

# Crowd modelling: aggregating non-expert views as a method for theorizing

Crowd  
modelling

239

Octavio González Aguilar  
*Felizmente Verde, Tepic, Mexico*

Received 17 April 2021  
Revised 29 July 2021  
Accepted 31 July 2021

## Abstract

**Purpose** – This paper aims to introduce a crowd-based method for theorizing. The purpose is not to achieve a scientific theory. On the contrary, the purpose is to achieve a model that may challenge current scientific theories or lead research in new phenomena.

**Design/methodology/approach** – This paper describes a case study of theorizing by using a crowd-based method. The first section of the paper introduces what do the authors know about crowdsourcing, crowd science and the aggregation of non-expert views. The second section details the case study. The third section analyses the aggregation. Finally, the fourth section elaborates the conclusions, limitations and future research.

**Findings** – This document answers to what extent the crowd-based method produces similar results to theories tested and published by experts.

**Research limitations/implications** – From a theoretical perspective, this study provides evidence to support the research agenda associated with crowd science. The main limitation of this study is that the crowded research models and the expert research models are compared in terms of the graph. Nevertheless, some academics may argue that theory building is about an academic heritage.

**Practical implications** – This paper exemplifies how to obtain an expert-level research model by aggregating the views of non-experts.

**Social implications** – This study is particularly important for institutions with limited access to costly databases, labs and researchers.

**Originality/value** – Previous research suggested that a collective of individuals may help to conduct all the stages of a research endeavour. Nevertheless, a formal method for theorizing based on the aggregation of non-expert views does not exist. This paper provides the method and evidence of its practical implications.

**Keywords** Crowdsourced research, Crowdsourcing, Crowd science, Theory building, Theorizing, Citizen science

**Paper type** Case study

## 1. Introduction

Social phenomena are more complicated than it was in the 1960s when the foundations of the current scientific research were grounded. [Feldman and Orlikowski \(2011\)](#) discuss the

---

© Octavio González Aguilar. Published in *International Journal of Crowd Science*. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

The author want to thank Professor Christian Wagner for all the meaningful guidance and advice. The author also want to thank Professor Carlos Jimenez for providing the space and access to participants during the data collection.



necessity of thinking in new ways of theorizing to understand contemporary organizations which nowadays are more complex, dynamic and distributed than before. It is required a systemic view of practice focused on a broader range of stakeholders (Brodie *et al.*, 2017) and path-breaking research (Möller, 2017). Since the 1960s, at least, researchers have put particular attention to how a theory must or must not look (Bacharach, 1989; Dublin, 1976; Feldman, 2004). Nevertheless, the distinction between theory as a product and theorizing as a process is less discussed. Theorizing is the process conducted to produce a theory (Folger and Turillo, 1999).

For Weick (1989) theorizing is a process of discovery, grounded on an artificial selection guided by the intentionality of the theorist. For Weick, a theory is a justified product rooted in a natural selection guided by validation and empirical evidence. This distinction is important because products improve by improving the process that produces them and a process can be standardized, taught and learned (Swedberg, 2016). Shepherd and Suddaby (2017) and Jaccard and Jacoby (2010) illustrate some practices and methods for theorizing. Unfortunately, all of them put at the centre of the theorizing process an expert researcher. This study contributes by providing a method for theorizing led by the aggregation of non-expert views. The purpose is not to achieve a scientific theory. On the contrary, the purpose is to achieve a model that may challenge current scientific theories or lead research in new phenomena.

The shift from a discourse of theory-praxis gap to theoretically path-breaking research leads to more exciting theories (Möller, 2017). Nevertheless, this shift is not easy to achieve. To be concerned with justification at the time of discovery can be counterproductive to theory generation (Weick, 1989). Academic institutions advance theory by relying on experts but, unfortunately, experts tend to perform familiar tasks (Hoffman, 1998). Thus, experts may omit variables that matter and include variables that do not (Page, 2008). This omission or inclusion may lead to the problem of “homogeneous theorizing”. Homogeneous theorizing is the unconscious selection of constructs, relations, explanations, theoretical frameworks or research methods closer to the cognition of the theorist mainly because it is cognitively more comfortable to fit the problem to those elements, not because the elements are the most suitable for the problem.

The collaboration with peers may diminish the problem of homogeneous theorizing; nevertheless, homophily, social influence or institutional incentives may lead to similar results. Homophily affects when the members of a group lock to common perspectives (Page, 2008) – for example, a school of thought. Social influence happens when people considered experts provide comments that change the decisions of others (Bolger and Wright, 2011). For example, in a publication process. The third affects when an academic considers a few sources of information. For example, when a few journals are deemed worthy of being studied. Consequently, the integration of non-experts in the research process is an alternative. Internet facilitates integration (Linstone and Turoff, 2011); especially following the principles of crowdsourced research (Love and Hirschheim, 2017). The remaining introduction is divided into two sections. Firstly, the literature review about crowdsourcing and crowd research is discussed. Secondly, the literature review associated with the aggregation of mental models is elaborated.

### *1.1 Crowdsourcing and crowd science*

In 2004, James Surowiecki (2004) writes his book “The wisdom of crowds” to points out that a crowd of random people, in some circumstances, may outperform the best individuals. He explains that the crowd leads to a kind of collective intelligence that may achieve better results than the smartest individuals. In this strand of thinking, in 2006, Jeff Howe coined

---

the term “crowdsourcing” to highlight a phenomenon: a distributed network of external collaborators may help business enterprises to solve their problems. Since then, a theory to explain collaborative platforms enabling successful crowdsourcing endeavours is required. Because of more than 10 years of research, the fundamental pieces of a theory of wise crowds already exist and are described as follow:

