimensions might not be useful when information needs to be repurposed, requiring resource-intensive changes, usually involving new recruitment campaigns, restructuring and redesigning user interfaces and databases (e.g., see Lukyanenko et al., 2014), and possibly losing information because of the impermanent nature of contributors and observed events.

2.2.2 The Problem of Control

When integrative crowdsourcing projects focus on traditional dimensions of information quality (accuracy and completeness), design decisions such as crowd recruitment policies, task design, and system design strategies are guided by these dimensions. For instance, when developers of citizen science systems focus on accuracy, their systems design enforces tight controls on the types of data that can be contributed (Burgess et al., 2017; Ellwood, Crimmins, & Miller-Rushing, 2017). In some cases, this restrictive design limits the contributors who can participate to those who are familiar with the task. For instance, eButterfly and eBird are prominent citizen science platforms that require contributors to report their sightings of butterflies and birds they observed (see Figures 2.1 and 2.2).

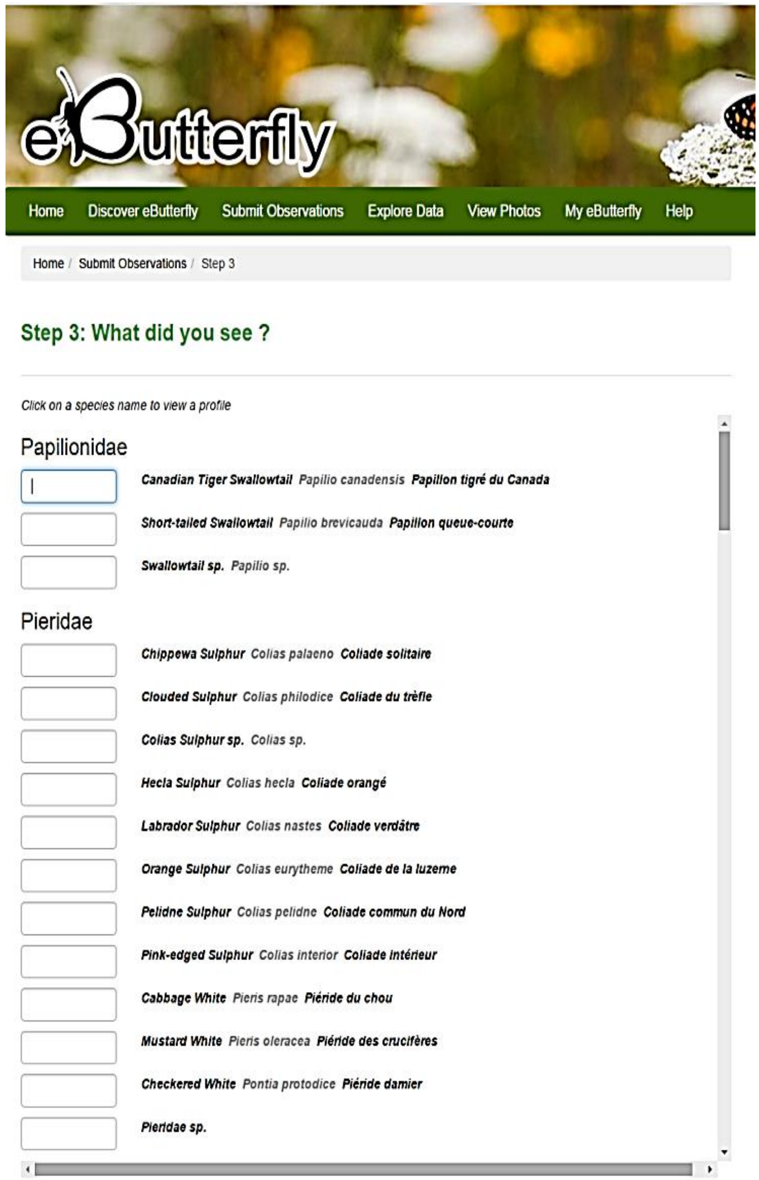


Figure 2.1. **Example of eButterfly's data reporting page**

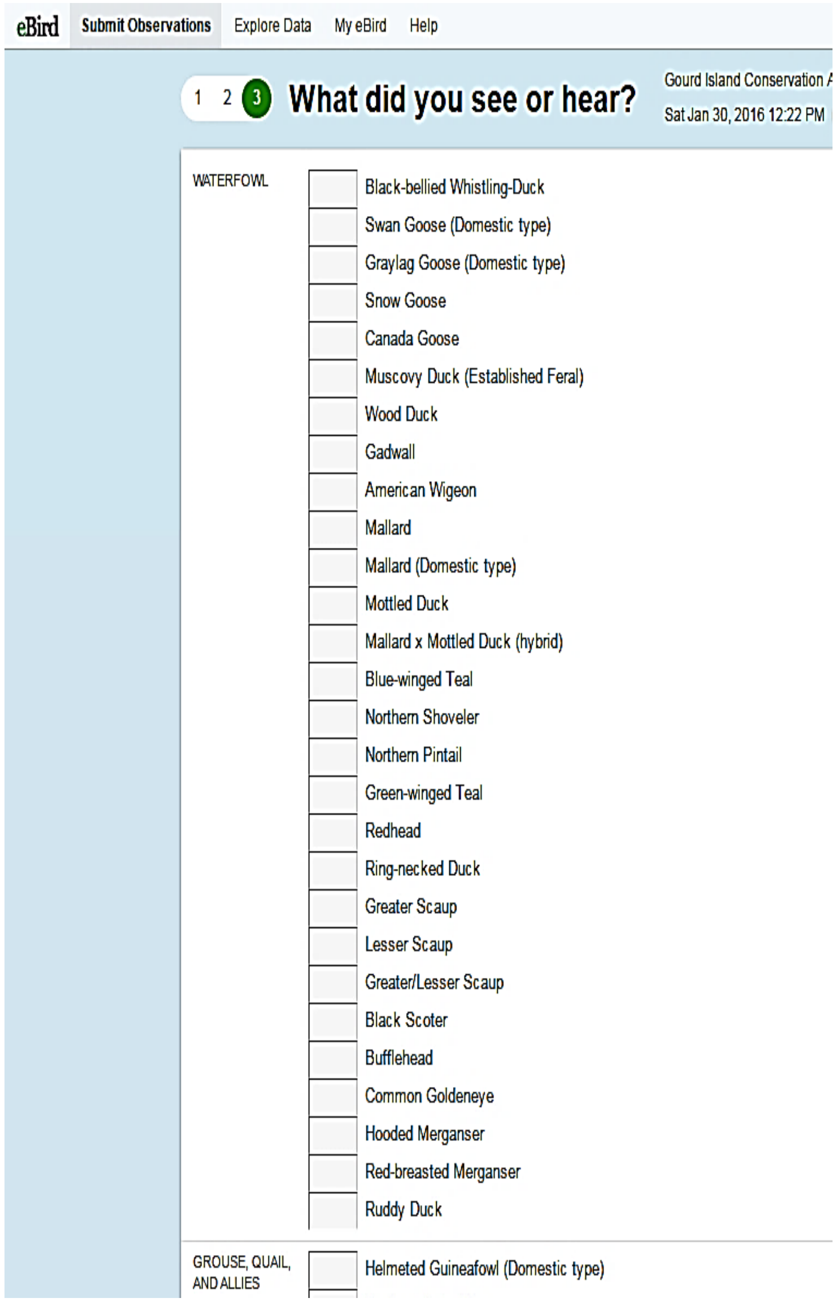


***Figure 2.2.* Example of eBird's data reporting page**

Both platforms (shown in Figures 2.1 and 2.2) require contributors to enter the number and species of the organism they have observed. These systems require contributors to know the scientific or common names of the organisms they are reporting. If an observed organism is unfamiliar, a contributor might guess or abandon the attempt to report a sighting (Parsons, Lukyanenko, & Wiersma, 2011). Crowdsourcers, like the designers of these citizen science platforms, prefer contributors with relevant knowledge or experience, under the assumption that knowledge and experience are positively related to information quality (Salk et al., 2016). For example, they may recruit people who have previously participated (or are presently participating) in a similar project (Bonter & Cooper, 2012; Burgess et al., 2017; Gura, 2013; Wiggins, Newman, Stevenson, & Crowston, 2011). We see contributor experience prioritized on platforms like Amazon Mechanical Turk, where the crowdsourcers can pay a premium to recruit "master" crowd workers for their tasks (Paolacci & Chandler, 2014; Peer, Brandimarte, Samat, & Acquisti, 2017). Alternatively, some crowdsourcers train participants to perform a task at an acceptable level (Cox et al., 2012; Hunter et al., 2013; Yang, Xue, & Gomes, 2018). Finally, active recruitment may be stopped when critical mass is reached while ensuring that current crowd members are retained (Nov, Arazy, & Anderson, 2011; Rotman et al., 2014). All these strategies assume that knowledgeable contributors provide more accurate and complete data, without considering the impact on repurposability.

Although preference for knowledgeable contributors is evident in practice and the crowdsourcing literature, several studies have reported that expert crowds (i.e., highly knowledgeable contributors) did not provide higher quality information than novice crowds. For example, three studies in an ecological context found that knowledgeable contributors did not provide more accurate or complete data than less knowledgeable contributors (Austen, Bindemann, Griffiths, & Roberts, 2016; Bloniarz & Ryan, 1996; Lukyanenko et al., 2014). Similar results have been reported in classifying damaged buildings (Staffelbach et al., 2015) and predicting movie marketing success (Escoffier & McKelvey, 2015). In the latter example, novices even outperformed knowledgeable contributors in terms of accuracy. Moreover, more accurate data are obtained when contributors are allowed to describe an entity they have observed in greater detail, stating its attributes (Lukyanenko et al., 2014). Collectively, these studies show that, contrary to conventional wisdom, a higher level of knowledge in a crowd does not necessarily result in improved information quality. This contention between conventional wisdom and the results of empirical tests in the literature is acknowledged by Ellwood, Crimmins, and Miller-Rushing (2017).

Studies by van der Velde et al. (2017) have argued that crowd members with limited knowledge can provide high-quality information even though experts are more precise (Lukyanenko et al., 2019). Experts also use fewer attributes to make classification decisions (Shanteau, 1992) and think more alike (McAuley & Leskovec, 2013) than non-experts. These characteristics of experts can work against the gathering of repurposable data and may instead facilitate homogeneity in crowdsourced information.

Crowdsourcers may attempt to ensure the reporting of only the presence of a set of attributes by implementing any of the control strategies already discussed. In contrast, we argue that data contribution should be structured such that contributors can create different sets of attributes of the entity that they consider relevant, from which different crowdsourcers can choose attributes that align with their current data requirements. In this case, the ability to choose desired attributes (which may be all or some attributes) from different sets of attributes is what we have referred to as repurposability, and the presence of different sets of attributes based on different contributor perspectives on what attributes are relevant is what we have termed information diversity. In other words, while data with a unitary view may be repurposable, diverse data can support greater repurposing.

In many cases, therefore, a focus on traditional dimensions of information quality inherently leads crowdsourcing systems to restrict the participation of interested crowd members based on their level of task-relevant knowledge (Burgess et al., 2017). Because traditional information quality dimensions are highly context-specific or use-specific, not generalizable to all types of integrative crowdsourcing projects, they cannot sufficiently guide the collection of repurposable data. Consequently, we propose the dimension of information diversity as a solution to the shortcomings of traditional information quality dimensions.

2.3 Defining and Measuring Information Diversity

Data is a crucial component of information systems, constituting "a perceptible representation of the real world from which a [consumer] can infer a view of the real-world system" (Wand & Wang, 1996, p. 89). In order to understand and adequately measure data, it is necessary to understand its structure. Like Wand and Wang (1996), we view data as a representation of real-world things. We, therefore, view data from the perspective of ontology.

Ontology helps us understand and describe things. The world is made of things, which are described in terms of their states and laws (Bunge, 1977; Wand & Wang, 1996). Humans understand and distinguish between things using attributes. For example, we assign a value to the colour attribute to distinguish rubies (red) from sapphires (blue). Things can also be composed of other things. Attributes of a thing help us define the state of the thing. Therefore, information about a thing may contain details of the attributes of the thing, the state of the thing, or other things that a thing is interacting with. Bunge's ontology posits that things can be described in terms of attributes which can either be *intrinsic*, i.e., solely depending on the thing, or *mutual*, i.e., dependent upon more than one thing (Bunge, 1977; Parsons & Wand, 2000; Wand & Weber, 1990).

Information about an entity can be expressed in terms of the attributes it possesses and the values of these attributes. In the example of emeralds, the shape "square" is an intrinsic attribute value. However, attribute values such as "precious" or "big" depend on the observer's prior experience, as well as the gemstone; they may differ from one contributor to another. The attributes of entities constitute the data in information systems. Attributes provide information about the properties of the entity (Wand & Weber, 1995). While people may use many different words when communicating, what provides relevant information about the observed entity are the attributes of the entity they report. Reported attributes about an entity can be analyzed to accurately determine the entity (Wand &

Wang, 1996). Examining information from the perspective of its constituent attributes not only helps us estimate the completeness dimension but also addresses the accuracy dimension of information quality.

The number of attributes and the types of attributes indicates the *diversity* within a dataset. Essentially, two or more pieces of (textual) information may be different from one another in terms of their diversity. Assuming that a piece of information can be broken down to sets of attributes, if A and B are two pieces of information contributors provide about an observed entity and A has attributes in common with B but more total attributes about the observed entity, then A is *more diverse*[2] than B. In other words, A is more diverse than B when the conditions in equations 1 and 2 hold.

$$|A| > |B| \quad \ldots (1)$$

$$|A \cup B - A \cap B| > |A \cap B| \quad \ldots (2)$$

Information diversity describes specifically the number and types or unique attributes reported in information about entities taking into consideration the similarity of the terms used (Ogunseye & Parsons, 2018). However, the number of attributes reported about an entity may differ among contributors depending on their perception of the requirements of the task and the differences in their knowledge of the task. For example, in

---

[2] *In these equations, we have assumed that all attributes contribute equally to information diversity*

a task requiring the reporting of any sighting of gemstones such as emeralds and diamonds in pictures containing any of three gemstones: diamonds, emeralds, and rubies, one contributor may report the following information: "the emerald is big." Another contributor may report their sighting of the same gemstone as "the square-shaped emerald looked big attached to the gold pendant." The latter contribution is richer than the former, containing more attributes about the entity emerald. The overall number of distinct attributes of an entity mentioned in contributed information can indicate the diversity of the information in a contribution.

Nevertheless, equation 1 does not necessarily mean two pieces of information A and B are diverse. For instance, if one contributor reports that they observed "green precious crystals" while two others report "green emeralds" and "emeralds," the observations have been reported at different levels of precision, but they convey similar amounts of information. Information consumers, only interested in the presence or absence of emeralds, would decipher the same amount of insight from each contribution (e.g., emeralds imply "precious green crystals"), and the contributions are thus equally diverse, even though some contained more attributes than others. To understand information diversity, we, therefore, need to consider more than just the number of attributes and include the type of attributes contributed about the entity, i.e., whether they are mutual[3] or intrinsic attributes.

It is also important to consider the meaning of the attributes and how different words may mean the same thing or describe the same attribute value. For instance, a reviewer may describe an item as “rare” or “scarce,” which means the same thing in this context. The degree of diversity between two pieces of information can be assessed by checking that the attributes they contain are dissimilar. Similarity has been defined as “the ratio between the amount of information in the commonality and the amount of information in the description of two objects” (Lin, 1998, p. 3). In this thesis, we consider this to mean the ratio between the number of attributes two datasets have in common, and the number of attributes available in total. Similarity has been measured by comparing the meanings of the words (attributes in our case) in contributed data (Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Lintean & Rus, 2012; Resnik, 1995; Tversky, 1977).

[3] Our reference to mutual attributes would focus on non-binding mutual attributes. “Non-binding mutual [attributes] are those [attributes] shared by two or more things that do not ‘make a difference’ to the things involved; for example, order relations or equivalence relations. By contrast, binding mutual [attributes] are those [attributes] shared by two or more things that do ‘make a difference’ to the things involved” (Rosemann & Green, 2002, p. 82). Kiwelekar & Joshi (2010, p. 4) further explains that non-binding mutual [attributes] are relational [attributes] that occur when “no interaction is involved between two related things. For example, younger than relationship between two persons does not show any kind of interaction”.

In general, two contributors are said to have provided similar contributions when the number of terms used to describe an entity, the frequency of terms used, the distance between terms, and the semantic properties of the terms used are the same (Gupta & Montezemi, 1997; Pirró, 2009; Tversky, 1977). Two contributions are considered diverse when these characteristics (the number of attributes, the types of attributes, and the degree of semantic diversity of these attributes) differ. We illustrate this for a case where information diversity is computed automatically, particularly in the case of large datasets. Nevertheless, information diversity can be estimated for very small datasets using simple statistics.

Calculating the information diversity index involves extracting attributes from textual data and classifying them as either mutual or intrinsic. Two contributions, A and B, with respective sets of mutual attributes, A_M and B_M, and intrinsic attributes, A_I and B_I, may have several attributes. Comparing their diversity involves determining how different the attributes are, considering all of the available attributes in both texts, i.e., $\frac{A \cup B - A \cap B}{A \cup B}$. If we assume that $A_{M3} = B_{M3}$ and $A_{I2} = B_{I3}$, then $(A \cup B - A \cap B) = \{A_{M1}, A_{M2}, A_{I1}, A_{I3}, B_{M1}, B_{I2}\}$, and the diversity index would be $\{A_{M1}, A_{M2}, A_{I1}, A_{I3}, B_{M1}, B_{I2}\}/\{A_{M1}, A_{M2}, A_{M3}, A_{I1}, A_{I3}, B_{M1}, B_{I2}, B_{I3}\}$. This gives

$$1 - \frac{A \cap B}{A \cup B} \quad (3)$$

where $\frac{A \cap B}{A \cup B}$ estimates the similarity in the number of attributes and their meaning, providing a similarity index. More appropriately, equation 3 can be represented as 1-SIM(A,B) where

SIM(A, B) is a function that maps the degree of similarity of entity attributes between information A and B to an index between 0 and 1 where 0 implies absolutely no similarity exists between the attributes and 1 implies the attributes in the two pieces of information are the same. There are several ways to determine in numeric terms how similar two attribute sets are (Gomaa & Fahmy, 2013). One example is to compute the cosine similarity of both attribute sets A and B. Cosine similarity is used to measure the similarity of texts (which are converted to vectors) based on the cosine of the angle between them (Dehak, Dehak, Glass, Reynolds, & Kenny, 2010; Mihalcea, Corley, & Strapparava, 2006). Cosine similarity places more emphasis on the meaning of the text rather than the length of texts. For the computations to take place, attributes are changed to numeric values (that is, vectorized) using word vectorization libraries that retain their contextual meanings, such as Word2Vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013) or GloVe (Pennington, Socher, & Manning, 2014).

To illustrate, consider the two reviews in Table 2.2 below. We extract the attribute values of the reviewed entity as adjectives – i.e., words that describe a noun, classifying the attributes into Mutual and Intrinsic using machine learning. To determine which attributes are mutual and which are intrinsic to an entity, we built on a polarity detection algorithm from spaCy (www.spacy.io) used in sentiment analyses. Intrinsic attributes – or adjectives – would not show significant polarity. E.g., three legs (three is neutral); purple coat (purple is neutral). In contrast, mutual attributes would show polarity, e.g., expensive ring (expensive may be negatively or positively polar and is dependent on the contributor); weak handles (weak is negatively polar and is a judgment dependent on the contributor).

We also compute the similarity of the attributes from which we calculate the percentage diversity, i.e., *percentageID = ((1 – SIM(A, B)) *100)*. However, even though the example in Table 2.2 addresses the determination of diversity between two pieces of contributed information, information diversity can be extended to two or more large datasets.

Table 2.2: Computing information diversity

Review	Unique Mutual and Intrinsic Attributes	Number of Intrinsic Attribute values	Number of Mutual Attribute values	Percentage Similarity	Percentage Information Diversity
I like these tools They're plastic and run on batteries but they work. Not as powerful as your gardener's gas powered machines but for the homeowner who doesn't plan on any heavy duty shrubbery cleaning these will do the trick nicely	power, plastic, heavy	1	2	11.418	88.582
Pretty low tech item but that's what I wanted. This one is small version which means if you have lots of leaves you'll be cleaning often.	low, small	0	2		

2.4 The Information Diversity Framework

We identify three factors that are essential for information diversity: the data model, the nature of the crowdsourced task, and the differences in contributors. These factors are built upon the collective intelligence genomes proposed by Malone, Laubacher, and Dellarocas (2010). They posit that four crucial questions need to be asked when making decisions about how to design collective intelligence projects—for which crowdsourcing is often used (Lukyanenko et al. forthcoming). These questions are: Who do we want in our crowds? What should be the problem that we pose to the crowd? Why would the crowd want to participate in our project? How should we structure the task? We separate these building blocks into human factors and system design factors and extend their model to include available information technology (IT) infrastructure. We explore these three building blocks in greater detail below, showing how they serve as a framework for information diversity (see Figure 2.3).

Available IT infrastructure. First, conceptual modeling literature has long emphasized the limitations of context-based or view-based data modeling. Parsons and Wand (2000) proposed an instance-based data model that described the need to represent things and their properties independently of predefined classes, enabling data to be used by different consumers with different views. They showed that it is possible to repurpose (reclassify) stored data that are not tethered to any classification scheme (Asgari, Parsons, & Wand, 2017; Saghafi, Wand, & Parsons, 2016). A data modeling strategy that is inconsistent with the principles of modeling data independent of a schema is seen in relational databases, where data schema are fixed and their evolution constrained (Codd, 1989). The relational data model and systems based on it have successfully allowed users to provide information that is congruent with the existing conceptualization of the data requirements of the system (Codd, 1989). High costs may be incurred when an existing database schema is altered.

However, the continued need for crowd-facing systems has necessitated the implementation of data models that are schema-free, allowing contributors to provide unstructured and structured data, upon which users can create need-based schemas at the application level (Leavitt, 2010). These database architectures—known as non-relational databases—are more in line with the data modeling approach proposed by Parsons and Wand (2008) than relational models, allowing systems built on them to inherently accommodate diverse information. Non-relational database architectures are used by major collectors of crowd data (user-generated content) like Facebook, Google, Twitter, and Amazon (Cattell, 2011). They are faster and scale better than traditional relational databases (Moniruzzaman & Hossain, 2013). Non-relational databases, facilitated by Web 2.0, allow for the collection of data from distributed groups of people and facilitate the collection and storage of diverse data.

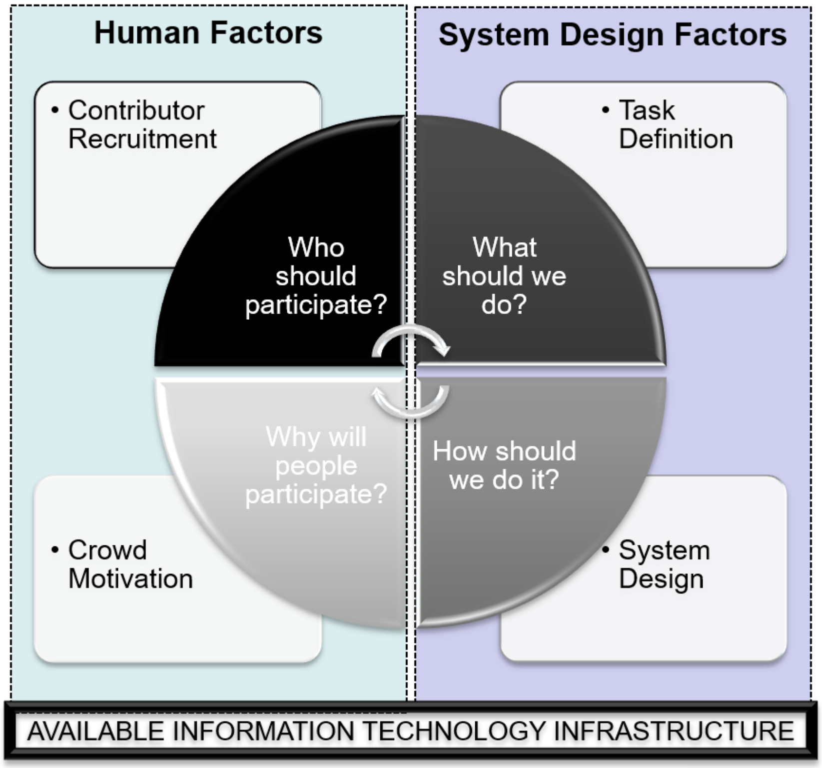


Figure 2.3. **The Information Diversity Framework.** Building blocks for information diversity (adapted from Malone et al., 2010).

Human Factors. Contributor differences, which may be spatiotemporal (an extrinsic factor) or cognitive (an intrinsic factor), play a role in the ability of crowdsourcing systems to collect diverse data. Extrinsic factors, like location and time, may be indicated in the information provided by contributors. These factors provide context to information about entities and are usually measured by the completeness dimension of information quality. However, beyond spatiotemporal differences in crowd members, a significant source of information diversity is cognitive diversity—differences between people resulting from differences in their knowledge and experiences (Sauer, Felsing, Franke, & Rüttinger, 2006), which can result from different training (Piven et al., 2006) or differences in professional and personal backgrounds (Colón-Emeric et al., 2006).

The ability of cognitive diversity to positively affect productivity and idea generation has been emphasized in the literature (Polzer, Milton, & Swarm Jr, 2002; Wu, Chen, Hui, Zhang, & Li, 2015), and harnessed for online review platforms (Mudambi & Schuff, 2010) and open innovation (Brabham, 2008). Cognitive diversity is the differences in how people frame and approach problems, organize and use information, and communicate, and the information that they produce (Kloos & Sloutsky, 2008; Mello & Delise, 2015). For example, Best et al. (2013) showed that infants and adults focus on different aspects (attributes) of the same phenomenon, because of differences in their knowledge. Similarly, Hoffman and Rehder (2010) and Spence and Brucks (1997) argued that people differ in the type of information they use or produce because of cognitive diversity. In agreement with Polzer, Milton, and Swarm Jr (2002) and Matzler, Füller, Hutter, Hautz, and Stieger (2016), we argue that cognitive diversity is an antecedent of information diversity and a necessity for repurposable data. We, therefore, argue that the cognitive diversity of crowds is a foundation for information diversity.

Cognitive diversity may also affect the motivation of crowd members (Frey, Lüthje, & Haag, 2011). Motivation may be different for individuals with different personality factors (Lee, Crowston, Harandi, Østerlund, & Miller, 2018). The quantity of effort that contributors commit to a crowdsourcing project may be motivated intrinsically due to the level of enjoyment of the task, or extrinsically – because of incentives or pressures external to the task (Liang, Wang, Wang, & Xue, 2018). Antecedents of intrinsic motivation like cognitive diversity affect the type of information collected in crowdsourcing projects (Crowston & Prestopnik, 2013; Ogunseye, Parsons, & Lukyanenko, 2017). Therefore, as crowdsourcers make decisions about who to recruit into crowds to ensure information diversity, it is necessary to consider the cognitive diversity of crowd members.

System Design Factors. The last building block of our information diversity framework is task design. Task design encompasses the two categories of collective intelligence "key questions" espoused by Malone 2010. These include what is to be done and how it should be done. Research on the impact of system design on information quality reveals that the design of systems can be restrictive, limiting the ability of crowd members to report information freely, based on their perspectives, and thereby discouraging information diversity (Lukyanenko et al., 2019). In addition, the nature of the crowdsourced task will affect the diversity of information contributors can provide. The degree of structure of tasks, the number and complexity of the decision inputs, the ease with which inputs can be evaluated, the amount of noise present in the inputs, and the ease with which the task can be decomposed are among the factors that can impact the type of data collected (Spence & Brucks, 1997). A well-structured, well-defined task would create a level playing field for cognitively diverse contributors, whereas tasks that require some structuring before they can be addressed may be more suitable for contributors with prior knowledge.

Nonetheless, the absence of this framework does not preclude the possibility of collecting diverse information. Systems built on relational databases could still collect diverse data; likewise, even though unlikely, cognitively similar crowd members may provide diverse data. However, we argue that the presence of one or more of these building blocks would impede the collection of homogeneous data.

2.5 Theoretical Support for the Information Diversity Dimension

The collection of diverse data is already technologically permissible through non-relational databases, which serve as a framework on which we can begin to build crowd-facing applications that harness the benefits of information diversity and repurposability. These new information systems environments, in which organizations can directly access crowds as data sources, have been dubbed by Parsons and Wand (2014) as *Open Information Environments*—information systems environments that accommodate diverse contributors' perspectives, users, and uses of data, including unanticipated uses (Parsons & Wand, 2014). According to Parsons and Wand (2014), information systems that operate in open environments should accommodate semantic diversity, ensure information quality, and allow for flexibility. We argue that crowd-facing information systems that can collect diverse information will meet these requirements.

Ability to accommodate semantic diversity. Today's open information systems must accommodate information contributors and information consumers who may have different views. Cognitive psychology literature shows a relationship between cognitive diversity and ability in a group, resulting from differences in experience and training (Colón-Emeric et al., 2006; Martins, Schilpzand, Kirkman, Ivanaj, & Ivanaj, 2013; Piven et al., 2006). Best, Yim, and Sloutsky (2013), Hoffman and Rehder (2010), and Kloos and Sloutsky (2008) all show that groups with cognitive diversity (people with different training or experience) provide data containing more distinct attributes of an entity. The presence of different attribute types in crowdsourced data is evidence of different perspectives, and such data can accommodate multiple views (Barsalou & Sewell, 1984; Parsons & Wand, 2000). The capacity for multiple views, usually through multidisciplinary teams, using data high in information diversity, is a critical success factor for big data (Günther et al., 2017).

Ability to ensure information quality. When information is to be repurposed, there is a high probability that previous information quality standards will no longer apply. One problem identified in (Woodall & Wainman, 2015) is that the required data for the new task may not be available because crowdsourcers were not aware that they would need the data for the new task. We know that a lack of diverse data may negatively affect future tasks; a valid question then is, what about current tasks? How does diversity affect known uses of data? Parsons and Wand (2014) identify this as a requirement for open information environments. Information diversity should, therefore, not impede the information system's ability to meet traditional requirements of information quality, such as accuracy and completeness. We explore the relationship between diverse data and the information quality dimensions identified in Wand and Wang (1996).

Information Diversity and Accuracy. According to Wand and Wang (1996), accuracy is an operation on correctly identified attributes of an entity that maps back to the correct entity and its state in the real world. For example, if in a citizen science task, contributors are required to report the types of precious stones in a given location, the attributes identified about the precious stones observed should be sufficient to correctly determine the type of precious stone available. In this case, it is expected that humans or machines can correctly identify the type of precious stone when provided with correct information about the state of the entity observed. Information diversity measures and encourages the reporting of different attributes of an entity in contributed information. Since contributors would be reporting from different perspectives, more attributes about the entity would be reported overall, than if the contributors had the same perspective. Most of the attributes required to identify an entity correctly would, therefore, be available in information high in diversity, more so than information pooled from only people with a singular perspective. We can, therefore, conclude that information diversity would support accuracy. The information diversity dimension does not replace the accuracy dimension but improves it, providing more details about the state of the real world that would lead to the correct operationalization of the data by the consumer.

Information Diversity and Completeness. Information consumers assess completeness from the perspective of their needs and not the actual completeness of the properties of a thing in terms of its intrinsic properties or its mutual properties. Complete data means data containing all attributes required by a specific data consumer for a particular use. Information diversity encourages the collection of information that meets multiple views of the entity and will support many of these views completely, providing the attributes needed to make decisions from those views. Therefore, information diversity will support completeness.

Figures 2.4a and 2.4b illustrate the consequence of the completeness dimension based on ontology. Figure 2.4a shows that, if a contributor C1 reports Attributes 1 and 2 about an entity in the real world (RW), and if these attributes are considered sufficient for the task at hand by information consumer U1, the information is complete according to the traditional definition of information completeness and quality. However, if Attribute 4 of the entity becomes relevant in the future, a new information-gathering process would have to be initiated, or novel insights may need to be forfeited (Bonter & Cooper, 2012). However, Figure 2.4b shows that if information diversity is encouraged, such that contributors C1 and C2 provide different perspectives to the information source, consumers U1 and U2 can derive multiple views from the data.

