Finding the future: Crowdsourcing versus the Delphi technique

Andrew Flostrand

Division of Industrial Marketing, Luleå University of Technology, Luleå, Sweden

Abstract When managers are unable to use quantifiable time series data to make forecasts or decide on uncertainties, they can either rely on their own intuition and judgment or resort to the insights of others. The Delphi technique is a well-known forecasting technique that relies on the pooled perspectives of experts to predict uncertain quantities or the outcomes of events. This relies on polling the opinions of experts, aggregating these opinions, feeding them back to the responding experts along with their own estimates, and having them repeat their judgment calls until some level of consensus is reached. More recently, however, the opinions of many others who are not experts have been sought on a range of topics in a loose assembly of similar techniques bundled under the title of crowdsourcing. This article compares Delphi and crowdsourcing as prediction and estimation tools for managers. It notes their differences and similarities, and provides a simple tool for executives to use in deciding whether or not to use these tools, and if so, which tool or combination of them will work best in a given situation.

© 2016 Kelley School of Business, Indiana University. Published by Elsevier Inc. All rights reserved.

1. Finding the future: Crowdsourcing versus the Delphi technique

The source of the expression, “It’s difficult to make predictions, especially about the future,” is uncertain. This saying has been attributed variously to Niels Bohr, Samuel Goldwyn, Yogi Berra, and Mark Twain. Its uncertainty notwithstanding, the maxim is undoubtedly true. Forecasting or predicting what will happen in the future is one of the most difficult tasks humans attempt to accomplish. Yet forecasting is an integral part of a manager’s role and executives are called upon constantly to make decisions and take positions now that will only come to fruition at some time in the future. Will markets develop and grow or wilt and decline? Will a new product succeed or fail? What effects will an advertising campaign have on customer perceptions? What will be the best name for a new product be? Will an investment now pay off in 5 or 10 years? What are the most important issues that an organization should focus on and how should these be prioritized?

Some forecasting methods extrapolate data from the past to predict what that same data will be like...
in the future. Time series analysis relies on various statistical techniques ranging from moving averages and simple linear regression to more sophisticated tools, such as the Box-Jenkins method (Box & Jenkins, 1970). Users of these methods presume that data patterns of the past will extend beyond the present. Whether they will remain stationary, fluctuate, increase, decline, or maintain can be determined by mathematical manipulation. When hard data is unavailable or when environmental conditions are very dynamic, forecasting resorts to opinions and judgments. Traditionally, experts have been sought after for these opinions. After all, who better to foretell the future within a particular domain than those who have truly experienced its past and live in its present? One of the best known and most widely used tools for polling the opinions of experts is the Delphi technique. However, more recently it has been contended that, in some contexts, a very large group of lay people without recognised expertise are collectively as adept as many experts at coming up with insightful answers. This approach, variously known as the wisdom of crowds (Surowiecki, 2005) or crowdsourcing (Howe, 2006), has recently gained prominence because large crowds of individuals can be easily accessed through technology. Within the broad concept of crowdsourcing, the subcategories of prediction markets (Graefe, Luckner, & Weinhardt, 2010; Wolfers & Zitzewitz, 2004) and innovation contests (Boudreau & Lakhani, 2013) provide familiar tools that countless organizations have used profitably. A simple request on a website to provide suggestions or answers, often fueled by tweets or reminders on a Facebook page, can rapidly lead to hundreds or even thousands of responses.

In this article, I describe and compare the Delphi technique and crowdsourcing activities, noting where they are similar and where they differ. I also discuss instances and opportunities where they might be used as forecasting tools and provide a decision tool for managers to aid their choice between these two useful techniques.