- *Disciplines*: After analysing the evolution of crowdsourcing from 2006 up to 2015, [Ghezzi et al. \(2018\)](#) conclude that the main disciplines studying this topic are information systems (e.g. technologies, platforms, mechanisms and algorithms), marketing (e.g. advertising and promotion), strategy (e.g. business models) and organizational studies (e.g. behaviours, motivations and performance).
- *Theories*: [Ikediego et al. \(2018\)](#) conduct a general review and states that the main theories trying to explain the phenomenon of crowdsourcing are the strategic theory (e.g. knowledge-based theory and resource dependency theory), the economic theory (e.g. transaction cost theory) and relational theory (e.g. institutional theory, social capital theory and social exchange theory).
- *Who*: [Ikediego et al. \(2018\)](#) argue that both organizations and individuals may conduct crowdsourcing. [Yu et al. \(2018\)](#) complement this idea by indicating that it is better to add diversity to the crowd instead of adding expertise.
- *What*: For [Schenk and Guittard \(2009\)](#) two types of the task may be crowdsourced: creative and non-creative task. For them, the non-creative task has two categories: routine tasks and complex tasks. In this strand of thinking, [Ikediego et al. \(2018\)](#) classify the tasks in micro and macro projects. According to them, micro-projects are repetitive, do not require domain-specific knowledge, are completed in few days and do not require special organization or incentives. Macro projects are just the opposite.
- *Why*: [Ikediego et al. \(2018\)](#) indicate that companies may have incentives such as the reduction of cost in the innovation process, the reduction of the time to market, the increment in the quality of the results or the increment in the diversity of both inputs and outputs. On the other hand, [Schenk and Guittard \(2009\)](#) notice that individuals are motivated mainly for ego gratification, reputation or economic incentives.
- *When*: [Boudreau and Lakhani \(2013\)](#) identified four scenarios when it is recommended to use a crowdsourcing schema. Firstly, when there is not a definition of what a good solution will look like. Secondly, when the participations will be accumulated and recombined into a valuable whole. Thirdly, when the bidder needs solutions to multiple problems. Fourthly, when there is the need to match offer and demand.
- *How*: [Brabham \(2008\)](#) estates and [Page \(2008\)](#) explains that the “wisdom of crowds” is not an average of the solutions provided. The wisdom comes from the aggregation of their diversity.
- *Quality*: For [Schenk and Guittard \(2009\)](#) the variety (or diversity) of options provided by the crowd is part of the quality of the crowdsourcing endeavour. The authors also indicate the originality and the fit with the expectations.

The research in crowdsourcing and collective intelligence also led to illustrations of crowds outperforming industry experts ([Poetz and Schreier, 2012](#)). Some of them, in academia. These exercises in parallel with the movement of citizen science, which searches for the

inclusion of non-experts in the research process, led to a new strand of research: crowd science. According to [Scheliga et al. \(2018\)](#), there are four strands of research associated with crowd science. Firstly, those cases where individual projects are described. Secondly, those where potential fields of application may exist. Thirdly, the stream arguing whether the borders of scientific knowledge are getting blurred. Finally, the stream arguing that crowd science will democratize academia. Nevertheless, there is a lack of practices to integrate a crowd of non-experts into existing scientific practices.

According to [Franzoni and Sauermann \(2014\)](#), one way to include scientific rigour in crowd science projects is to embed them in the software or platforms tools supporting crowd science. Still, there are no mechanisms that ensure group diversity, independence of participants and decentralization ([Yu et al., 2018](#)). This paper proposes a mechanism with such characteristics for the theorizing process. This mechanism is based on the aggregation of non-expert views. The reasoning behind their aggregation and the expectation of being a useful method for theorizing is argued in the following section.

### *1.2 Aggregation of non-expert views*

The concept of “expert” is not widely accepted. Relational theories argue that someone is an expert because others say so and realist/substantive theories say that performance is the main reason to justify that someone is an expert ([Collins and Evans, 2007](#)). According to the latter, this study defines an expert as someone who provides a richer set of alternatives and interpretations than the average person during the theorizing process. Experts present a broader set of other options and interpretations than novices ([Foley and Hart, 1992](#)). Still, the aggregation of diverse non-expert views may achieve better solutions than the expert because the cognition of one individual does not restrict the aggregation. In other words, adding experts trained with similar methods, theories or outlets preferences to the theorizing process improve the process less than adding diversity.

The Lattice theory of mental models for complex systems ([Moray, 1988, 1990, 1998, 1996](#)) explains why a non-expert view serves as a source of evidence during the theorizing process. This theory postulates that human beings construct models of causal relationships to explain how a particular system works. [Folger and Turillo \(1999\)](#) recognize that perception, imagination or discourse builds mental models and [Hoffman \(1998\)](#) indicates that those associations represent cause-effect relationships among objects or events. [Page \(2008\)](#) states that mental models are predictive models used by human beings to make decisions. Therefore, during this paper, the reasoning is as follows: Firstly, a non-expert view is:

- (1) A personal theory used by a non-expert to explain a phenomenon.
- (2) An elicited mental model from a person who is not an expert but has perceived the phenomenon.

Secondly, the person builds mental models based on real-world evidence with information lost. Thirdly, the aggregation of multiple mental models may reduce the loss. Fourthly, the variability conveys the relevance of those elements.

The aggregation of non-expert views is a plausible method for theorizing. Still, this method must avoid problems such as interaction among views and homogeneity of participants. The former promotes fear of disagreement, waste of time in meetings, the difficulty of schedules, social loafing, group polarization and dominant participants ([Nadkarni and Nah, 2003](#)). Therefore, [Rowe and Wright \(2011\)](#) recommend removing any indicator of the majority or minority opinions and confidence levels during information exchange. The latter maintain events and objects inside of the boundaries of the owners of the project ([Lukyanenko et al., 2017](#)). Consequently, to organize the methods that may

aggregate mental models, this study proposes a  $2 \times 2$  matrix (Table 1). based on the criteria of heterogeneity of members and Information exchange.

The first quadrant includes methods that do not limit group interaction or homogeneity of inputs (Ossadnik *et al.*, 2013). The second quadrant includes methods that avoid group interaction but remain the homogeneity of inputs (Kwahk *et al.*, 2007). The third quadrant includes methods that pursue the heterogeneity of inputs but do not avoid group interaction (Cunha *et al.*, 2016). The fourth quadrant includes methods that explicitly avoid group interaction and foster the heterogeneity of inputs (Nadkarni and Nah, 2003). This study is different because:

- It proposes a method that receives and aggregates the inputs of non-experts. That includes the leading researcher.
- It may consider a few modellers or huge crowds.
- It promotes crowd science.

This study proposes a method for the fourth quadrant. The case study introduced in Section 2 reports the results of designing and implementing a method for theorizing which follows the characteristics explained above. In particular, the case study answers: To what extent the aggregation of non-expert views may lead to an expert-level research model?

## 2. Case study

Since 1995 the Dolphin programme has connected undergrads with professors around the country or overseas. The selected undergrads have the best grade point averages (GPAs) in each department. The programme runs yearly. Each undergrad collaborates with a Professor by doing research for around eight weeks. The aim is to foster the undergrads' research spirit. In 2018, Colombia and Costa Rica integrated their students into the programme. A total of 6,309 undergrads participated just in México. The number of Professors involved was around 1,900 worldwide. The Professors became tutors entirely for free. For many undergrads, it is the first time they leave their home and they are involved in a formal research project. The Dolphin programme organizes a conference at the end of the summer. During the meeting, the undergrads discuss their results.