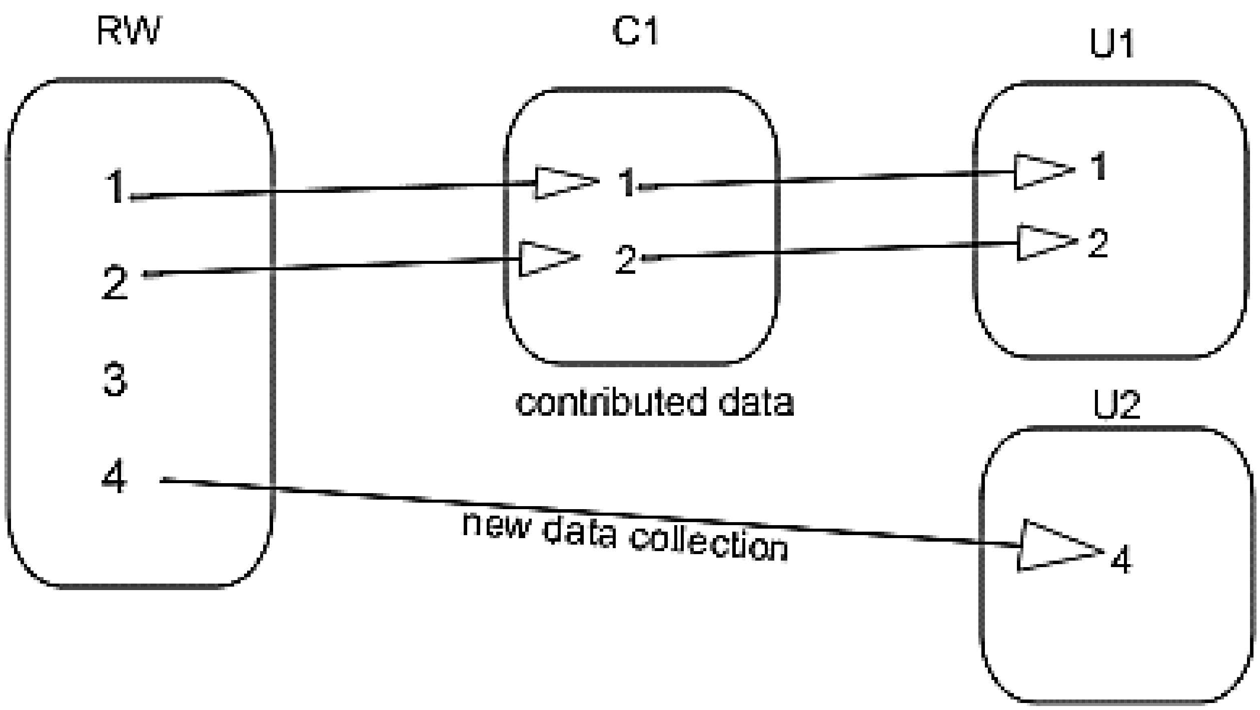


Figure 2.4a. **Completeness *without* information diversity.**

When completeness for a predetermined purpose is the focus of an integrative crowdsourcing task, contributors can report only attributes of the entity in the real-world (RW) that meet the requirements of the task; in this case, attribute 1 and 2 (i.e., C1) for user U1. If a new user U2 ever needs to repurpose the data to get insights that involve RW attribute 4, U2 will need access to the original entity, which may require repeating the information crowdsourcing task or may be impossible if the original phenomenon cannot be repeated.

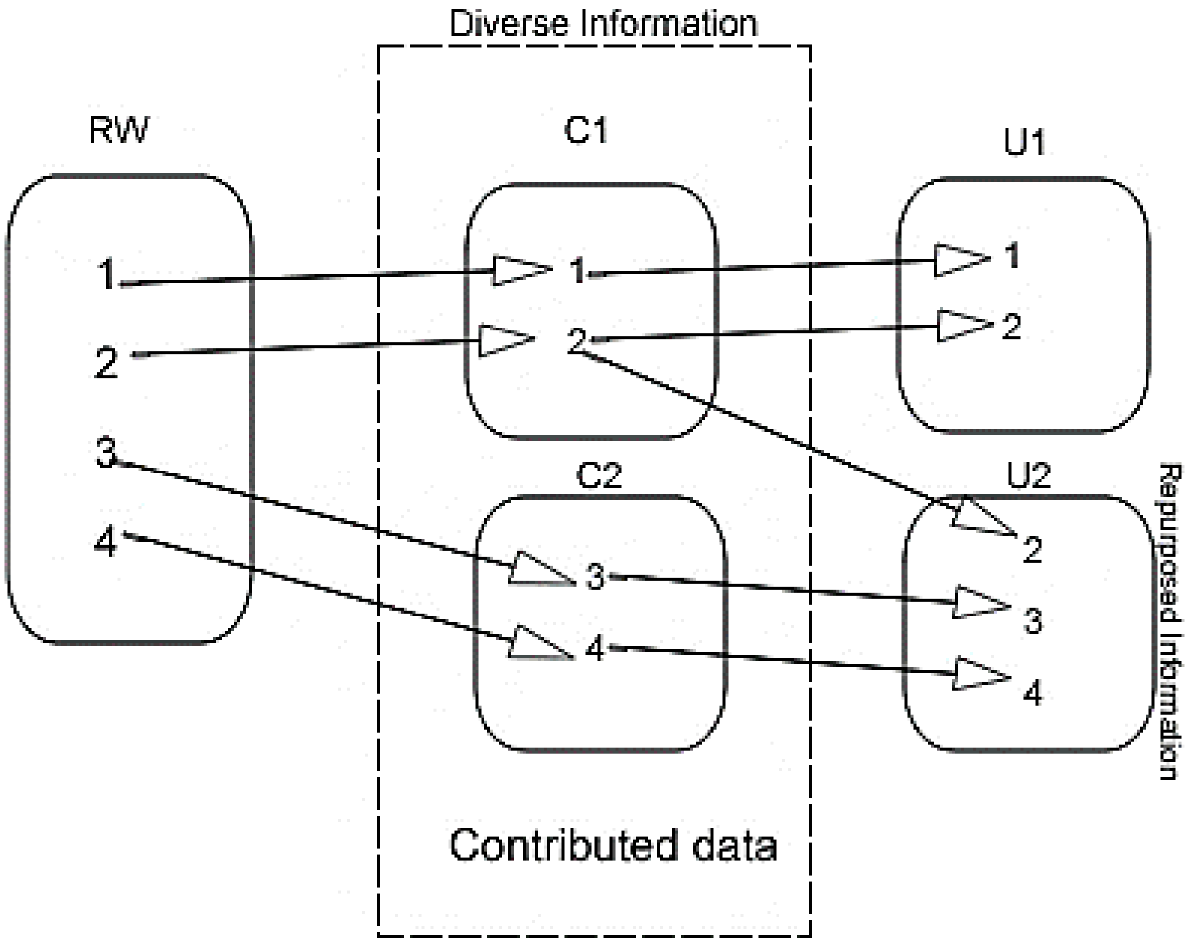


***Figure 2.4b.* Completeness *with* information diversity.**

When information diversity is the focus of an integrative crowdsourcing task, contributors can report any attribute of the entity in the real-world (RW) and not only those that are relevant for an immediate task (e.g., C1 and C2 report different attributes of the entity). If a user U2 ever needs to repurpose the data to gain insight about attribute 2, 3 and 4 of the entity, the information crowdsourcing task does not need to be repeated, and the user does not need access to the original RW phenomena because contributors in the crowd will have provided ample information about it.

Focusing on Completeness, which is use-oriented, reduces the likelihood of collecting repurposable information, whereas focusing on information diversity ensures that complete information is collected. Therefore, information diversity subsumes the completeness dimension. Moreover, because accuracy is mainly an operation on the complete attributes of an entity, information diversity also supports the collection of accurate data.

Ability to ensure information flexibility. Parsons argued, "different people (or the same person at different times) may organize knowledge about things according to a different set of classes or categories" (Parsons, 1996, p. 1436). Therefore, data that fits a single view may be useless soon after the view evolves, requiring more information that was not initially collected. What is complete at one moment, or usage instance, may not be complete at another. This also applies to the problem of insufficient information. The level of granularity applied to a thing can change, and attributes that did not matter earlier may later become important. Imagine that a contributor reports that a given insect can fly and that this information is considered enough for classification today. If we learn later that there are two species of this insect and that one flies with its body facing downward while another flies with its body facing upward, the single attribute recorded becomes insufficient and incapable of providing the required insight. Information diversity supports flexibility, allowing for emergent uses of data. This is shown in Lukyanenko et al. (2019), which reported the results of data collection using a citizen science system designed to allow people to report data freely. They found that the diverse data collected in their prototype open citizen science system was also more complete than the less diverse information collected in a more restrictive system.

2.6 Application of the Information Diversity Framework

Existing literature has shown the usefulness and value of repurposing crowdsourced data. To illustrate the viability of information diversity, we review two articles from two separate research domains: ecology and agriculture. We have not set out to discuss the quality of the research but to discuss how they provide evidence for the benefits of diverse data for insight and information quality.

The papers we discuss both employ data from Twitter—a microblogging site that allows registered users to post media and text of not more than 280 characters in length. Twitter uses a variant of the NoSQL database and IT infrastructure that allows it to collect and process unstructured data in petabytes per year (Lai, 2010). User-generated contributions to Twitter (tweets) are unrestricted and can be flexibly categorized by the user using hashtags. Twitter also allows anyone, regardless of their level of knowledge, to contribute data on any topic of interest to them.

The first paper that we examined titled: *Testing the potential of Twitter mining methods for data acquisition* (Hart, Carpenter, Hlustik-Smith, Reed, & Goodenough, 2018), compared mined Twitter data to the results of three previously published studies that used traditional citizen science methods to collect data. The first study reportedly used citizen science to quantify the spatiotemporal distribution and environmental triggers of ant mating flights (Hart, Hesselberg, Nesbit, & Goodenough, 2018). The second study used citizen science to investigate the “geographical patterns, seasonal peaks daily rhythms, and location of spiders … within houses during the autumn” (Hart, Nesbit, & Goodenough, 2018, p. 2195). In the third study, the focus was on monitoring the behaviour of starlings using citizen science to understand how predators and temperature impact them. The published results of these three studies, when compared to mined tweets from Twitter, showed that repurposed Twitter data accurately replicated the results, including spatial and spatiotemporal findings of the published citizen science studies. In particular, the study on winged ants revealed that very few tweets (5 of 597) provided unambiguous information identifying the species of the ants. Most of the tweets contained attributes describing the ants. However, twitter-derived data on the temporal patterns of the ants showed “remarkable agreement” with national scale temporal patterns described in existing research. Similarly, there was significant similarity in the location and direction of movement of ants as reported in twitter data and previous research.

For spiders, there was also no significant difference in the temporal distribution of recorded sighting and from tweets. The time of day in which the spiders were spotted in research and reported on twitter differed, with twitter reports being made later than research results. However, the location of spider sighting was similar. The sex of the spiders observed showed a male bias for both twitter data and research data 75.4% and 82.3% male, respectively. Nonetheless, there was no significant association between both results.

Finally, for starlings, the spatial patterns reported in previous research were also detected in tweet data. The key hotspots reported on twitter coincided with those identified in research by Goodenough, Little, Carpenter, and Hart (2017).

A second paper, by Zipper (2018) titled "Agricultural Research Using Social Media Data," investigated the utility of social media for monitoring spatiotemporal patterns in agriculture. The study used Twitter data to map state-level corn and soy planting progress, comparing the result of their analysis with traditional survey-based monitoring mechanisms. Specifically, this research compared the result of their repurposing of Twitter data to data acquired from the US Department of Agriculture National Agriculture Statistics Service (NASS). They found that Twitter data was significantly similar to the NASS data. The discrepancy between the results was stated to be attributable to the incompleteness of the NASS data or some inadequacy in any of the datasets.

Furthermore, the Twitter data provided additional contextual insights not available with the NASS dataset on the causes and indicators of replanting—a difficult but sometimes necessary decision that farmers must make to sustain their farming operations. Twitter data provided insights into the "extent, causes, and decision-making process related to replanting decisions" and agricultural management practices. Twitter data also provided above NASS data, contextual information regarding farmer sentiments about agricultural products and their shifting beliefs about agricultural practices. It allows for the tracking of adoption of agricultural practices and can be a source of guidance to agricultural extension services on which parts of the country to direct their efforts and what information or training need to be intensified. Generally, Twitter data was sufficient to track planting progress across states, helping to measure the spatiotemporal dynamics of crop planting and would be useful for monitoring emerging issues in agriculture. Nonetheless, the author recognizes that the quantity of useful data in a purely social media approach to citizen science can be limited and thus problematic.

In conclusion, both studies show that diverse information is repurposable high-quality information, which matches that gathered using rigorous scientific and citizen science processes. Also, Zipper (2018) showed that the repurposing of diverse data could lead to insights not readily available in more targeted, non-diverse data. These studies, the success of online review systems, and several other studies already described here are indicators of the viability of information diversity not only to support information quality for known and predetermined uses of data but also to support high-quality decision making for unanticipated uses of data.

2.7 Discussion

The world of information systems is changing. The climate of the era is that of crowdsourcing, repurposable data for analytics, and unconstrained contribution. While there is merit in crowdsourcers instituting and maintaining stringent controls on data contribution for integrative crowdsourcing systems that seek to collect data for some purposes, other integrative crowdsourcing projects that seek to facilitate novel discoveries would benefit from allowing information diversity. Moreover, even when data collection needs to follow strict protocols, a hybrid approach, in which the contributor is also allowed to contribute freely after contributing data that fit the crowdsourcer’s immediate requirements, may allow for the collection of crowdsourced data that can adapt to changes in hypotheses and business needs.

This study contributes to the discussion on the need to go beyond traditional data quality measures (Lukyanenko, Parsons, & Wiersma, 2016), and to the literature on open information environments. The information diversity dimension introduced in this chapter can guide the measurement of variety and insight, both for big data research and smaller-scale crowdsourcing projects. While humans, with their limited cognitive resources, prefer precision in collected data, so that they are easier to analyze, the future of data usage is machine-driven, with various automated analytics approaches that have been created. Our discussion of information diversity, therefore, seeks to extend traditional information quality measurements to cater to the need to easily process large unstructured data with no negative impact on our ability to determine accuracy or completeness for known and emergent data uses.

Nevertheless, several studies already provide useful guidance on the different aspects of our proposed framework. For instance, Bonney et al. (2009), Cooper, Dickinson, Phillips, & Bonney (2007), and Wiggins & Crowston (2012) explore design decisions relating to *the goal of the crowdsourcing project.* Specifically, Wiggins and Crowston (2012) described the typologies of citizen science projects based on their goals. This typology was determined from the projects' "characteristics and needs." The result of their work reveals a "relationship between resources, geographic scale of projects, and the relative emphasis on different combinations of goals in citizen science projects." It may serve as a framework, helping sponsors in their formulation of project goals. Similarly, *the* *motivation to contribute or reasons why crowd members will participate in a crowdsourcing project are discussed in* (Lee, Crowston, Østerlund, & Miller, 2017; Nov, Arazy, & Anderson, 2011; Raddick et al., 2009; Rotman et al., 2012). For example, Lee et al.'s (2017) work showed that recruitment "messages appealing to learning, contribution and social proof were more effective than a message appealing to altruism" (p 227). Guidance on *how crowdsourcing systems should be designed* is also provided in Lukyanenko et al. (2017), and Lukyanenko, Parsons, Wiersma (2014), with Lukyanenko *et. al*. 2014 showing that "the practice of modeling information requirements in terms of fixed classes unnecessarily restricts the IQ of user-generated data sets" (p 669). This desing research sheds light on the implications of system design choices for the accuracy and dataset completeness of crowdsourced data.

However, for insights pertaining to contributor recruitment, empirical and practical guidance are based on the assumption that a crowd of knowledgeable contributors will provide better quality data than a crowd of less knowledgeable contributors. Consequently, contributor selection is primarily based on the "… knowledge of contributing individuals" as this helps sponsors "feel comfortable with data quality" (Wiggins et al. 2011 p.17). Research and practice, therefore, favour recruiting knowledgeable contributors over less knowledgeable ones (Wiggins, Newman, Stevenson, & Crowston 2011; Budescu & Chen 2014). For example, Budescu & Chen provide a strategy for testing and **"**eliminating poorly performing individuals from the crowd" by identifying experts in the crowd "who consistently outperform the crowd." Nevertheless, an exclusively positive association between task proficiency or experience and data quality is unsubstantiated in several research including Austen, Bindemann, Griffiths, & Roberts, (2016); Kallimanis et al. (2017); & Crall et al. (2011) which report that both knowledgeable and less knowledgeable contributors provide crowdsourced data that are accurate and usable. Therefore, current guidance on recruitment for crowdsourcing based on this assumption is at best questionable.

Our focus will, therefore, be on human factors as it concerns the recruitment of crowds, and in the subsequent chapters of the thesis, we dive deeper into the effect of cognitive diversity on information diversity. We address how the type and level of knowledge contributors possess, and the longitudinal effect of contributor knowledge affects information diversity. In addition, we investigate the dependencies between information diversity and the traditional dimensions of information quality, particularly accuracy and completeness. Understanding how cognitive diversity impacts information diversity and quality would help inform crowdsourcers on who to recruit into their crowds, and researchers on how to design systems that harness the strengths and mitigate the weaknesses of contributors based on their level and type of knowledge.

2.8 Conclusion

In traditional organizations, decision-makers can control processes for information creation and management, choosing who will be allowed to provide data and how it will be used. In this setting, information consumers can assess the accuracy and completeness of contributed information (Parsons & Wand, 2014). Because the use of information is predetermined, it is sensible to define information quality as the fitness of the information for its use, and more information can be requested when necessary. However, advances in web technology allow organizations access to data sources outside their control. Since the possible uses of data may be emergent and data collection is terminal, data is most valuable if it is applicable for purposes beyond the original intent at the time it is collected. Therefore, the quality of data is no longer based solely on its ability to meet anticipated requirements, but also on its ability to meet unanticipated needs.

However, the definition and measurement of information quality have not evolved to reflect the change in information quality required for open information environments. The measurement of information quality has been guided by information consumers' classifications, which implicitly use a consumer's view of *what information is needed for a task* to determine who is recruited and how the crowdsourcing system and the task are designed. This measurement of information quality is targeted toward data that is suitable for predetermined uses. However, both conceptual modeling and data quality research emphasize that data are more valuable and provide more insights to users when it can be repurposed by different users, for both anticipated and sometimes unanticipated uses.

Repurposable data can answer a variety of questions from the same or different users. It is useful for integrative crowdsourcing systems that operate in open information environments, pooling information from disparate, spatially, and temporally distributed volunteers about a phenomenon of interest to the crowdsourcer. Because the hypothesis motivating data collection may not be fully formed at the time of data collection, crowdsourcers interested in repurposability prefer rich datasets, adaptable to emerging user requirements and unanticipated needs. Moreover, now that data are increasingly traded and purchased by different organizations as a resource, the need to estimate repurposability is even more pressing.

Repurposability is essentially what data science is about: transforming data from one form to answer new questions or to provide new insights for decision-making purposes. Therefore, we extend the measurement of information quality to cover the need for business insight from crowdsourced data. While the literature provides insight into how to measure and improve traditional dimensions of information quality, particularly accuracy and completeness, insights from repurposing data can give organizations competitive advantages. More knowledge is therefore needed about how to measure the repurposability of crowdsourced data. We posit that repurposability is a direct consequence of information diversity and improving the value of crowdsourced data implies encouraging information diversity.

By including the information diversity dimension in the information quality dimension, information quality can be used to assess the value of data for repurposability. At the same time, the addition of an information diversity dimension will make information quality generalizable as a measure of information quality to all types of integrative crowdsourcing systems. The information diversity dimension will provide insights into the quality of crowdsourced information where traditional measures fall short and will encourage the design of inclusive crowdsourcing systems.

2.8.1 Limitations

Four core impediments to information quality exist: (a) incompleteness, when there are states in the real world that are not represented in the data; (b) insufficiency, when several states of the real world are represented by the same attributes in the data; (c) meaninglessness, when data contain attributes that do not exist in the real world; and (d) inaccuracy, where attributes in data cannot be mapped to real-world states correctly. Based on ontology theory, incompleteness, insufficiency, and meaninglessness are design deficiencies of information, and they can lead to operational deficiencies such as inaccuracy. Wand and Wang describe accuracy to be a result of the user's interpretation of the data. Inaccuracy results when operationalization of data are incorrect or based on a deficiency in the representation of states of the real world. We consider incompleteness and insufficiency to be variations of the same dimension, defined by missing states of the real world and have treated them as the same.

We also do not address the meaninglessness dimension. We contend that if data collection is goal-directed, contributors will provide information about the states of the real world that they observe, which they consider relevant, matching the real world, and leading to a more faithful and detailed representation of the real world. Moreover, Nelson et al. (2005) argued that accuracy also means meaningfulness. Since diversity can support accuracy, it follows that diversity can support meaningfulness. Here, we have considered the dimension of meaningfulness (or the problem of meaninglessness) as self-evident.

References

Aceves-bueno, E., Adeleye, A. S., Bradley, D., Tyler Brandt, W., Callery, P., Feraud, M., … Tague, C. (2015). Citizen Science as an Approach for Overcoming Insufficient Monitoring and Inadequate Stakeholder Buy-in in Adaptive Management: Criteria and Evidence. *Ecosystems*, *18*(3), 493–506. https://doi.org/10.1007/s10021-015-9842-4

Ardagna, D., Cappiello, C., Samá, W., & Vitali, M. (2018). Context-aware data quality assessment for big data. *Future Generation Computer Systems*, *89*, 548–562.

Asgari, M., Parsons, J., & Wand, Y. (2017). Attribute Lattices: A Schema-free Unified Framework for Data Semantics. *ER Forum/Demos*, 142–155.

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. *Scientific Reports*, *6*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5028888/

Ballard, H. L., Dixon, C. G., & Harris, E. M. (2017). Youth-focused citizen science: Examining the role of environmental science learning and agency for conservation. *Biological Conservation*, *208*, 65–75.

Ballesteros, J., Carbunar, B., Rahman, M., Rishe, N., & Iyengar, S. S. (2014). Towards safe cities: A mobile and social networking approach. *IEEE Transactions on Parallel and Distributed Systems*, *25*(9), 2451–2462.

Barsalou, L. W., & Sewell, D. R. (1984). *Constructing representations of categories from different points of view*. Retrieved from http://eprints.gla.ac.uk/112570/

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119.

Bloniarz, D. V., & Ryan, H. D. P. (1996). The use of volunteer initiatives in conducting urban forest resource inventories. *Journal of Arboriculture*, *22*, 75–82.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Bonter, D. N., & Cooper, C. B. (2012). Data validation in citizen science: A case study from Project FeederWatch. *Frontiers in Ecology and the Environment*, *10*(6), 305–307.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, *14*(1), 75–90.

Bunge, M. (1977). Treatise on basic philosophy: Ontology I: the furniture of the world (Vol. 3). Springer Science & Business Media.

Burgess, H. K., DeBey, L. B., Froehlich, H. E., Schmidt, N., Theobald, E. J., Ettinger, A. K., … Parrish, J. K. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation*, *208*, 113–120. https://doi.org/10.1016/j.biocon.2016.05.014

Cattell, R. (2011). Scalable SQL and NoSQL data stores. *Acm Sigmod Record*, *39*(4), 12–27.

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaivete among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–130. https://doi.org/10.3758/s13428-013-0365-7

Cheung, C. M.-Y., Sia, C.-L., & Kuan, K. K. (2012). Is this review believable? A study of factors affecting the credibility of online consumer reviews from an ELM perspective. *Journal of the Association for Information Systems*, *13*(8), 618.

Codd, E. F. (1989). Relational database: A practical foundation for productivity. In *Readings in Artificial Intelligence and Databases* (pp. 60–68). Elsevier.

Colón-Emeric, C. S., Ammarell, N., Bailey, D., Corazzini, K., Lekan-Rutledge, D., Piven, M. L., … Anderson, R. A. (2006). Patterns of medical and nursing staff communication in nursing homes: Implications and insights from complexity science. *Qualitative Health Research*, *16*(2), 173–188.

Cox, T. E., Philippoff, J., Baumgartner, E., & Smith, C. M. (2012). Expert variability provides perspective on the strengths and weaknesses of citizen-driven intertidal monitoring program. *Ecological Applications*, *22*(4), 1201–1212.

Crall, A. W., Renz, M., Panke, B., & Newman, G. J. (2011). Is there a role for the public in monitoring invasive species? *CAB Reviews: Perspectives in Agriculture, Veterinary Science, Nutrition and Natural Resources*, *6*, 1–7. https://doi.org/10.1079/PAVSNNR20110023

Crowston, K., & Prestopnik, N. R. (2013). Motivation and data quality in a citizen science game: A design science evaluation. *Proceedings of the Annual Hawaii International Conference on System Sciences*, 450–459. https://doi.org/10.1109/HICSS.2013.413

De la Calzada, G., & Dekhtyar, A. (2010). On measuring the quality of Wikipedia articles. *Proceedings of the 4th Workshop on Information Credibility*, 11–18. ACM.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391.

Dehak, N., Dehak, R., Glass, J. R., Reynolds, D. A., & Kenny, P. (2010). Cosine similarity scoring without score normalization techniques. *Odyssey*, 15.

Dellarocas, C., Gao, G., & Narayan, R. (2010). Are consumers more likely to contribute online reviews for hit or niche products? *Journal of Management Information Systems*, *27*(2), 127–158.

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95.

Ellwood, E. R., Crimmins, T. M., & Miller-Rushing, A. J. (2017). Citizen science and conservation: Recommendations for a rapidly moving field. *Biological Conservation*, *208*, 1–4. https://doi.org/10.1016/j.biocon.2016.10.014

Escoffier, N., & McKelvey, B. (2015). The Wisdom of Crowds in the Movie Industry: Towards New Solutions to Reduce Uncertainties. *International Journal of Arts Management*, *17*(2), 52.

Frey, K., Lüthje, C., & Haag, S. (2011). Whom Should Firms Attract to Open Innovation Platforms? The Role of Knowledge Diversity and Motivation. *Long Range Planning*, *44*(5), 397–420. https://doi.org/10.1016/j.lrp.2011.09.006

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems*, *26*(3), 10–14.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57*(2), 321–326.

Gobinath, J., & Gupta, D. (2016). Online reviews: Determining the perceived quality of information. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 412–416. https://doi.org/10.1109/ICACCI.2016.7732080

Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, *68*(13), 13–18.

Goodenough, A. E., Little, N., Carpenter, W. S., & Hart, A. G. (2017). Birds of a feather flock together: Insights into starling murmuration behaviour revealed using citizen science. *PloS One*, *12*(6), e0179277.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*.

Gupta, K. M., & Montezemi, A. R. (1997). Empirical evaluation of retrieval in case-based reasoning systems using modified cosine matching function. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *27*(5), 601–612.

Gura, T. (2013). Citizen science: Amateur experts. *Nature*, *496*(7444), 259–261. https://doi.org/10.1038/nj7444-259a

Harrison, C., Jorder, M., Stern, H., Stavinsky, F., Reddy, V., Hanson, H., … Balter, S. (2014). Using online reviews by restaurant patrons to identify unreported cases of foodborne illness—New York City, 2012–2013. *MMWR. Morbidity and Mortality Weekly Report*, *63*(20), 441.

Hart, A. G., Carpenter, W. S., Hlustik-Smith, E., Reed, M., & Goodenough, A. E. (2018). Testing the potential of Twitter mining methods for data acquisition: Evaluating novel opportunities for ecological research in multiple taxa. *Methods in Ecology and Evolution*, *9*(11), 2194–2205.

Hart, A. G., Hesselberg, T., Nesbit, R., & Goodenough, A. E. (2018). The spatial distribution and environmental triggers of ant mating flights: Using citizen-science data to reveal national patterns. *Ecography*, *41*(6), 877–888.

Hart, A. G., Nesbit, R., & Goodenough, A. E. (2018). Spatiotemporal variation in house spider phenology at a national scale using citizen science. *Arachnology*, *17*(7), 331–335.

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.

Hunter, J., Alabri, A., & van Ingen, C. (2013). Assessing the quality and trustworthiness of citizen science data. *CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE*, *25*(4, SI), 454–466. https://doi.org/10.1002/cpe.2923

Jacobs, C., & Zipf, A. (2017). Completeness of citizen science biodiversity data from a volunteered geographic information perspective. *Geo-Spatial Information Science*, *20*(1), 3–13. https://doi.org/10.1080/10095020.2017.1288424

Jensen, M. L., Averbeck, J. M., Zhang, Z., & Wright, K. B. (2013). Credibility of anonymous online product reviews: A language expectancy perspective. *Journal of Management Information Systems*, *30*(1), 293–324.

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.

Korfiatis, N., GarcíA-Bariocanal, E., & SáNchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*(3), 205–217.

Lai, E. (2010, February 23). Twitter growth prompts switch from MySQL to "NoSQL" database. Retrieved April 4, 2019, from Computerworld website: https://www.computerworld.com/article/2520084/twitter-growth-prompts-switch-from-mysql-to--nosql--database.html

Leavitt, N. (2010). Will NoSQL databases live up to their promise? *Computer*, *43*(2), 12–14.

Lee, T. K., Crowston, K., Harandi, M., Østerlund, C., & Miller, G. (2018). Appealing to different motivations in a message to recruit citizen scientists: Results of a field experiment. *Journal of Science Communication*, *17*(1), A02.

Li, X., Hitt, L. M., & Zhang, Z. J. (2011). Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems*, *27*(4), 9–42.

Liang, H., Wang, M.-M., Wang, J.-J., & Xue, Y. (2018). How intrinsic motivation and extrinsic incentives affect task effort in crowdsourcing contests: A mediated moderation model. *Computers in Human Behavior*, *81*, 168–176. https://doi.org/10.1016/j.chb.2017.11.040

Lin, D. (1998). An Information-Theoretic Definition of Similarity. *In Proceedings of the 15th International Conference on Machine Learning*, 296–304. Morgan Kaufmann.

Lintean, M. C., & Rus, V. (2012). Measuring Semantic Similarity in Short Texts through Greedy Pairing and Word Semantics. *FLAIRS Conference*.

Lu, B., Li, B., Wang, L., Yang, J., Liu, J., & Wang, X. V. (2014). Reusability based on life cycle sustainability assessment: Case study on WEEE. *Procedia Cirp*, *15*, 473–478.

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.

Lukyanenko, R., Parsons, J., & Wiersma, Y. F. (2016). Emerging problems of data quality in citizen science. *Conservation Biology*, *30*(3), 447–449. https://doi.org/10.1111/cobi.12706

Lukyanenko, R., Parsons, J., Wiersma, Y. F., & Maddah, M. (2019). Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Quarterly*, *43*(2), 623–647.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. *MIT Sloan Management Review*, *51*(3), 21.

Martins, L. L., Schilpzand, M. C., Kirkman, B. L., Ivanaj, S., & Ivanaj, V. (2013). A contingency view of the effects of cognitive diversity on team performance: The moderating roles of team psychological safety and relationship conflict. *Small Group Research*, *44*(2), 96–126.

Matzler, K., Füller, J., Hutter, K., Hautz, J., & Stieger, D. (2016). Crowdsourcing strategy: How openness changes strategy work. *Problems and Perspectives in Management*, *14*(3), 450–460.

McAuley, J. J., & Leskovec, J. (2013). From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. *World Wide Web*, 897–908. Retrieved from http://dl.acm.org/citation.cfm?id=2488466

McKenzie, L. J., Long, L., Coles, R. G., & Roder, C. A. (2000). Seagrass-Watch: Community based monitoring of seagrass resources. *Biologia Marina Mediterranea*, *7*(2), 393–396.

Mello, A. L., & Delise, L. A. (2015). Cognitive diversity to team outcomes: The roles of cohesion and conflict management. *Small Group Research*, *46*(2), 204–226.

Mihalcea, R., Corley, C., & Strapparava, C. (2006). Corpus-based and knowledge-based measures of text semantic similarity. *AAAI*, *6*, 775–780.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 3111–3119.

Mittal, A., & Goel, A. (2012). Stock prediction using twitter sentiment analysis. Standford University, CS229 (2011 Http://Cs229. Stanford. Edu/Proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. Pdf), 15.

Moniruzzaman, A. B. M., & Hossain, S. A. (2013). Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. *ArXiv Preprint ArXiv:1307.0191*.

Mudambi, S. M., & Schuff, D. (2010). Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185–200. https://doi.org/10.2307/20721420

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, *21*(4), 199–235.

Nerbonne, J. F., & Nelson, K. C. (2008). Volunteer macroinvertebrate monitoring: Tensions among group goals, data quality, and outcomes. *ENVIRONMENTAL MANAGEMENT*, *42*(3), 470–479. https://doi.org/10.1007/s00267-008-9103-9

Nisar, T. M., & Yeung, M. (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. *The Journal of Finance and Data Science*, *4*(2), 101–119. https://doi.org/10.1016/j.jfds.2017.11.002

Nov, O., Arazy, O., & Anderson, D. (2011). Dusting for science: Motivation and participation of digital citizen science volunteers. *Proceedings of the 2011 IConference*, 68–74. Retrieved from http://dl.acm.org/citation.cfm?id=1940771

Nsoesie, E. O., Kluberg, S. A., & Brownstein, J. S. (2014). Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports. *Preventive Medicine*, *67*, 264–269.