2. What is Delphi?

The Delphi technique, a method for polling a pool of experts, gets its name from the ancient oracle of Delphi. People brought questions to Delphi from all over ancient Greece in search for answers from the priestesses of Apollo. The usually cryptic answers predicted future events—whether a farmer should plant his crop or not or whether an empire should go to war and what the outcome would be. As a forecasting tool, the Delphi technique had its origins in the RAND Corporation in the 1950s, when the U.S. government was trying to estimate Soviet nuclear capacity. The corporation’s objective was to develop a technique to obtain the most reliable consensus of a group of experts (Helmer & Dalkey, 1962). It focused on tasks where reliable objective data was not available and must be estimated and trend analysis techniques could not be applied.

Linstone and Turoff (1975) describe Delphi as a method for structuring a group communication process, allowing a group of individuals to address a complex problem collectively. In a typical Delphi process, a group of experts in a particular field is polled on an issue of interest. A convenience sample of experts is used rather than an inferential survey; the emphasis is on expertise rather than on statistical representation. These individuals are asked to give their best estimates of uncertain events and phenomena on criteria, such as its importance in the future; the likelihood of its occurrence; and the potential impact on an organization, an industry, a country, or even the world. Their responses are then pooled and an aggregate response is calculated on each issue. The Delphi technique’s strength lies in providing a mechanism for these experts to converge on what might have begun as a diverse range of opinions to a narrower and more actionable range of consensus opinions so that in a second round, each expert is informed of the pooled group estimate on each issue and reminded of their own prior answer. They are then invited to reassess their judgment in light of the group perspective. This process is then reiterated in following rounds. Typically, groups achieve a broad consensus over three rounds, with some exercises not converging until the fifth round.

The Delphi technique has been used to poll and aggregate information from experts in a number of management disciplines. For example, Brancheau and Wetherbe (1987) and Brancheau, Janz, and Wetherbe (1996) polled chief information officers in large organizations to determine what they perceived to be the key issues in information systems management in the next 10 years. In three rounds, a clear prioritization ranking of the top 20 issues resolved into focus. In marketing, Ramaseshan and Pitt (1990) identified the most important distribution issues facing business-to-business marketers in Australia using a three-round mailed Delphi study. These authors showed how general consensus was achieved over three rounds by calculating a standard deviation on the importance rating of each issue over each round, as well as a total standard deviation for all of the issues over each of the three rounds. Between rounds one and two, the standard deviations decreased
noticably but there was only a small change between rounds two and three. This indicated that it was unlikely that moving to a fourth round of polling would result in noticeable differences. In the human resources field, Milkovich, Annoni, and Mahoney (1972) used Delphi to forecast short-term labor requirements by engaging eight experts to contribute predictions for the number of work hours that would be needed for an outlined project. In this study, the mean value of the estimates shifted substantially between the first and fifth rounds, and the range of estimates fell by half.

3. What is crowdsourcing?

In his book, Surowiecki (2005) tells the story of the TV game show Who Wants to be a Millionaire? Contestants have three lifelines in their bid to win a million dollars. When they are in doubt about the correct answer to a question, they can phone a friend, reduce the number of possible answers from four to two, or poll the audience for the right answer. Cynics might doubt the wisdom of asking a crowd of people who attend TV game shows for help and would prefer to call a friend at home who has good general knowledge. As it turns out, according to Surowiecki, calling a friend gives the correct answer 65% of the time, while polling the studio audience provides the right answer 91% of the time. The individual person in the audience may not be very smart. However, all the individuals in the crowd acting in concert turn out to be very wise indeed. Outsourcing wisdom to the crowd, also known as crowdsourcing, can be a very profitable strategy.

Bruggen, Lilien, and Kacker (2002) give an excellent explanation for why crowd wisdom works. Imagine that an individual’s opinion of a value is a combination of some amount of actual knowledge and some amount of random error (the ratio is not important, even if the random error dominates). An aggregation of a sufficient number of these individuals will result in the random error components cancelling each other out, permitting the true knowledge elements to remain.