### 2.1 Pre-study

**2.1.1 Participants.** Ten undergrads (henceforth “modeller”) were selected out of 138 to participate in the pre-study. The modellers invited covered the five areas of knowledge defined by the coordination of the Dolphin programme. Those areas are physics, mathematics and science, biology, chemistry and agricultural science, medicine and health, humanities and behavioural science and social, law and economic science. Nevertheless, only nine modellers appeared. The modellers were selected based on their availability. Their age was around 20 and 22 years old. The leading researcher called them one by one. During the call, the leading researcher explained the relevance of the project, indicated the show-up reward and listed the free slots in the schedule. Each modeller selected the best time for him/her and was a first-time applicant. For each modeller, the show-up award was two movie tickets.

| Diversity intention | Group interaction              |                                   |
|---------------------|--------------------------------|-----------------------------------|
| Homogeneous         | I. Homogeneous interaction     | II. Homogeneous non-interaction   |
| Heterogeneous       | III. Heterogeneous interaction | IV. Heterogeneous non-interaction |

**Table 1.**  
Matrix of diversity intentionality and group interaction

The main researcher invited five assistants to avoid the intervention of the researcher during the data collection. The assistants helped to prepare the materials and were trained one day before the pre-study. The training consisted of reading the protocol and practice with their peers. They were students from a university with no relation to the study. The assistants were eager to participate because this activity helped to fulfil part of their graduation requirements. The assistants were psychology students in the last term. They were three women and two men who have never participated in the Dolphin programme. They assisted for 2 h every day. The principal researcher assigned one male and one female to lab 1 and one male and one female to lab 2. Another female helped to capture the data into the system and prepare the materials.

*2.1.2 Sampling.* The following step was to sample modellers and concepts. Modellers sampling must assure their cognitive differences. This type of sampling is theoretical rather than statistical (Glaser and Strauss, 1967). Page (2008) highlights that more than the number of crowd members what matters is the cognitive diversity of them. Therefore, the question is not about if there are enough modellers, but if there is enough diversity. Onwuegbuzie and Collins (2007) suggest collecting information from three to six focus groups formed by six to nine participants. Therefore, a pool of diverse modellers from 18 ( $3 \times 6$ ) to 54 ( $6 \times 9$ ) shall be enough.

Concept sampling also follows this strand of thinking. The researcher may argue the following reasons to consider a pool of concepts:

- a single theory does not fit the phenomena but concepts from multiple theories may help to theorize about it;
- concepts of competing theories can actually complement each other;
- alternative causal reasoning is required for current theories; and
- the concepts challenge a current theory, among others relevant for theoretical purposes.

The set of concepts must be grounded on previous knowledge and must vary from five to nine. Ideally, seven, which is the average number of concepts that, cognitively, a human being can work with (Miller, 1956).

*2.1.3 Materials.* Each modeller worked with four theories “A”, “B”, “C” and “D”. Nevertheless, the order differs according to the data collection design. Consequently, to provide the appropriate material in each session, the material of each theory was organized inside an independent yellow folder. The yellow folder had a tag associated with the theory (i.e. “A”, “B”, “C” and “D”) and a legend. The legend stated, “do not open it. Wait until the researcher says so”. Each yellow folder contained three components. Firstly, a grey folder tagged “Training”. Secondly, a grey folder tagged “modelling” plus the name of the theory (i.e. “A”, “B”, “C” and “D”). Finally, a half-page indicating the word “activity” plus the name of the activity (i.e. “A”, “B”, “C” or “D”), the context of the activity and the objective.

The training material comprehends a set of paper-based components. The set includes:

- Two independent concepts.
- The effect.
- Arrows to connect the concepts among them and with the effect.
- “+” and “-” symbols to indicate the type of relation.
- “Relevant” and “irrelevant” tags to indicate whether the concepts are necessary for the theory.
- “Point” tags to rank the concepts based on their relevance to the theory.

Independently of the number of theories to work with, the modeller only uses the training material once. A set of four videos were recorded and played during the training session to guide the modeller. The working material is inside the grey folder tagged “Modelling” and it differs based on the theory associated with the folder (i.e. “A”, “B”, “C” and “D”).

The data collection took place in a University in Mexico that participates and organizes the Dolphin programme. It was required two labs for data collection. The labs were in a closed section of the University’s library. Each lab used its own set of materials (i.e. yellow folder “A”, “B”, “C” and “D”). The lab equipment consisted of a personal desk, a chair, two tablets, a headphone and a microphone. The microphone and the first tablet recorded the explanations of the modeller. The second tablet played the videos with the instructions for the modeller. Both labs were far enough from each other. Consequently, the modellers did not hear the explanations of another modeller. All the material was white and black to avoid distractions (Figure 1) and the assistants used a printed protocol to conduct the sessions.

**2.1.4 Procedures.** The data collection conducts a within-subjects experiment design to answer the research question. The within-subject design is the best option because all the modellers participate in the four treatments (Creswell, 2014). Consequently, it is possible to measure the results of a modeller after receiving each treatment. Figure 2 depicts the within-subjects repeated measures design. The objective of the design is to avoid the risk that the order of the treatments affects the results. Consequently, modellers in Group A models the theories in the following order A-B-C-D. Modellers in Group B models the theories in the following order B-C-D-A. Modellers in Group C models the theories in the following order C-D-B-A and finally, Modelers in Group D models the theories in the next order D-A-B-C. Then all the results are evaluated as a single group.

The purpose of the research is not to evaluate the effects of the order in any variable. Nevertheless, if all the modellers received the same treatment, it could be a cofounded variable. Consequently, the order (i.e. A, B, C and D) changed to reduce the possibility of bias. Therefore, the design considers four groups so that each one starts with a different

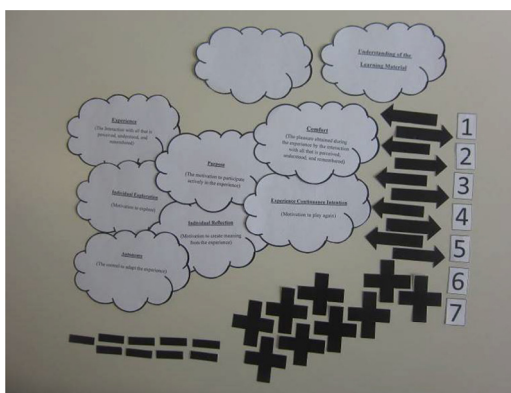


Figure 1. Materials for modelling

|         |                                   |
|---------|-----------------------------------|
| Group A | X1---O---X2---O---X3---O---X4---O |
| Group B | X2---O---X3---O---X4---O---X1---O |
| Group C | X3---O---X4---O---X1---O---X2---O |
| Group D | X4---O---X1---O---X2---O---X3---O |

Figure 2. Data collection design

theory. An exhaustive evaluation shall include 24 groups instead of four. That is the permutation of the four theories (i.e.  $4 \times 3 \times 2 \times 1$ ). Still, as explained above, an exhaustive evaluation of the order is not the purpose of this study. Because in a typical setting, a crowd of modellers will work with only one theory.