Ogunseye, S., & Parsons, J. (2018). Designing for Information Quality in the Era of Repurposable Crowdsourced User-Generated Content. *International Conference on Advanced Information Systems Engineering*, 180–185. Springer.

Ogunseye, S., Parsons, J., & Lukyanenko, R. (2017). Do Crowds Go Stale? Exploring the Effects of Crowd Reuse on Data Diversity. WITS.

Oldekop, J. A., Bebbington, A. J., Berdel, F., Truelove, N. K., Wiersberg, T., & Preziosi, R. F. (2011). Testing the accuracy of non-experts in biodiversity monitoring exercises using fern species richness in the Ecuadorian Amazon. *Biodiversity and Conservation*, *20*(12), 2615–2626.

Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016). Sentiment analysis of Twitter data for predicting stock market movements. *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 1345–1350. IEEE.

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*(3), 184–188.

Parsons, J. (1996). An information model based on classification theory. *Management Science*, *42*(10), 1437–1453.

Parsons, J., Lukyanenko, R., & Wiersma, Y. (2011). Easier citizen science is better. *Nature*, *471*(7336), 37–37. https://doi.org/10.1038/471037a

Parsons, J., & Wand, Y. (2000). Emancipating instances from the tyranny of classes in information modeling. *ACM Transactions on Database Systems (TODS)*, *25*(2), 228–268.

Parsons, J., & Wand, Y. (2014). *A foundation for open information environments*. Retrieved from http://aisel.aisnet.org/ecis2014/proceedings/track17/7/

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, *70*, 153–163. https://doi.org/10.1016/j.jesp.2017.01.006

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211–218.

Pirró, G. (2009). A semantic similarity metric combining features and intrinsic information content. *Data & Knowledge Engineering*, *68*(11), 1289–1308.

Piven, M. L., Ammarell, N., Bailey, D., Corazzini, K., Colón-Emeric, C. S., Lekan-Rutledge, D., … Anderson, R. A. (2006). MDS coordinator relationships and nursing home care processes. *Western Journal of Nursing Research*, *28*(3), 294–309.

Polzer, J. T., Milton, L. P., & Swarm Jr, W. B. (2002). Capitalizing on diversity: Interpersonal congruence in small work groups. *Administrative Science Quarterly*, *47*(2), 296–324.

Ransbotham, S., Gerbert, P., Reeves, M., Kiron, D., & Spira, M. (2018). Artificial Intelligence in Business Gets Real. *MIT Sloan Management Review, September*, *17*.

Rehm, G., Schonefeld, O., Witt, A., Chiarcos, C., & Lehmberg, T. (2008). *SPLICR: A sustainability platform for linguistic corpora and resources*.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. *ArXiv Preprint Cmp-Lg/9511007*.

Rotman, D., Hammock, J., Preece, J. J., Boston, C. L., Hansen, D. L., Bowser, A., & He, Y. (2014). Does motivation in citizen science change with time and culture? *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 229–232. Retrieved from http://dl.acm.org/citation.cfm?id=2556492

Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. *International Journal of Information Management*, *37*(3), 150–154.

Saghafi, A., Wand, Y., & Parsons, J. (2016). *Querying Instances–A Protocol Analysis Study*. AMCIS, Systems Analysis and Design

Salk, C., Sturn, T., See, L., & Fritz, S. (2016). Local Knowledge and Professional Background Have a Minimal Impact on Volunteer Citizen Science Performance in a Land-Cover Classification Task. *Remote Sensing*, *8*(9), 774. https://doi.org/10.3390/rs8090774

Sauer, J., Felsing, T., Franke, H., & Rüttinger, B. (2006). Cognitive diversity and team performance in a complex multiple task environment. *Ergonomics*, *49*(10), 934–954.

Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*, (1), 93–107.

Schomberg, J. P., Haimson, O. L., Hayes, G. R., & Anton-Culver, H. (2016). Supplementing public health inspection via social media. *PloS One*, *11*(3), e0152117.

Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, *81*(1), 75–86.

Spence, M. T., & Brucks, M. (1997). The moderating effects of problem characteristics on experts' and novices' judgments. *Journal of Marketing Research*, *34*(2), 233–247.

Staffelbach, M., Sempolinski, P., Kijewski-Correa, T., Thain, D., Wei, D., Kareem, A., & Madey, G. (2015). Lessons learned from crowdsourcing complex engineering tasks. *PloS One*, *10*(9), e0134978.

Surowiecki, J. (2005). *The wisdom of crowds*. Retrieved from https://books.google.com/books?hl=en&lr=&id=_t2KDQAAQBAJ&oi=fnd&pg=PA224&dq=suriowiecki+wisdom+of+the+crowd&ots=yxDnKFz_3u&sig=Lm8zzs4JmSE2PbBwWWLG99KByvw

Tamm, T., Seddon, P., & Shanks, G. (2013). *Pathways to value from business analytics*. Thirty Fourth International Conference on Information Systems, Milan 2013

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327.

van der Velde, T., Milton, D. A., Lawson, T. J., Wilcox, C., Lansdell, M., Davis, G., … Hardesty, B. D. (2017). Comparison of marine debris data collected by researchers and citizen scientists: Is citizen science data worth the effort? *Biological Conservation*, *208*, 127–138.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–96.

Wand, Y., & Weber, R. (1990). Mario Bunge's Ontology as a formal foundation for information systems concepts. *Studies on Mario Bunge's Treatise*, *18*.

Wand, Y., & Weber, R. (1995). On the deep structure of information systems. *Information Systems Journal*, *5*(3), 203–223.

Wang, L., Gopal, R., Shankar, R., & Pancras, J. (2015). On the brink: Predicting business failure with mobile location-based checkins. *Decision Support Systems*, *76*, 3–13.

Wang, R. Y., Reddy, M. P., & Kon, H. B. (1995). Toward quality data: An attribute-based approach. *Decision Support Systems*, *13*(3), 349–372. https://doi.org/10.1016/0167-9236(93)E0050-N

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. *2011 IEEE Seventh International Conference on E-Science Workshops*, 14–19. https://doi.org/10.1109/eScienceW.2011.27

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16*(1), 85–102.

Woodall, P. (2017). The Data Repurposing Challenge: New Pressures from Data Analytics. *Journal of Data and Information Quality (JDIQ)*, *8*(3–4), 11.

Woodall, P., & Wainman, A. (2014). Re-purposing manufacturing data: A survey. *Proceedings of the International Data and Information Management Conference*.

Woodall, P., & Wainman, A. (2015). Data quality in analytics: Key problems arising from the repurposing of manufacturing data.

Wu, T., Chen, L., Hui, P., Zhang, C. J., & Li, W. (2015). Hear the Whole Story: Towards the Diversity of Opinion in Crowdsourcing Markets. *Proc. VLDB Endow.*, *8*(5), 485–496. https://doi.org/10.14778/2735479.2735482

Xu, J. D., Benbasat, I., & Cenfetelli, R. T. (2013). Integrating service quality with system and information quality: An empirical test in the e-service context. *Mis Quarterly*, *37*(3), 777–794.

Yang, R., Xue, Y., & Gomes, C. (2018). Pedagogical Value-Aligned Crowdsourcing: Inspiring the Wisdom of Crowds via Interactive Teaching.

Zipper, S. C. (2018). Agricultural research using social media data. *Agronomy Journal*, *110*(1), 349–358.

Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89.

Chapter Three: **How Knowledge Affects the Diversity of Crowdsourced Data**

Abstract

Studies of information quality in crowdsourcing explore how to ensure that contributors can provide data that is fit for the intended use. These studies, to a large extent, proceed from a tacit assumption that knowledgeable crowds are better for data quality than less knowledgeable crowds. The question they ask regarding information quality in crowdsourcing centers around how to ensure crowdsourcing projects are accessible to only knowledgeable contributors who will provide data that is fit for some intended use.

One recommended approach to ensure participants in a crowdsourcing task have requisite knowledge is to train potential contributors on the requirements of the task. However, several examples of crowdsourced data acquired from untrained crowds have provided high-quality information for decision-making that have met predetermined and unanticipated requirements. For example, Yelp's data have been repurposed to track the spread of food-borne diseases. Similarly, data from twitter were repurposed to accurately predict the yield of crops. These real-world examples raise a different question: how likely are trained contributors to report high-quality repurposable information that can meet not just the anticipated requirements but also the unanticipated requirements of crowdsourcers?

In this chapter, we simulate a citizen science crowdsourcing task using artificial stimuli to test the effect of implicit and explicit training on the diversity of contributed data. We also investigate the effect of the level of contributors' knowledge on the diversity of information they provide. Using 84 participants in a controlled laboratory experiment, we compared the results of trained and untrained contributors and found that untrained contributors reported more diverse data than trained contributors. In addition, we found that

implicitly trained contributors provided more diverse data than explicitly trained contributors. Finally, we found that knowledge is negatively associated with the reporting of diverse data.

Since information diversity is an indicator of the repurposability of data, our results suggest that recruiting primarily trained contributors may actually be hurting the acquisition of repurposable data than opening up crowdsourcing projects to everyone.

3.1 Background

Advances in information technology and the web have provided opportunities to collect and access information from spatiotemporally distributed groups of people on topics of interest to both contributors and information consumers. Crowdsourcers tap into the availability and willingness of crowds to gather information that helps in decision making. Access to such external information is revolutionizing industry and research, and has been successfully used in diverse contexts for understanding customers, developing new products, improving service quality, and supporting scientific research (Castriotta & Di Guardo, 2011; Hosseini, Phalp, Taylor, & Ali, 2014; Tarrell et al., 2013; Tripathi, Tahmasbi, Khazanchi, & Najjar, 2014).

However, organizations must consider the quality of data they can collect when leveraging undefined crowds as data sources. Unlike when organizations source data internally and can design their information management processes to generate high-quality data from known contributors, sourcing external data limits an organization’s ability to manage the data collection process and ensure the quality of crowdsourced information.

Because the level of expertise and motivation of members of the crowd that contribute data are typically unknown, crowdsourcers tend to recruit contributors who are knowledgeable in the domain of the phenomenon of interest as a mechanism for ensuring the quality of crowdsourced data (Wiggins & He, 2016; Wiggins et al., 2011; Surowiecki, 2005). When knowledgeable contributors are scarce, which is usually the case, crowdsourcers mitigate this scarcity by training potential contributors to attain desired levels of proficiency before allowing them to participate in crowdsourcing endeavours.

However, although crowdsourcers assume that a knowledgeable contributor will provide higher quality data, repurposable data must be sourced from contributors with varied views, guided by the need for the collected data to meet "multiple different fitness for use requirements" (Woodall, 2017, p 11). In other words, repurposability is achievable when the crowdsourced data is diverse, containing information about different dimensions of the observed phenomena. Diverse data is data gathered from people with different perspectives about the phenomenon of interest, but people can share the same or similar perspectives (Barsalou & Sewell, 1984) through training or shared experiences (Chen, 1990). Therefore, for crowdsourcers who seek high quality and repurposable data, pertinent questions arise about the use of knowledgeable contributors: a) How does training affect the ability to collect diverse data? b) Will seeking data diversity result in a trade-off of accuracy and completeness?

Consequently, we take a step towards better understanding the effect of contributor knowledge on the diversity and quality of collected crowdsourced data. We consider training by both explicit rules and implicit rules. Experimentally, we examine how the type of training provided affects the diversity of crowdsourced data collected from crowd members.

At the same time, we recognize that even when knowledgeable contributors are available, crowdsourcers may recruit contributors based on their level of knowledge. Crowdsourcers may screen out potential contributors who do not have a certain level of education or score a specific point on a qualifying test (see Budescu & Chen, 2014). Tacit assumptions about the benefits of expert knowledge, rather than empirical facts, inform crowdsourcer expectations around the impact of contributors' level of knowledge on the quality of crowdsourced data (Ogunseye & Parsons, 2016). However, two studies in the citizen science crowdsourcing context found that experts did not report higher quality data – as defined in the context of the studies – than novices (Austen, Bindemann, Griffiths, & Roberts, 2016; Lukyanenko, Parsons, & Wiersma, 2014).

In addition to investigating how training contributors to become knowledgeable in a crowdsourcing task affects information quality, there is also a need for empirically validated theoretical insights into how contributors' levels of knowledge affect the quality of information contributed, including the diversity of information they contribute. Our findings will be of benefit to crowdsourcers and developers of crowdsourcing systems who are interested in the repurposability of collected data, and those who make recruitment decisions intended to ensure the collection of high quality, diverse information.

3.2 Theoretical Foundation and Development of Hypotheses

Humans acquire knowledge through an assimilation-accommodation cycle (Piaget & Inhelder, 1969). When we encounter a new instance of a phenomenon, we examine the instance, comparing it to previously encountered instances from memory. If we determine that the new instance is sufficiently similar to previous instances (i.e., it is a member of an existing class we have), we assimilate it, ascribing to the new instance our expectations from previous encounters with similar instances. On the contrary, if we find that the new instance is dissimilar to all other instances of phenomena we have previously encountered, we accommodate the new instance by creating a new schema in memory to store the attributes of the novel instance. In other words, we create a new class to store instances of novel phenomena.

Classification (or categorization) is the process of assimilating and accommodating instances into classes. According to classification theory, when humans seek to identify an instance of a phenomenon of interest (entity), they consider its attributes and compare the observed attributes with the attributes they already know (Goldstone & Kersten, 2003; Harnad, 2005; Piaget & Inhelder, 1969; Rosch, 1973). The way we identify instances of phenomena is dependent on our knowledge of the phenomena. When there is existing knowledge, humans compare specific attributes of an observed phenomenon with their learned attributes from previous exposures to the phenomenon to draw inferences or classify it (Piaget & Inhelder, 1969). In contrast, when we do not have any prior knowledge of a phenomenon or if we do not have any relevant attributes to which to compare the entity, we tend to pay attention to those attributes of the phenomenon that stand out[4] (Katsuki & Constantinidis, 2014; Wolfe, 1994). Research findings on how infants, young children, and adults classify entities provide further evidence for how we allocate attention in the presence and absence of relevant knowledge. Infants (six to eight months), and young children who lack prior knowledge, tend to pay attention to more of an instance's attributes than adults, who pay attention to a few specific attributes because of their familiarity with the instances (Best, Yim, & Sloutsky, 2013; Gelman & Markman, 1986; Kloos & Sloutsky, 2008). This has also been indicated in adults who visit places for the first time and try to absorb as much of the new environment as they can (Gopnik, 2009).

Classification is also how we manage our limited cognitive resources (Goldstone & Kersten, 2003). The amount of sensory information that exists in typical human environments is significantly higher than what humans can process. Because of our limited cognitive resources, we naturally pay selective attention to particular entities and critical attributes of those entities that help in classifying them (Bjorklund and Harnishfeger, 1990). As we attend to stimuli (or a few attributes of a stimulus) for classification purposes, we ignore or suppress other stimuli we do not use (Prat-Ortega & de la Rocha, 2018). This phenomenon is called selective attention – "the differential processing of simultaneous sources of information" (Johnston & Dark, 1986, p. 44). There are two broad paradigms of selective attention: early selection and late selection (Awh, Vogel, & Oh, 2006; Huang-Pollock, Carr, & Nigg, 2002; Johnston & Dark, 1986). Early selection theories argue that

[4] "attributes" used here can be replaced by stimulus or location (Katsuki & Constantinidis, 2014)

sensory information about a stimulus are held in a register, where they undergo pre-attentive analysis based on any existing knowledge the contributor has about the stimulus. Following this analysis, selected information passes a fixed cognitive channel into consciousness, where semantic analysis takes place, while information that is not selected is filtered out (Broadbent, 1958; Lachter, Forster, & Ruthruff, 2004).

In contrast, late selection theories argue that selectively attending to aspects of an information source takes place at a later stage of information processing. Attention allocation takes place after a message has been semantically analyzed, and during the response preparation stage (Deutsch & Deutsch, 1963). Proponents of the late-stage selection paradigm argue that we choose aspects of information about a stimulus we attend to based on our existing knowledge of the stimulus (Awh et al., 2006).

Both the early and late-stage theoretical perspectives agree that the existence of prior knowledge shapes attention allocation. When attributes that have been committed to memory (i.e., have become a part of our knowledge-base) are used to guide attention, then attention is directed from the top-down or is knowledge-driven, i.e., the "internal guidance of attention based on prior knowledge…" (Katsuki & Constantinidis, 2014, p 509). Similarly, Buschman and Miller (2007, p. 1860) described top-down attention allocation as depending on "volitional shifts of attention," which are "derived from knowledge about the current task (e.g., finding your lost keys)."

On the other hand, if we have not previously committed attributes about an entity to memory, or we have a first-time encounter with an entity, the attributes of the entity solely direct our attention, and thus, our attentional allocation is bottom-up or stimulus-driven. In bottom-up attentional allocation, “target stimuli ‘pop out’ if they differ sufficiently from their background in terms of features such as colour or orientation” (Katsuki & Constantinidis, 2014, p 509). Bottom-up attention is “automatic” and driven by “properties inherent in stimuli … (e.g., a flashing fire alarm)” i.e., the salience of an entity’s attributes can direct our attention (Buschman & Miller, 2007).

Specific factors that affect bottom-up and top-down attention allocation identified in the literature (Wickens & McCarley, 2008) are:

1. *Salience:* Stimuli or attributes of stimuli that are prominent in a contributor’s visual space can capture the contributor’s attention and are thus said to be salient. Attributes of stimuli, such as their color, size, and shape, affect their capacity to attract an observer’s attention (Theeuwes, 2010) and are the default attention capture mechanism when the contributor has no prior knowledge or insufficient prior knowledge guiding their attention allocation (Buschman & Miller, 2007; Katsuki & Constantinidis, 2014).

2. *Effort:* Some properties of the visual field and stimuli determine how easy it is for stimuli to capture attention. These properties include the organization of stimuli in the visual field and the need to either assimilate or accommodate an observed entity. Accommodating an entity may be easy, like in the case where an observer has only seen the entity for the first time. It may also be taxing, requiring the scrutinizing of an entity’s attributes to identify how they differ from known attributes. A lesser

effort is required to focus on the similarities between the attributes of a stimulus and those of previously observed stimuli stored in memory than is needed to focus on differences. Also, less effort is required to focus on salient attributes, or just familiar attributes, than to search for attributes that may lead to accommodation (Wickens & McCarley, 2008).

3. *Expectancy:* If contributors have been cued either through experience or explicit instructions to expect specific attributes to be present in an observed entity, this expectation will inform their attentional allocation. Expectancy is, therefore, the provision of guiding information to contributors, which may be a description of the expected entity or the context of an identification task that influences the attentional distribution of contributors.

4. *Value:* This is the utility that can be derived or lost from knowing the attributes that are necessary to identify a stimulus. Contributors ascribe value to diagnostic attributes (i.e., attributes that help classify an entity, and efficiently perform an identification task).

Expectancy and value form the top-down mental factors that drive attention allocation, whereas effort and salience are bottom-up attentional allocation influencers. Bottom-up attention allocation is stimulus-driven, while top-down attention allocation is knowledge-driven.

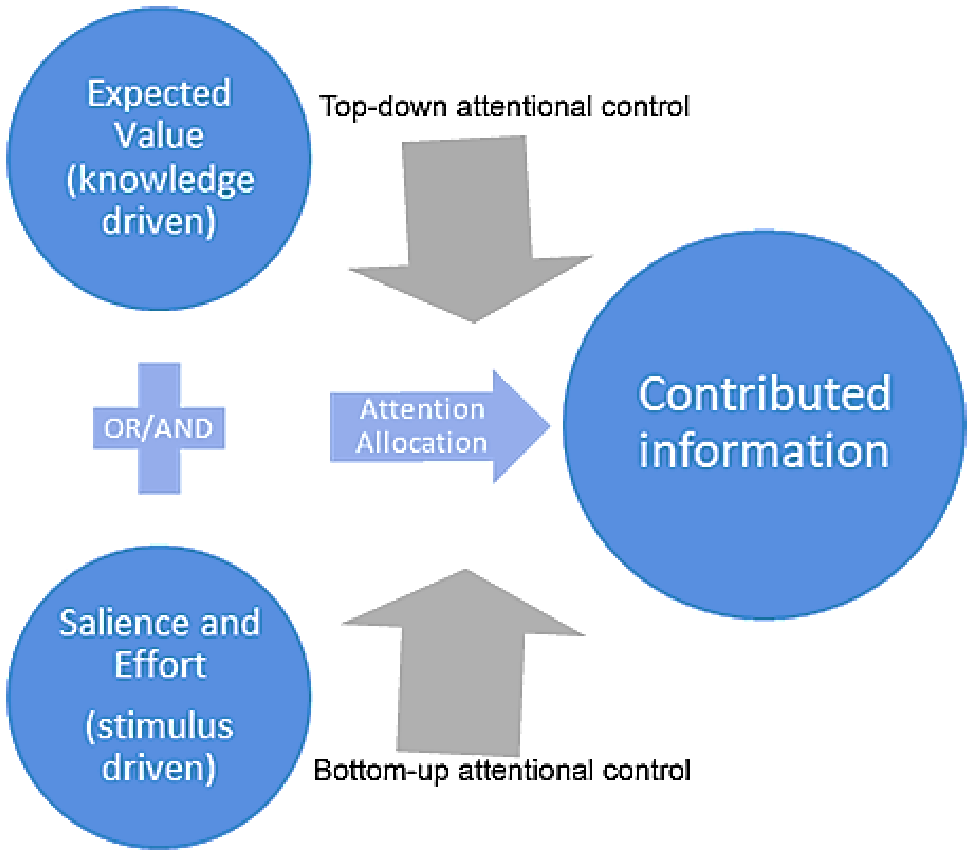


Figure 3.1: Factors affecting contributed information

Figure 3.1 illustrates that attentional allocation can be bottom-up, top-down, or a combination of the two and determines the type of information that crowds contribute. Factors responsible for top-down attention allocation would manifest in contributors who have been trained explicitly on the attributes needed to classify an entity. On the one hand, explicit training – teaching contributors attributes and rules about attributes (i.e., inclusion rules) that can be used to identify an entity – would cause trained contributors to look out for familiar attributes when they observe an entity for the purpose or classifying it. On the other hand, factors responsible for bottom-up attention distribution such as salience and effort could lead to the automatic formation of inclusion rules by contributors who are exposed to multiple instances of a stimulus. Such exposure to instances of a stimulus would allow contributors to derive inclusion rules implicitly. Implicit training thus implies teaching contributors inclusion rules by continued exposure to instances of a stimulus and allowing them to learn through inferencing leading to an autonomously determined inclusion rule[5].

When crowdsourcers who are interested in ensuring the collection of quality data train contributors implicitly, the contributors are tasked with learning unsupervised and inferring their own classification rules, whereas, the learning of explicitly trained contributors is rule-based. Nevertheless, whether explicit or implicit, training helps crowdsourcers transfer inclusion rules to contributors because humans can learn the

[5] Implicit training is similar to unsupervised learning (Kloos & Sloutsky, 2008) and inference learning (Hoffman & Rehder, 2010)

attributes other people consider pertinent and perform classification tasks using these learned attributes (Chen, 1990). Barsalou and Sewell (1984) show this ability for humans to learn and use the schema of others in experiments that revealed that humans with diverse social or cognitive backgrounds can adopt points of view other than their own and accurately provide data according to the owners of the adopted point of view. For instance, students were able to accurately answer questions about professors like the type of alcoholic beverages that professors drink, their athletic activities, birthday presents, cars, famous people admired, important goals in life, and people to get advice from. Students learned the schema of professors by observing their professors; contributors can learn from crowdsourcers and adopt their views through either explicit or training. GalaxyZoo exemplifies the use of training to share crowdsourcers' inclusion rules with contributors. The GalaxyZoo project trains potential contributors on how to identify stars and galaxies and tests them before participation.

Nonetheless, since explicit and implicit training requires attention to be allocated in different ways, the resulting contributed information will differ between contributors who have been exposed to these two training approaches. Training contributors to perform crowdsourcing tasks may, therefore, have unintended consequences for the type of data they contribute. Unlike trained contributors, the salience of an observed entity's attributes directs the attention of untrained contributors. The information provided by untrained contributors who are unconstrained by inclusion rules would, therefore, differ from the information provided by explicitly and implicitly trained contributors.

When explicitly trained, contributors in a crowd are expected to focus on (i.e., value) the same set of diagnostic attributes. We expect trained contributors will direct attention based on their knowledge of diagnostic attributes rather than attribute salience. On the other hand, we expect the salience of observed attributes guides the allocation of attention for implicitly trained contributors. Because the amount of available cognitive resources of contributors differ, the number of attributes implicitly trained contributors consider salient, and the amount of effort they put into searching out and observing attributes will differ. Similarly, untrained contributors will also commit different amounts of cognitive resources to searching-out and processing attributes. We, therefore, seek to understand how training or the lack thereof, affects the quality of information reported in crowdsourcing tasks.

Using the theory of selective attention, we focus on three themes of hypotheses: (a) we hypothesize about how training will affect information diversity. To more fully explore the predicted effect of knowledge on information diversity, we develop hypotheses about key attribute types that indicate diversity in contributor perspective. These information diversity components include the number of mutual and behaviour attributes reported and the number of attributes reported about the secondary entities present in a contributor's visual space. (b) we hypothesize about how training will affect the reporting of variability in observed entities. And finally (c), we hypothesize about the effect of training on information quality dimensions, including information diversity. We also explore how these dimensions relate to one another.

3.2.1 *The Effect of Training on Information Diversity*

How contributors acquire the knowledge needed to classify entities they observe will affect the diversity of data they report about the entity. Explicit training equips contributors to apply top-down attentional control. Explicit training leads contributors to expect specific attributes of a stimulus to be present when they observe the stimulus and helps them value or prioritize these expected attributes. Being exposed to explicit classification rules will lead contributors to focus on diagnostic attributes (i.e., attributes that help in classifying the entity) (Hoffman & Rehder, 2010).

In contrast, implicit training leads contributors to attend to as many salient attributes as possible and may lead to more attributes being used in inclusion rules than would be used by explicitly trained contributors. When crowd members are required to report diagnostic attributes, we expect implicitly trained contributors to report salient attributes in their self-determined inclusion rules. At the same time, contributors who have learned the same explicit rules would focus mainly on these rules and therefore report similar diagnostic attributes. Nonetheless, the attributes of an entity considered salient by different contributors should be highly similar because salience of attributes is inherent in the entity (Buschman & Miller, 2007; Katsuki & Constantinidis, 2014). In cases where inclusion rules consist of salient attributes of an entity, we predict that even though implicitly trained contributors formulate inclusion rules themselves, they would report a similar number of diagnostic attributes, i.e., attributes that constitute inclusion rules, about an observed entity as explicitly trained.

In the same vein, untrained contributors are expected to apply a bottom-up, stimuli-driven approach to attention allocation without a goal in mind or any prior knowledge of the diagnostic attributes. Contributors who have not received any cues about the task or explicit rules are therefore expected to be more likely to focus on the most salient attributes of the primary stimulus and other stimuli in their field of vision (Itti & Koch, 2000; Niebur & Koch, 1996; Wolfe, 1994). While not all diagnostic attributes may be salient, when most diagnostic attributes are salient, untrained contributors are therefore expected to also report similar numbers of diagnostic attributes as the implicitly and explicitly trained contributors. There will be no significant difference in the number of diagnostic attributes reported by untrained, implicitly trained, and explicitly trained contributors.

> ***H1a:*** *Explicitly trained contributors will report a similar number of diagnostic attributes of a target entity as implicitly trained contributors and untrained contributors*

Contributors who have been trained to perform a specific task have a greater tendency than untrained contributors to attend selectively to attributes that fit their training and ignore other aspects of the phenomenon under consideration (Hoffman & Rehder, 2010). Knowledge of the diagnostic attributes of an entity helps to reduce the cognitive resources expended on identification tasks. Hence, it is more cognitively economical for an explicitly trained contributor to focus on these attributes when observing an entity and ignore other non-diagnostic attributes. Implicitly trained contributors would also be expected to decipher which attributes of the target entity are diagnostic and which are not by identifying and learning which attributes repeatedly occur in all the instances of a stimulus to which they are exposed. This is possible because people can learn to classify entities unsupervised by studying the statistical frequency of entity attributes from repeated exposure to stimuli (Barlow, 1989; Kloos & Sloutsky, 2008). This process of identifying diagnostic attributes for implicitly trained contributors would entail first paying attention to the salient attributes of the entity and then revising the list of relevant attributes with each exposure to the stimulus until they are confident about the valuable attributes and those that are irrelevant to a task. The extent to which implicitly trained contributors have learned diagnostic attributes will be evident in the accuracy of the information they provide. Notwithstanding this, implicitly trained contributors, who use a bottom-up approach to arrive at their top-down knowledge, will attend to more attributes than explicitly trained contributors. Implicitly trained contributors will, therefore, be more conversant with the non-diagnostic attributes of a primary entity than explicitly trained contributors.

Conversely, when not implicitly or explicitly trained, contributors will not selectively attend to specific attributes but will instead observe salient attributes. As explicitly and implicitly trained crowd members use the knowledge from their training, we posit that they will ignore attributes of the stimulus that are outside the scope of their inclusion rules. However, implicitly trained contributors have a bottom-up approach to attentional allocation and have been cued on the objectives of the task. They will, therefore, pay attention to more of a primary entity's attributes, whether diagnostic or not. Implicitly trained contributors will, therefore, use a top-down approach but with a broader set of attributes, including non-diagnostic attributes, at their disposal for deciding class membership. Untrained contributors will also use a bottom-up stimulus-driven approach to attention allocation.

H1b: *Untrained contributors will report fewer non-diagnostic attributes of a target entity than implicitly trained contributors but more non-diagnostic attributes than explicitly trained contributors*

Trained contributors who are sensitized to the goals of a crowdsourced task and the attributes of the entity that are relevant to successfully performing a classification task, will commit their cognitive resources to determining whether a target stimulus possesses these attributes of the target stimulus and ignore other stimuli present in their visual field. Their attention is, therefore, goal-directed, aimed at expected and valuable diagnostic attributes. Implicitly trained contributors will differ from explicitly trained contributors in their ability to pay attention to the attributes of other stimuli in their visual field. Implicitly trained contributors will be more inclined to distribute their attention among multiple stimuli and attributes than explicitly trained contributors.

However, untrained contributors will have a greater tendency to pay attention to other stimuli when they are present in the contributor's field of vision because they are less task-directed and are more salience-driven than implicitly trained contributors. We expect that if any other stimuli in the contributor's visual field are salient, then the untrained contributor who has not been primed to focus on any stimulus would report more of these stimuli' attributes than other groups. Untrained contributors are not sensitized to attributes needed to perform a classification task, what the task is about, or what expected, and acceptable responses are. They are more likely to pay attention to salient attributes and will report information about these. Untrained contributors are therefore expected to provide more diverse data about all entities in a visual field than trained contributors.

> ***H1c:*** *Untrained contributors will report more data about secondary stimuli attributes than implicitly trained contributors who will, in turn, report more of these attributes than explicitly trained contributors*

Altogether, trained contributors will know more about the diagnostic attribute of the primary entity and will accurately report more of these attributes than untrained contributors. Whereas, untrained contributors will report more attributes in general about every entity in their visual field. Untrained contributors would, therefore, show less selective attention, reporting more diverse data in general than trained contributors. Again, because of their lower selective attention, implicitly trained contributors will report more diverse data than explicitly trained contributors who will show more selective attention to mainly the diagnostic attributes of a target stimulus than other attributes of stimuli present in their visual field.

Based on these arguments, we hypothesize:

> ***H1d:*** *Untrained contributors will report more diverse data than implicitly trained contributors who in turn will report more diverse data than explicitly trained contributors*

3.2.2 The Effect of Training on the Reporting of Variability in Instances of Stimulus

Attributes of instances of a phenomenon can vary from one instance of the phenomenon. This variability may be because of differences in the number of attributes present from one instance to another or differences in particular attributes from one instance of an entity to another.

Training can sensitize or desensitize contributors to variability in attributes. In addition, training can make contributors selectively attend to a specific set of attributes improving their capacity to notice and report variability in instances of an entity based on those attributes. On the other hand, inattentional blindness or change blindness can limit the reporting of variability in the observed instances of an entity in trained contributors. Inattentional blindness occurs when a contributor fails to see some visible attributes of an entity in their visual field because they are selectively attending to other attributes of the entity (Simons, 2000). Trained contributors who attend to only attributes they were exposed to during their training may not report information about the presence or absence of other attributes (or stimuli) not covered in their training. Similarly, change blindness is said to occur when participants do not notice changes to attributes or stimuli because they are attending to other attributes or stimuli (Rensink, O'Regan, & Clark, 1997).