Crowdsourcing is defined as the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people rather than from traditional employees or suppliers. This has been enhanced in the internet age by the ability to access large crowds of people easily, rapidly, and at almost no cost (Kietzmann, 2017). Prpić, Shukla, Kietzmann, and McCarthy (2015) distinguish between four different types of crowdsourcing. First, there is crowd voting. In its simplest form, an organization requests choices between alternatives from a crowd. The confidence in these choices can be captured using devices called decision markets. Second, there is idea crowdsourcing, in which an organization invites opinions for big or small questions and then evaluates the proposed ideas. The crowd can be encouraged to provide constructive input to an idea crowdsourcing effort by offering the contributors of the best ideas some form of reward, be it a tangible prize or simple recognition. Such activities are called innovation contests. Third, there is microtasking, in which a large task is broken down into very small components that can be easily and quickly performed by an individual and then combined into a completed task. Finally, there is solution crowdsourcing, where an organization invites and tests contributions for very specific problems and then adopts the best answer.

It is obvious that there are many differences between Delphi and crowdsourcing, but there are also considerable overlaps in how these techniques function. The next section compares Delphi and crowdsourcing as possible techniques organizations can use to forecast uncertain futures or to resolve problems or opportunities that might occur in the future.

4. Comparing Delphi and crowdsourcing

Managers in an organization can obviously make forecasts themselves, based on their own judgment. However, it is unlikely that an individual manager will have access to all the knowledge surrounding a particular issue and it is doubtful that a manager can make an unbiased judgement (Best, 1974). Much of the research into human judgment has shown that we process information poorly and are equally bad at aggregating different predictors (Slovic & Lichtenstein, 1971). Furthermore, while a good deal of the data required for decision making and forecasting is available from internal organizational sources, such as archives and accounting reports, this may not be relevant to some of the tasks at hand. Where the data is not available or is not entirely appropriate, managers might need to obtain it in the form of responses from a group. The group might be made up of experts or it may simply be a large crowd of individuals. In short, managers might need to choose between some form of Delphi technique and crowdsourcing. Table 1 illustrates the similarities and differences between the two techniques, which are discussed in the next two sections.
Table 1.  Delphi and crowdsourcing: Similarities and differences

<table>
<thead>
<tr>
<th>Similarities</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants are anonymous to one another</td>
<td>Relies on experts with explicit knowledge</td>
</tr>
<tr>
<td>Incentive to participate is offered in some form</td>
<td>Relies on non-expert lay people</td>
</tr>
<tr>
<td>Less expensive than using consultants</td>
<td>Ideally 10–18 experts (Okoli &amp; Pawlowski, 2004)</td>
</tr>
<tr>
<td>Impartiality and objectivity</td>
<td>Open to typically very large groups</td>
</tr>
<tr>
<td>Neither have checks for reliability</td>
<td>Multiple rounds of estimation, with feedback to the experts between rounds</td>
</tr>
<tr>
<td></td>
<td>Single round of estimation, no feedback</td>
</tr>
<tr>
<td>High quality data</td>
<td>Results of study is of interest to participants</td>
</tr>
<tr>
<td>Low non-response</td>
<td>Highly variable data quality</td>
</tr>
<tr>
<td>Participants not anonymous to researcher</td>
<td>Self-selected</td>
</tr>
<tr>
<td>Participants can’t be competitors</td>
<td>Participants are generally anonymous to researcher</td>
</tr>
<tr>
<td>Duration at least 3–5 days</td>
<td>Open ended</td>
</tr>
</tbody>
</table>

4.1. Similarities

There are significant similarities between the Delphi technique and crowdsourcing. Individual participants are anonymous to each other in both methods. This is an important advantage because it overcomes the problem of having too much communication—which can cause the group to be less intelligent (Surowiecki, 2005)—at a point where participants can be influenced in their judgments and decision making by other members of the group.