The theories selected as conditions represent two factors (Table 2). The first factor is the number of concepts (including the effect). This number may be fewer or greater than seven concepts. The second one is the distance from the farthest independent concept to the effect. This distance is  $\leq 3$  or  $> 3$ . This distance is considered based on the longest path. One more criterion also influenced. All of them must be from information systems contexts. The reasoning behind this criterion is that none of the modellers studied information systems or computer science or a related field. Still, all of them have used information systems. Consequently, the modellers are not experts, but they have lived the experience, or, at least, they have the cognitive tools required to suggest a plausible research model in the area.

Lin and Bhattacharjee (2010) propose Theory A. This theory has six concepts and the “distance” is three concepts. The distance represents the path technical quality  $\rightarrow$  perceived enjoyment  $\rightarrow$  attitude. Laumer *et al.* (2016) propose Theory B. This theory has eight concepts and the distance is two. The distance is calculated with the path routine seeking  $\rightarrow$  perceived usefulness. Davis *et al.* (1989) proposed Theory C. This theory has six concepts and the distance is five. The distance represents the path external variables  $\rightarrow$  perceived ease of use  $\rightarrow$  perceived usefulness  $\rightarrow$  attitude towards using  $\rightarrow$  behavioural intention to use. Finally, Chen *et al.* (2012) propose Theory D. This theory has eight concepts and the distance is four. The distance is calculated with the path system quality  $\rightarrow$  confirmation  $\rightarrow$  perceived usefulness  $\rightarrow$  satisfaction.

During the data collection stage, the procedure was as follows: Firstly, the modeller arrived. Then, the researcher verified the identity, the lab assigned and finally, provided a code. After that, the modeller went to the indicated lab and provided the code to the assistant (Table 3). Thus, the assistant contrasted the code with the protocol and, based on that, the assistant assigned the first yellow folder. Afterwards, the assistant played the training videos and the modeller used the training material. The modeller asked whether the modeller has any doubt. If there was no doubt, the modeller continued with the working materials. If there was a doubt, the modeller repeated the training videos. The training videos helped to standardize the indications. Finally, the assistant photographed the research model and provided, according to the order of the treatments, another yellow folder.

**Table 2.**  
Theories modelled

| Concepts/Distance | $\leq 3D$        | $> 3D$           |
|-------------------|------------------|------------------|
| $< 7C$            | 3D6C<br>Theory A | 5D6C<br>Theory C |
| $> 7C$            | 2D8C<br>Theory B | 4D8C<br>Theory D |

**Table 3.**  
Modellers

| Groups  | Modellers codes |      |      |      |      |
|---------|-----------------|------|------|------|------|
| Group A | A1M1            | A2M1 | A3M1 | A4M1 | A5M1 |
| Group B | A1M2            | A2M2 | A3M2 | A4M2 | A5M2 |
| Group C | A1M3            | A2M3 | A3M3 | A4M3 | A5M3 |
| Group D | A1M4            | A2M4 | A3M4 | A4M4 | A5M4 |



The modelling process was as follows. Firstly, the modeller identified independent concepts. Secondly, the modeller identified the effect. To differentiate the independent concepts and the effect, the main researcher printed the formers with lower case letters and the latter with upper case letters. Thirdly, the modeller used the arrows to indicate which concepts are causes of another. The effect never is a cause of another concept. Fourthly, for each relation, the modeller shows whether the relation is positive (i.e. “+”) or negative (i.e. “-”). Fifthly, the modeller added an “irrelevant” or “relevant” tag to each of the independent concepts. Sixthly, the modeller added a “point” tag to each of the independent concepts to rank them. Seventhly, the modeller argued the logic of the model and how the relationships depicted by the model explain the effect.

The design of the pre-study avoids cognitive fatigue. Each modeller work with two activities per day. Therefore, each modeller had to come for two consecutive days. Each day the modeller came for 1 h. Each modeller should model the four theories in two sessions. This pre-study provided evidence of three things. Firstly, the modellers learnt the method by following the 2.53 min of video training. Secondly, the modellers required around 30 min to complete two tasks. Then, it is possible to finish the four tasks in just 1 h and a half without cognitive fatigue. Therefore, this avoids the risk of losing data for mortality. Thirdly, it is possible to reduce the reward from two movie tickets (one per session) to only one without affecting the results. It reduces the cost of the data collection by half.

## 2.2 Study

**2.2.1 Participants.** The researcher and assistants from the pre-study participated in the study. Nevertheless, the modellers from the pre-study did not participate in the study.

**2.2.2 Sampling.** During the study, 20 new modellers participated. They represented the five fields of knowledge described above. The objective of this stage was to collect the data of four modellers from each domain. The characteristics of the participants were the same as those from the pre-study. Nevertheless, the study considered neither the participants nor the data provided by them during the pre-study. The inclusion of the five areas provided diversity. This diversity refers to mental models and cognitive tools. The inclusion of four modellers from each field was necessary to give a different condition for each modeller. [Table 3](#) conveys the organization of the modellers according to the data collection design. The study required the modeller for only 1 h and a half, only one day. Consequently, the incentive for the modeller changed to only one movie ticket. The sample of concepts was the same as in the pre-study.

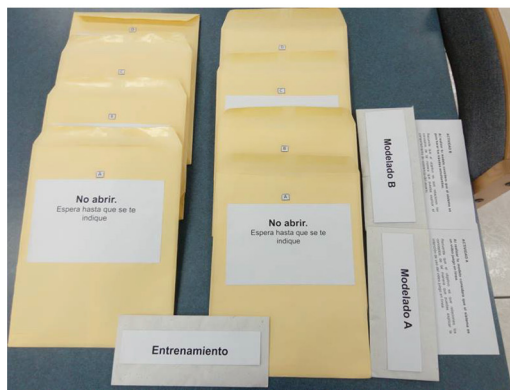
**2.2.3 Materials.** The materials were the same as described during the pre-study. [Plate 1](#) illustrates the lab, the materials organized by treatments and the materials once a modeller has used them to build a model.