The difference between change blindness and inattentional blindness is that for inattentional blindness, the contributor fails to attend to an attribute so they cannot notice its absence or presence in subsequent instances of the stimulus (Mack, 2003). For change blindness, the contributor may have attended to the attribute but have not permanently committed the information to memory, and so would notice if the attribute is missing but would not notice if the attribute has been modified. In other words, "[c]hanges to attended objects frequently go unnoticed (Wheeler & Treisman, 2002; Williams & Simons, 2000) particularly when the changes are unexpected" (Simons & Rensink, 2005, p17). This is because even though contributors may attend to an object, only the attributes of that object needed for their tasks are committed to consciousness (Simons & Rensink, 2005; Triesch, Ballard, Hayhoe, & Sullivan, 2003).

Trained contributors focus on diagnostic attributes. Therefore, there is a propensity to report variability that involves these diagnostic attributes. This effect is stronger for explicitly trained contributors than implicitly trained contributors. Explicitly trained contributors will not suffer inattentional or change blindness when the attributes involved are diagnostic. On the other hand, explicitly trained contributors will be prone to these types of blindness when the attributes involved are non-diagnostic or concern other stimuli in their visual field because they will mainly selectively attend to diagnostic attributes of the target stimulus. However, unlike explicitly trained contributors, implicitly trained contributors will report more variability caused by non-diagnostic attributes. They will also notice more variability affecting diagnostic attributes than untrained contributors. This is because implicitly trained contributors will distribute their attention between diagnostic and non-diagnostic attributes of the target stimulus as they learn diagnostic attributes from ground-up.

> ***H2a:*** *Explicitly trained contributors will report more variability involving the diagnostic attributes of a target stimulus than will implicitly trained contributors and untrained contributors.*
>
> ***H2b:*** *Implicitly trained contributors will report more variability involving the non-diagnostic attributes of a target stimulus than will explicitly trained contributors and untrained contributors.*

Untrained contributors will be most susceptible to change and inattentional blindness as they have not learned which attributes are pertinent to the classification task or which attributes to expect. They may attend to different attributes at the same time because they are not selectively attending to any attributes, so do not commit any attributes to memory. It is, therefore, highly probable that the differences in the states of attributes from one instance of a stimulus to another will go unreported by untrained contributors.

3.2.3 The Effect of Training on Information Quality Dimensions

Crowdsourcers train contributors to ensure they provide accurate, complete, or both accurate and complete data (Wiggins et al., 2011). Accuracy and completeness are the two crucial dimensions of information quality most pertinent to information consumers (Wang & Strong, 1995). While training may help contributors acquire relevant knowledge for the crowdsourcing task, crowdsourcers also view the levels of knowledge possessed by contributors as relevant to information quality (Gura, 2013; Wiggins, Newman, Stevenson, & Crowston, 2011). Because information diversity can be a desirable outcome of an information crowdsourcing process, there is a need to understand how information diversity relates to accuracy and completeness in different training conditions.

Accuracy is an operation on the number of attributes correctly analyzed by contributors, that is, contributors perceive the attributes of an entity and analyze those attributes matching it to diagnostic attributes in their memory to correctly classify the entity. When this observation of attributes, analyses, and pattern matching (i.e., operation on attributes) is successful, the contributor will be accurate. Otherwise, the contributor will report inaccurate data. Accuracy is, therefore, evidenced by the number of correct identifications made about an entity (Wand & Wang 1996). Explicitly trained contributors will provide more accurate classifications than implicitly trained contributors. We do not expect untrained contributors to be able to classify entities as they have no knowledge to guide such a classification. Nonetheless, we already know from prior research that people can accurately report the attributes of entities, whether they are trained or not (Lukyanenko et al., 2014).

Explicitly trained contributors will do better than implicitly trained contributors in the classification task because of their possession of specific rules to guide their inclusion and exclusion of a stimulus from the target category. Implicitly trained contributors arrive at an inclusion rule guided by salience and impeded by the amount of cognitive effort required to elicit the right sets of inclusion rules. They may or may not elicit the correct rule or the complete set of diagnostic attributes needed to classify a stimulus and will, therefore, be less accurate than explicitly trained contributors.

> ***H3a:*** *Explicitly trained contributors would report more accurate data than implicitly trained contributors*

Completeness has been defined in the literature as the presence of information about an entity that is sufficient for a particular use (Nelson et al., 2005). Completeness includes the breadth and depth of information (or attributes) reported about an entity (Wang & Strong, 1996). Breadth refers to the number of unique attributes reported about a stimulus, while the depth refers to the amount of information provided about each attribute. However, completeness is contextual, depending on the crowdsourcing task. Information that is complete in the context of one task may not be complete for another task (Wang & Strong, 1996).

We predict that the completeness of attributes reported in crowdsourced information will be affected by top-down attentional allocation such that explicitly trained contributors will focus on the diagnostic attributes to which they have been introduced and ignore attributes that are not diagnostic, providing incomplete information about the observed stimuli. Implicitly trained contributors will have used a bottom-up approach to learn attributes during training; thus, they would have attended to non-diagnostic attributes as well as diagnostic attributes of the primary entity. Therefore, even though their attention will be allocated top-down during a classification task, the attributes they have attended to and committed to memory would include some of the non-diagnostic attributes they have previously been exposed to. Implicitly trained contributors will, therefore, report more complete data about the target entity than explicitly trained contributors.

At the same time, untrained contributors have not had the opportunity to learn about the task or which entity is the primary entity and will distribute their attention broadly across all salient entities in the visual field, including the salient attributes of secondary entities. Why we expect untrained contributors to report more attributes in general than implicitly or explicitly trained contributors, the number of attributes they report per attribute type (e.g., diagnostic and non-diagnostic attributes) for the target entity will be lower than some of the trained contributors as they trade-off focusing on the target entity alone for focusing on all the entities in their visual field. They will, therefore, report fewer attributes of the target entity, whether diagnostic or non-diagnostic, than contributors who have learned to selectively attend to a particular attribute type. Consequently, we predict the following:

> ***H3b:*** *Implicitly trained contributors will report more complete data about a stimulus than untrained and explicitly trained contributors*

Beyond the differences in accuracy between explicitly trained and implicitly trained contributors, accuracy can also be used to estimate contributors' level of knowledge of a crowdsourcing task. That is, accuracy can be used to assess a contributor's expertise, competence, and familiarity with a task, all of which have been used to operationalize "level of knowledge" in the literature (Schultze & Leidner, 2002; Stein, 1992). Performance-based assessment of level of knowledge such as the number of accurate classification reported is a more reliable measure of a contributor's knowledge than subjective methods such as self-reporting (Clemen, 2008; Davis-Stober, Budescu, Dana, & Broomell, 2014; Lin & Cheng, 2009, Bouillard, White, Jackson, Austen, & Schroeder, 2019),

The ability to learn rules that help classify by selectively attending to attributes of a stimulus comes 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 individuals who have low levels of knowledge about a task. Like infants, non-experts also lack *a priori* class-forming rules. Infants can, therefore, help us understand how non-experts and "expert amateurs" – people with incomplete knowledge – perceive instances (Keil, 2011; Kloos & Sloutsky, 2008). Gopnik explained (in an interview available at bigthink.com) from her research findings that adults can "functionally … 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 posit that a non-expert contributors' exposure to an instance in a citizen science project is also an example of a situation that activates the default bottom-up attentional allocation. Conversely, the tendency of adults to employ rule-based classification can help us understand knowledgeable contributors and expect them to selectively attend to a target stimulus and report only aspects of the stimulus that is related to their existing knowledge. Rule-based classification allows knowledgeable contributors to focus on relevant features for identifying instances of classes, producing cognitive economy (efficiency of classification). Thus, they are less likely to attend to non-diagnostic attributes than will novices and will report less non-diagnostic information and less secondary entity information. The more knowledgeable a contributor is about a crowdsourcing task, evidenced by their level of accuracy, the lesser the diversity of the data they will report.

Therefore, when crowdsourcers are interested in traditional information quality, data contributors may tradeoff information diversity for accuracy as they selectively attend to only attributes in their inclusion rules at the expense of all other non-diagnostic attributes and attributes describing the behaviour of the entity. On the other hand, the more complete the data a contributor reports are, the higher their tendency to report diverse data. Contributors who report complete data can distribute their attention to non-diagnostic attributes and secondary entity attributes. Thus, contributors who report complete data will most likely be contributors who are not strongly affected by selective attention and do not only focus on diagnostic attributes. When contributors have been trained, contributors who do not apply selective attention would be mostly less knowledgeable contributors.

> ***H3c:*** *Information diversity will be negatively associated with accuracy while being positively related to completeness across the implicitly trained and explicitly trained groups*

3.3 Study Design

We designed an experiment in the context of citizen science using artificial stimuli. Citizen science is a "partnership between volunteers and scientists to address research questions" (Crall et al., 2011, p. 433) usually culminating in citizens assisting with data collection and/or analysis, defining the research question, or even designing a study while gaining scientific knowledge through their involvement in the research. Wiggins et al. (2011, p. 17) argue that "most [citizen science] projects show greater concern over the lack of contributor expertise than the lack of analysis methods suited to the type of data generated in citizen science." Many citizen science projects, therefore, seek knowledgeable contributors and can train contributors to acquire the desired level of knowledge as a means of ensuring data quality.

The target and distractor artificial stimuli used in this study are called tyrans and non-tyrans, respectively. These stimuli were designed following Kloos and Sloutsky's

(2008) artificial stimuli. Tyrans are a class (species) of artificial insects whose members meet an inclusion rule (a set of attributes and values of these attributes). Stimuli that do not meet this rule are non-tyrans. The inclusion rule is that *tyrans have a short tail, two or three buttons on their light blue bodies, blue wings, and either one or two rings on each blue wing.* Non-tyrans may look like tyrans but will fail at least one of these requirements. Each image was presented to participants in Powerpoint Slides. Figure 3.2 shows a sample tyran and a sample non-tyran used in the experiment.

We tested the materials with 12 students from the Department of Biology who are familiar with observing, classifying, and reporting organisms. We tested for the suitability of the prompt to elicit unbiased responses from contributors. We found that asking contributors a non-leading question like "what do you see?" was less biasing than asking contributors to identify the entity they have observed. So we used the prompt "What do you see?" in this study. We also tested for the complexity of the task and the ease of learning the inclusion rule. We carried out another pretest to examine the effect of changes made based on our initial pretest. The participants in the second pretest were fifteen business students who participated for course credit. All participants recorded their sightings on an answer sheet. Based on our findings from the pretests described in Appendix A, we set the display time for each image presented to participants to forty seconds. We also modified the inclusion rule to consist of five of the seven attributes of the target entity. The complete experimental material is available in Appendix B.

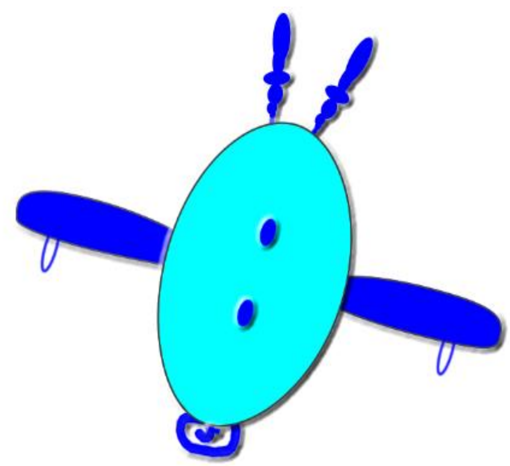
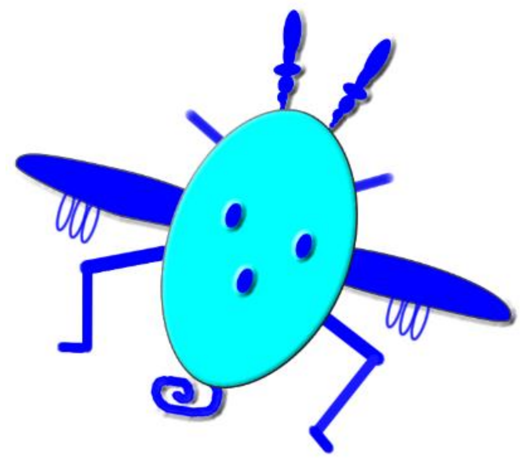
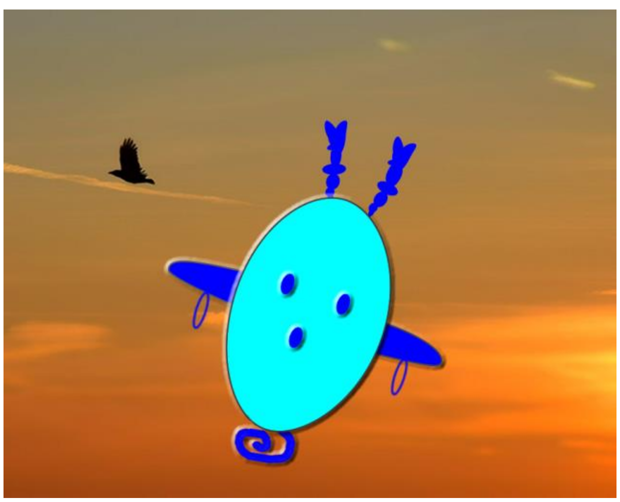
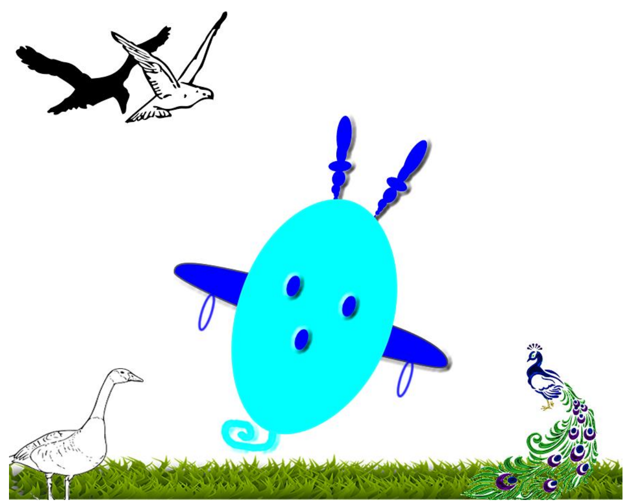
	
Tyran. Follows the inclusion rule: Two blue wings, short tail, light blue body, two or three buttons on the light blue body, and one or two rings on each blue wing	Non-tyran because it has three rings on each wing. The number of legs is not diagnostic

Figure 3.2: Sample Tyran and Non-Tyran Images

Several variations of tyrans and non-tyrans were created to test each of the three hypotheses specifically. We presented sixteen images (a mixture of tyrans and non-tyrans) to participants. All sixteen images test the capacity of contributors to report accurate and diverse information. However, six images were selected to be examined for variations of diagnostic attributes and non-diagnostic attributes – three for each attribute type. For example, the antennae, even though non-diagnostic, are shorter in some of the images of tyrans presented than the ones presented in the training/orientation phase of the experiment. The presence of patterns on the wings of some of the tyrans, the number and shape of antennae, and the number of legs on the insect are additional manipulations present in the images.



Tyran (two-lobed antennae and shorter wings)
Here a change to a non-diagnostic attribute has occurred



Tyran (shorter wings and different coloured tail)
Here a change to a diagnostic attribute has occurred

Figure 3.3: Variations in Diagnostic and Non-Diagnostic Attributes

Four slides containing catch items were placed intermittently among the test item slides (tyran and non-tyran insects) to check if participants paid attention and were alert to the experiment. The catch items were different shapes/coloured stimuli that are not insects, and the participant was expected to correctly report them as non-tyrans or provide specific descriptions of their attributes. The image slides were presented in a non-randomized order to all groups.

3.3.1 Participants

After approval from the University's Ethics Review Board, 93 participants recruited for the study were assigned randomly to 3 groups: untrained, implicitly trained, and explicitly trained groups. Upon preliminary examination, one report was excluded from the implicitly trained group for incompleteness and another report for inaccurate reporting of catch item. Two other reports were excluded from the untrained group for inaccurate reporting of catch items and one for illegibility. To make the number of reports equal across the groups, we excluded the last report from the implicitly trained group and the last three reports from the explicitly trained group, leaving a total of 84 participants across the three groups whose reports were used for our analysis.

Consequently, each group had 28 participants who were all students of Memorial University of Newfoundland. Fourteen of the students participated for donations to their class graduation. In addition to these fourteen students, ten students participated solely for the chance to win one of two $100 gift cards. Sixty students participated for course credit. Thirty-six of the participants were male, and forty-eight were female.

Participants in the explicitly trained group were provided with an inclusion rule with which to classify stimuli as either tyrans or non-tyrans. They went through a training phase in which they were taught the rule and shown five sample tyrans to allow them to become familiar with applying the criteria in the classification task. Participants were also tested on their knowledge and received feedback on their ability to identify tyrans. This was achieved by presenting them with images and verbally inquiring if they thought it was a tyran or not, and why. After they provided their answers, we showed them the correct response and explained how they satisfied the inclusion rule.

The implicitly trained group was briefed on the task to be performed and shown the same five target stimuli used to teach the explicitly trained group, to allow them to infer the inclusion criteria. However, we did not provide explicit rules to this group, nor did we give them feedback on their ability to determine if a stimulus is a tyran or not. Also, we did not show the Untrained Group any sample images. However, like the other groups, they were informed that we were interested in examining how people report things. More information about the experimental procedure is presented in Appendix B.

3.4 Results

Two members of the Thesis Supervisory Committee and I developed the coding scheme that accounts for attributes of the target entity and attributes of other stimuli reported by the contributor. The objective of the coding scheme is to help measure contributors' degree of selective attention due to their treatment by identifying which entity attributes they report and which they ignore. The attributes we coded for are presented in Table 3.1. We counted the number of attributes reported about the stimuli in the presented images.

We used a one-way analysis of variance (ANOVA) and Tukey's HSD[6] test for post-hoc comparison of the group averages (excluding the catch item images used for screening purposes only) to compare the variables described in Table 3.1 below, across the groups.

[6] Tukey's Honestly Significant Difference (Tukey's HSD) corrects for multiple comparisons (Homack, 2001)

Table 3.1: Variables coded for in contributed data

Codes	Description
Accuracy	Accuracy of primary entity (tyran or non-tyran)
Diagnostic_Attr	Number of primary entity diagnostic attributes mentioned
Diagnostic_Values	The number of values reported for each diagnostic attribute of the primary entity. i.e., the amount of information reported for each attribute. E.g., values for the diagnostic attribute blue wings may be "short," or "curvy."
Non-diagnostic_Attr	Number of primary entity non-diagnostic attributes mentioned
Non-diagnostic_Values	Number of attribute values for non-diagnostic attributes (e.g., the colour of the tail, where the presence of a tail is a diagnostic attribute, and length of the tail is a diagnostic value, but the colour of the tail is a non-diagnostic attribute value even though the tail is diagnostic)
Behavior_Attr	Entity behaviour: descriptions provided for the behaviour or perceived activity of the entity
Mutual_Attr	Entity mutual attribute: descriptions provided for the relation of the primary entity in terms of other entities, including its environment
Secondary_Ent	Number of secondary entities provided
Secondary_Ent_Attr	Secondary entity attribute (attributes of secondary entities)
Secondary_Ent_Value	Secondary entity attribute value
Secondary_Ent_Mutual	Secondary entity mutual attribute: description of the relationship between secondary entities
Secondary_Ent_Behavior	Number of descriptors of secondary entity behaviour reported
Diagnostic_Attr_Variance	Variability in diagnostic attributes reported
Non-Diagnostic_Attr_Variance	Variability in non-diagnostic attributes reported

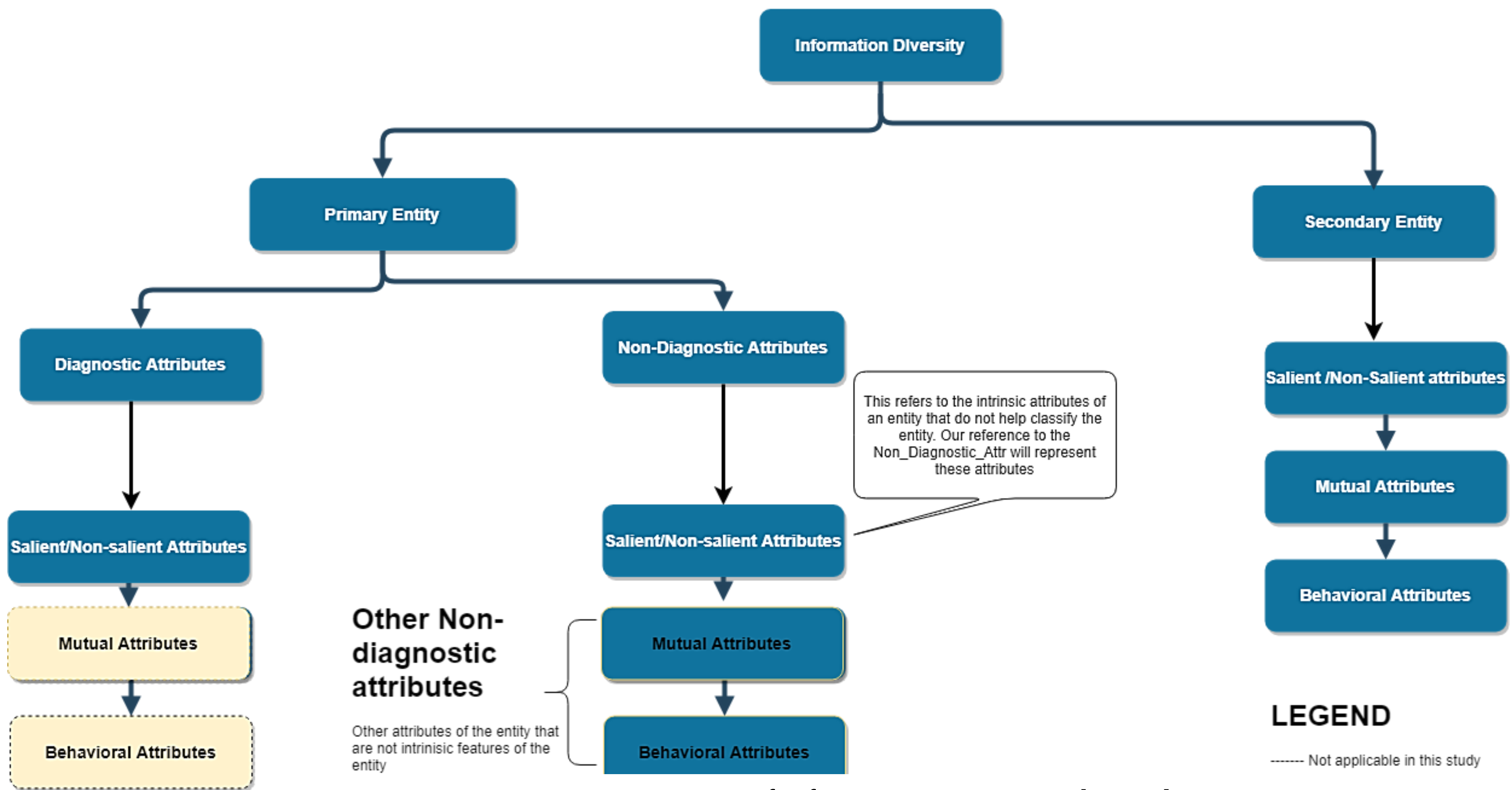


Figure 3.4: Components of Information Diversity in this study

The components of information diversity are attributes of primary entities and secondary entities present in the visual field (presented image) of the contributor. As depicted in Figure 3.4, each image has one primary entity (except in catch images used to screen out inattentive participants), and either no secondary entity or one or more secondary entities. Salient attributes of a primary entity that can be used to classify it as a tyran or non-tyran are diagnostic attributes. Other salient or non-salient attributes of an entity that are not important for classification are non-diagnostic attributes. Also, mutual attributes and behavioural attributes of an entity can be diagnostic attributes or non-diagnostic. For instance, the statement that the female *Aedes egypti* mosquito swims in water may help in the classification of the insect as male or female, but swimming, even though diagnostic is not an attribute that is a part of the features inherent in the mosquito's body.

However, in this study, all mutual and behaviour attributes are non-diagnostic. We, therefore, have two categories of non-diagnostic attributes – non-diagnostic attributes that are inherent in the stimulus (simply referred to as non-diagnostic_attr in this thesis) and other non-diagnostic attributes not inherent in the stimulus, e.g., mutual and behaviour attributes. Secondary entities are common organisms like birds and insects, and we do not separate their inherent attributes into diagnostic and non-diagnostic attributes.

To ensure the results of our analyses are not due to the inherent differences in the images presented to participants, we standardized the data for each variable across the presented images using the Robust Scaler. The Robust Scaler is a standardization and variance scaling technique provided in the Scikit-learn machine learning package of python, and it is the most accommodating of outliers since it uses data in the 1st quartile and 3rd quartile to center and scale the entire data set, extremely high values do not have any effect on the results (www.scikit-learn.org).

3.4.1 Results for the Effect of Training on Information Diversity (H1)

To determine the difference in diversity between participants who have received different types of training about the entity, we compared the number of diagnostic attributes (Diagnostic_Attr) and non-diagnostic attributes of the primary entity (Non-diagnostic_Attr) between the treatment conditions. These comparisons address H1, i.e., there will be no significant difference in the number of diagnostic attributes reported between trained and untrained contributors; and H2, i.e., untrained contributors will report fewer non-diagnostic attributes of a target entity than implicitly trained contributors but more non-diagnostic attributes than explicitly trained contributors. We also compared other non-diagnostic attributes, such as attributes describing the state of the primary entity (Behavior_Attr) and attributes describing the primary entity's interaction with other entities or its environment (Mutual_Attr), for each image presented to the participants. The results are presented in Tables 3.2 and 3.3.

Table 3.2: ANOVA Results for Primary Entity Attributes*

Variable	Group Mean					
	Untrained	**Explicit**	**Implicit**	***F***	**p-value**	η^2
Diagnostic_Attr	1.272	1.514	1.440	0.92	0.399	0.001
Non-diagnostic_Attr	**2.009**	**0.580**	**2.801**	**60.405**	**0.000**	**0.083**
Behavior_Attr	**0.491**	**0.045**	**0.022**	**36.820**	**0.000**	**0.052**
Mutual_Attr	**1.763**	**0.681**	**0.725**	**32.843**	**0.000**	**0.047**

*Significant differences are bolded (p=0.05)

Table 3.3: Post-Hoc Test Results for Primary Entity Attributes*

	A	**B**	**mean(A)**	**mean(B)**	**Mean Diff.**	**Std. Err**	**T**	**p-value**	η^2
Diagnostic_Attr	Explicit	Implicit	1.514	1.440	0.074	0.183	0.407	0.900	0.000
	Explicit	Untrained	1.514	1.272	0.242	0.183	1.324	0.383	0.002
	Implicit	Untrained	1.440	1.272	0.167	0.183	0.917	0.718	0.001
Non-diagnostic_Attr	**Explicit**	**Implicit**	**0.580**	**2.801**	**-2.221**	**0.205**	**-10.844**	**0.001**	**0.116**
	Explicit	**Untrained**	**0.580**	**2.009**	**-1.429**	**0.205**	**-6.975**	**0.001**	**0.052**
	Implicit	**Untrained**	**2.801**	**2.009**	**0.792**	**0.205**	**3.869**	**0.001**	**0.016**
Behavior_Attr	Explicit	Implicit	0.045	0.022	0.022	0.062	0.362	0.900	0.000
	Explicit	**Untrained**	**0.045**	**0.491**	**-0.446**	**0.062**	**-7.243**	**0.001**	**0.055**
	Implicit	**Untrained**	**0.022**	**0.491**	**-0.469**	**0.062**	**-7.605**	**0.001**	**0.061**
Mutual_Attr	Explicit	Implicit	0.681	0.725	-0.045	0.151	-0.295	0.900	0.000
	Explicit	**Untrained**	**0.681**	**1.763**	**-1.083**	**0.151**	**-7.162**	**0.001**	**0.054**
	Implicit	**Untrained**	**0.725**	**1.763**	**-1.038**	**0.151**	**-6.867**	**0.001**	**0.050**

*Significant differences are bolded (p=0.05)

From Table 3.2, the number of diagnostic attributes reported, Diagnostic_Attr, is not significantly different across the groups with $F(2,1341) = 0.92$, $p = 0.399$ at a 5% level of significance. We can, therefore, conclude that the number of diagnostic attributes reported is equal across the groups, supporting H1a. However, Non-diagnostic_Attr is significantly different across the three groups with $F(2,1341) = 60.405$, $p < 0.0000$ at a 5% level of significance. From the post-hoc tests, we observe that all group means are significantly different from each other, with the average Non-diagnostic_Attr for the Implicitly Trained Group being the maximum and that for the Explicitly Trained Group being the minimum, supporting H1b.

The number of attributes reported that describe the primary entity's behaviour Behavior_Attr is also significantly different across the groups. The post-hoc test results suggest that the group means for the Explicitly Trained Group, and the Implicitly Trained Group are significantly lower than the average for the Untrained Group. However, the Explicitly Trained Group and the Implicitly Trained Group are not significantly different. The number of mutual attributes is also significantly different across the groups. From the post-hoc tests, we again observe that the group means of the Explicitly Trained Group and the Implicitly Trained Group are significantly lower than that of the Untrained Group. However, there is no significant difference between the Explicitly Trained Group and the Implicitly Trained Group.

For the secondary entities, we analyze the variables Secondary_Ent, Secondary_Ent_Attr, Secondary_Ent_Behavior, and Secondary_Ent_Mutual. Further, since we define information diversity as the number of unique attributes reported about an entity, information diversity is the sum of all the attributes reported for each image, given as:

Information Diversity = Diagnostic_Attr + Non-diagnostic_Attr + Behavior_Attr + Mutual_Attr + Secondary_Ent_Attr + Secondary_Ent_Mutual + Secondary_Ent_Behavior + Diagnostic_Attr_Variance + Non-Diagnostic_Attr_Variance

We present the results of our analyses in Tables 3.4 and 3.5.

Table 3.4: ANOVA Results for Information Diversity*

Variable	**Group Mean**					
	Untrained	Explicit	Implicit	F	p-value	η^2
Secondary_Ent_Attr	**1.730**	**0.246**	**1.261**	**43.953**	**0.000**	**0.062**
Secondary_Ent_Behavior	**0.491**	**0.056**	**0.257**	**13.193**	**0.000**	**0.019**
Secondary_Ent_Mutual	**1.105**	**0.826**	**0.547**	**8.640**	**0.000**	**0.013**
Information Diversity	**8.984**	**4.070**	**7.433**	**85.967**	**0.000**	**0.114**

*Significant differences are bolded (p=0.05)

Table 3.5: Post-hoc Results for Information Diversity*

	A	B	mean(A)	mean(B)	Mean Diff.	Std. Err	T	p-value	η^2
Secondary_Ent_Attr	**Explicit**	**Implicit**	**0.246**	**1.261**	**-1.016**	**0.162**	**-6.274**	**0.001**	**0.042**
	Explicit	**Untrained**	**0.246**	**1.730**	**-1.484**	**0.162**	**-9.170**	**0.001**	**0.086**
	Implicit	**Untrained**	**1.261**	**1.730**	**-0.469**	**0.162**	**-2.896**	**0.011**	**0.009**
Secondary_Ent_Behavior	**Explicit**	**Implicit**	**0.056**	**0.257**	**-0.201**	**0.085**	**-2.368**	**0.047**	**0.006**
	Explicit	**Untrained**	**0.056**	**0.491**	**-0.435**	**0.085**	**-5.131**	**0.001**	**0.029**
	Implicit	**Untrained**	**0.257**	**0.491**	**-0.234**	**0.085**	**-2.763**	**0.016**	**0.008**
Secondary_Ent_Mutual	Explicit	Implicit	0.826	0.547	0.279	0.134	2.078	0.094	0.005
	Explicit	Untrained	0.826	1.105	-0.279	0.134	-2.078	0.094	0.005
	Implicit	**Untrained**	**0.547**	**1.105**	**-0.558**	**0.134**	**-4.157**	**0.001**	**0.019**
Information Diversity	**Explicit**	**Implicit**	**4.070**	**7.433**	**-3.363**	**0.383**	**-8.777**	**0.001**	**0.079**
	Explicit	**Untrained**	**4.070**	**8.984**	**-4.914**	**0.383**	**-12.825**	**0.001**	**0.155**
	Implicit	**Untrained**	**7.433**	**8.984**	**-1.551**	**0.383**	**-4.048**	**0.001**	**0.018**

*Significant differences are bolded (p=0.05)

Table 3.4 shows that Secondary_Ent_Attr is significantly different across the groups. The post hoc test results in Table 3.5 suggest that all group means are significantly different from one another, with the Untrained Group being the maximum and the Explicitly Trained Group being the minimum. Secondary_Ent_Behavior is also significantly different across the three groups. The post hoc tests show that the average Secondary_Ent_Behavior for Untrained Group is the maximum and is significantly higher than that of the Explicitly Trained Group, which is the minimum. However, there is no statistically significant difference between the number of Secondary_Ent_Behavior reported by the Untrained

Group and the Implicitly Trained Group. Secondary_Ent_Mutual is also significantly different across the groups. The post hoc tests show that the Untrained Group's mean for Secondary_Ent_Mutual is highest and significantly higher than that for the Explicitly Trained Group and the Implicitly Trained Group. However, there is no significant difference between the Explicitly Trained Group and the Implicitly Trained Group for Secondary_Ent_Mutual.