With both Delphi and crowdsourcing, some form of compensation or reward is offered. This may not necessarily be a monetary reward. For example, Delphi panel members can be motivated to participate by being granted access to the aggregated views of their peers or by being promised reading lists on the important issues they have considered (Ramaseshan & Pitt, 1990; Watson, 1990). Crowd members are often motivated simply by identifying with a community. This is the case of Threadless, an online t-shirt vendor that relies on a crowd to help improve and select designs for the firm to produce and sell on its website (Brabham, 2010).

In addition, both of these methods can be argued as superior to working with a consultant. Using Delphi or relying on a crowd will be less costly than engaging with consultants to address a problem in most cases. Allied to this point is that on some tasks, panels of experts or a crowd will produce better judgments than local wisdom and would also not have the agenda that some consultants bring to the table.

Finally, Delphi and crowdsourcing have the advantages of impartiality and objectivity. In both cases, the participants are removed from the problem or the issues. They are generally safe from the risk of confirmation bias or from the prospect of favoring judgments solely based on preconceptions (Kahnemann & Tversky, 1974).

4.2. Differences

While there are similarities between Delphi and crowdsourcing, there are also a number of important differences. First, there are distinct differences between the participants. In the case of Delphi, participants are carefully selected individuals who have explicit knowledge and experience of the phenomenon to be studied. Put simply, they are experts. In some instances this will be extremely important, especially where some kind of technical expertise is required to fully understand questions and make judgments. In crowdsourcing, the level of explicit knowledge and expertise is not considered important. Like panning for gold, somewhere out in that large crowd there are valuable flakes of knowledge that crowdsourcing will separate from the grains of irrelevance.

The second difference between Delphi and crowdsourcing is the survey size requirement. In the former, one relies on a small group, while in the latter the group must be very large in order to gain the knowledge the firm is looking for. In each technique, the opinion of an individual participant is an unknown mix of knowledge and random error.
However, these ratios are distinctly different, as is shown in Figure 1. In the case of Delphi, the large knowledge contribution per expert permits a small number to be used in a study. In contrast, the large degree of random error in the case of an individual crowdsourcing participant’s opinion necessitates that a very large group of participants be used in a study.

Third, the typical Delphi study is conducted over at least two, usually three, and sometimes more iterations or rounds. This happens so experts will adequately understand, consider, and reconsider the aggregated opinions of their peers. In the case of crowdsourcing, the study is typically over a single round only and no further feedback from participants is sought. Fourth, the results of a Delphi study will inevitably be of interest to an expert participant. This is not necessarily the case in crowdsourcing, where the study might either be so short as to not provide detailed information to participants, or it may not be of great interest to them due to their lack of expertise.

Fifth, while in the case of Delphi there will be no bad data or incentives to mislead or corrupt the research, the same cannot be said for crowdsourcing. Exogenous factors can sometimes cause crowdsourcing participants to deliberately hijack an exercise. The recent crowdsourcing project designed to name a British research ship is a case in point. The National Environment Research Council invited people to suggest and vote for names for the new vessel. A BBC Radio presenter jokingly suggested on air that the £200 million Arctic explorer be named Boaty McBoatface. When the results of the exercise came in, the overwhelming majority of votes were for that name, by a margin of four to one over the next choice. Sadly, in the opinion of the author, this name was not ultimately selected to name the ship but it did survive by being given to a remote control submersible on board. Other recent examples (Ellis-Peterson, 2016) include crowdsourcing exercises by Greenpeace to name a humpback whale (Mr. Splashy Pants), and the city of Austin’s Solid Waste Services Department (The Fred Durst Society of the Humanities and Arts).

Sixth, there is usually a low non-response rate to Delphi studies, because some methodologies solicit participation before the study even starts and others encourage those who did not participate in the first round to participate in the second, and so forth (Watson, 1990). In the case of crowdsourcing, participants self-select—they either decide to click on a website or not. A consequence of this is that researchers do not really know beforehand how many participants they will get, whereas in Delphi studies there is a reasonable indication.