**2.2.4 Procedures.** The leading researcher contacted all the modellers to arrange the meetings. The modellers indicated which session fitted better for them. Then, the process continued as described in the pre-study section. Nevertheless, the modellers came one instead of two days and the assistant reviewed each model. The model must comply with the following specifications. Firstly, all the independent concepts had a connection with at least another one; on the contrary, the independent concept must have an “irrelevant” tag. Secondly, the effect has at least one independent concept as a cause. Thirdly, all the independent concepts have a “relevance” tag. Fourthly, all the relationships have only one direction symbol. Fifthly, all the independent concepts have only one “point” tag. Finally, the assistant took pictures of all the models.

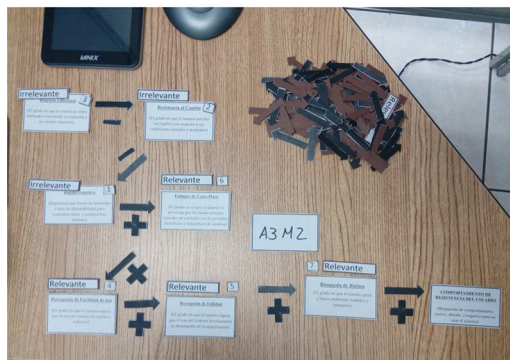
During the pre-study, some modellers indicated that all the independent concepts were direct causes of the effect. Their models did not include any mediator. This situation



(a)



(b)



(c)

Plate 1.  
Infrastructure for  
data collection

appeared at the beginning of the study; consequently, when a modeller did not theorize any mediator, the assistant called the leading researcher. Then he asked the modeller whether he/she understood the task. If so, the leading researcher disregarded the four models provided by that modeller. Therefore, the study only considers models with at least one

mediator. The modellers explicitly said to the leading researcher that they understood the task. Furthermore, they look motivated, asking when the results of the study will be available. Finally, the aggregation of the data started.

### 3. Aggregation

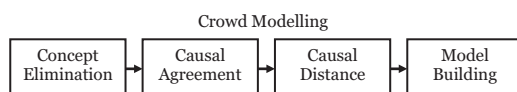
#### 3.1 Execution

Figure 3 depicts the algorithm for the aggregation of non-expert views (henceforth, “crowd modelling”). The following sections report the implementation of the algorithm throughout four phases: concept elimination, causal agreement, causal distance and model building. Approval voting is the base for the first two phases. In approval voting:

- The decision maker votes for as many options as desired.
- The winner is the option with more votes.
- The researcher collects more information from the decision maker than from plurality voting.
- The performance is superior to existing procedures with respect to fairness (Fishburn and Little, 1988). Furthermore:
- Approval voting is simple to use.
- Is simple to use.
- It reflects the decision maker preferences and frequently achieve similar results than competing methods such as Borda winner and Condorcet winner (Regenwetter and Grofman, 1998).

**3.1.1 Concept elimination.** Each non-expert view provided a certain number of points for each concept. The algorithm calculated the total number of points for each independent concept and dropped those below the mean. The formula:  $(k)(n/2)$  indicates the cut-off, where “n” is the number of non-expert views and “k” is the number of independent concepts. For example, there were 20 non-expert views for each theory. Consequently, the value of “n” was 20 for all the theories. Tasks A and C had five independent constructs; therefore, the value of “k” for these theories was 5. The tasks B and D had seven independent constructs; therefore, the value of “k” for these theories was 7. Then the cut-off for theories A and C was  $\leq 50$  and the cut-off for theories B and D was  $\leq 70$ .

The leading researcher dropped the following constructs. The independent concept social image in Theory A. The independent concept external variables, in Theory C and finally, the independent concept hedonic value, in Theory D. In Theory B, all the independent concepts remained. This study evaluates “relevant” and “irrelevant” tags. The formula was the same to evaluate the cut-off for the relevance criterion. Nevertheless, “k” always had the value of 1, given that the possible values were binary (1 = relevant, 0 = irrelevant). Considering this criterion, all the concepts, of all the theories, were above the mean. The evidence suggests that the “relevant” and “irrelevant” tags are not required. The aggregated model orders the concepts that remain with consecutive numbers based on their aggregated ranking. The numbers start from one (the least important) and finalizing with the most important (Figure 4).



**Figure 3.**  
Algorithm for the aggregation of non-expert views

3.1.2 *Causal agreement.* The leading researcher defined a language for the aggregated model. The purpose of this language is to convey the degree of the agreement. This language defines four causal-relationships connectors and Figure 5 depicts them. The singular agreement means that only one non-expert view indicates the causal relationship. The formula:  $m = 1$  defines it. The shared agreement means that, at least, two non-expert views suggest the relationship. It is defined by the formula:  $m > 1$ . The majority agreement means that more than half of non-expert views indicate the relationship. The formula:  $m > (n/2)$  defines it. The consensus agreement means that all the non-expert views indicate the causal relationship. The formula:  $m = n$  defines it. In the formulas, the variables are “n” which represents the total number of non-expert views and “m” which represents the total number of non-expert views sharing the same interpretation.

3.1.3 *Causal distance.* Figure 6 depicts a problem in the aggregation. The non-expert views 1 and 2 agree that the independent concept X is a cause of the effect. Nevertheless, given that the first view adds a mediator, the aggregation will not indicate this agreement. This situation may lead to a loss of information. The algorithm only calculates the values for those relationships in the aggregated model. Nevertheless, the values come from all the non-expert views. It is necessary to divide one by the number of causal relationships among two concepts to calculate the causal distance. Figure 7 depicts the formula to calculate the aggregated causal distance between two concepts. In the formula, “j” indicates the number of non-expert views suggesting the causal relationship. Furthermore, “x” is the causal distance calculated in view “j” for that relationship.

3.1.4 *Model building.* Each model depicts the aggregations by using the language for crowd modelling. The models present the independent concepts above the mean, their

Figure 4.  
Ranking

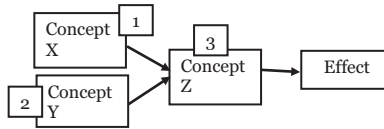


Figure 5.  
Type of agreements

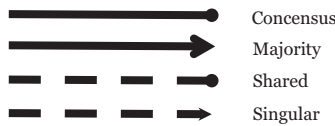


Figure 6.  
Causal distance

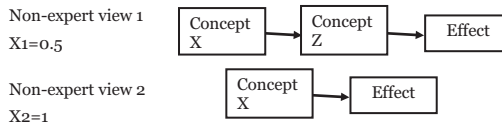


Figure 7.  
Aggregated causal  
distance

$$\left( \sum_{i=1}^j x_i \right) / j$$

relations with the highest causal agreement, the directions of the relationships, the ranks based on the points and the effect. There are two critical considerations. Firstly, no cyclical relation must exist (i.e. a cause is a cause of its cause). Each causal path must start from the effect and prioritize the highest causal agreement, if equal, prioritize the causal distance, if equal, prioritize the ranking. The algorithm follows that causal path and drops the causal relationship that creates a cyclical relation. Secondly, each independent concept must be the cause of at least another independent concept or effect. Thus, Figures 8, 9, 10 and 11 represent the aggregated models for theories A, B, C and D.