Furthermore, as shown in Table 3.4 information diversity is significantly different across the groups with $F(2{,}81) = 85.967$, $p = 0.000$. Post hoc test results in Table 3.5 shows that the group mean for the Untrained Group is significantly higher than the Explicitly Trained Group and the Implicitly Trained Group, while the group mean for the Implicitly Trained Group is significantly greater than the mean for the Explicitly Trained Group.

3.4.2 Results for Hypotheses on Ability to Report Variability (H2)

Variability in target stimulus is measured using the variables Diagnostic_Attr_Variance and Non-Diagnostic_Attr_Variance. To compare the difference in these variables across the groups, we use one-way ANOVA. For post-hoc comparison of the group means, we use Tukey's HSD test. The results are presented in Tables 3.6 and 3.7.

Table 3.6: ANOVA Results for Variability*

Variable	Group Mean					
	Untrained	Explicit	Implicit	F	p-value	η^2
Diagnostic_Attr_Variance	0.056	0.112	0.100	0.890	0.411	0.001
Non-Diagnostic_Attr_Variance	**0.067**	**0.011**	**0.279**	**14.757**	**0.000**	**0.022**

*Significant differences are bolded (p=0.05)

Table 3.7: Post-hoc test Results for Variability

	A	B	mean(A)	mean(B)	Mean Diff.	Std. Err	T	p-value	η^2
Diagnostic_Attr_Variance	Explicit	Implicit	0.112	0.100	0.011	0.044	0.252	0.900	0.000
	Explicit	Untrained	0.112	0.056	0.056	0.044	1.261	0.419	0.002
	Implicit	Untrained	0.100	0.056	0.045	0.044	1.008	0.664	0.001
Non-Diagnostic_Attr_Variance	**Explicit**	**Implicit**	**0.011**	**0.279**	**-0.268**	**0.052**	**-5.150**	**0.001**	**0.029**
	Explicit	Untrained	0.011	0.067	-0.056	0.052	-1.073	0.610	0.001
	Implicit	**Untrained**	**0.279**	**0.067**	**0.212**	**0.052**	**4.077**	**0.001**	**0.018**

The ANOVA results in Table 3.6 show that Diagnostic_Attr_Variance is not significantly different for the three groups ($F(2,221) = 0.154, p = 0.064$) at the 5% level of significance. The post hoc tests also show that there are no significant differences in the pairwise group means. However, Non-Diagnostic_Attr_Variance is significantly different for the comparison groups with $F(2,249) = 18.196$, $p < 0.0001$. The post hoc tests (Table 3.7) show that the Implicitly Trained Group is significantly higher than the Untrained Group and the Explicitly Trained Group, but there is no significant difference in the group means of Untrained Group and the Explicitly Trained Group.

While all groups reported variability in diagnostic attributes, the implicitly trained group who have attended to the attributes of the tyran reported more variability in the non-diagnostic attributes.

3.4.3 Results for Accuracy and Information Quality Dimensions (H3)

We analyzed the data from our experiment to understand the relationship between training and information quality dimensions, including information diversity. We also investigate the relationship between information quality dimensions in the presence or absence of training.

Firstly, we investigated how different types of training affects accuracy and completeness. The response to accuracy is 0-1 valued, and there is no response to accuracy for the Untrained Group. Accuracy measures whether or not a contributor was able to correctly classify the primary entity as either a tyran or a non-tyran. When contributors correctly classify an entity, we enter 1 for accuracy, and when they do not, we record 0. We compare the proportion of Accuracy = 1 in the Explicitly Trained Group and the Implicitly Trained Group using a chi-square test.

Table 3.8: Accuracy for Explicitly and the Implicitly Trained Groups

	Accuracy		
Group	0	1	Total
Explicit	62	386	448
Implicit	162	286	448
Total	224	672	895

From Table 3.8, the chi-squared statistic of independence is 58.339, with a p-value of **0.0000**. The proportion of accuracy in the Explicitly Trained Group is 0.861, and the proportion accuracy in the Implicitly Trained Group is 0.638. Thus, we can conclude that the proportion of accuracy in the Explicitly Trained Group is significantly higher than the Implicitly Trained Group.

Secondly, to understand how training impacts traditional information quality dimensions, we examine the effect of training on completeness and accuracy. We operationalize completeness in the context of the study's task – which is the classification of a target stimulus as either tyran or non-tyran – as the reporting of sufficient breadth and depth of diagnostic and non-diagnostic attributes about the target entity. We, therefore, compare the number of unique attributes reported (i.e., breadth) and the number of attribute values reported about each unique attribute (i.e., depth) across the groups. The results are presented in Table 3.9 and Table 3.10.

Table 3.9: ANOVA Results for Differences in Completeness*

Variable	Group Mean					
	Untrained	Explicit	Implicit	F	p-value	η^2
Completeness (Breadth)	**3.281**	**2.094**	**4.241**	**22.327**	**0.000**	**0.032**
Completeness (Depth)	**2.835**	**1.016**	**2.773**	**25.507**	**0.000**	**0.037**

*Significant differences are bolded (p=0.05)

Table 3.10: Post-hoc Results for Differences in Completeness*

	A	B	mean(A)	mean(B)	Mean Diff.	Std. Err	T	p-value	η^2
Completeness (breadth)	**Explicit**	**Implicit**	**2.094**	**4.241**	**-2.147**	**0.322**	**-6.670**	**0.001**	**0.047**
	Explicit	**Untrained**	**2.094**	**3.281**	**-1.187**	**0.322**	**-3.688**	**0.001**	**0.015**
	Implicit	**Untrained**	**4.241**	**3.281**	**0.960**	**0.322**	**2.982**	**0.008**	**0.010**
Completeness (Depth)	**Explicit**	**Implicit**	**1.016**	**2.773**	**-1.758**	**0.289**	**-6.077**	**0.001**	**0.040**
	Explicit	**Untrained**	**1.016**	**2.835**	**-1.819**	**0.289**	**-6.289**	**0.001**	**0.042**
	Implicit	Untrained	2.773	2.835	-0.061	0.289	-0.212	0.900	0.000

*Significant differences are bolded (p=0.05)

Table 3.8 shows that the breadth of attributes reported is significantly different across the groups with $F(2{,}1341) = 22.327$, $p = 0.000$. Post hoc test results from Tables 3.9 and 3.10 show that the group means for the Explicitly Trained Group report significantly fewer attributes about the target entity than the Implicitly Trained Group and the Untrained Group, but the Untrained Group reports fewer attributes about the target entity than the Implicitly Trained Group. Depth is also significantly different across the groups with

$F(2{,}1341) = 25.507$, $p < 0.0000$. The post-hoc test results show that the mean for the Untrained Group is significantly greater than the mean for the Explicitly Trained Group. The Implicitly Trained Group also has a mean that is greater than that of the Explicitly Trained Group. However, the means of the Implicitly Trained Group is lesser than those for the Untrained Group.

Finally, we report the combined effect of completeness and accuracy on the reporting of secondary entities and diverse data. We have used multivariate linear regression to determine the relationship between these variables. Table 3.11 shows the regression coefficients and their p-values. Multiple R^2 values to determine the combined effect is also reported together with p-values. For the Explicitly Trained Group, accuracy and completeness both affect the diversity of attributes reported. However, accuracy has a negative relationship with diversity, whereas completeness has a positive one. However, accuracy and completeness are not associated with the reporting of secondary entities for explicitly trained contributors.

If we consider the Implicitly Trained Group only, accuracy has no significant relationship to the diversity of contributed information, while completeness is negatively associated with the reporting of secondary entities but positively associated with the reporting of diverse data.

Table 3.11: Traditional Information Quality Dimensions and Information Diversity[7]

Variable	**Explicit**			**Implicit**		
	Accuracy (p-value)	Completeness (p-value)	R^2(p-value)	Accuracy (p-value)	Completeness (p-value)	R^2 (p-value)
Secondary_Ent	2.4702 (0.096)	**-0.244 (0.038)**	0.192(0.07)	0.2197 (0.837)	**-0.3057 (0.001)**	**0.371(0.031)**
Information Diversity	**-1.7671 (0.042)**	**0.8522 (0.00)**	**0.899(0.00)**	-0.1365 (0.833)	**0.857 (0.00)**	**0.915 (0.00)**

3.5 Discussion

The results of this study show that training does not affect the capacity of crowds to report diagnostic attributes accurately. Both untrained and trained contributors were able to accurately report diagnostic attributes, which can be used by humans or machines to determine the class of a stimulus. Crowdsourcers whose projects mainly require the accurate classification of stimuli should therefore not have any problems using untrained or trained contributors when they can automate the classification of stimulus based on the reported attributes. For instance, machine learning algorithms can classify stimuli based on reported diagnostic attributes. Since there are usually more untrained contributors than trained contributors, using untrained contributors to collect diagnostic data may be a more efficient use of resources, allowing crowdsourcing projects to collect more data by

[7] Coefficients are listed in the columns for both accuracy and completeness. Combined R^2 values are provided in a separate column and p-values are in parentheses for each variable.

involving more people. However, when humans are required to carry out classification (pattern matching) tasks, then explicitly trained contributors are more accurate.

Crowdsourcers may also be interested in diverse data, which is more amenable to repurposing and may yield more insight than uniform data (Ogunseye & Parsons, 2018). This study reveals that implicitly trained contributors are better at reporting complete attributes because they have been primed through training to attend to both diagnostic and non-diagnostic attributes of an entity through bottom-up attentional allocation. However, untrained contributors provide more information about the attributes they report than trained contributors (i.e., greater depth). The depth of attributes untrained contributors report can yield more insight into an entity. The capacity for contributors to report a description of attributes is important since it may be difficult for contributors or crowdsourcers to revisit the exact state of a phenomenon after it has occurred. Therefore, crowdsourcers will want to capture as much detail the first time. Furthermore, our results show that, although implicitly trained contributors report more inherent non-diagnostic data than other groups, untrained contributors report more non-diagnostic attributes in general (i.e., combining the primary entity's inherent non-diagnostic attributes with its other non-diagnostic attributes such as mutual attributes and intrinsic attributes). In the same vein, untrained contributors report more attributes about secondary entities in their visual space than any other group.

Moreover, even though we expected explicitly trained contributors who had been sensitized to diagnostic attributes to report more variability in these attributes, we found that explicitly trained contributors did not have any advantage over other groups as they all reported a significantly similar amount of variability in diagnostic attributes. Without learning explicit or implicit rules, untrained contributors were able to identify salient attributes and report variations to these attributes when they occur. Apparently, despite their distributed attention, untrained contributors commit enough salient attributes to memory and detect when these attributes changed between presented images. They are not blind to changes in diagnostic attributes, nor are they distracted by the presence of other stimuli in their visual field.

Also, Untrained contributors perform as well as Explicitly Trained contributors when it comes to reporting variability in non-diagnostic attributes. Implicitly Trained contributors, however, report more variability in non-diagnostic attributes than any other group mainly because they have been sensitized to pay attention to all inherent attributes of a stimulus, i.e., both diagnostic and non-diagnostic and are at an advantage when there is a need to report changes to these attributes. Implicitly trained contributors, therefore, have a lower tendency to suffer from change blindness or inattentional blindness, unlike the other groups who have not committed sufficient non-diagnostic attributes to memory.

Nonetheless, if we assume that crowdsourcers are interested in variability in diagnostic attributes as that may in some cases imply the existence of a new species or another subclass of an entity, then from our results, we can state that untrained contributors perform as well as trained contributors. The negative effect of a lack of selective attention on untrained contributors becomes obvious when it comes to reporting the inherent attributes of a stimulus that are non-diagnostic. Although untrained contributors perform better than contributors with explicit task knowledge, they do not do as well as contributors who have learned these attributes implicitly during training. This limitation in their capacity to learn non-diagnostic (and usually non-salient) attributes also manifests in their inability to detect changes in these attributes when they occur.

At the same time, the level of knowledge that contributors possess, evidenced by their level of accuracy, is, to a large extent, negatively correlated with the reporting of diverse data, while information diversity is positively related to completeness. Contributors who report more complete attributes are more likely to report diverse data and are less likely to report accurate classifications. Less knowledgeable contributors are more likely, therefore, to report diverse data than more knowledgeable contributors. This contradicts the widespread assumption of a positive relationship between knowledge and information quality, that motivate studies such as Budescu & Chen (2014) and Yang, Xue, & Gomes (2018). We view this result to be a consequence of selective attention. For one, we posit that accurate classification – a proxy for "level of knowledge" – is an outcome of the ability to selectively attend to the pertinent diagnostic attributes at the expense of other attributes of an entity in the visual space of a contributor. Contributors who can report accurate data, i.e., classify phenomena based on accurately identified attributes, need to tradeoff reporting information diversity to do so. However, accuracy is not a significant factor in predicting the amount of diverse data that Implicitly Trained contributors will report. Again, accuracy is not tested for the Untrained Contributors who have not been intimated on the purpose of the task or the classifications of the entities presented to them.

Altogether, information diversity promotes discoveries as it enables different users and uses of data, which can lead to both anticipated and unanticipated insights. Many crowdsourcing projects require the flexibility that diverse data affords. Since attributes ignored today may become diagnostic in the future (Hoffman & Rehder 2010), if there is ever a need for particular information from crowdsourced data, data sourced from untrained contributors will be better suited to provide such unanticipated insights, whereas, data acquired from trained contributors will be inadequate. Table 3.13 summarizes our findings.

Table 3.12: Summary of Hypotheses and Findings

Hypotheses	**Comments on Findings**	**Supported**
H1: Explicitly trained contributors will report a similar number of diagnostic attributes of a target entity as implicitly trained contributors and untrained contributors	There was no significant difference in the groups	Yes
H1b: Untrained contributors will report fewer non-diagnostic attributes of a target entity than implicitly trained contributors but more non-diagnostic attributes than explicitly trained contributors	Implicitly Trained Group reported more than the Untrained Group and the Explicitly Trained Group. Combined, the untrained group reported more non-diagnostic attributes than any other group	Yes
H1c: Untrained contributors will report more data about secondary stimuli and their attributes than implicitly trained contributors who will, in turn, report more of these data than explicitly trained contributors	True in all cases	Yes
H1d: Untrained contributors will report more diverse data than implicitly trained contributors who will in-turn report more diverse data than explicitly trained contributors	Untrained Contributors reported more diverse data than Implicitly Trained Contributors who in turn reported more diverse data than Explicitly Trained contributors	Yes
H2a: Explicitly trained contributors will report more variability involving the diagnostic attributes of a target stimulus than will implicitly trained contributors and untrained contributors.	Equal across all groups	No
H2b: Implicitly trained contributors will report more variability involving the non-diagnostic attributes of a target stimulus than will explicitly trained contributors and untrained contributors.	The implicitly trained group reported more variability in non-diagnostic attributes than explicitly trained and untrained contributors who reported a statistically similar amount of variability.	Yes
H3a: Explicitly trained contributors would report more accurate data than will implicitly trained contributors	True	Yes
H3b: Implicitly trained contributors will report more complete data about a stimulus than untrained and explicitly trained contributors	True for breadth, while untrained contributors reported more information about each attribute (i.e., Depth)	Yes
H3c: Information diversity will be negatively associated with accuracy while being positively related to completeness across the implicitly trained and explicitly trained groups	True for the explicitly trained group. However, accuracy does not affect diversity in the implicitly trained group	Yes

3.6 Conclusion

Repurposable data is adaptable to both anticipated and previously unanticipated needs. The collection of repurposable data requires that crowdsourced data be complete, accurate, and diverse. Because knowledge of some subject matter is widely assumed necessary if one is to provide high-quality data about that subject, knowledgeable contributors are typically preferred over novice contributors in many data crowdsourcing applications. Training potential participants on the crowdsourcing task to be performed, therefore, provides a way for crowdsourcers to ensure that the data they collect is of high quality. However, because information diversity is a requirement for repurposability, there is a need to understand how training affects the collection of diverse data. Using an experiment in the context of citizen science involving 84 participants reporting sightings of an artificial insect, we examined the effect of two training approaches on the diversity of contributed information.

Furthermore, we investigated the relationship between traditional information quality dimensions of accuracy and completeness and the new information quality dimension – information diversity. We found that teaching contributors explicit inclusion rules encourages knowledge-driven attentional allocation, which results in less diverse data. Allowing contributors to discover inclusion rules implicitly results in more diverse data, but not as diverse as if they are not trained at all. From this study, we can conclude that there is no significant advantage to restricting participation in crowds based on the type of training received by a contributor or the level of knowledge possessed by potential contributors. Every benefit to be derived from recruiting explicitly trained contributors can be derived from untrained contributors when classification is automated. However, when crowdsourcers must train, we find that it is better to train contributors implicitly as this leads to the reporting of more diverse data than contributors with explicit task knowledge. At least, there is a possibility that data sourced from crowds who have been implicitly trained can be further analyzed by an expert to correct for classification deficits that may occur due to a lack of explicit inclusion rules. The possibility of recovering non-diagnostic attributes if explicitly trained contributors are used may be next to none.

3.6.1 Limitations

The study described in this chapter uses an experimental design and therefore inherits the constraints inherent in such designs. In favour of control over aspects of our experiment, we have sacrificed realism. For example, we assume that contributors have similar levels of motivation, which may not be the case in the real world. We also assumed that the attributes and interactions depicted in the artificial images provided are highly similar to what is available in nature. The experiment also suffers from selection bias seeing as we only used university students, particularly business students. Using university students or strictly business students may already create an artificial knowledge-level of contributors. As with all experiments, there is, therefore, a possibility that the results obtained in our controlled setting may differ from the result that would be obtained in the real world.

References

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. *Scientific Reports*, *6*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5028888/

Awh, E., Vogel, E. K., & Oh, S.-H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201–208.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation*, *1*(3), 295–311.

Barsalou, L. W., & Sewell, D. R. (1984). *Constructing representations of categories from different points of view*. Retrieved from http://eprints.gla.ac.uk/112570/

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119.

Broadbent, D. E. (1958). *Perception and communication. Elmsford, NY, US*. Pergamon Press. http://dx. doi. org/10.1037/10037-000.

Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science*, *61*(2), 267–280.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860–1862.

Castriotta, M., & Di Guardo, M. C. (2011). Open Innovation and Crowdsourcing: The Case of Mulino Bianco. In *Information Technology and Innovation Trends in Organizations* (pp. 407–414). Retrieved from http://link.springer.com/chapter/10.1007/978-3-7908-2632-6_46

Chen, X. (1990). Local incommensurability and communicability. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1990*, 67–76. Philosophy of Science Association.

Clemen, R. T. (2008). Comment on Cooke's classical method. *Reliability Engineering & System Safety*, *93*(5), 760–765.

Davis-Stober, C. P., Budescu, D. V., Dana, J., & Broomell, S. B. (2014). When is a crowd wise? *Decision*, *1*(2), 79.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80.

Gelman, S. A., Collman, P., & Maccoby, E. E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, 396–404.

Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, *23*(3), 183–209.

Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. *Handbook of Psychology*.

Gopnik, A. (2009). How to Think Like a Baby—Big Think. Retrieved May 18, 2019, from https://bigthink.com/videos/how-to-think-like-a-baby

Gura, T. (2013). Citizen science: Amateur experts. *Nature*, *496*(7444), 259–261. https://doi.org/10.1038/nj7444-259a

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. *Handbook of Categorization in Cognitive Science*, 20–45.

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.

Homack, S. R. (2001). Understanding What ANOVA Post Hoc Tests Are, Really.

Hosseini, M., Phalp, K. T., Taylor, J., & Ali, R. (2014). *Towards crowdsourcing for requirements engineering*. Retrieved from http://eprints.bournemouth.ac.uk/21904/

Huang-Pollock, C. L., Carr, T. H., & Nigg, J. T. (2002). Development of selective attention: Perceptual load influences early versus late attentional selection in children and adults. *Developmental Psychology*, *38*(3), 363.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10–12), 1489–1506.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Review of Psychology*, *37*(1), 43–75.

Katsuki, F., & Constantinidis, C. (2014). Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136

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.

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review*, *111*(4), 880.

Lin, S.-W., & Cheng, C.-H. (2009). The reliability of aggregated probability judgments obtained through Cooke's classical model. *Journal of Modelling in Management*, *4*(2), 149–161.

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.

Mack, A. (2003). Inattentional blindness: Looking without seeing. *Current Directions in Psychological Science*, *12*(5), 180–184.

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of information and system quality: An empirical examination within the context of data warehousing. *Journal of Management Information Systems*, *21*(4), 199–235.

Niebur, E., & Koch, C. (1996). Control of selective visual attention: Modeling the" where" pathway. *Advances in Neural Information Processing Systems*, 802–808.

Ogunseye, S., & Parsons, J. (2016). Can Expertise Impair the Quality of Crowdsourced Data? *Proceedings of the 15th AIS SIGSAND Symposium, Lubbock, Texas*. Retrieved from https://www.researchgate.net/profile/Shawn_Ogunseye/publication/311573602_Can_Expertise_Impair_the_Quality_of_Crowdsourced_Data/links/584e154708aeb98925264f37.pdf

Ogunseye, S., & Parsons, J. (2018). Designing for Information Quality in the Era of Repurposable Crowdsourced User-Generated Content. *International Conference on Advanced Information Systems Engineering*, 180–185. Springer.

Piaget, J., & Inhelder, B. (1969). *The psychology of the child* (Vol. 5001). Retrieved from https://books.google.com/books?hl=en&lr=&id=GSXRDQAAQBAJ&oi=fnd&pg=PR1&

dq=piaget++learning+assimilation+and+accomodation&ots=Ku0KljePRv&sig=yP1SFQbR3Sx-WKN51pKPZLy88xA

Rensink, R. A., Kevin O'Regan, J., & Clark, J. J. (2000). On the failure to detect changes in scenes across brief interruptions. *Visual Cognition*, *7*(1–3), 127–145.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To See or not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science*, *8*(5), 368–373. https://doi.org/10.1111/j.1467-9280.1997.tb00427.x

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Schultze, U., & Leidner, D. E. (2002). Studying knowledge management in information systems research: Discourses and theoretical assumptions. *MIS Quarterly*, 213–242.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20.

Stein, E. W. (1992). A method to identify candidates for knowledge acquisition. *Journal of Management Information Systems*, *9*(2), 161–178.

Tarrell, A., Tahmasbi, N., Kocsis, D., Tripathi, A., Pedersen, J., Xiong, J., … de Vreede, G.-J. (2013). *Crowdsourcing: A snapshot of published research*. Retrieved from http://aisel.aisnet.org/amcis2013/EndUserIS/GeneralPresentations/2/

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*(2), 77–99.

Triesch, J., Ballard, D. H., Hayhoe, M. M., & Sullivan, B. T. (2003). What you see is what you need. *Journal of Vision*, *3*(1), 9–9.

Tripathi, A., Tahmasbi, N., Khazanchi, D., & Najjar, L. (2014). Crowdsourcing typology: A review of is research and organizations. *Proceedings of the Midwest Association for Information Systems (MWAIS)*. Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1018&context=mwais2014

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, *131*(1), 48.

Wickens, C. D., & McCarley, J. M. (2008). *Applied Attention Theory*. Boca Raton, FL: CRC Press.

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. *2011 IEEE Seventh International Conference on E-Science Workshops*, 14–19. https://doi.org/10.1109/eScienceW.2011.27

Williams, P., & Simons, D. J. (2000). Detecting changes in novel, complex three-dimensional objects. *Visual Cognition*, *7*(1–3), 297–322.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Yang, R., Xue, Y., & Gomes, C. (2018). Pedagogical Value-Aligned Crowdsourcing: Inspiring the Wisdom of Crowds via Interactive Teaching.

Chapter Four: **Understanding the Effect of Experience on Crowdsourced Data**

Abstract

Organizations and individuals who own or use crowdsourcing platforms implicitly value experience, expecting experienced contributors to report higher quality data than inexperienced contributors. Guided by selective attention theory from cognitive psychology, we examine this assumption in two types of crowdsourcing platforms – an online review platform and a citizen science platform. Using a machine-learning-based classification algorithm on datasets from these two crowdsourcing platforms, we find that the diversity of information reported in contributed data and the usefulness of contributed data decreases as contributors gain experience in a crowdsourcing task. Since usefulness is an outcome of information quality, we see from our sampled datasets that increasing experience from participation in crowdsourcing tasks is, in the long run, detrimental for the collection of diverse data. We, therefore, make recommendations for how owners and users of crowdsourcing platforms can keep their crowds from getting stale.

4.1 Background

Organizations and individuals (collectively, crowdsourcers) use crowdsourced information to make decisions. Integrative crowdsourcing, i.e., crowdsourcing that seeks to pool information about a phenomenon of interest from a distributed group of people, is a growing source of such decision-making information. For information consumers, the quality of crowdsourced information has a significant influence on the quality of insights it can produce. For instance, in online shopping, shoppers usually cannot evaluate products before purchase and must depend on crowdsourced information about products in the form of reviews and product descriptions (Rust, Inman, Jia, & Zahorik, 1999; Weathers, Sharma, & Wood, 2007). Decisions made by shoppers from inadequate information costs the online shopping industry between $100 and $260 billion annually in product returns (Minnema, Bijmolt, Gensler, & Wiesel, 2016; Sahoo, Dellarocas, & Srinivasan, 2018). In the same fashion, research results derived from low-quality crowdsourced information may lead to invalid conclusions or bad decisions. Therefore, like practitioners, many researchers are skeptical in their use of information, particularly crowdsourced information (Forbes, 2018; Weigelhofer & Pölz, 2016). Information consumers, who may be everyday online shoppers, researchers in academia, or decision-makers in the industry, would, therefore, benefit from higher quality information collected through crowdsourcing.

Crowdsourcers, i.e., owners of crowdsourcing projects, can proactively prevent the collection of low-quality information in their projects by first deciding who will be allowed to participate as members of the crowd. Also, crowdsourcers can ensure the quality of contributed data by employing assessors to evaluate contributions or by using automated data validation techniques after the fact (Gura, 2013; Malone, Laubacher, & Dellarocas, 2010; Wiggins, Newman, Stevenson, & Crowston, 2011). Proactive measures of quality assurance, such as crowd recruitment decisions, inform all other design decisions about the crowdsourcing project, such as how to simplify the definition of their task to suit the level of knowledge of potential contributors, how to motivate potential contributors, and how to design the task. In addition, proactive prevention is less resource-intensive than measures taken after data acquisition, saving crowdsourcers the cost of recruiting contributors, and collecting data that would later be classified as low quality and discarded.

The literature gives us insights about how to recruit the best contributors to engender a collection of high-quality data. For example, Budescu & Chen (2014) evaluated the knowledge of crowd members through knowledge-test questions interspersed in the task to determine the level of related knowledge a crowd member has, which determines whether their contribution should be permitted. Yang et al. (2018) promoted the training of crowd members who lacked knowledge about the crowdsourced task. These studies mainly focus on the accuracy dimension of information quality while seeking to ensure the quality of crowdsourced information, and they show a preference for contributors with knowledge of the task. However, the literature emphasizes that data is more valuable and provides more insights to users when it is diverse, allowing different data consumers to use it for both anticipated and unanticipated uses (e.g., see Hunter et al., 2013; Parsons, 1996). It would, therefore, be useful also to improve our understanding of how the choice of crowd members affects information diversity, i.e., the number of unique attributes represented in data.

The key questions that arise, therefore, are two-fold: (a) is it better for information diversity to only recruit people with prior experience in the data collection activity or to allow (or even encourage) participation by any contributor regardless of their level of relevant experience? (b) if there may be new uses for the collected data, should crowdsourcers actively recruit new participants throughout the life cycle of their projects (i.e., continuous recruitment), or should recruitment be a singular event at the start of a project? This chapter takes a step towards better understanding the impact of experience on the diversity of crowdsourced information. We explore the potential limitations of relying on the same crowd, particularly for projects that engage crowds in discoveries or evolve to encompass uses of data that were not anticipated when the project was designed. Furthermore, we investigate the relationship between information diversity and perceived information quality to gain insight into how information diversity affects the usefulness of data. Understanding the shortcomings of engaging with the same crowds in crowdsourced tasks will guide crowdsourcers in the development of targeted strategies in the design of their projects, improving their potential to collect high-quality data.

4.2 Crowd Member Knowledge and Information Diversity

Crowdsourcers prefer highly experienced contributors who have a greater knowledge of crowdsourcing tasks, over novices or amateurs. This preference influences crowd recruitment (Wiggins et al. 2011, Austen et al., 2016) and is based on the assumption that experienced crowds will provide higher quality data than unexperienced crowds. Experience is a source of knowledge (Leonard & Sensiper, 1998), which is defined in the business context as “information that is relevant, actionable and based at least partially on experience” (Leonard & Sensiper, 1998, p. 112). Although explicit knowledge and other aspects of our cognition may remain the same over time, tacit knowledge changes with increasing experience. In this study, we focus on experience gained through participation in a crowdsourcing task.

In integrative crowdsourcing, participants may have no prior tacit or explicit knowledge, some tacit knowledge but no explicit knowledge, some explicit knowledge but no tacit knowledge, or, more commonly, some combination of both tacit and explicit knowledge (Argote & Miron-Spektor, 2011). Through continued participation, contributors acquire experience, which may or may not refine their explicit knowledge in cases where they had some previous participation. Experience is the part of a contributor's knowledge that is guaranteed to change with participation, regardless of the composition or amount of a contributor's prior knowledge. It is measured in terms of the "cumulative number of tasks" performed (Argote & Miron-Spektor 2011, p 1124) and may include successful and unsuccessful task performances (Denrell & March, 2001; Kim & Rehder, 2009; Sitkin, 1992).

Experience results from continued participation in a task or from participation in a novel task (Katila & Ahuja, 2002; March, 1991). At the same time, experience may vary in frequency and pace from one individual to another. It may be gained before a task (Carrillo & Gaimon, 2000), that is, from prior participation in a similar or related task. Experience may also be gained during or after a task (Ellis & Davidi, 2005; Morris & Moore, 2000; Roese & Olson, 1995). Experience is an antecedent for selective attention. For instance, (Schwartzstein, 2014, p. 1424) argues that an "agent needs to learn which variables are worth attending to through experience." We explore the relationship between experience and selective attention in the next section and synthesize literature to develop hypotheses on how selective attention from experience will affect the diversity of information provided by contributors.

4.3 Hypotheses Development

The tendency for selective attention and classification occurs naturally in humans as we gain experience by observing regularities about entities in our world (Perruchet & Pacton, 2006; Saffran, Aslin, & Newport, 1996; Turk-Browne, Scholl, Chun, & Johnson, 2009; Zhao, Al-Aidroos, & Turk-Browne, 2013). As time passes, humans have more opportunities to be exposed to stimuli and observe these regularities, leading to the development of an attentional set (a set of attributes about a stimulus considered salient and co-occurring). Attentional sets are what guide selective attention. When the attributes of stimuli are encoded into memory as an attentional set, subsequent exposure to the attributes of a stimuli activate an attentional set, maintaining those attributes in memory and increasing their relevance for selective attention (Awh & Jonides, 2001; Downing, 2000; Postle, Awh, Jonides, Smith, & D'Esposito, 2004; Bradley R. Postle, 2006). Statistical frequency of exposure to stimuli, thus, impacts selective attention (Sloutsky, 2003), which employs recognition memory to direct our attention (Cosman & Vecera, 2013).

The capacity of humans to pay attention increases with time as they develop from infancy to adulthood (Richards & Turner, 2001). In other words, as we gain experience, we become more open to selectively attending to information to manage our limited cognitive resources. Infants are exemplars of how humans respond when they lack enough information about stimuli. Since many stimuli are new to infants, they cannot selectively attend to the attributes of those stimuli. They reason about entities by observing the salient features of individual stimuli and are therefore naturally comparable to novice contributors in an integrative crowdsourcing context (Keil, 1989; Kloos & Sloutsky, 2008). For this reason, the ability to accurately and efficiently classify stimuli using their key attributes is a distinguishing factor between adults and infants (Best, Yim, & Sloutsky, 2013).

Conversely, the tendency for adults to selectively attend to attributes of phenomena with which they have prior experience helps us understand how experienced contributors report data in a crowdsourced task. As experience increases, the tendency for selective attention increases correspondingly. Adult humans decide which attributes of stimuli to which to attend based on their prior experience with similar stimuli, and they continue to value the usefulness of those attributes the more they are exposed to similar stimuli (Gazzaley & Nobre, 2012).

4.3.1 Hypothesis on Number of Attributes in Contributions

Generally, we expect experienced contributors to use a top-down attentional distribution and, therefore, selectively attend to specific attributes of stimuli, reporting only data they consider pertinent to a task from their experience. Less experienced contributors are less inclined to attend selectively, and therefore consider more attributes of a stimulus with which they lack prior experience (Corbetta & Shulman, 2002; Itti & Koch, 2001). Experience also determines what attributes knowledgeable contributors prioritize when observing future instances of a class (Kim & Rehder, 2011).

