Assessing Delphi panel composition for strategic foresight — A comparison of panels based on company-internal and external participants

Bernadette Förster *, Heiko von der Gracht

Institute for Futures Studies and Knowledge Management (IFK), EBS Business School, Konrad-Adenauer-Ring 15, 65187 Wiesbaden, Germany

A R T I C L E   I N F O

A B S T R A C T

Decision makers seek advice from others in order to make more accurate decisions, justify these decisions, and share responsibility. The Delphi survey technique finds broad acceptance as a decision support and forecasting tool. Recent research has discussed the composition of Delphi panels and whether company internal or external panelists should be consulted for strategic foresight. We make a contribution to this discussion by investigating whether internal and external participants of Delphi studies lead to differing results and how the differences can be utilized by decision makers. We consider differences that might be inherent not only to quantitative but also to qualitative Delphi data. Results of our research reveal that there are several significant differences between the two panels' evaluations, which lead to varying consultation practices for different strategic purposes.

1. Introduction

When managers evaluate long-term strategic choices for their firms, the possible development of various issues related to the company and its environment have to be considered, such as changing customer demand, technological processes, or demographic change. In such contexts, uncertainty is often high. Furthermore, developments outside the company's focus might influence the business area [1]. That is why in situations of perceived environmental uncertainty, decision makers frequently consult other knowledgeable persons to increase decision accuracy, justify their decisions, and spread the responsibility [2–4]. The higher the uncertainty of a future development and the more events can affect the outcome of an event, the more the evaluations regarding the event will differ and the opinions of the individuals will diverge [5]. An instrument to provide decision makers with orientation is the Delphi method [6]. It is particularly applied in judgmental forecasting and corporate foresight where companies strive to generate forecasts about relevant issues in order to establish a more profound basis for strategic decisions [7]. The Delphi method is sometimes referred to as a crowdsourcing technique [8]. However, in contrast to most other crowdsourcing techniques, such as prediction markets [9], which aim at surveying a great number of people including laypersons, the Delphi survey method aims at surveying a limited group of knowledgeable people from a certain field [10,11].

While the Delphi method has proven its validity in many research and business contexts [12,13], the appropriate panel composition remains a controversial issue for critical reflection [14,15]. In our research, we focus on what types of experts should participate in a Delphi survey, and thereby provide strategists with advice, in order to orient decisions to a particular business context. In general, previous literature has distinguished between internal advice, or guidance from persons within the same company, and external advice, or assistance from persons outside the company [16].

In addition to this issue, research has shown that a lack of diversity among the participants in a Delphi panel could induce biased results (e.g. [14,17,18]).[1] More precisely, in

* Corresponding author. Tel.: +49 611 7102 2101; fax: +49 611 7102 10 2101.
  E-mail addresses: bernadette.foerster@web.de (B. Förster),
Heiko.vonderGracht@ebs.edu (H. von der Gracht).

---

[1] For the scope of this paper, we use the terms diversity and heterogeneity as synonyms.

heterogeneous panels, in which panelists are likely to have contradictory opinions, for instance concerning desirability, some individuals provide evaluations that are above the real value of estimation, while others make judgments in the opposite direction below the real value. Therefore, the estimations "bracket" the real value: averaging the values of the estimations leads to a value closer to the real value. Such bracketing diminishes errors [19,20]. In homogeneous panels, participants are likely to have similar opinions and therefore biases. Their evaluations are either above or below the mean, and do not bracket the real value.

The dimensions of heterogeneity for the purposes of a Delphi survey are manifold. For instance, individuals can differ in their age, gender, cultural and educational background, knowledge base, profession, values, attitudes, or tenure [21–23]. In our research, we particularly focus on the diversity that arises from selecting participants from different companies, where individuals work in diverse fields and perform differing tasks, resulting in professional background and knowledge base heterogeneity [24–26]. We explicitly do not consider the degree of diversity within one company to result from the versatility of tasks that the employees perform.

We conducted two separate Delphi surveys on an identical set of projections. The thematic scope of our studies was the European logistics service industry concerning possible future developments, such as collaborations, technological advancements, innovations, and customer expectations. The projections were quantitatively assessed along three dimensions: probability of occurrence, likely impact on the industry, and desirability. Moreover, optional written arguments of qualitative justification could be provided.

Managers from one large international company composed the internal panel. Whereas experts from different companies offering the same services as the reference company of the internal panel, as well as the two primary stakeholder groups "customer companies" and "suppliers", and panelists from academia comprised the external panel. Therefore, the two Delphi panels had different levels of heterogeneity with respect to their professional backgrounds.