3.2 Evaluation

It is necessary to evaluate the similitude among the aggregated models and those published by previous research. This evaluation will help to answer the research question. Table 4 reports the results. The percentage of similitude are 80% for theories A and C, 60% for Theory B and 70% for Theory C. For example; Theory A has five independent concepts and one effect. The effect cannot be a cause and an independent concept cannot connect to itself.

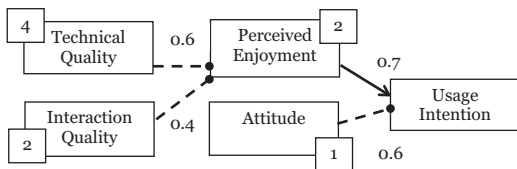


Figure 8. Theory A

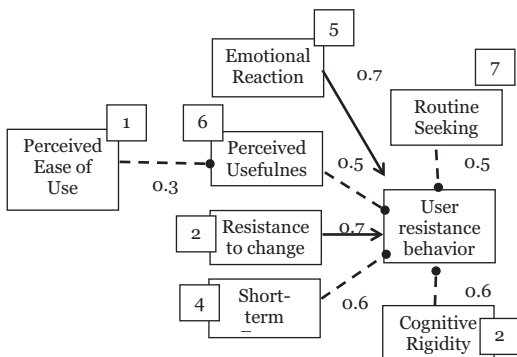


Figure 9. Theory B

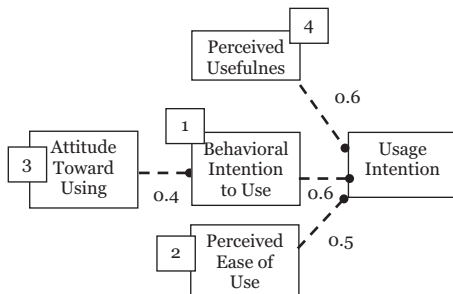


Figure 10. Theory C

Therefore, there are 25 possible causal relationships to indicate in the aggregated model. The aggregated model did not suggest 18 unpublished causal relationships and suggested 2 published causal relationships. Thus, the aggregated model was correct in 20 out of the 25 predictions. That is 80% of effectiveness. The same logic applies to the rest of the aggregated models.

A match between the aggregated model and the published theory is a prediction of a causal relationship. The prediction may have two values: exist or do not exist. Therefore, in the example, Theory A states that technical quality leads to perceived enjoyment. Technical quality does not indicate a causal relationship with the other four concepts. There is no causal relationship predicted to interaction quality, social image, attitude or usage intention. Therefore, Theory A only depicts one out of five possible causal relationships. The aggregated model for Theory A indicates the same. Therefore, there are five matches. That is the correct prediction that one causal relationship exists and the correct prediction that four causal relationships do not exist. The process is the same for the rest of the concepts.

### 3.3 Discussion

Some of the results are: Theories A and D are closer than B and C to the published theories. Thus, it does not matter whether the number of concepts involved is six or eight. The distance from the farthest independent concept to the effect is also not relevant, as long as it varies from two to five. Consequently, the method for theorizing worked for all the theories evaluated. The aggregation of non-expert views is a plausible means to theorize. The differences between the aggregated models and the published models may be due to cultural differences, the problem of homogeneous theorizing, lack of incentives to be right or another reason. To determine why theories A and D were closer to the theories published than theories B and C is a topic of another study.

The purpose of the method is not to provide evidence to validate existing theories. If there is no theory for a particular phenomenon, the purpose is to theorize a new theory. If there are current theories, the purpose is to theorize alternative explanations or challenge the causal relationships. Therefore, Table 5 helps to guide the theorist. If the aggregated model

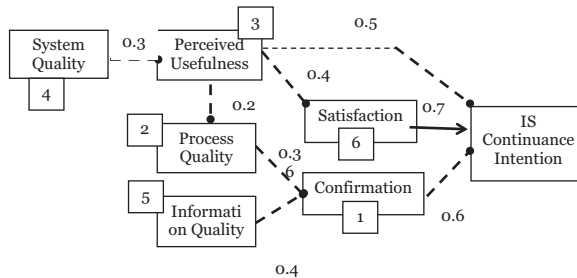


Figure 11.  
Theory D

| Theories | Equal (#) | Equal (%) | Different (#) | Different (%) |
|----------|-----------|-----------|---------------|---------------|
| A        | 20        | 80        | 5             | 20            |
| B        | 31        | 63        | 18            | 37            |
| C        | 18        | 72        | 7             | 28            |
| D        | 39        | 80        | 10            | 20            |

Table 4.  
Results

suggests a causal relationship that is already published, then the theorizing process search for alternative explanations. If the aggregated model suggests a causal relationship that is not already published, then the theorizing process tries to challenge the theory. If the aggregated model does not indicate a causal relationship that the published theory does, then the theorizing process tries to challenge the theory. In this process, it is particularly important to consider the causal agreement and the causal distance.

For example, in Theory A, the theorist must test the causal relationship among perceived enjoyment and usage intention without the mediation suggested in the published theory especially because the aggregated model indicates a majority causal agreement and a causal distance of 0.7. The theorist should test the causal relationship between interaction quality and perceived enjoyment. The reason is that the aggregated model depicts a relationship that the theory published does not. The causal relationships among technical quality and perceived enjoyment, as well as attitude and usage intention is in both models. Thus, the theorist should search for alternative explanations. After that, the published theory indicates the causal relationship between perceived enjoyment and attitude, whereas the aggregated model does not. That means that the theorist should try to challenge that.

#### 4. Conclusion, limitations and future research

The results of the case study answer the research question and provide evidence that the aggregation of non-expert views may achieve expert-like research models with a similitude that may vary from 60% to 80%. These expert-like research models may serve as an input for both positivist and interpretative researchers. It is possible to organize the lessons learnt in terms of crowd collaboration, diversity, contribution and performance. It is also possible to organize the lessons learnt in terms of the theorizing purpose that may be to find different cause-relationships or to find alternative explanations.