Several studies have tested the effect of experience on selective attention. For example, Pick and Frankel tested second graders and sixth graders' capacity for selective attention, revealing that the capacity to attend selectively increased for both groups when they were exposed to the task before they were tested compared to those who had no prior exposure to the task. In another study, Smith, Kemler, & Aronfreed (1975) tested the capacity of kindergarteners, second graders, and fifth graders to focus their attention on a task in the presence of distractors. Like Strutt, Anderson, and Well, (1975); Well, Lorch, and Anderson, (1980), and Best et al. (2013), who tested the ability of children and adults to classify in the presence of distractors, Smith et. al. found that young children are inexperienced at selectively attending to relevant attributes in the presence of irrelevant attributes.

These studies reveal that experience in a task acquired by contributors will lead to an increase in the reporting of attributes they learned to be relevant to the task (Harnad, 2005); thus, they are expected to be less likely to attend to irrelevant attributes compared to novices (Katsuki & Constantinidis, 2014; Plebanek & Sloutsky, 2017). Experienced contributors would also be more inclined to ignore variability in non-salient attributes of an entity when they occur. They are more resistant to learning something new (Plebanek & Sloutsky, 2017), increasing their tendency towards change blindness and impeding their ability to provide data that can lead to discoveries. The use of an attentional set is therefore expected to inhibit contributors' ability to report minor variations not present in these encoded attributes. On the other hand, novices and less experienced contributors employ a bottom-up attentional distribution strategy and are expected to report more information about stimuli they observe, compared to experienced contributors.

Entities have a finite number of attributes. These attributes may be intrinsic – an inherent part of the entity, or mutual – attributes that describe a relationship between two or more entities (Wand, Storey, & Weber, 1999). Our reference to mutual properties refers to Non-binding mutual properties. "Non-binding mutual properties are those properties shared by two or more things that do not 'make a difference' to the things involved; for example, order relations or equivalence relations. By contrast, binding mutual properties are those properties shared by two or more things that do 'make a difference' to the things involved" (Rosemann & Green, 2002, p. 82). Kiwelekar & Joshi (2010, p. 4) further explain that non-binding mutual properties are relational properties that occur when "no interaction is involved between two related things. For example, younger than relationship between two persons does not show any kind of interaction".

In many cases, mutual attributes are irrelevant to the identification of the entity within a class, i.e., the classification task, but they may aid the diagnosis of what the entity is and its state. Mutual attributes are mainly adjectives that describe an entity's relation to other entities. Adjectives are functions that map the meaning of a noun phrase to the meaning of another noun phrase, whether or not both nouns are explicitly stated (Kamp, 2013). As intrinsic attributes are the physical attributes of an entity, mutual attributes are the main source of diversity as they are dependent on the contributor and can represent state-related information about the entity. Because the diagnostic attributes of a class are mostly intrinsic, experienced contributors will report fewer mutual attributes about entities they observe compared to inexperienced contributors. Inexperienced contributors will provide attributes that cover both intrinsic and mutual properties of the observed entity, reporting more mutual attributes compared to experienced contributors. Consequently, we predict that as contributors gain experience in a task, the number of mutual attributes reported will decrease.

> ***H1:*** *The number of mutual attributes provided will decrease as contributors gain experience in a task.*

Besides, contributors report data at different rates. Even when contributors report data at similar frequencies, the number of entities they report about may differ. Factors external to a crowdsourced task can cause differences in the rates at which people contribute data. Such external factors (e.g., online shopping systems) may include direct marketing pushes by online stores or seasonality. For example, people shop more and review more products during the holiday seasons (Smith, 1999). In citizen science projects, active recruitment campaigns can increase crowd member turnout during the campaign periods. Internal factors (those inherent to the task itself) may also be responsible for variation in the frequency of participation by crowd members. Such internal factors may include the design of the crowdsourcing platform to restrict the frequency of participation. For example, crowdsourcing systems that apply gamification may require control of the frequencies at which their participants contribute data. The nature of the crowdsourced task may also dictate the frequency of participation for crowd members. Crowdsourcing projects that involve reporting about stars in the sky at night or insects pollinating flowers in spring are accessible exemplars.

To investigate how the diversity of data contributed by crowds change, we assume that the level of experience that crowd members have about the task increases monotonically. Contributors' knowledge of a task will increase as they gain experience

(Harnad, 2005). An increase in experience in a task implies an increase in the tendency for learned inattention to attributes of entities considered by the contributor to be trivial to the task. Experienced contributors will only report attributes to which they have selectively attended and consider salient from repeated observations of an entity. As Cosman and Vecera (2014) argue, the frequency of exposure to a stimulus and the relationships between its attributes (statistical learning) is encoded into memory, contributing to the creation of an attentional set that inhibits the distribution of attention to other attributes of the stimulus considered less salient. The more contributors use their attentional set, the more likely they are to allocate attention to the attributes in the set alone (Awh & Jonides, 2001; Downing, 2000; Olivers, Meijer, & Theeuwes, 2006; Ryan, Althoff, Whitlow, & Cohen, 2000; Woodman & Luck, 2007). However, less experienced contributors will report more attributes as they lack the capacity for top-down attention allocation and are less likely to attend selectively to specific attributes because of prior knowledge (Zhao et al., 2013).

H2: *Contributors will provide less diverse data with increasing experience*

4.3.2 Hypothesis on Usefulness of Contribution

Precision can be a desired dimension of information quality for some crowdsourcing tasks. Nonetheless, when the goal of a crowdsourced task is to collect not readily accessible information about phenomena which may be used by more than one consumer for different purposes, every detail and perspective that can be represented in the crowdsourced dataset is potentially pertinent. Online shopping, where shoppers with different informational needs may access reviews and use the insights garnered from those reviews to make decisions about the purchase or non-purchase of a product, is a primary example of such a crowdsourcing task. The primary purpose of online reviews is to guide shoppers in their decision-making endeavours. Therefore, reviews are helpful when they can provide guidance and inform decisions of users who may have similar or very different criteria (requirements) for their decision outcomes (Mudambi & Schuff, 2010; Poston & Speier, 2005). When more helpful product reviews are available, the likelihood of goods being purchased increases, and the likelihood that they are returned after purchase decreases (Sahoo et al., 2018). It is thus beneficial to crowdsourcers to provide shoppers or data consumers with their most helpful reviews as these increase sales (Duan, Gu, & Whinston, 2008) and reduces decision-making time and cognitive stress from searching (Todd & Benbasat, 2000). We predict that contributions with more mutual attributes will be perceived by data consumers to be more helpful than contributions with fewer mutual attributes.

Other online review platforms such as Yelp use "usefulness" ratings to mean the same thing as "helpfulness" (McAuley & Leskovec, 2013). We also frame helpfulness as an operationalization of usefulness, i.e., helpful contributions are contributions that information consumers consider useful.

H3: *The usefulness of contributed data will be negatively related to experience*

4.4 Empirical Approach

4.4.1 Dataset Description

We use datasets from two integrative crowdsourcing domains to test the developed hypotheses. The first dataset is from a publicly available online review dataset, while the second dataset is from a citizen science system developed by members of the supervisory committee to collect data about flora and fauna in the province of Newfoundland and Labrador. Using these two different types of datasets helps cover two primary types of integrative crowdsourcing datasets, that is, integrative crowdsourcing that considers accuracy to be important (e.g., citizen science crowdsourcing), and integrative crowdsourcing that primarily focuses on informativeness (e.g., online reviews). The use of both datasets makes our findings more generalizable to other integrative crowdsourcing systems.

Furthermore, the datasets used also differ in the following ways: first, the review dataset is based on *abstract* categorization (Goldstone & Kersten, 2003), that is, the similarities between the entities in the category are not physical or concrete. An example is the Baby Products dataset from Amazon, which contains data about different types of baby products such as feeding bottles, toys, and clothes. These products have very few attributes in common except their use for babies. The Amazon dataset lacks information that can be used to subcategorize or decipher the similarity of products. Further, contributors rarely provide more than one review for the same product because they can edit previous reviews.

When people make purchases on Amazon.com, they are prompted to provide reviews on the purchase shortly after they receive it. Contributors are also able to rate the product they have purchased on a scale of 1 to 5 stars, where 1 star is the lowest possible rating, indicating that the crowd member rates the product as being of the poorest quality, and 5 stars imply that the crowd member considers the product to be of excellent quality.

Shoppers access reviews about products they are considering and use reviews to guide their purchase decisions. Shoppers are also able to rate the reviews based on their helpfulness in the decision-making process. The Amazon datasets used were collected over 18 years (1996 to 2014) by researchers from the University of California, San Diego (He & McAuley, 2016). Because the Amazon dataset contains different products, the effect of selective attention would be lesser than if the contributors had reviewed the same product multiple times. For each new product they encounter, the tendency to create new attention sets (to act as novices) and report more attributes occurs. We use the Amazon dataset about products that would be used in a Patio, in the discussions in this chapter. The dataset contains 993,490 records.

In contrast, datasets from citizen science projects are usually about entities that are of the same natural kind or more concretely similar (for more information on category types see Goldstone & Kersten, 2003). For this study, we use NLNature's data. NLNature collects data about fauna and flora in Newfoundland from contributors around the province, allowing contributors to provide data about different instances of the same type of organism more than once. The NLNature dataset allows us to investigate the changes in the diversity of the data that crowd members contribute as they gain experience. The data used from this project was collected from 2009 to 2013 and has 12,175 records.

Like McAuley & Leskovec (2013), we restrict the Amazon dataset contributors used in the analyses to those who have participated at least 50 times in the eighteen-year dataset. For the NLNature dataset, we analyze data from all contributors. The number of contributors with 50 or more reviews in the Amazon dataset is 24, with 1749 observations. The maximum number of observations provided by any contributor in the Amazon dataset is 161. For the 12,175-record NLNature dataset, 637 contributors provided the contributions used in the analyses.

4.5 Analyses

To analyze the data, we broke them down to attributes using machine learning. First, we syntactically parsed every contributed textual data item into parts of speech using the spaCy library in order to extract the adjectives (attributes) in the text. spaCy has a 92% accuracy rate in parsing and producing relevant parts of speech (Honnibal & Johnson, 2015). Following the extraction of attributes, we classify these attributes into intrinsic attributes and mutual attributes using a classification algorithm we developed based on the spaCy framework, adapted from sentiment analyses algorithms, that checks for the polarity of the attributes. Intrinsic attributes, attributes that are inherent in a stimulus should not show any polarity; that is, they should not reflect any positive or negative sentiments but be neutral. E.g., red, round, three.

On the other hand, mutual attributes show polarity — for example, beautiful, cheap, or full. Finally, we compare the similarity of the attributes arranged in chronological order for each contributor. To compare different attribute sets, we use Word2Vec to generate word vectors for the attributes. Word2vec "is a two-layer neural net that processes text. Its input is a text corpus, and its output is a set of vectors…turning text into numerical form" (www.skymind.ai). We then used spaCy's neural network model, which we trained with over one million unique vectors we compare vectors of each piece of contribution with a previous one by the same contributor. Our new similarity comparison algorithm was tested against the STSBenchmark – "[a] shared training and evaluation [data]set carefully selected from [an existing and already standardized] corpus of English shared task data" from 2012-2017 (Cer, Diab, Agirre, Lopez-Gazpio, & Specia, 2017). Our model achieved a 71% accuracy in determining the similarity of attributes. Figure 4.1 illustrates the process of comparing the diversity of two or more contributions. The information diversity score is the inverse of the similarity score.

Table 4.1 presents all variables used in the analyses. We analyzed the datasets using a Linear Mixed Model Regression method to account for the longitudinal nature of the data, making reviewerID a random factor in our model.

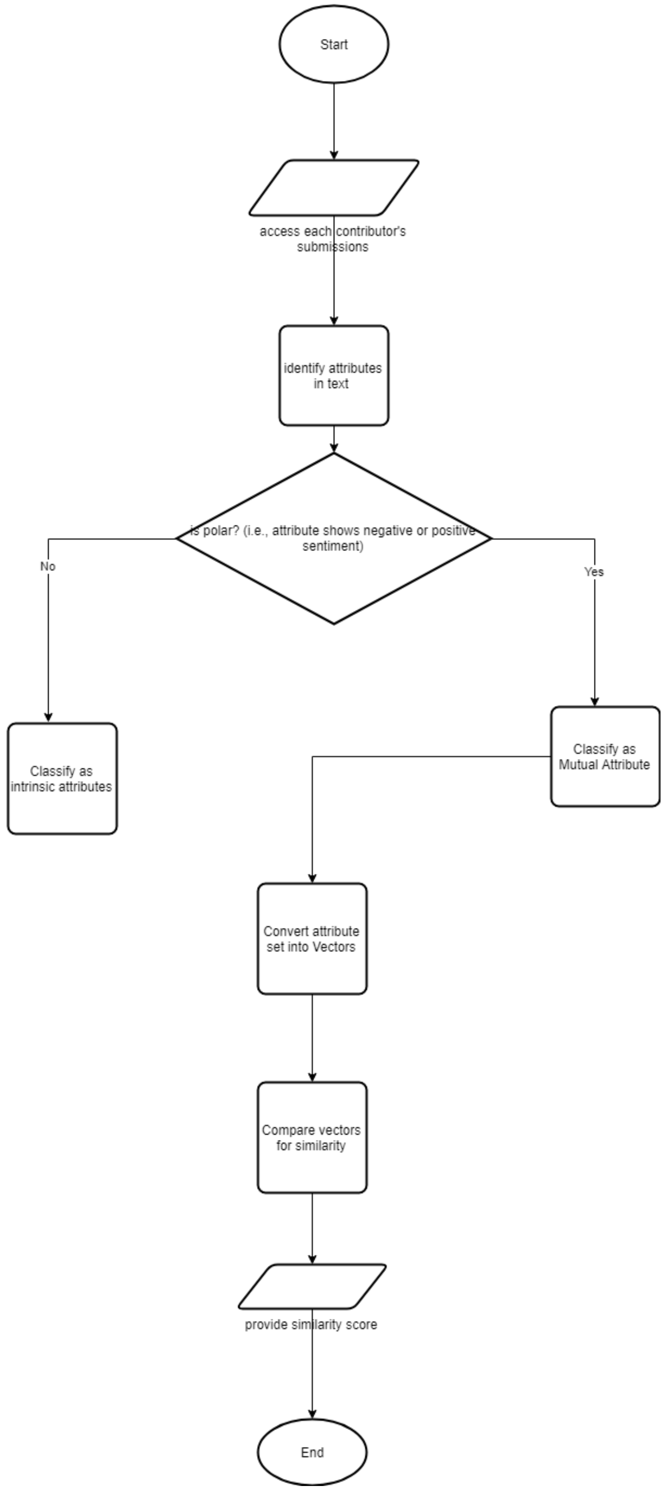


Figure 4.1: The information diversity comparison process

Table 4.1 – Names and descriptions of variables used in the analyses

Name	Description
reviewerID	Reviewer subject identifier
usefulness	Derived from a tuple of values showing positive review responses and the total number of review views. We compute usefulness as the number of upvotes for helpfulness/number of people who viewed the contribution.
reviewTime	Date of review submission
Intrinsic	Number of objective attribute responses
Mutual	Number of subjective attribute responses
textCount	Length of text contained in a review
Attr_Count	Total number of attributes in review
%Intrinsic	Percentage of objective attribute responses
%Mutual	Percentage of subjective attribute responses
Adj_Mutual	Average number of mutual attributes reported by the same contributor per day
experience	Number of reviews submitted by a contributor is used to measure the contributor's experience

We computed the average percentage of mutual attributes (%Mutual) across each day for contributors who participated more than once per day in the online review crowdsourcing task and recorded the results as Adj_Mutual.

4.6 Results

In our analyses, we take into consideration that multiple responses from the same contributor are interdependent. We describe the results of our text and statistical analyses below.

4.6.1 Results for Experience and Mutual Attributes

Analyzing the Amazon dataset containing reviews about products used in the Patio dataset and the NLNature dataset, we found that, as experience increased, mutual attributes decreased for both Amazon and NLNature datasets (See Table 4.2 and Table 4.3).

Table 4.2: Regression Results for Percentage of Mutual Attributes (Amazon)

```
=========================================================
Model:            MixedLM Dependent Variable: Adj_Mutual
No. Observations: 1749    Method:             REML
No. Groups:       24      Scale:              148.7841
Min. group size:  49      Likelihood:         -6895.2215
Max. group size:  161     Converged:          Yes
Mean group size:  72.9
---------------------------------------------------------
              Coef.  Std.Err.   z    P>|z| [0.025 0.975]
---------------------------------------------------------
Intercept     87.643    1.402 62.533 0.000 84.896 90.390
Experience   -0.025    0.011 -2.248 0.025 -0.047 -0.003
=========================================================
```

Table 4.3: Regression Results for Percentage of Mutual Attributes (NLNature)

Model:	MixedLM	Dependent Variable:	Adj_Mutual
No. Observations:	12175	Method:	REML
No. Groups:	637	Scale:	502.4432
Min. group size:	1	Likelihood:	-1246.6490
Max. group size:	2225	Converged:	Yes
Mean group size:	19.1		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	70.148	4.441	15.794	0.000	61.443	78.853
Experience	0.162	0.080	2.031	0.042	0.006	0.318

From the results in Table 4.2 and Table 4.3, we see that for the Amazon dataset, the z-stats are highly significant for the experience coefficient at z= -2.248, p=0.025. The percentage of mutual attributes reported in data is negatively associated with experience. The slope for experience is -0.025. An increase in experience by one more review results in a reduction of the percentage mutual attributes reported by -2.5% (Supporting H1).

For the NLNature data, experience has a coefficient of 0.162. A unit increase in crowd experience results in a 16.2% increase in the percentage of mutual attributes reported (Not supporting H1). The amount of variability accounted for by the experience variable in the Amazon dataset is 1% and 8% the NLNature dataset.

4.6.2 Results for Experience and Information Diversity

To analyze the data from contributors who have reported an entity more than once, we compared the diversity of attributes they report as their experience increases. Experience has a slope of -0.023 significant at $p<0.1$[8], z= -1.773. For NLNature data, we find that experience has a slope of -0.333. The z-value of experience is -2.835 at p=0.005. A unit increase in experience results in a 2.3% decrease in diversity for Amazon data and a 3.3% decrease in diversity for NLNature data. The amount of variability accounted for by the experience variable in the Amazon dataset is 1.3%, and NLNature is 11.7%. Both these results support H2a. We show the results in Table 4.4 and Table 4.5 below.

Table 4.4: Regression Results for information Diversity (Amazon)

Model:	MixedLM	Dependent Variable:	info_diversity
No. Observations:	1749	Method:	REML
No. Groups:	24	Scale:	201.0307
Min. group size:	49	Likelihood:	-7144.8050
Max. group size:	161	Converged:	Yes
Mean group size:	72.9		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	91.017	1.060	85.837	0.000	88.939	93.095
Experience	-0.023	0.013	-1.773	0.076	-0.047	0.002

Table 4.5: Regression Results for Information Diversity (NLNature)

Model:	MixedLM	Dependent Variable:	info_diversity
No. Observations:	12175	Method:	REML
No. Groups:	637	Scale:	1221.5859
Min. group size:	1	Likelihood:	-1350.4995
Max. group size:	2225	Converged:	Yes
Mean group size:	19.1		

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
Intercept	80.602	4.216	19.119	0.000	72.339	88.864
Experience	-0.333	0.117	-2.835	0.005	-0.563	-0.103

4.6.3 Results for Experience and Usefulness

To understand how experience affects the usefulness of contributed data, we use the helpfulness score provided in Amazon.com datasets. We compute a usefulness score using the number of people who upvoted the review.

Table 4.6: Regression Results for Usefulness (Amazon)

```
============================================================
Model:            MixedLM Dependent Variable: usefulness
No. Observations: 1749    Method:             REML
No. Groups:       24      Scale:              302.0023
Min. group size:  49      Likelihood:         -7501.1510
Max. group size:  161     Converged:          Yes
Mean group size:  72.9
-------------------------------------------------------------
              Coef.   Std.Err.    z     P>|z|  [0.025  0.975]
-------------------------------------------------------------
Intercept      9.892     1.323   7.477  0.000   7.299  12.485
Experience    -0.110     0.016  -7.023  0.000  -0.141  -0.080
============================================================
```

We see from the results that the variability accounted for in the Amazon dataset is 1.6%. z-stat is -7.023 for experience with a slope of -0.11.

4.7 Discussion

The level of experience of contributors to crowdsourcing projects will continue to increase as they participate in the project or other related projects. Using selective attention theory, we predicted that experience will negatively affect the number of mutual attributes reported, and the diversity of information contributed to integrative crowdsourcing projects. We considered two different types of integrative crowdsourcing: citizen science and online reviews.

For the datasets we analyzed, we found that the percentage of mutual attributes reported by contributors decreased in the Amazon data as contributors gained experience. However, the percentage of mutual attributes in the data increased for the citizen science project as contributors gained experience providing multiple reports of organisms. This outcome may be due to the differences in the entities reviewed. Perhaps, reporting a largely varied sub-classes of organisms under a general class (i.e., all animals or fauna in Newfoundland) such as the sightings of birds, foxes, and moose and in their unique environments prompted an increase in the number of mutual attributes reported.

Nevertheless, despite the increase in mutual attributes in the NLNature dataset, we find that contributed data in both datasets became more homogeneous with increased experience. This decrease in diversity is indicative of the effect of knowledge-driven attention allocation. Contributors focus more on the same set of attributes and report these attributes, having learned them through repeated exposure. Interestingly, even when contributors reported a higher percentage of mutual attributes in their data, the meaning of the attributes was increasingly similar as they gained experience, which may mean they are focusing on the same attributes across different entities. Also, decreasing diversity in the attributes reported is a possible indication of blindness to the variability in attributes that may exist in different instances of similar or dissimilar entities.

In online reviews, such negative effects of selective attention may be indicated by a reduction in the amount of subjective detail provided in reviews and the focus on only certain aspects of reviewed entities. If most of the reviews provided on an online review site come from contributors who are strongly affected by selective attention, potential shoppers may not have access to a sufficient breadth of perspectives beyond what the seller can provide (i.e., objective information). This decline in the subjectivity of contributed data hurts the ability for shoppers to make informed purchase decisions (Gobinath & Gupta, 2016; Korfiatis, GarcíA-Bariocanal, & SáNchez-Alonso, 2012; Li, Hitt, & Zhang, 2011). Furthermore, the results showed that experience negatively affects the helpfulness (usefulness) of crowdsourced data.

Generalizing this result, we posit that it would be difficult for experienced contributors to report data that can lead to discoveries or novel insight. The diminishing quality of crowdsourced data implies crowds do go stale, and if the rate of decline in diversity is not met or surpassed by the rate of recruitment, then crowdsourced data may eventually become misleading and harmful to potential shoppers, or erroneous and delimiting for researchers due to the tunnel vision of experienced contributors. The notion that the quality of data decreases with increasing experience, as espoused in this study, therefore, necessitates a re-evaluation of crowd hiring practices that favour experience or that suggests onetime recruitments. Table 4.7 summarizes our findings.

Table 4.7: Summary of Findings

Hypotheses	**Comments**	**Supported**
H1: The number of mutual attributes provided will decrease as contributors gain experience in a task.	While this was the case in the review dataset, in our citizen science dataset, we found that the percentage of mutual attributes reported increased. This may have something to do with the type of entity being reported about.	Partial
H2: Contributors will provide less diverse data with increasing experience	True in all cases.	Yes
H3: The usefulness of contributed data will be negatively related to experience	This was true for the Amazon dataset used to test it.	Yes

4.7.1 Recommendations

The results of this study suggest that crowdsourcers need to be put in place measures that prevent their crowds from becoming selectively attentive to only certain attributes if the crowd is expected to continue to report quality data. To guide crowdsourcers on this, we make the following recommendations:

a. Test regular for homogeneity in crowdsourced data and the need to take corrective actions to refresh the crowd when necessary, such as recruiting more contributors or assigning contributors to the reporting of information about other entities they have not previously reported.

b. Encourage the contribution of diverse data through the design of integrative crowdsourcing systems. Integrative crowdsourcing systems can be designed to be more accommodating of diverse data reducing rather than constraining contributors to provide only data that meets the crowdsourcers immediate requirements. Principles to guide such open designs are discussed in Parsons and Wand (2014).

c. Encourage contributors to participate in crowdsourcing projects that differ from their current or previous projects to reduce or eliminate the formation of inclusion rules and limit the effect of learned inattention to attributes. The literature discusses the effect of redundancy in the formation of inclusion rules and selective attention tendencies, so stymieing this tendency through non-redundancy of tasks will be beneficial to information quality.

d. Use innovative technologies like conversational agents that can ask follow-up questions from contributors about the data being provided, helping them expand on the initial answers. Such conversational agents would behave like recommendation agents or customer service bots, parsing texts entered about an entity and asking follow-up questions based on an evolving knowledgebase about the entity.

4.8 Conclusion

The goal of this study is to understand how experience affects data diversity and at the same time, investigate if contributed data becomes homogeneous over time. In this study, we answered the following questions: does the diversity of crowdsourced data decline as crowds gain experience through participation in different projects or long-term participation in a single project? If so, how does this decrease in diversity affect the quality of crowdsourced data? To answer these questions, we showed through empirical tests how increasing experience might diminish the tendency for crowd members to provide diverse data. We also showed that experience does indeed affect the perceived quality of data (using the usefulness rating as a proxy).

This study, therefore, explored the potential limitations of relying on the same crowd, particularly for projects that seek to engage crowds in discoveries or evolve to encompass unanticipated uses of data. Furthermore, since usefulness is a consequence of information quality, decreased usefulness is, therefore, an outcome of decreased information quality. However, because mutual attributes are not attributes captured by key traditional information quality dimensions like accuracy and completeness, these traditional dimensions would be inadequate in estimating the loss of quality we have identified here. These results, therefore, further validate the need for the information diversity dimension as an antecedent of information quality.

Understanding the benefits and shortcomings of engaging with the same crowds will guide crowdsourcers and crowdsourcing organizations in the development of targeted incentive strategies and more effective data collection implementations that are sensitive to the nature of the crowds involved in their projects.

4.8.1 Limitations

The limitations of this study revolve mainly around the data used. The amount of data that can be processed to test the hypotheses was constrained by hardware resources. It is difficult to make generalizations about the Amazon data for a particular group of product. Comparisons made between the Amazon and NLNature datasets are also limited in construct validity as the number of contributions used to estimate experience in Amazon data is different for NLNature. Improvement of this study will focus on using more generalizable data and cloud computing and AutoML resources to analyze large datasets for insight extensively.

References

Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science*, *22*(5), 1123–1137.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Best, C. A., Yim, H., & Sloutsky, V. M. (2013). The cost of selective attention in category learning: Developmental differences between adults and infants. *Journal of Experimental Child Psychology*, *116*(2), 105–119.

Budescu, D. V., & Chen, E. (2014). Identifying expertise to extract the wisdom of crowds. *Management Science*, *61*(2), 267–280.

Carrillo, J. E., & Gaimon, C. (2000). Improving manufacturing performance through process change and knowledge creation. *Management Science*, *46*(2), 265–288.

Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., & Specia, L. (2017). Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation. *ArXiv Preprint ArXiv:1708.00055*.

Collins, H., & Evans, R. (2007). Rethinking expertise University of Chicago Press. *Chicago IL*.

Communications, F. C. (n.d.). Poor-Quality Data Imposes Costs and Risks on Businesses, Says New Forbes Insights Report. Retrieved February 13, 2019, from Forbes website: https://www.forbes.com/sites/forbespr/2017/05/31/poor-quality-data-imposes-costs-and-risks-on-businesses-says-new-forbes-insights-report/

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201.

Cosman, J. D., & Vecera, S. P. (2013). Context-dependent control over attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(3), 836.

Cosman, J. D., & Vecera, S. P. (2014). Establishment of an attentional set via statistical learning. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(1), 1.

Denrell, J., & March, J. G. (2001). Adaptation as information restriction: The hot stove effect. *Organization Science*, *12*(5), 523–538.

Downing, P. E. (2000). Interactions between Visual Working Memory and Selective Attention. *Psychological Science*, *11*(6), 467–473.

Duan, W., Gu, B., & Whinston, A. B. (2008). The dynamics of online word-of-mouth and product sales—An empirical investigation of the movie industry. *Journal of Retailing*, *84*(2), 233–242.

Ellis, S., & Davidi, I. (2005). After-event reviews: Drawing lessons from successful and failed experience. *Journal of Applied Psychology*, *90*(5), 857.

Erickson, L., Petrick, I., & Trauth, E. (2012). *Hanging with the right crowd: Matching crowdsourcing need to crowd characteristics.* Retrieved from http://aisel.aisnet.org/amcis2012/proceedings/VirtualCommunities/3/

Forbes. (2018). 3 Barriers To Data Quality And How To Solve For Them. Retrieved May 31, 2019, from https://www.forbes.com/sites/steveolenski/2018/04/23/3-barriers-to-data-quality-and-how-to-solve-for-them/#44ded9b629e7

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135.

Gobinath, J., & Gupta, D. (2016). Online reviews: Determining the perceived quality of information. *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 412–416. https://doi.org/10.1109/ICACCI.2016.7732080

Goldstone, R. L., & Kersten, A. (2003). Concepts and categorization. *Handbook of Psychology*.

Gopnik, A., Griffiths, T. L., & Lucas, C. G. (2015). When Younger Learners Can Be Better (or at Least More Open-Minded) Than Older Ones. *Current Directions in Psychological Science*, *24*(2), 87–92. https://doi.org/10.1177/0963721414556653

Gura, T. (2013). Citizen science: Amateur experts. *Nature*, *496*(7444), 259–261. https://doi.org/10.1038/nj7444-259a

Harnad, S. (2005). To cognize is to categorize: Cognition is categorization. *Handbook of Categorization in Cognitive Science*, 20–45.

He, R., & McAuley, J. (2016). Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. *Proceedings of the 25th International Conference on World Wide Web*, 507–517. International World Wide Web Conferences Steering Committee.

Honnibal, M., & Johnson, M. (2015). An Improved Non-monotonic Transition System for Dependency Parsing.

Hunter, J., Alabri, A., & van Ingen, C. (2013). Assessing the quality and trustworthiness of citizen science data. *CONCURRENCY AND COMPUTATION-PRACTICE & EXPERIENCE*, *25*(4, SI), 454–466. https://doi.org/10.1002/cpe.2923

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194. https://doi.org/10.1038/35058500

Kamp, H. (2013). Meaning and the Dynamics of Interpretation: Selected Papers of Hans Kamp. BRILL.

Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal*, *45*(6), 1183–1194.

Katsuki, F., & Constantinidis, C. (2014). Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. *The Neuroscientist*, *20*(5), 509–521. https://doi.org/10.1177/1073858413514136

Keil, F. C. (1989). *Concepts, kinds, and conceptual development*. Cambridge, MA: MIT Press.

Kim, S., & Rehder, B. (2009). Knowledge effect the selective attention in category learning: An eyetracking study. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, 230–235. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.176.100&rep=rep1&type=pdf

Kim, ShinWoo, & Rehder, B. (2011). How prior knowledge affects selective attention during category learning: An eyetracking study. *Memory & Cognition*, *39*(4), 649–665.

Kiwelekar, A. W., & Joshi, R. K. (2010). An Object-Oriented Metamodel for Bunge-Wand-Weber Ontology. *ArXiv Preprint ArXiv:1004.3640*.

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.

Korfiatis, N., GarcíA-Bariocanal, E., & SáNchez-Alonso, S. (2012). Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, *11*(3), 205–217.

Leonard, D., & Sensiper, S. (1998). The role of tacit knowledge in group innovation. *California Management Review*, *40*(3), 112–132.

Li, X., Hitt, L. M., & Zhang, Z. J. (2011). Product reviews and competition in markets for repeat purchase products. *Journal of Management Information Systems*, *27*(4), 9–42.

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. *MIT Sloan Management Review*, *51*(3), 21.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, *2*(1), 71–87.

McAuley, J. J., & Leskovec, J. (2013). From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews. *World Wide Web*, 897–908. Retrieved from http://dl.acm.org/citation.cfm?id=2488466

McAuley, J., & Leskovec, J. (2013). Hidden factors and hidden topics: Understanding rating dimensions with review text. *Proceedings of the 7th ACM Conference on Recommender Systems*, 165–172. ACM.

Minnema, A., Bijmolt, T. H., Gensler, S., & Wiesel, T. (2016). To keep or not to keep: Effects of online customer reviews on product returns. *Journal of Retailing*, *92*(3), 253–267.

Morris, M. W., & Moore, P. C. (2000). The lessons we (don't) learn: Counterfactual thinking and organizational accountability after a close call. *Administrative Science Quarterly*, *45*(4), 737–765.