Our aim was to substantiate recommendations for conducting internal or external Delphi surveys when the results are used for strategic foresight and decision making. Therefore, we formulated the research questions: *Do company-internal Delphi panels and external Delphi panels lead to differing quantitative and qualitative results? If so, how do they differ? Which panel should be consulted for which foresight activity?*

In the remainder of the paper, we review current literature on both advice taking and diversity in group decision making processes, explain the two Delphi surveys conducted in more detail, illustrate the data analysis of quantitative and qualitative data, and discuss the significance of the results in the context of the previously reviewed literature. We conclude with recommendations for selecting Delphi panel participants for different purposes, point out limitations that are inherent to our research, and indicate possible future research directions.

2. Theory and proposition development

The rationality in decision making is bounded because the information available is limited and asymmetries in information might exist [27]. We aim to provide recommendations

for decision makers, as to whether they should consult internal or external sources for guidance in Delphi surveys. Power and social identity theories have demonstrated that decision makers prefer either one or the other type of advice [28,29]. According to previous literature, various pros and cons exist for the two sources of information, as depicted in Table 1.

The choice between internal and external advisors becomes relevant when a Delphi panel should be recruited for a survey to support decision making. Many researchers argue that the panel of a Delphi survey should be composed of participants with great expertise in order to achieve accurate forecasts [36,37]. However, social psychology has demonstrated that expertise itself may not be sufficient to obtain accurate judgment but should be combined with diversity [21,38]. Diversity in a panel proves to be beneficial for obtaining useful results in several ways: the advice comes from multiple independent sources; the panelists have different skills and points of view. The definition of diversity, or heterogeneity, usually involves demographic characteristics as well as aspects related to the individuals' professional experience [21–23].

The fact that heterogeneity and its benefits are also important for Delphi studies has already been outlined by several researchers [14,18,21,39,40]. Heterogeneous Delphi panels ensure a reduction in polarization [21] and desirability bias [18], and enable bracketing. The latter is especially useful for Delphi studies as the feedback which the participants receive includes average values [41], a reasonable way to consider multiple sources of information when probability forecasts are made [19,42]. Such average values become more accurate through bracketing.

Since the results of a Delphi study are frequently used for advice in the strategic decision process, the exchange of diverse perspectives is important. Previous literature highlights that although taking advice increases the decision accuracy in general [43,44], when the advice comes from multiple sources, this effect is even greater. Therefore, many scholars recommend using multiple independent sources [19,42,45]. Moreover, more detailed information, such as the reasons for a certain evaluation, especially if it provides ambivalent perspectives, further improves the accuracy of judgments. From the authors' perspective, these elements are also core characteristics of the Delphi technique.

In Delphi panels, the composition can be varied by the scope and degree of heterogeneity in professional background, among others. The panel can either have an internal or external scope and the companies from which the panel is composed of can either be specialized in one operational area or have multiple operational foci. We developed a framework for the differentiation of heterogeneity in Delphi panels. For our purposes, heterogeneity is judged by professional background and the knowledge base that can be associated with the company's field of operation. Thus, we narrow our research to these two dimensions for the purpose of clarity and conciseness. Further dimensions that induce heterogeneity, such as age, tenure, or education, are excluded and beyond the scope of this research. The potential combinations of scope and field of operation are represented in Fig. 1.

We conducted two Delphi surveys in the European logistics service industry: one survey included an internal panel solely consisting of managers from one large international company, a highly specialized logistics service provider

Table 1
Comparison of internal and external advice.

Internal advice		External advice	
Pro	Contra	Pro	Contra
• Internal sources are aware of company's capabilities: strengths and weaknesses [30]	• Competition within company [29]: Seeking advice of a colleague might be beneficial to reputation of advisor rather than the consulting party [16]	• Insights gained beyond borders of firm might be novel to company [28]	• Not frequently practiced since some decision makers perceive internal knowledge to be superior to external knowledge [28]
• Seeking internal advice might foster an atmosphere of trust that encourages employees to share information and to develop new ideas [28]	• Decision makers are more critical: internal information can be better evaluated than external information [29]	• Scarce and therefore valuable information [29]	• False information harder to identify [29]
• Easy to obtain and analyze [16,29]	• In general, regarded to be less valuable than external advice [16]	• Persons seeking outside information rewarded for efforts with improved status [16]	• Danger that information from external sources is assumed to be true knowledge: accuracy is rarely challenged [31]
• Generated information is kept internal [32]	• Adhering to internal sources reinforces existing mindset and blind areas [1]	• "*Organizational myths*" and "*artificial and illusory level of comfort*" are challenged by external sources of information [33, p. 66] • Involve stakeholders [34,35] • External stakeholders have a birds-eye perspective and focus on different aspects because of different backgrounds [1,21].	

(corresponding to quadrant I); the other survey consulted an external panel consisting of panelists from the entire focus industry, therefore including not only logistics service providers but also three different stakeholders: customers, suppliers, and academics (corresponding to quadrant IV). Even though the external Delphi panel can be divided into its subgroups (each subgroup corresponding to quadrant III), we focus on the comparison of the internal and external panels in order to compare how Delphi panel constitution leads to varying survey results of the same projections.