Seventh, and allied to the previous point, in the case of Delphi, the experts are not anonymous to the researcher, who needs to be able to give them intra-round feedback. This means that the researcher can contact individual experts directly for further insights should they be required. In crowdsourcing, participants typically are anonymous to the researcher, and so the researcher is not able to contact them for further feedback. This is compounded by the fact that participants can indulge in multiple-voting behavior, and even when they are required to disclose names and email addresses, they can use fake contact details.

Eighth, the ultimate use of the information gained constrains who can participate. In Delphi studies, for example, competitors as experts probably will not be included in a study in order to protect proprietary information and generate a competitive advantage. The public at large are the participants in crowdsourcing studies, and these people in general do not have vested interests in the study or its results. Finally, with regard to the duration of the studies, a Delphi study will take between 3 and 5 days at least in order for experts to think carefully about the issues, assess them, give feedback, and then absorb and respond to the feedback on the subsequent round. In crowdsourcing, the study is of indeterminate duration. A quick poll can take a day or less, while other exercises are ongoing, such as the movie rating website Rotten
The similarities and the significant differences between Delphi and crowdsourcing mean that managers might need assistance in choosing the most appropriate method. The next section suggests a tool that will assist this decision making.

5. Deciding between Delphi and crowdsourcing

Having established the functional definitions of the two tools and the differences between the two, and having defined a specific question to be addressed, a manager has four options: (1) use neither Delphi nor crowdsourcing, (2) use only Delphi, (3) use only crowdsourcing, or (4) use both Delphi and crowdsourcing. The manager makes the decision that is appropriate for the character of the question and the environmental context of the forecasting problem to be addressed. Figure 2 shows the four paths and the research characteristics that make each appropriate.

5.1. Neither Delphi nor crowdsourcing

The first path—using neither Delphi nor crowdsourcing—may be the appropriate choice for a number of reasons, such as when values of interest can be estimated using straightforward statistical forecasting techniques. In cases where the information already available is as good as the knowledge scattered across either an expert or lay community outside of the firm, using either of these two methods would gain no advantage. Instead, tools such as time series analysis may be appropriate.

In addition, neither the Delphi technique nor crowdsourcing would work well when the value of the answer does not justify the investment in time to use either tool or when proprietary of strategically sensitive information could be revealed to inappropriate recipients in the course of the exercise.

5.2. Delphi alone

The second path is to use only the Delphi technique. This is best when the question is ‘messy,’ in the sense that there are many—some potentially unknown—factors contributing to the values or rankings being sought by the forecaster. When the sophistication of the question requires significant levels of expertise in the subject matter or if it is not reasonable to assume that a lay person would be able to fully comprehend or meaningfully form an opinion or estimation, then Delphi is appropriate and crowdsourcing is not.

Delphi works best when the forecaster has reasonable access to a community of experts who are willing participants. It may be easy to imagine a group of experts in a given field, but Delphi is only an available option when these experts individually have some interest in knowing the results of the research the forecaster is proposing and have no aversions to sharing the knowledge they have. Delphi is also optimal when the researcher is seeking to identify and prioritize issues and when the forecaster requires estimations of certainty. The first use of Delphi by the Rand Corporation (Helmer & Dalkey, 1962) is an example of this type of goal setting, as it sought to accurately assess the specific sources of nuclear security threats and their relative priority.

Delphi requires a significant time investment. While the data collection phase itself is just a matter of mail exchanges and value consolidating so it passes quickly, the preliminary actions take much more time. Sourcing a group of 10–18 qualified experts who are willing to contribute time on multiple rounds of questioning may take days, even weeks, of effort and that is before the data collection phases begin. If the knowledge the researcher is seeking would only be valuable if known very soon, Delphi would not be viable and the researcher may have to resort to crowdsourcing.