Mudambi, S. M., & Schuff, D. (2010). Research Note: What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com. *MIS Quarterly*, *34*(1), 185–200. https://doi.org/10.2307/20721420

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243.

Parsons, J. (1996). An information model based on classification theory. *Management Science*, *42*(10), 1437–1453.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, *10*(5), 233–238.

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of Selective Attention: When Children Notice What Adults Miss. *Psychological Science*, 956797617693005. https://doi.org/10.1177/0956797617693005

Postle, B. R., Awh, E., Jonides, J., Smith, E. E., & D'Esposito, M. (2004). The where and how of attention-based rehearsal in spatial working memory. *Cognitive Brain Research*, *20*(2), 194–205.

Postle, Bradley R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience*, *139*(1), 23–38.

Poston, R. S., & Speier, C. (2005). Effective use of knowledge management systems: A process model of content ratings and credibility indicators. *MIS Quarterly*, 221–244.

Richards, J. E., & Turner, E. D. (2001). Extended visual fixation and distractibility in children from six to twenty-four months of age. *Child Development*, *72*(4), 963–972.

Roese, N. J., & Olson, J. M. (1995). Functions of counterfactual thinking. *What Might Have Been: The Social Psychology of Counterfactual Thinking*, 169–197.

Rosemann, M., & Green, P. (2002). Developing a meta model for the Bunge–Wand–Weber ontological constructs. *Information Systems*, *27*(2), 75–91. https://doi.org/10.1016/S0306-4379(01)00048-5

Rust, R. T., Inman, J. J., Jia, J., & Zahorik, A. (1999). What you don't know about customer-perceived quality: The role of customer expectation distributions. *Marketing Science*, *18*(1), 77–92.

Ryan, J. D., Althoff, R. R., Whitlow, S., & Cohen, N. J. (2000). Amnesia is a deficit in relational memory. *Psychological Science*, *11*(6), 454–461.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Sahoo, N., Dellarocas, C., & Srinivasan, S. (2018). The impact of online product reviews on product returns. *Information Systems Research*.

Schwartzstein, J. (2014). Selective Attention and Learning: Selective Attention and Learning. *Journal of the European Economic Association*, *12*(6), 1423–1452. https://doi.org/10.1111/jeea.12104

Sitkin, S. B. (1992). Learning through failure: The strategy of small losses. *Research in Organizational Behavior*, *14*, 231–266.

Sloutsky, V. M. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences*, *7*(6), 246–251.

Smith, L. B., Kemler, D. G., & Aronfreed, J. (1975). Developmental trends in voluntary selective attention: Differential effects of source distinctness. *Journal of Experimental Child Psychology*, *20*(2), 352–362. https://doi.org/10.1016/0022-0965(75)90109-5

Smith, M. F. (1999). Urban versus suburban consumers: A contrast in holiday shopping purchase intentions and outshopping behavior. *Journal of Consumer Marketing*, *16*(1), 58–73.

Strutt, G. F., Anderson, D. R., & Well, A. D. (1975). A developmental study of the effects of irrelevant information on speeded classification. *Journal of Experimental Child Psychology*, *20*(1), 127–135.

Todd, P., & Benbasat, I. (2000). Inducing compensatory information processing through decision aids that facilitate effort reduction: An experimental assessment. *Journal of Behavioral Decision Making*, *13*(1), 91–106.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.

Wand, Y., Storey, V. C., & Weber, R. (1999). An Ontological Analysis of the Relationship Construct in Conceptual Modeling. *ACM Trans. Database Syst.*, *24*(4), 494–528. https://doi.org/10.1145/331983.331989

Weathers, D., Sharma, S., & Wood, S. L. (2007). Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. *Journal of Retailing*, *83*(4), 393–401.

Weigelhofer, G., & Pölz, E.-M. (2016). Data quality in citizen science projects: Challenges and solutions. *Front. Environ. Sci. Conference Abstract: Austrian Citizen Science Conference 2016*.

Well, A. D., Lorch, E. P., & Anderson, D. R. (1980). Developmental trends in distractibility: Is absolute or proportional decrement the appropriate measure of interference? *Journal of Experimental Child Psychology*, *30*(1), 109–124.

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. *2011 IEEE Seventh International Conference on E-Science Workshops*, 14–19. https://doi.org/10.1109/eScienceW.2011.27

Wood, S. L., & Lynch, J. G. (2002). Prior knowledge and complacency in new product learning. *Journal of Consumer Research*, *29*(3), 416–426.

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *33*(2), 363.

Yang, R., Xue, Y., & Gomes, C. (2018). Pedagogical Value-Aligned Crowdsourcing: Inspiring the Wisdom of Crowds via Interactive Teaching.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychological Science*, 0956797612460407.

Chapter Five: **Conclusion and Expected Contributions**

The value of data is in the insight it provides. Insights derived from data have become a significant source of competitive advantage for organizations today (Chen, Chiang, & Storey, 2012; LaValle, Lesser, Shockley, Hopkins, & Kruschwitz, 2011). Organizations derive insight for data-driven decision-making by repurposing data using analytics (Woodall, 2017). Crowdsourced repurposable data is "diverse data" consisting of different contributor views and lends itself to being used in different contexts (Ogunseye & Parsons, 2018). However, collecting diverse data does not just stand to provide useful insights; it can also reduce the resources expended on repeat data collection or acquisition due to changing data requirements.

In the words of Peter Drucker, "What gets measured, gets managed" (Willcocks & Lester, 1996, p. 280); managing crowdsourcing processes to generate repurposable data begins with being able to measure the amount of diversity in data. Research and practice favour measuring the quality of data using key dimensions such as accuracy and completeness. This thesis theoretically explored the limitations of conventional information quality in measuring diversity and the insightfulness of data. We identified the limitations of traditional information quality as non-generalizability, over-dependence on contributor knowledge, and the lack of a metric for diversity.

Consequently, we extended the dimensions of information quality to include information diversity – the number of unique attributes contained in a dataset. Furthermore, we developed a framework for collecting diverse data through integrative crowdsourcing systems. The framework was validated using two cases from the literature with the integral ingredients for information diversity identified as accommodating IT infrastructure, system design, and human factors.

However, because crowdsourcers may be wary of strategies that allow contributors to provide data without restrictions for concerns about accuracy and completeness, we proceeded to explore the consequences of seeking diverse data on traditional information quality and vice versa. Since crowdsourcers rely on the knowledge of contributors for information quality (Wiggins et al., 2011), we investigated how training crowds or the recruitment of experienced contributors affect information quality dimensions, including information diversity.

Recruitment based on knowledge implies smaller crowds, fewer data sources, and a consistent effort to keep these contributors motivated through different stages of participation in a crowdsourcing project (Lee, Crowston, Østerlund, & Miller, 2017). It may also result in fewer perspectives being represented in crowdsourced data and fewer people getting the chance to learn about crowdsourcing projects, especially citizen science projects, which have education as one of their core tenets. In this thesis, we questioned the necessity of knowledge for the collection of high-quality data. By synthesizing existing literature on classification and selective attention, we showed that while this strategy is expedient for survival and the management of our mental resources, it makes it difficult to learn and make discoveries as knowledge increases, as we would rather assimilate than accommodate. In fact, "as our knowledge grows, we become less open to new ideas"

(Gopnik et al. 2015 p. 87), which means we are also less likely to produce data that leads to new ideas.

We extend these psychology theories to the domain of data quality in crowdsourcing. These studies have targeted two questions: (1) How do the different levels and types of contributor knowledge affect the quality of crowdsourced data? and (2) How does experience affect the quality of crowdsourced data? To answer these questions, we conducted two studies using a laboratory experiment and existing real-life datasets, respectively.

The results from our experiment strongly suggest that restricting participation in crowds through training has adverse consequences for the diversity and quality of information contributed to crowdsourcing projects. Trained contributors have a greater tendency to only focus on aspects of a stimulus that are congruent with their existing knowledge. In contrast, untrained contributors not only report accurate data; they also report diverse data about both primary and secondary stimuli in their visual fields. Chiefly, training did not advantage trained contributors in terms of the accuracy of attributes reported about a target entity but disadvantaged them when it came to reporting diverse data about. The level of a contributor's knowledge also negatively affected the completeness and diversity of contributed data.

In our analyses of secondary data from an online review system and a citizen science platform, we found that increasing experience resulted in selective attention to only specific attributes of diverse entities for which data was reported. Contrary to widespread assumptions about the benefits of experience as seen in Amazon Mechanical Turk, which pays experienced high performing contributors a premium over new crowd-members, increasing experience hurts informativeness and the usefulness of crowdsourced data.

We conclude that recruitment operations should ensure that people with different types and levels of knowledge can participate in crowdsourcing tasks, bringing their diverse attention allocation capabilities and prior knowledge (or lack thereof) to bear for the capturing of multidimensional, repurposable, and high-quality data.

5.1 Contribution to Theory

In the IS literature, the interaction among dimensions of data quality has been mainly investigated from the perspective of presented data, as seen on websites and e-commerce platforms (DeLone & McLean, 1992; Wixom & Todd, 2005; Xu, Benbasat, & Cenfetelli, 2013). This thesis increases our knowledge of the interactions between data quality dimensions from the contributed data perspective, which is more relevant to crowdsourcing and other crowd-facing systems than to e-commerce platforms. Unlike previous studies, this thesis considers the multidimensionality of contributor knowledge in crowdsourcing by looking at the levels and types of contributor knowledge and how they affect the goal of integrative crowdsourcing – the collection of high-quality data.

This thesis extends the theory of selective attention to the explanation and prediction of the effect of knowledge on contributed data. To the best of our knowledge, it is the first study to examine the effect of selective attention on crowds and on the diversity of contributed data. It uses the components described in Wickens & McCarley (2008) to explain how expectations contributors have about the attributes of a phenomenon of interest and the value ascribed to these attributes can inform how they allocate their attention and consequently the information reported about the phenomenon. At the same time, when contributors are not guided by knowledge, they are more open to learning the attributes of the phenomenon. The characteristics of the attributes of a phenomenon, including the amount of effort required to observe these attributes, dictate what is reported about the phenomenon.

Furthermore, we learned from the studies in this thesis that while attribute salience and effort may control the allocation of attention when contributors have little or no knowledge of the phenomenon , these salient attributes themselves eventually become a source of selective attention, as contributors begin to expect and value them for the classification of future instances of the phenomenon.

The thesis, therefore, provides predictive theory (Gregor, 2006) about how contributors will perform tasks related to providing information about a phenomenon, in the short term and long term. It emphasizes that crowdsourcers will need to counterbalance the tendency for knowledgeable contributors to report about only attributes of phenomenon aligning with their prior knowledge with the tendency for less knowledgeable contributors to report salient attributes requiring minimal effort to observe and vice versa.

In the context of online reviews, selective attention can skew what knowledgeable contributors focus on an entity. If the attribute of an entity aligns with their expectations, then they may provide information mainly concerning that aspect of the phenomenon. Also, if the attributes of an entity cause cognitive dissonance in the contributor, where changes in the attributes of a phenomenon are completely tangential to expected values, then contributors may also focus on reporting about theses disconcerting attributes.

Finally, findings from this thesis creates empirically justified descriptive knowledge (Gregor & Hevner, 2013) that describes the theoretical factors that lead to the unexplained results in the literature about why less knowledgeable contributors report data that is as accurate as that reported by more knowledgeable contributor (see examples in Austen, Bindemann, Griffiths, & Roberts, 2016; Escoffier & McKelvey, 2015).

5.2 Contribution to Practice

This thesis goes beyond existing studies and seeks to provide theory-driven empirical evidence for how and why contributors differ and what to expect in terms of the quality of the data they provide. Insights from this thesis will help guide crowdsourcers on how to design crowdsourcing processes, especially the recruitment decisions suitable for particular project conditions. For example, when classification tasks can be automated, then crowdsourcers would be better served if they open their projects to everyone as untrained contributors like trained contributors can provide diagnostic attributes needed for classification. However, when classification is to be done by contributors, then implicitly trained contributors provide more diverse data than explicitly trained contributors with minimal sacrifice of accuracy. The thesis thus provides prescriptive knowledge about how to acquire high-quality repurposable data, providing an empirically validating a framework to accomplish this (Gregor & Hevner, 2013). It can serve as a source of kernel theory for the development of integrative crowdsourcing information systems.

On the whole, the thesis answers questions about why information diversity is needed, how knowledge acquired through training or experience can affect information diversity, and how information diversity fits into the information quality framework. The thesis will increase the inclusiveness of crowdsourcing, motivating the consideration of humans' natural tendency for error, and selective attention in the design crowdsourcing system (Reason, 1990).

To ensure the dissemination of the theory in the thesis to practitioners and lay users of crowdsourcing projects, we hope to publish the findings of this research in practitioner-focused outlets. We may also seek collaboration with organizations around their data collection tasks to showcase the benefits of the information diversity dimension for information quality and data repurposing.

References

Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. *Scientific Reports*, *6*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5028888/

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, *36*(4).

DeLone, W. H., & McLean, E. R. (1992). Information systems success: The quest for the dependent variable. *Information Systems Research*, *3*(1), 60–95.

Escoffier, N., & McKelvey, B. (2015). The Wisdom of Crowds in the Movie Industry: Towards New Solutions to Reduce Uncertainties. *International Journal of Arts Management*, *17*(2), 52.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly*, 611–642.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2).

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, *52*(2), 21.

Lee, T. K., Crowston, K., Østerlund, C. S., & Miller, G. (2017). Recruiting Messages Matter: Message Strategies to Attract Citizen Scientists. *CSCW Companion*, 227–230. Retrieved from https://citsci.syr.edu/sites/crowston.syr.edu/files/cpa143-leeA.pdf

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.

Ogunseye, S., & Parsons, J. (2018). Designing for Information Quality in the Era of Repurposable Crowdsourced User-Generated Content. *International Conference on Advanced Information Systems Engineering*, 180–185. Springer.

Reason, J. (1990). *Human Error* (1 edition). Cambridge England; New York: Cambridge University Press.

Wiggins, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. *2011 IEEE Seventh International Conference on E-Science Workshops*, 14–19. https://doi.org/10.1109/eScienceW.2011.27

Willcocks, L., & Lester, S. (1996). Beyond the IT productivity paradox. *European Management Journal*, *14*(3), 279–290.

Wixom, B. H., & Todd, P. A. (2005). A theoretical integration of user satisfaction and technology acceptance. *Information Systems Research*, *16*(1), 85–102.

Woodall, P. (2017). The Data Repurposing Challenge: New Pressures from Data Analytics. *Journal of Data and Information Quality (JDIQ)*, *8*(3–4), 11.

Xu, J. D., Benbasat, I., & Cenfetelli, R. T. (2013). Integrating service quality with system and information quality: An empirical test in the e-service context. *Mis Quarterly*, *37*(3), 777–794.

APPENDIX A: PRETEST

Twelve students from the Department of Biology at Memorial University of Newfoundland participated in the pretest exercise. The pretest was designed using a 2x3 factorial design. First, participants were randomly assigned to either a complex rule condition or a simple rule condition to help us understand the potential effect of the complexity of the inclusion rule on the participants' ability to report accurate and complete information about the artificial stimuli. Participants in both conditions were presented one of two different inclusion rules to enable them to perform the identification task. Participants in the simple rule condition were presented with a rule that only had two attributes: A tyran is an insect that has 2 or 3 buttons on a light-blue body. The participants in the complex rule condition were presented with a five-attribute inclusion rule: A tyran is an insect with 2 or 3 buttons on its light blue body, 1 or 2 rings on each of its blue wings, a short curly tail.

Secondly, we tested three different question phrasing:

1. Report your sighting
2. Is this a tyran? Yes___ No___ What is the reason for your response?
3. What do you see?

We asked for written and verbal feedback. We found through the exercise that the four participants who had question type 2 regardless of the complexity of the condition described only the target stimuli and ignored the other stimuli present in some of the images they saw. These four participants reported the inclusion rules they have learned. We considered that the question might be too direct, priming the participants to fixate on whether a stimulus is a tyran or not. All four participants, who got question 1, reported about the tyran and its interaction with other entities. Entities that were not interacting with the tyran was ignored. The question, though not as direct as question 2, appeared to cause the participant to fixate on the tyran and its activities. Three participants interpreted the question as "do you see a tyran?" and responded as such. For example, one participant stated, *"tyran with button aligned vertically."* They viewed the other entities in the presented images as distractors, put there to impede their sighting of the target stimulus. Only one participant reported the other entities in the images. Nonetheless, this one participant reported only animate entities and ignored the inanimate entities in the images. Such entities like fence, sky or table were ignored and not considered a part of "their sighting" whether or not the tyran and other insects were interacting with it. However, the third question, "what do you see" was the most inclusive. Of the four participants who got question 3, three of them listed every other stimulus present and one the last one listed the inclusion rules they have learned. One of the participants wrote in the feedback question about the clarity of the instruction, *"I was unsure if I was being asked to describe the variations in the tyran or the entire scene*." Nonetheless, this participant reported all the entities available in the picture and their interaction.

Finally, we found that the complexity of the rule did not have any effect on the participants' ability to identify the stimulus. However, the simple rule had a negative effect on the participants' perception of the task. Some participants in the simple rule condition searched for additional diagnostic attributes as they considered the rule too simple and unrealistic. A participant in the simple rule condition stated

"It was unclear as to whether the orientation of the buttons mattered. If the orientation of the button doesn't matter, it might be a little too easy … Insect ID in the field can be complicated."

Another participant in the simple rule condition stated,

"Possibly test if other features of the insect affects whether or not they get marked as a Tyran or not, such as wings, tail, and antennae, while still having the dots."

Participants in the simple rule condition also stated that the task took too long, and they were "bored."

Based on the outcome of this pretest, we adjusted the materials eliminating the simple rule condition.

To test these modifications, we carried out another test with fifteen students from the faculty of business as part of a Business Research Experience Course. Students participated for course credit. Using twenty images, including four catch items, we tested for a suitable time for completing the test. We tried 50 seconds, 40 seconds and 30 seconds and found 40 seconds to be the most suitable across the three groups.

Other changes made at this stage included improving the images to systematically test for accuracy, completeness as well as diversity, collecting biographic information from the participants and refining the recruitment information to be more attractive to our target audience.

The new experimental material is presented in Appendix B.

APPENDIX B: EXPERIMENT PROTOCOL

B.1. Description:

Participants will be randomly assigned to three groups: trained, implicit learning, and untrained. For the first group, the inclusion rule – how to identify a tyran – will be provided to them. The second group will be provided with sample tyrans so they can deduce the inclusion rule themselves

The third group will not be trained, and no sample will be provided to them

The total number of images that will be presented for the test is 20. This does not include images used in the learning stage by the untrained contributors. The current experiment will require 15 minutes to complete. 5 minutes for training or learning and 10 minutes for the test. For the untrained group, the experiment should take 12-13 minutes in total.

B.2. Images Presented and Coding Schemes

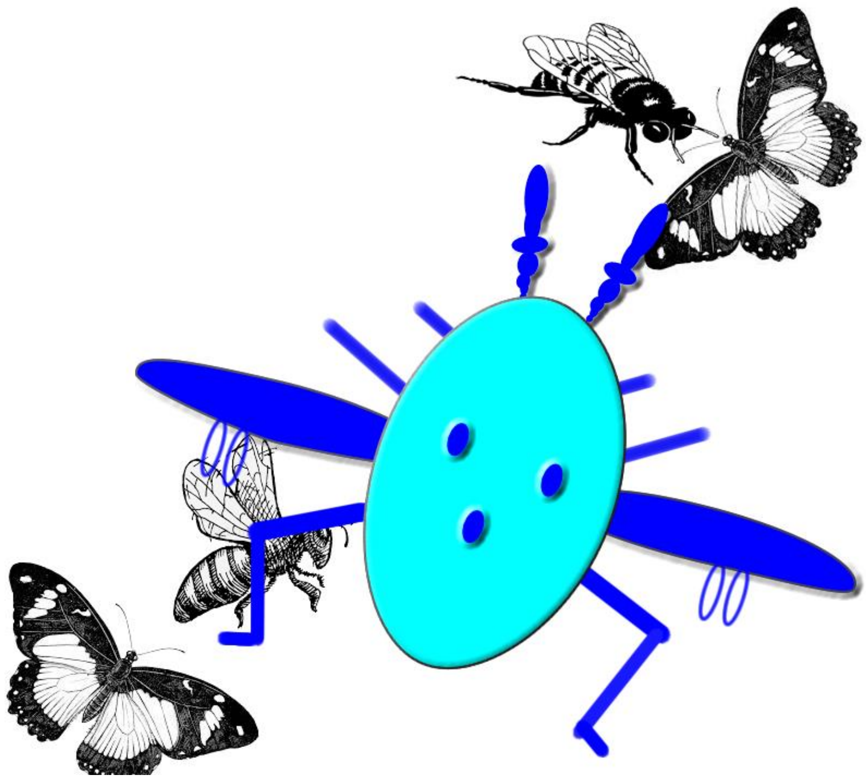


Figure B.1: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it lacks a (short) tail

The ideal contributor will report that it lacks a tail, report about the presence of 6 legs and give some details about the four insects in the picture. They may also choose to describe the intrinsic and mutual attributes of the secondary stimuli too

Coding Scheme:

Non-tyran   Reason: lacks a short tail

Attributes: 2 long blue wings, 2 antennae, 4 legs, 3 buttons, light blue body, slightly tilted, 2 rings each.

Housefly interacting with butterfly. A bee and a butterfly underneath the non-tyran's legs

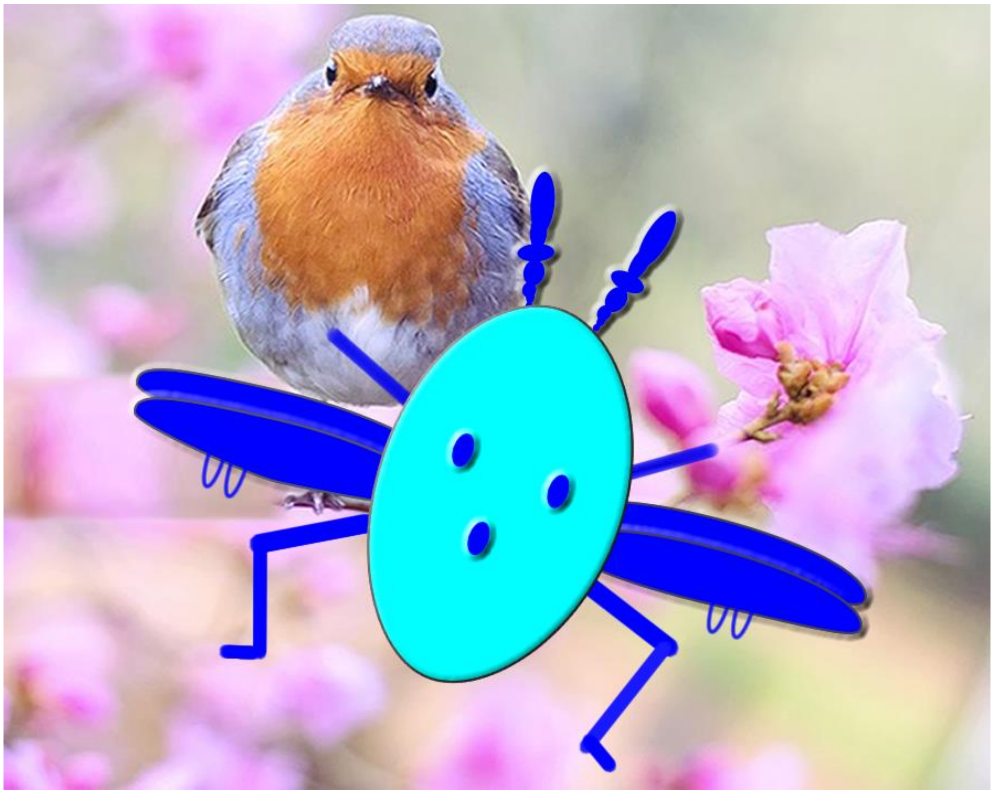


Figure B.2: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it lacks a short tail

The ideal contributor will report that it lacks a tail and mention all the attributes of the stimulus. They should also report the presence of 4 wings and mention the presence of a bird and flowers. They may also choose to describe the intrinsic attributes of the stimuli

Coding Scheme:

Non-tyran Reason: lacks a short tail

Attributes: 4 long blue wings, 2 antennae, 4 legs, 3 buttons, light blue body, slightly tilted. A grey and orange-coloured bird behind. In a bush of pink flowers.

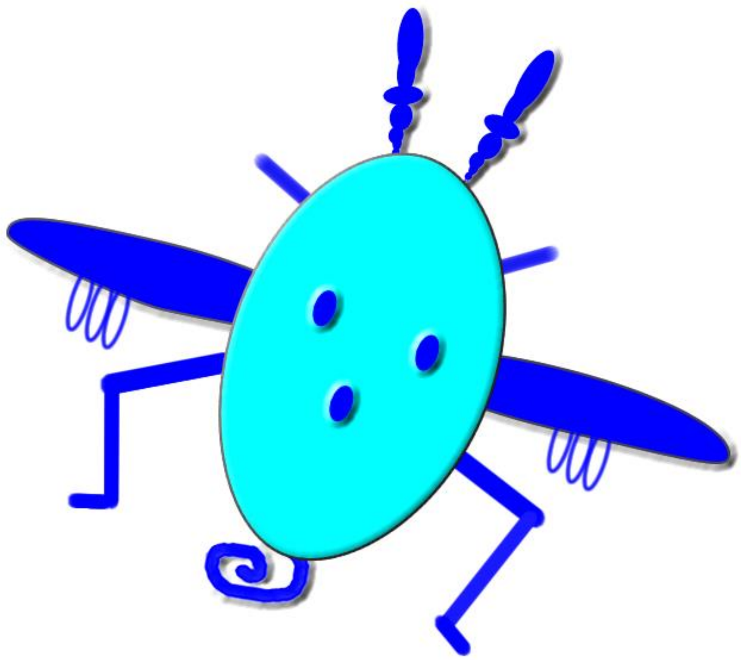


Figure B.3: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it has three rings. The ideal contribution will include details about the attributes of the stimulus.

Coding Scheme:

Non-tyran Reason: because it has three rings

Attributes: 2 long blue wings, 2 antennae, 4 legs, 3 buttons, light blue body, slightly tilted, short blue tail 3, rings.

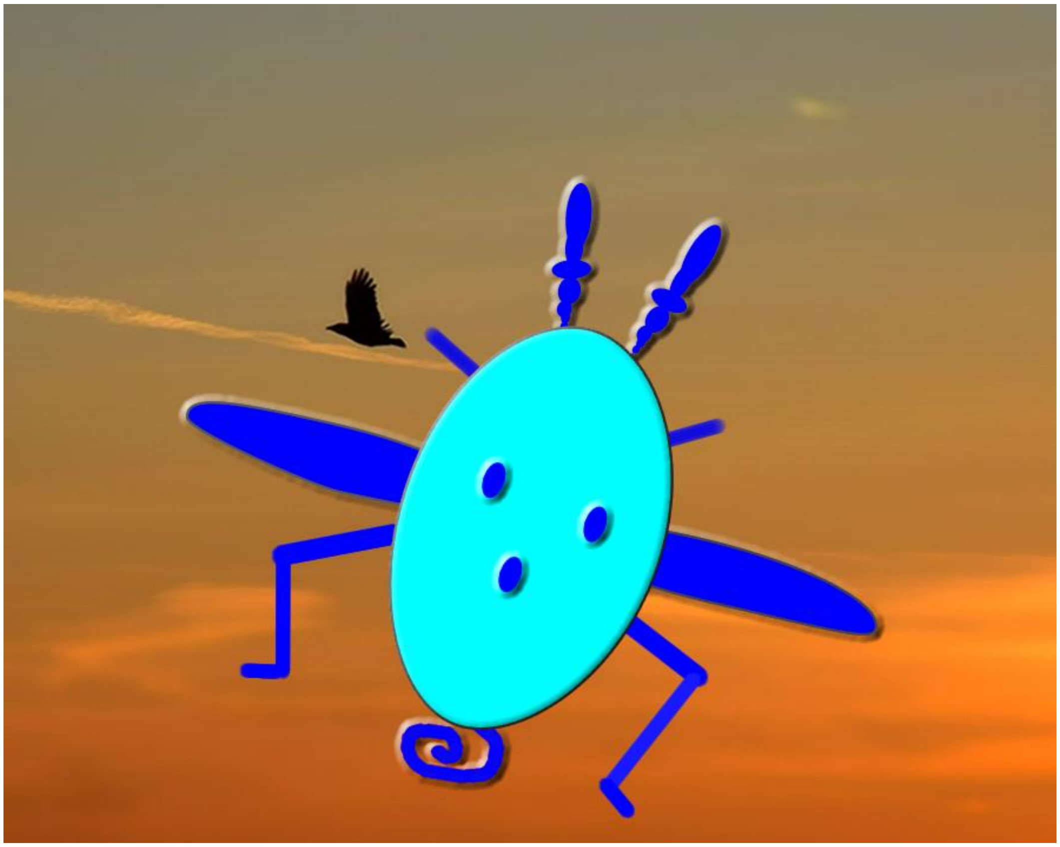


Figure B.4: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it lacks rings on its wings

The ideal contribution will report the lack of rings and mention the evening sky and the bird flying by. It would also provide details about the attributes of the stimulus, whether essential or not.

Coding Scheme:

Non-tyran Reason: because it has no rings

Attributes: 2 long blue wings, 2 antennae, 4 legs, 3 buttons, light blue body, slightly tilted, short blue tail 3, no rings. The non-tyran is flying, and a black bird is flying behind in the night sky.

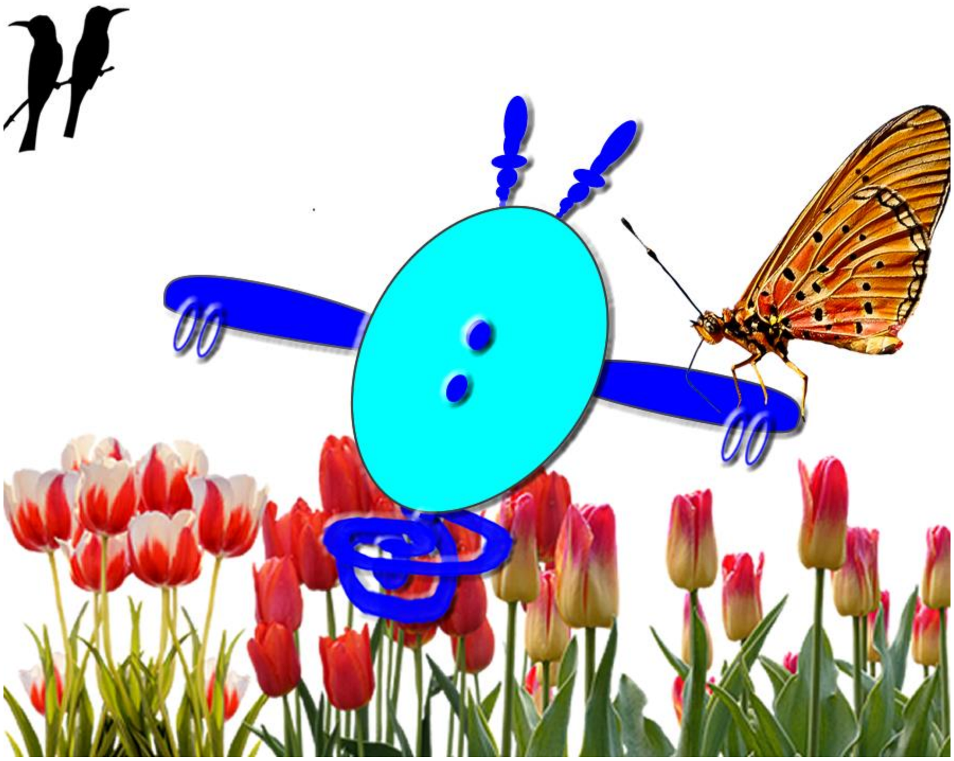


Figure B.5: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it lacks a short tail

The ideal contribution will report that it lacks a short tail. It should also report the presence of 2 birds on a twig/tree and mention the presence of a butterfly on the wing of the tyran and flowers

Coding Scheme:

Non-tyran   Reason: Long tail

Attributes: 2 long blue wings, 2 antennae, 0 legs, 2 buttons, light blue body, slightly tilted, short blue tail 3, rings.

The non-tyran is in a bush of red and white and pink and yellow flowers with an orange and black speckled butterfly on its right wing, apparently feeding. Two black birds (looking like ravens) are on a twig behind.