2.1. Quantitative comparison of homogeneous internal (quadrant I) versus heterogeneous external (quadrant IV) panels

Although our internal Delphi panel is highly diversified with regard to employment positions, departments, and geographical areas, it is homogenous according to our key criteria for heterogeneity, namely the operational focus of the employer. The external panel consists of numerous companies with multiple areas of operation, resulting in a higher degree of professional diversity. Due to the different professions, areas of operation, and levels of experience in the external panel, we expect the external panelists to have different kinds of knowledge in comparison to the internal panelists and to focus on different aspects of the presented projections [1,29]. As a result, we expect the quantitative estimations of the external panel to vary from those provided by the internal panel. Furthermore, the heterogeneity in the external panel is expected to lead to a broader range of evaluations, while the internal panelists will have similar estimations.

Proposition 1. *The quantitative estimates of the homogeneous internal and the heterogeneous external Delphi survey differ significantly from each other.*

Proposition 1a. *The shape of data distribution of the two panels differs.*

Proposition 1b. *The mean values of the evaluations differ.*

Proposition 1c. *The data of the external panel is more dispersed than the internal panel.*

Proposition 1d. *Due to the higher degree of novelty in the feedback which participants receive, the panelists of the external panel change their evaluations more often.*

2.2. Qualitative comparison of homogeneous internal (quadrant I) versus heterogeneous external (quadrant IV) panels

In the Delphi process, the participants have the opportunity to provide written arguments to support their evaluations. With an internal panel, the participants know that their arguments remain within the company: they can provide confidential information since it is kept internally [24]. One the one hand, information that is shared by the individuals in a group discussion process is introduced and repeated more readily than novel knowledge [46,47]. On the other hand, if the individuals are linked to each other socially, they are more likely to carry the costs for introducing novel knowledge [22]. Therefore, we expect the internal panelists to exchange knowledge that is common to the majority of employees of the company and to introduce confidential information or their work experiences in the discussion process. They are expected to be highly focused on the company's field of operation, while disregarding the points of view of other actors in the industry. Since employees regularly share knowledge, they might use catchwords frequently: they expect their colleagues to understand the meaning without lengthy explanations. However, at the same time, the commitment within such a panel is likely

		Field of operation	
		Homogeneous	Heterogeneous
Scope	Internal	I Single specialist company	II Single company operating in multiple industries
	External	III Single industry panel	IV Multiple companies with different operational foci

Fig. 1. Panel classification.

to be very high, so that they are expected to use detailed explanations of their reasoning when they find it necessary.

The external participants are expected to introduce insights from their professional area, even though they are not likely to reveal detailed information. We expect them to introduce a high variety of information and a high level of uniqueness. Their commitment can be anticipated to be lower than the commitment of the internal panelists. Therefore, we expect external participants to provide less and shorter arguments, and invest less effort in formulation. However, we do not regard the length of the argument to be an indicator of its quality; shorter statements might be formulated more precisely. External panelists will not likely share confidential information or work experiences but rather point out observable trends in their professional area.

Proposition 2. *The qualitative arguments of the homogeneous internal and the heterogeneous external Delphi survey differ significantly from each other.*

Proposition 2a. *The external panel provides a similar number of arguments for the different dimensions while the amount of arguments provided by internal panelists tends towards one dimension.*

Proposition 2b. *The arguments provided by the external panelists are more unique.*

Proposition 2c. *The kinds of arguments provided by the two panels differ.*

Proposition 2d. *The content of the arguments of the two panels differs.*

3. Methodology

3.1. Industry selection

The reference company we selected for our internal Delphi panel is a logistics service provider with headquarters in Germany, which coordinates business activities throughout Europe. It is one of the 20 largest contract logistics companies in Europe in terms of turnover. The logistics service sector has several characteristics that are typical for most industries today. Dynamic markets and a high degree of competition, environmental regulations, and economic efficiency demands necessitate flexible processes and quickly responding management [48,49].