5.3. Crowdsourcing alone

The third path is when crowdsourcing alone is the best forecasting tool. The sole use of crowdsourcing
can be optimal in several research situations. When the research question is straightforward, easily defined, and can be reasonably expected to be understood fully by the general public, crowdsourcing is the best option. It is also the best method when value is being placed on diversity of opinion, rather than the aggregated expertise of a few. An example of this would be a company inviting creative idea submissions to be considered by a policy-making body. Note that in the case of the British research ship, the suggestion of the name Boaty McBoatface was not the problem; it was the combined voting mechanism that created the incentive to choose it. A less stressful, if less humorous, process would have been to first solicit anonymous submissions of names and then curate a subset list of these that were judged appropriate for public voting in a second phase. This is similar to the process followed by Threadless.

Crowdsourcing is beneficial when the research question aims to measure and forecast public sentiment, such as inviting the public to vote for favorite contestants in a game show. It should be noted that if researchers are seeking to measure public sentiment accurately, a statistically rigorous survey will be preferable to crowdsourcing. But when the researcher needs answers within a few hours or days, a snapshot of public opinion will assist in this. For some cases, it might be appropriate to use ready-to-go crowds such as Amazon Mechanical Turk, where a client pays a small fee for each pre-registered participant’s contribution.

Crowdsourcing is optimal when field experts, like those used in the Delphi technique, may be expected to have a structural bias or inappropriately narrow interpretation of the question. Research conducted by Wood and Lynch (2002) demonstrates how prior knowledge can manifest as a negative impact to new learning. Likewise, prediction of consumer behaviors also showed that expertise showed no measured advantage over novices (Armstrong, 1991). When tiny bits of information that cohere into an integrated picture of knowledge are widely spread, crowdsourcing can be expected to gather these, tapping subtle insights and esoteric knowledge from a large and diverse public.

5.4. Both Delphi and crowdsourcing on same questions

The fourth and last path is to employ both a Delphi study and a crowdsourcing exercise for the same research questions. This is the correct path when the significant investment of time and labor to do both is justified by the value of the research. Both methods are also optimal when the insights of experts and the wisdom of crowds are equally important to decisions made with the research findings or the forecasts. This may often be the case for elected bodies when public inclusion is appropriate.

When part of the research question is to measure any differences that may exist between the two data collection options, both methods should be used. For example, if it were found that the predictions of a small group of experts were different from the lay public, it would be insightful for decision makers to explore both perspectives. What has caused people at large to be wrong about something, and what can be done about that? On the other hand, have a group of experts missed something important, and if so, why?

6. Conclusion

This article focused on the forecasting choices open to managers when they prefer not to, or do not have the option to, use quantitative data in some kind of time series analysis. Managers can obviously, and quite simple, rely on their own judgment. In many cases however, they might prefer to trust, or at least have their judgments supported by, expertise or mass public opinion. Delphi and crowdsourcing are two readily available and relatively simple tools managers can use to obtain this support. Delphi has the advantage of using experts to provide insights and is relatively simple to implement and administer. It has the advantage of condensing expert opinion on a matter that requires expertise.

Crowdsourcing is a newer phenomenon. Its use has been accelerated by the mass communication tools made available on the internet. It has accentuated the fact that while the average individual might not be very knowledgeable, a group of average individuals can hold exceptional wisdom. Numerous companies and websites have been established to exploit the wisdom that crowdsourcing brings to forecasting, including Cultivate Labs1 and Lumenogic2.

More recently, researchers have begun to explore the possibility of combining Delphi and crowdsourcing into a suite of prediction techniques.3 These suggestions should result in adding the tools to

---

2 http://www.lumenogic.com/www/technology.html
3 For example, https://yrjomyllya.wordpress.com/2013/05/14/the-crowdsourcing-delphi-combining-the-delphi-methodology-and-crowdsourcing-techniques/
harness the wisdom of experts and of the crowd to a manager’s toolkit, including how best to choose which method to use individually and how to combine the methods for optimal results.

References