In terms of crowd collaboration, the non-expert views may come from collectives or individuals. In the case study, 20 modellers worked independently. In terms of crowd diversity, the professional background was the criterion to sample modellers. The case study included modellers from multiple areas of knowledge. In terms of crowd contribution, the case study exemplified how modellers may participate in the theorizing of one or up to four theories at the time. In terms of crowd performance, the pool of concepts may vary from five to nine. In the case study, the modellers worked with theories with six or eight concepts.

The theorist may theorize causal relationships or reasonings. When there is no reference theory, crowd modelling leads to a causal-relationship diagram and a set of reasons to sustain the relationships. When there is a reference theory, crowd modelling leads to a set of causal relationships that may or not exist in the reference theory. If they do not exist, that challenge the reference theory. Especially when the causal distance of that relationship is higher than 0.5 and the causal agreement is majority or consensus. If they exist, the theorist

| Aggregated model | Theory published                     |                              |
|------------------|--------------------------------------|------------------------------|
|                  | Yes                                  | No                           |
| Yes              | Search for alternative explanations* | Try to challenge the theory* |
| No               | Try to challenge the theory          | No action                    |

**Note:** \*Specially if the causal distance is greater than 0.5 and the causal agreement is majority or consensus

**Table 5.**  
Analysing the aggregated model

may search for alternative explanations. Finally, if the reference theory depicts relationships that the crowd does not see, that also challenge the reference theory – the context and participants matter for any conclusions.

The case study provides evidence to argue that the aggregation of non-expert views is a plausible means to theorize expert-like research models. This evidence is especially important for institutions with limited access to experts or databases. It is also essential for those who want to avoid the homogeneous theorizing problem. The main limitation of the method is that, according to its rationality, it only works for social sciences. The case study provides evidence to answer the research question of the study. Furthermore, the algorithm provides the grounds for a standard execution, whereas integrates the principles discussed during Section 1.

The results of this study have three implications. Firstly, it provides a method for theorizing that requires less statistical and background knowledge, as well as fewer resources, than similar approaches to support crowd science. Furthermore, provide the theoretical argumentation to explain how and why the method works. Secondly, it defines the problem of homogeneous theorizing and suggests a method to reduce it. Finally, suggests pieces to complement the theory of wise crowds. In Section 1.1 the pieces already considered are discussed. Still, the case study provides evidence that both topics discussed in Section 1.2 must be considered in the theory of wise crowds: firstly, the aggregation of mental models from a cognitive perspective. Specially, to explain under which conditions the diversity of mental models performs better to convey the reasoning of social facts. Secondly, the design of algorithms and platforms that integrate the theory by design. This piece must formalize the design principles of artefacts focused on theorizing and theory building.

## References

- Bacharach, S.B. (1989), "Organizational theories: some criteria for evaluation", *Academy of Management Review*, Vol. 14 No. 4, pp. 496-515.
- Bolger, F. and Wright, G. (2011), "Improving the Delphi process: lessons from social psychological research", *Technological Forecasting and Social Change*, Vol. 78 No. 9, pp. 1500-1513.
- Boudreau, K.J. and Lakhani, K.R. (2013), "Using the crowd as an innovation partner", *Harvard Business Review*, Vol. 91 No. 4, pp. 62-69.
- Brabham, D.C. (2008), "Crowdsourcing as a model for problem solving: an introduction and cases", *Convergence: The International Journal of Research into New Media Technologies*, Vol. 14 No. 1, pp. 75-90, doi: [10.1177/1354856507084420](https://doi.org/10.1177/1354856507084420).
- Brodie, R.J., Nenonen, S., Peters, L.D. and Storbacka, K. (2017), "Theorizing with managers to bridge the theory-praxis gap", *European Journal of Marketing*, Vol. 51 NoS7/8, pp. 1173-1177.
- Chen, L., Meservy, T.O. and Gillenson, M. (2012), "Understanding information systems continuance for information-oriented mobile applications", *Communications of the Association for Information Systems*, Vol. 30 No. 1, pp. 127-146.
- Collins, H. and Evans, R. (2007), *Rethinking Expertise*, 1st edition, The University of Chicago Press, Chicago, IL.
- Creswell, J.W. (2014), *Research Design*, SAGE Publications.
- Cunha, A.A.R., Filho, J.L.S. and Morais, D.C. (2016), "Aggregation cognitive maps procedure for group decision analysis", *Kybernetes*, Vol. 45 No. 4, pp. 589-603.
- Davis, F.D., Bagozzi, R.P. and Warshaw, P.R. (1989), "User acceptance of computer technology: a comparison of two theoretical models", *Management Science*, Vol. 35 No. 8, pp. 982-1003.
- Dublin, R. (1976), "Theory building in applied areas", in Dunnette, M. (Ed.), *Handbook of Industrial and Organizational Psychology*, Rand McNally, Chicago, IL, pp. 17-39.