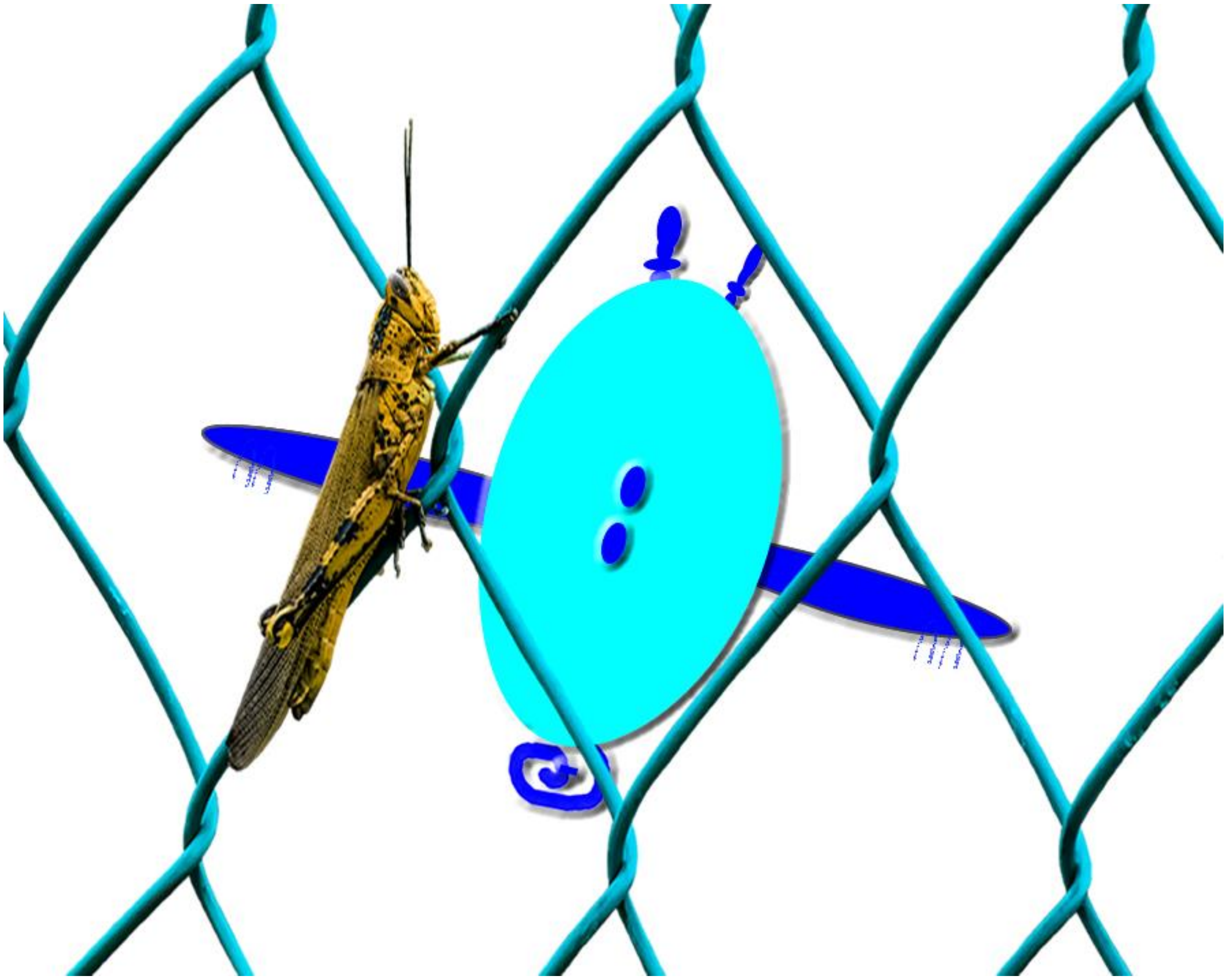


Figure B.6: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it lacks rings on its wings

The ideal contribution will report the lack of rings and discuss the fence and grasshopper in the picture. It should also mention the shortness of its antennae.

Coding Scheme:

Non-tyran Reason: No rings

Attributes: 2 long blue wings, 2 short antennae, 0 legs, 2 buttons, light blue body, slightly tilted, short blue tail 0 rings instead has some hairy features where rings should be.

The non-tyran is behind a light-blue fence. A yellow and black locust is on the fence left side of the non-tyran.

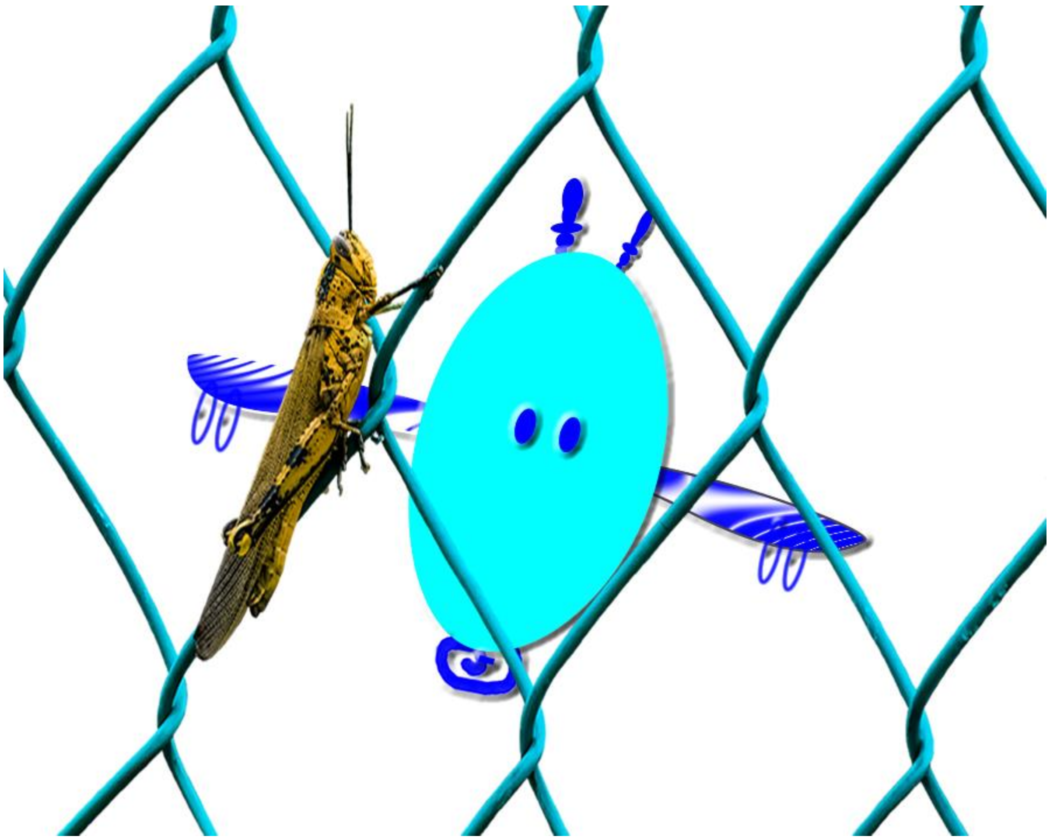


Figure B.7: Tests for knowledge of the inclusion rule

This stimulus is not a typical tyran because it lacks solid blue wings

The ideal contribution will report the colour of the wings and discuss the fence and grasshopper in the picture. It should mention all the essential and non-essential attributes of the stimuli

Coding Scheme:

Non-tyran Reason: Wings has white stripes

Attributes: 2 long white and blue striped wings, 2 short antennae, 0 legs, 2 buttons, light blue body, slightly tilted, short blue tail, and 2 rings.

The non-tyran is behind a light-blue fence. A yellow and black locust is on the fence left side of the non-tyran

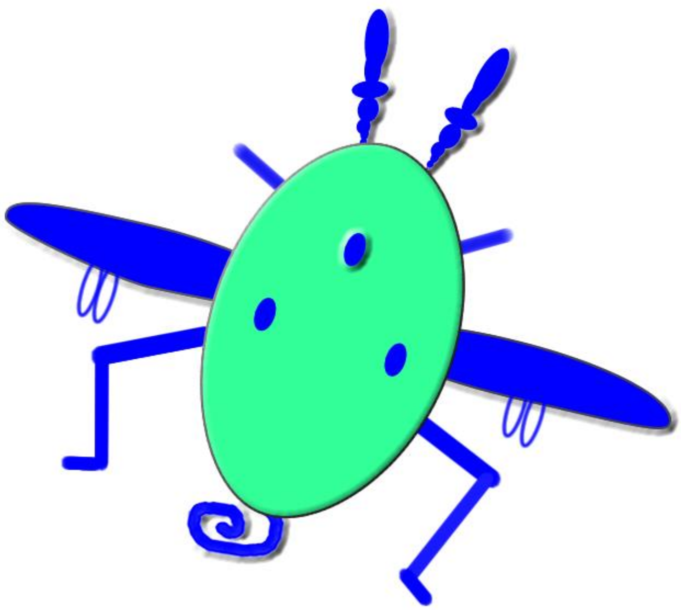


Figure B.8: Tests for knowledge of the inclusion rule

This stimulus is not a tyran because it lacks a light blue body

The ideal contribution will report the colour of the body, and mention the other attributes

Coding Scheme:

Non-tyran Reason: green body

Attributes: 2 long blue wings, 2 antennae, 4 legs, 2 buttons, green body, slightly tilted, short blue, and tail 2 rings.

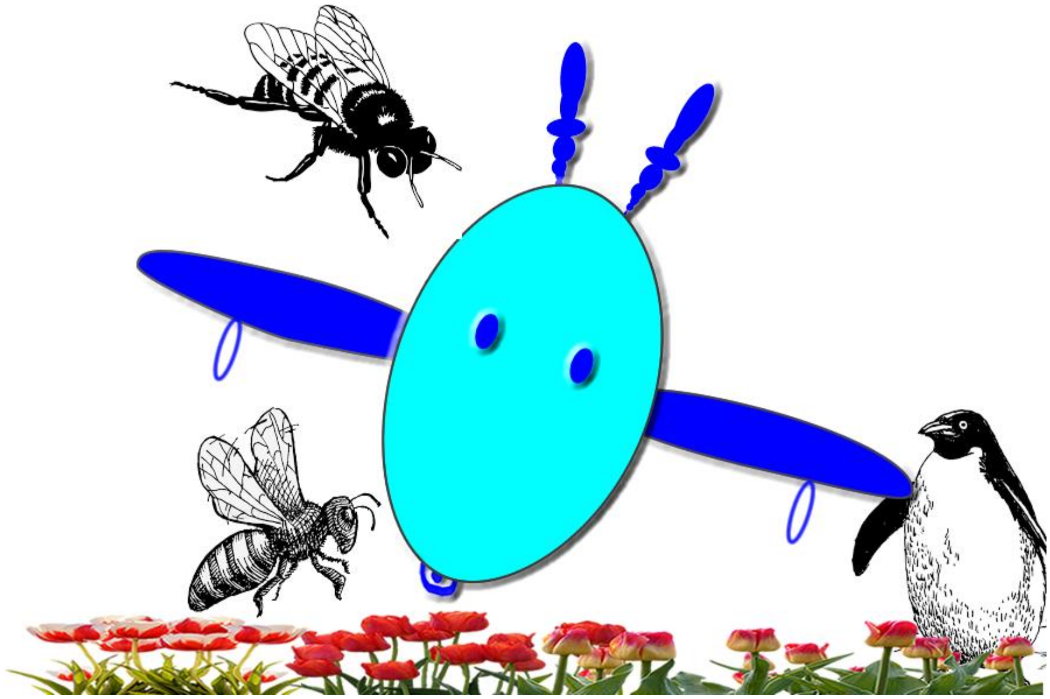


Figure B.9: Tests for selective attention to rules only

This stimulus is a tyran with a very short tail

The ideal contribution will report the shortness of its tail and discuss the other entities in the picture. It should also mention the attributes of the tyran

Coding Scheme:

Tyran Reason:

Attributes: 2 long blue wings, 2 antennae, 0 legs, 2 buttons, light blue body, slightly tilted, very short blue tail 1 rings. The tyran is hovering over a bush of flowers surrounded by a penguin, a bee and a housefly. The housefly is descending towards the tyran as if to attack it.red and yellow flowers

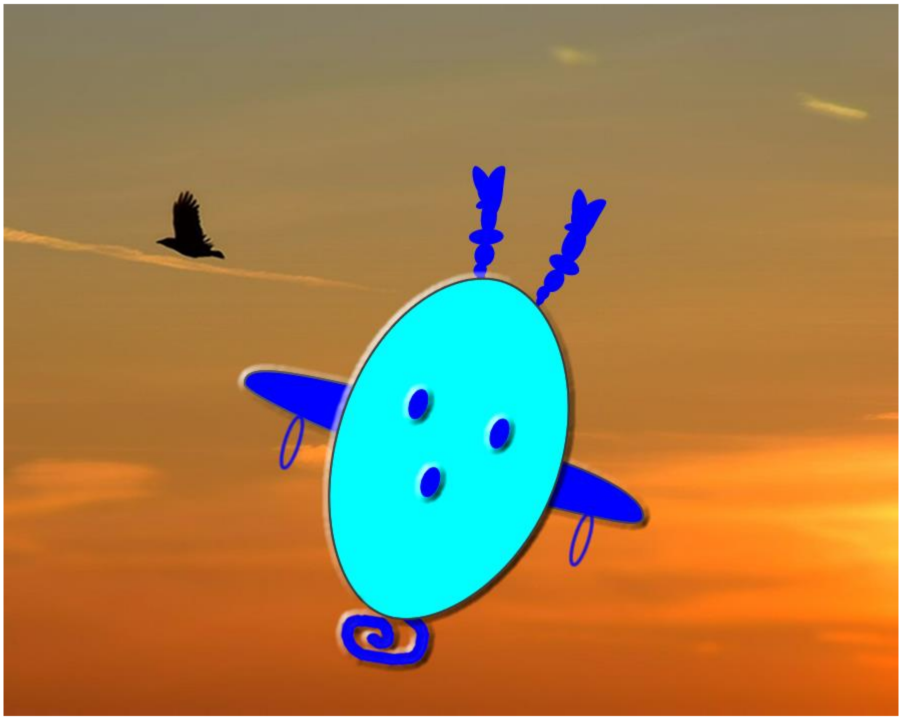


Figure B.10: Tests for selective attention to rules

This stimulus is a tyran with short wings

The ideal contribution will report the shortness of the tyran's wings and discuss the other entities in the picture. It should also mention the attributes of the tyran like the split-end antennae

Coding Scheme:

Tyran Reason:

Attributes: 2 short blue wings, 2 antennae with 2 lobbed end each, 0 legs, 3 buttons, light blue body, slightly tilted, short blue tail 1 ring each. The non-tyran is flying, and a black bird is flying behind in the night sky.

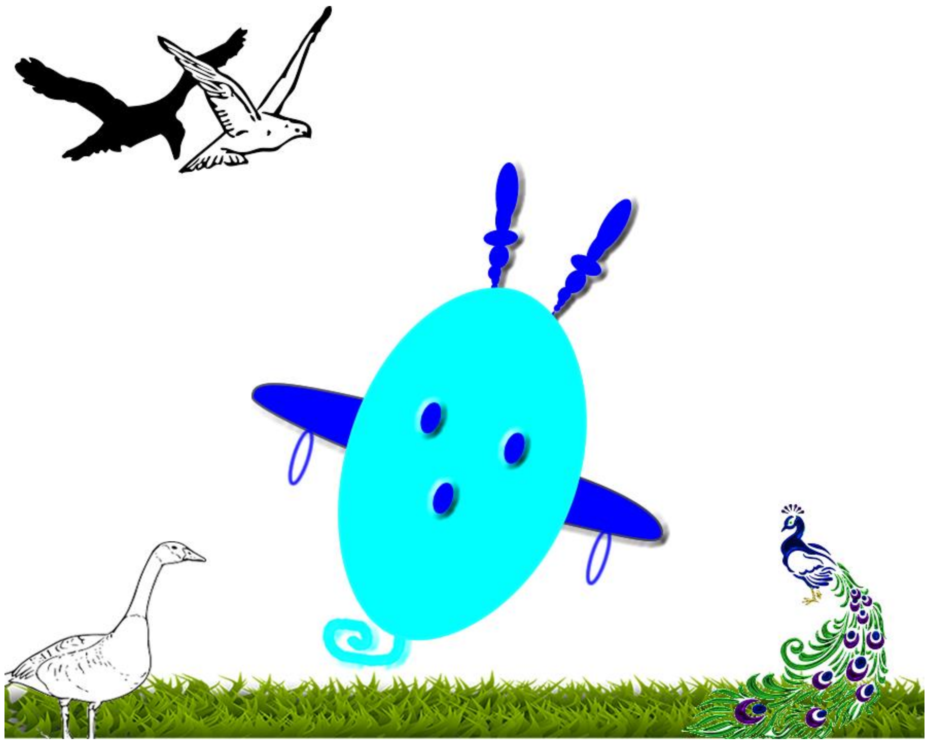


Figure B.11: Tests for selective attention to rules

This stimulus is a tyran with a light blue tail

The ideal contribution will report the colour of its tail and discuss the other entities in the picture. It should also mention the attributes of the tyran.

Coding Scheme:

Tyran Reason:

Attributes: 2 short blue wings, 2 antennae, 0 legs, 3 buttons, light blue body, slightly tilted, short light blue tail, 1 ring each. The tyran is flying over a field of grass. A peacock and a white goose approaching on its right and left, respectively. Two birds, one black and another white, appear to be descending or flying bye in the sky.

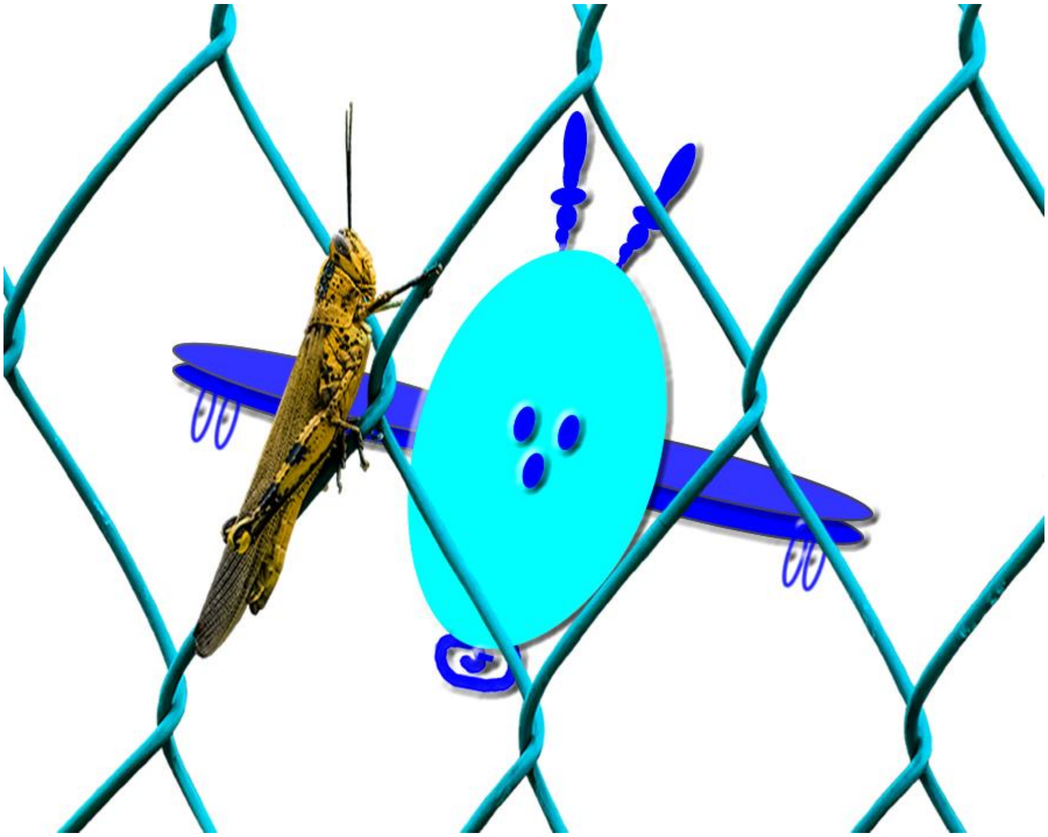


Figure B.12: Tests for selective attention to rules

This stimulus is a non-tyran with four wings

The ideal contribution will report the presence of 2 extra wings and discuss the other entities in the picture. It should also mention the attributes of the tyran

Coding Scheme:

Non-tyran Reason: insufficient rings for all the wings

Attributes: 4 blue wings, 2 antennae, 0 legs, 3 buttons, light blue body, slightly tilted, short blue tail 2 rings each. The tyran is behind a light-blue fence. A yellow and black locust is on the fence left side of the tyran

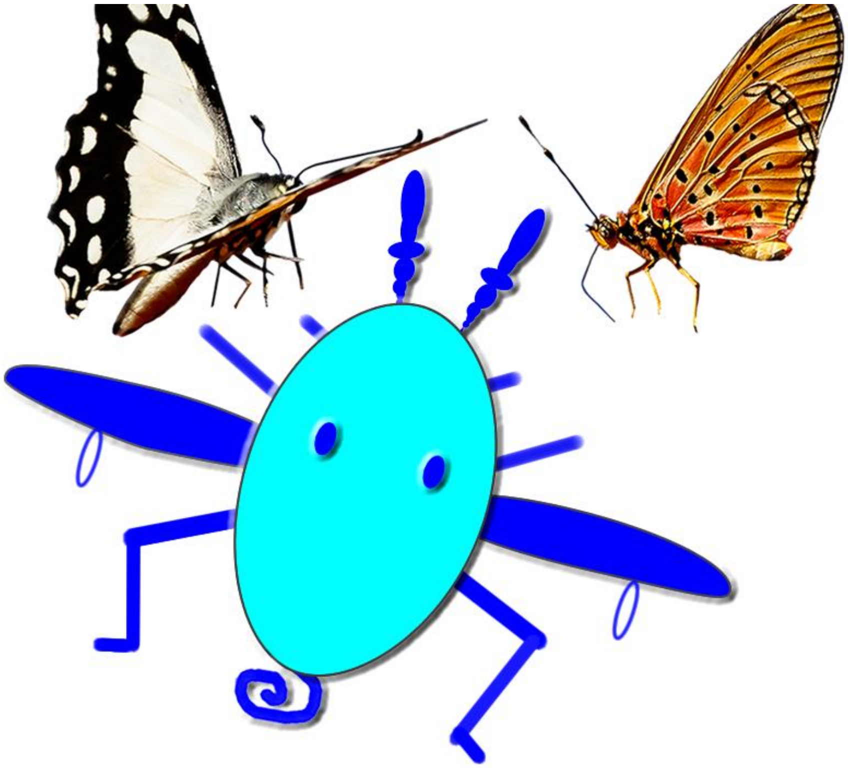


Figure B.13: Tests for selective attention to rules

This stimulus is a tyran with 6 legs (where the normal stimuli contributors will be exposed to during the training phase will only have 4 legs).

The ideal contribution will report the number of legs and discuss the other entities in the picture. It should also mention the attributes of the tyran

Coding Scheme:

tyran   Reason:

Attributes: 2 blue wings, 2 antennae, 6 legs, 2 buttons, light blue body, slightly tilted, short blue tail 1 ring each. The tyran is behind a light-blue fence. A yellow and black butterfly descending on it from the right and a black-and-white butterfly is attacking from the left.

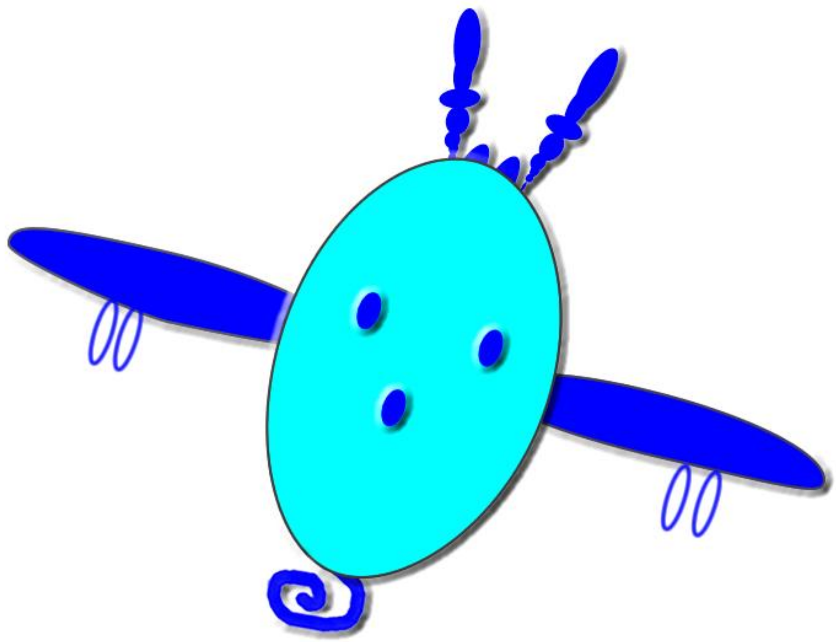


Figure B.14: Tests for selective attention to rules

This stimulus is a tyran with 2 extra antennae

The ideal contribution will report extra antennae and discuss the other entities in the picture. It should also mention the attributes of the tyran

Coding Scheme:

Tyran Reason:

Attributes: 2 blue wings, 2 long antennae, and 2 small antennae, 0 legs, 3 buttons, light blue body, slightly tilted, short blue tail, 2 rings each.

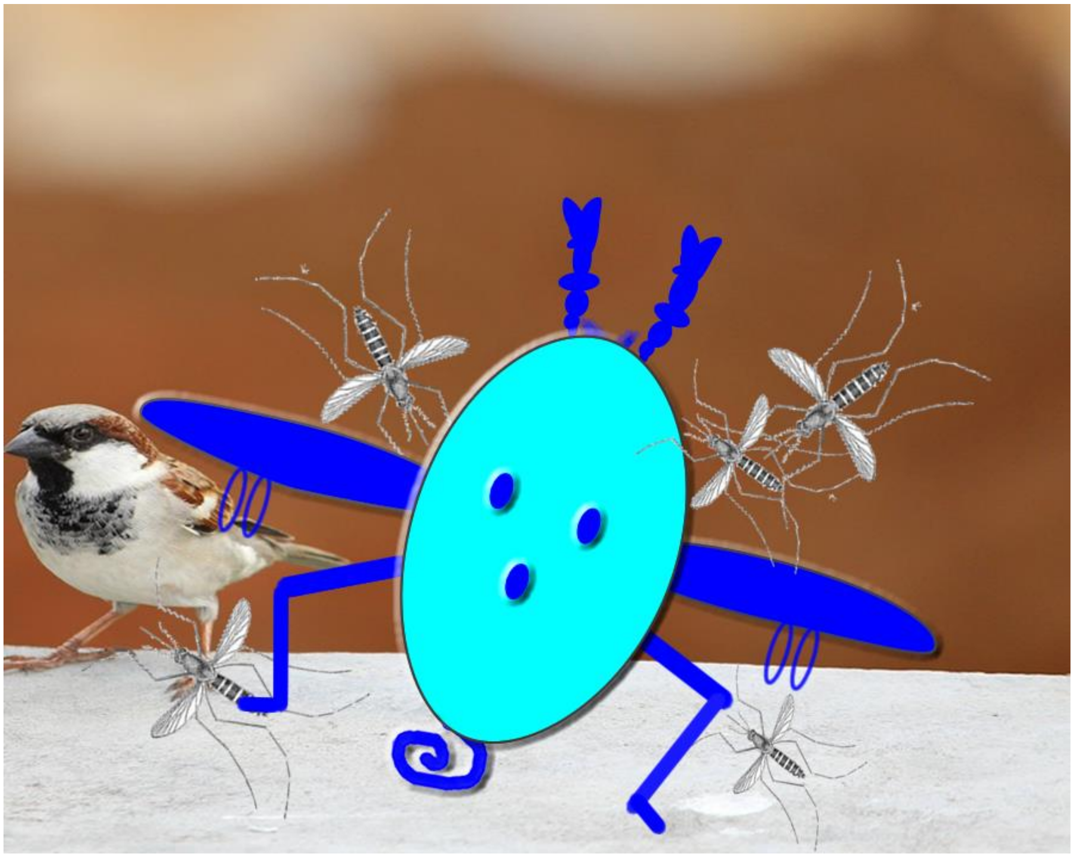


Figure B.15: Tests for selective attention to rules

This stimulus is a tyran with only two legs and a flower-like (split-end) antennae. It also has 2 tiny antennae. The ideal contribution will report these modifications to the non-diagnostic attributes and discuss the other entities in the picture. It should also mention the attributes of the tyran

Coding Scheme:

tyran Reason: insufficient rings for all the wings

Attributes: 2 blue wings, 2 long antennae with 2 lobbed ends and 2 small antennae,2 legs, 3 buttons, light blue body, slightly tilted, short blue tail, 2 rings each. It is surrounded by five mosquitoes. The tyran and small bird stands on a white snowing surface on the left.

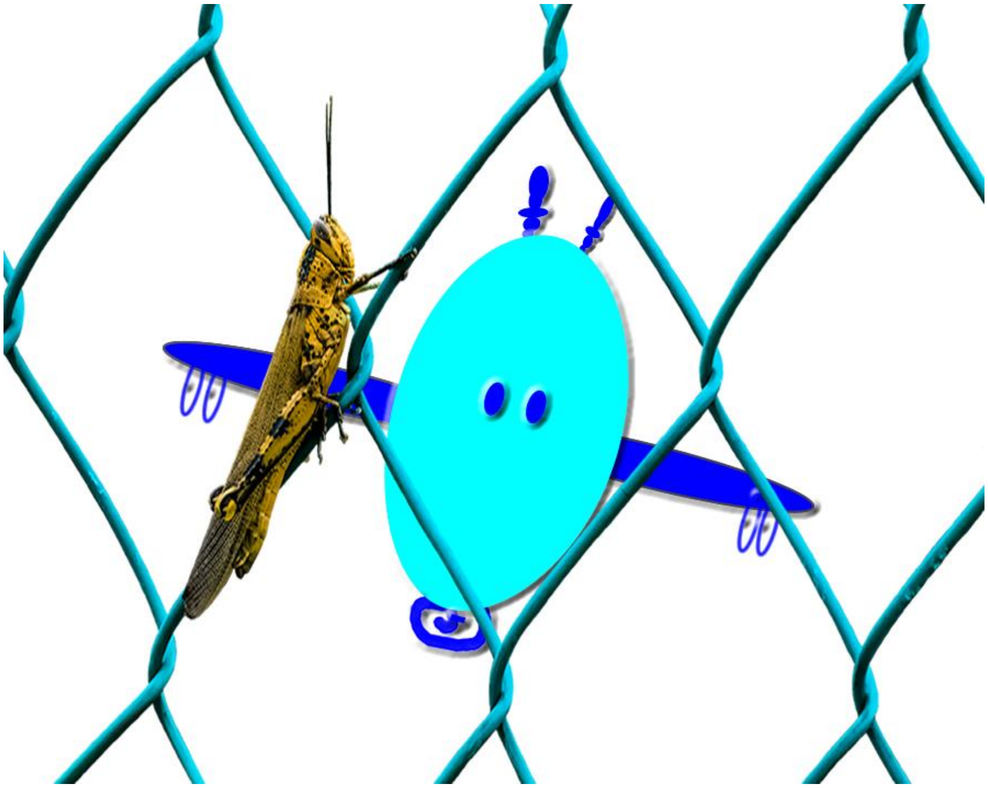


Figure B16: Tests for selective attention to rules

This stimulus is a tyran with short antennae

The ideal contribution will report the shorter antennae and discuss the other entities in the picture. It should also mention the attributes of the tyran

Coding Scheme:

Attributes: 2 blue wings, 2 short antennae, 0 legs,2 buttons, light blue body, slightly tilted, short blue tail, 2 rings each.

The tyran is behind a light-blue fence. A yellow and black locust is on the fence left side of the tyran

Catch Items

Catch Items will be placed after every 4th image. Participants should report all 4 catch items correctly for their data to be used in the analysis.

GLOSSARY

Term	Definition
Attribute	a quality or feature regarded as a characteristic or inherent part of someone or something
Classification	the action or process of classifying something according to shared qualities or characteristics
Cognitive diversity	the inclusion of people who have different styles of problem-solving and can offer unique perspectives because they think differently
Cognitive Dissonance	cognitive dissonance is used to describe the feelings of discomfort that result when your beliefs run counter to your behaviors and/or new information
Diagnostic Attributes	Attributes that can help classify an entity. Usually intrinsic
Mutual Attributes	Attributes that depend on two or more entities
Non-Diagnostic Attributes	Attributes of an entity that is not essential for classifying it
Selective Crowdsourcing	crowdsourcing that seeks to choose the best input(s) from a number of competing inputs provided by a crowd of people
Inclusion Rule	A set of rules about the attributes of an entity that help determine membership of a class
Integrative Crowdsourcing	Crowdsourcing that seeks to “pool complementary input from the crowd”
Intrinsic attributes	Attributes inherent in a thing. A part of a thing