- Feldman, D.C. (2004), "What are we talking about when we talk about theory?", *Journal of Management*, Vol. 30 No. 5, pp. 565-567.
- Feldman, M.S. and Orlikowski, W.J. (2011), "Theorizing practice and practicing theory", *Organization Science*, Vol. 22 No. 5, pp. 1240-1253.
- Fishburn, P.C. and Little, J.D.C. (1988), "An experiment in approval voting", *Management Science*, Vol. 34 No. 5, pp. 555-568.
- Foley, M. and Hart, A. (1992), "Expert-Novice differences and knowledge elicitation", in Hoffman, R.R. (Ed.), *The Psychology of Expertise: cognitive Research and Empirical AI*, 1st edition, Springer-Verlag, Berlin, pp. 233-244.
- Folger, R. and Turillo, C.J. (1999), "Theorizing as the thickness of thin abstraction", *Academy of Management Review*, Vol. 24 No. 4, pp. 742-758.
- Franzoni, C. and Sauermann, H. (2014), "Research policy", *Research Policy*, Vol. 43 No. 1, pp. 1-20, doi: [10.1016/j.respol.2013.07.005](https://doi.org/10.1016/j.respol.2013.07.005).
- Ghezzi, A., Gabelloni, D., Martini, A. and Natalicchio, A. (2018), "Crowdsourcing: a review and suggestions for future research", *International Journal of Management Reviews*, Vol. 20 No. 2, pp. 343-363, doi: [10.1111/ijmr.12135](https://doi.org/10.1111/ijmr.12135).
- Glaser, B.G. and Strauss, A.L. (1967), "Theoretical sampling", in *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Weidenfeld and Nicolson, pp. 45-77.
- Hoffman, R.R. (1998), "How can expertise be defined? Implications of research from cognitive psychology", in Williams, R., Faulkner, W. and Fleck, J. (Eds), *Exploring Expertise: Issues and Perspectives*, 1st edition, Macmillan Press, London, pp. 81-100.
- Ikediego, H.O., Ilkan, M., Abubakar, A.M. and Bekun, F.V. (2018), "Crowd-sourcing (who, why and what)", *International Journal of Crowd Science*, Vol. 2 No. 1, pp. 27-41, doi: [10.1108/IJCS-07-2017-0005](https://doi.org/10.1108/IJCS-07-2017-0005).
- Jaccard, J. and Jacoby, J. (2010), *Theory Construction and Model-Building Skills: A Practical Guide for Social Scientists*, The Guilford Press, New York, NY.
- Kwahk, K.Y., Kim, H.W. and Chan, H.C. (2007), "A knowledge integration approach for organizational decision support", *Journal of Database Management*, Vol. 18 No. 2, pp. 41-61.
- Laumer, S., Maier, C., Eckhardt, A., et al. (2016), "User personality and resistance to mandatory information systems in organizations: a theoretical model and empirical test of dispositional resistance to change", *J Inf Technol*, Vol. 31, pp. 67-82, available at: <https://doi.org/10.1057/jit.2015.17>
- Lin, C. and Bhattacharjee, A. (2010), "Extending technology usage models to interactive hedonic technologies: a theoretical model and empirical test", *Information Systems Journal*, Vol. 20 No. 2, pp. 163-181.
- Linstone, H.A. and Turoff, M. (2011), "Delphi: a brief look backward and forward", *Technological Forecasting and Social Change*, Vol. 78 No. 9, pp. 1712-1719.
- Love, J. and Hirschheim, R. (2017), "Crowdsourcing of information systems research", *European Journal of Information Systems*, Vol. 26 No. 3, pp. 315-332.
- Lukyanenko, R., Wiersma, Y., Huber, B., Parsons, J., Wachinger, G. and Meldt, R. (2017), "Representing crowd knowledge: guidelines for conceptual modeling of user-generated content", *Journal of the Association for Information Systems*, Vol. 18 No. 4, pp. 297-339.
- Miller, G.A. (1956), "The magical number seven, plus or minus two: some limits on our capacity for processing information", *The Psychological Review*, Vol. 63 No. 2, pp. 81-97.
- Möller, K. (2017), "Questioning the theory-praxis gap in marketing – types and drivers of research implementation", *European Journal of Marketing*, Vol. 51 Nos 7/8, pp. 1163-1172.
- Moray, N. (1988), "A lattice theory of mental models of complex systems", *Engineering Psychology Research Laboratory*, University of Illinois, EPRL.
- Moray, N. (1990), "A lattice theory approach to the structure of mental models", *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, Vol. 327 No. 1241, pp. 577-583.

- Moray, N. (1996), "A taxonomy and theory of mental models", *Proceedings of the human factors and ergonomics society 40th annual meeting*, pp. 164-168.
- Moray, N. (1998), "Identifying mental models of complex human – machine systems", *International Journal of Industrial Ergonomics*, Vol. 22 Nos 4/5, pp. 293-297.
- Nadkarni, S. and Nah, F.F. (2003), "Aggregated causal maps: an approach to elicit and aggregate the knowledge of multiple experts", *Communications of the Association for Information Systems*, Vol. 12, pp. 406-436.
- Onwuegbuzie, A.J. and Collins, K. (2007), "A typology of mixed methods sampling designs in social science research", *The Qualitative Report*, Vol. 12 No. 2, p. 36, available at: <https://doi.org/10.1016/j.bbi.2003.12.001>
- Ossadnik, W., Kaspar, R.H. and Schinke, S.M. (2013), "Constructing a tailor-made performance management system supported by knowledge elicitation tools and dynamic modeling", *International Journal of Business Research and Management*, Vol. 4 No. 4, pp. 75-98.
- Page, S.E. (2008), *The Difference: how the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*, Princeton University Press, Princeton, NJ.
- Poetz, M.K. and Schreier, M. (2012), "The value of crowdsourcing: can users really compete with professionals in generating new product ideas?", *Journal of Product Innovation Management*, Vol. 29 No. 2, pp. 245-256.
- Regenwetter, M. and Grofman, B. (1998), "Approval voting, borda winners and condorcet winners: evidence from seven elections", *Management Science*, Vol. 44 No. 4, pp. 520-533, doi: [10.1287/mnsc.44.4.520](https://doi.org/10.1287/mnsc.44.4.520).
- Rowe, G. and Wright, G. (2011), "The Delphi technique: past, present and future prospects – introduction to the special issue", *Technological Forecasting and Social Change*, Vol. 78 No. 9, pp. 1487-1490.
- Scheliga, K., Friesike, S., Puschmann, C. and Fecher, B. (2018), "Setting up crowd science projects", *Public Understanding of Science*, Vol. 27 No. 5, pp. 515-534, doi: [10.1177/0963662516678514](https://doi.org/10.1177/0963662516678514).
- Schenk, E. and Guittard, C. (2009), "Crowdsourcing: what can be outsourced to the crowd, and why?", *Halshs-00439256*, Vol. 1, pp. 1-29.
- Shepherd, D.A. and Suddaby, R. (2017), "Theory building: a review and integration", *Journal of Management*, Vol. 43 No. 1, pp. 59-86.
- Surowiecki, J. (2004), *The Wisdom of Crowds*, 1st edition, Doubleday, New York, NY.
- Swedberg, R. (2016), "Before theory comes theorizing or how to make social science more interesting", *The British Journal of Sociology*, Vol. 67 No. 1, pp. 5-22.
- Weick, K.E. (1989), "Theory construction as disciplined imagination", *Academy of Management Review*, Vol. 14 No. 4, pp. 516-531.
- Yu, C., Chai, Y. and Liu, Y. (2018), "Literature review on collective intelligence: a crowd science perspective", *International Journal of Crowd Science*, Vol. 2 No. 1, pp. 64-73, doi: [10.1108/IJCS-08-2017-0013](https://doi.org/10.1108/IJCS-08-2017-0013).

### Corresponding author

Octavio González Aguilar can be contacted at: [ogonzalez@felizmenteverde.com](mailto:ogonzalez@felizmenteverde.com)

---

For instructions on how to order reprints of this article, please visit our website:

[www.emeraldgroupublishing.com/licensing/reprints.htm](http://www.emeraldgroupublishing.com/licensing/reprints.htm)

Or contact us for further details: [permissions@emeraldinsight.com](mailto:permissions@emeraldinsight.com)