This reference company transports perishable groceries, among other items. Therefore, they are keen to fulfill additional hygiene and transport temperature requirements in order to avoid health risks and subsequent reputational damage. Since the company has a close cooperative relationship with its customer companies, it aspires to maintain transparent sustainability efforts and conducts innovative, customized activities to remain highly efficient. In order to become better prepared for the challenges of the future, this company is currently undergoing paradigm change: the company aims to become a learning organization and implement a scenario culture within the company (for learning organizations, see e.g. [50]). Due to this situation, it may be useful to engage in activities that broaden a company's mindset and question current world views to extend the tacit knowledge of the company [1,51]. Moreover, the results of the Delphi survey at hand can be utilized in scenario planning and are assumed to be successful as long as the stage of defensive avoidance behavior has not been reached by company management yet [52]. Since this company was in a stage of change, the panelists were highly motivated to participate in the research project and to collaborate with the research facilitators. We selected this specific company and environment since numerous stakeholders might have different points of view regarding the future of the logistics service sector, for example competitors, suppliers (i.e. truck manufacturers) as well as customers of the logistics service providers (i.e. restaurants), and related associations. The importance of hygienic transport, short delivery times, and environmentally friendly transport are important factors for stakeholders, but as soon as the cost factor is introduced the opinions differ. The variety of companies acting in the logistics service sector and the broad range of potential participants from which the panelists can be recruited for the external panel are additional benefits of the selected industry.

3.2. Panel selection

The internal panel of the Delphi survey consisted of European managers, representing different positions within the company: the President of the European headquarters, Vice President of freight management, Senior Vice Presidents for different logistic regions, Senior Managers of transport, or Managing Directors of different countries. Therefore, there was a high degree of diversity due to various cultural and professional backgrounds. The panelists were identified via the company's Intranet and job descriptions. In total, 65 managers at a minimum of senior management level were identified as

potential participants for the survey. These potential participants received advanced information about the survey from the project's steering committee and a personalized link to the online Delphi survey platform. Fifty-seven managers, equaling a response rate of 87.7%, completed the survey, which proves a high degree of interest in the research project. We can assume that the panel is representative due to its high response rate within a short survey period. Thus, the likelihood of a non-response bias is very low.

For the external panel of the Delphi survey, we investigated the value chain of the reference company's industry and identified potential categories of stakeholders. The case company provided information concerning customers, competitors, suppliers, associations, and academia based on internal contact databases. Moreover, our research team identified companies of interest via trade dictionaries and company websites. Participants of a Delphi survey should have a level of expertise in the topic, which can be determined, among others, by the level of management, the degree of job specialization, and the academic or educational background of the individual [53]. The positions of the invited participants depended on the company and its internal labels for management positions. For example, knowledgeable participants for our external panel from the subgroup of logistics service providers included CEOs, country managers, or senior directors, and for the subgroup of customers sales directors, supply chain directors, purchasing managers, or quality managers. Supplier participants comprised CEOs, managing directors, or supply chain directors. Furthermore, we invited consultants who worked either for logistics service providers or for their customer companies, as well as academics from the supply chain services and the logistics fields. Overall, 449 potential external participants were identified and contacted personally and were provided with a personalized invitation link to the Delphi study. Eighty-eight actually completed the survey. As the survey was market oriented, the majority of participants were customers (distribution of participants into the subgroups compare Table 2). The final response rate was 19.6%.

Compared to the internal panel, the response rate of the external panel was relatively low, possibly due to the target group being senior level managers of the industry that are frequently asked to participate in market research. Previous management research surveying senior managers yield comparable response rates [54,55]. In order to ensure that the low response rate does not limit generalizability of results [56], we checked for non-response bias, as recommended by scholars for mail and Internet surveys [57,58]. It is common to compare the evaluations of early respondents with those of late respondents [58]. The non-parametric Mann–Whitney-U-test applied to control for significant differences among early and late respondents for the initial assessments of estimated probability does not reveal significant differences. We therefore can assume that the respondents do not differ from the non-respondents.

3.3. Variables

The development of projections is a crucial process in Delphi studies since it directly impacts the validity and reliability of the study [59,60]. The Delphi facilitator team consisted of four senior researchers competent in Delphi research as well as in the reference company's industry. We followed a rigorous

process along 4 phases. First, the four members of the facilitator team, as well as a corporate team of three key decision makers of the reference company (including the European president), compiled an individual list of strategic issues for the industry's future. After consolidation, 134 strategic issues were identified, which were partly grounded in literature and partly based on experience in practice. Second, a workshop was conducted to further consolidate the list of issues. Redundancies were eliminated and the issues were further aggregated. The resulting list of 62 strategic issues was structured according to Porter's Five Forces Model [61] and prioritized by their expected impact in an open group discussion process. Prioritization limited the maximum number of projections to 20 in order to avoid research fatigue among Delphi panelists and thereby increase data validity [62]. Third, the facilitator team developed a draft of 25 projections following established guidelines for projection formulation and based on the 20 strategic issues [53]. The projections were short, concise and provoking, but allowed for broad discussion to cover all issues already eliminated in the consolidation process. In joint discussion with the corporate team, a final set of 18 projections was selected and the text formulation refined. Fourth, a test group of 6 senior managers, including representatives from the reference company's customers as well as suppliers, was formed in order to test the projection set for face and content validity as well as understanding. Their feedback resulted in further refinement of wording. During the Delphi survey process, we presented the final set of 18 Delphi projections to both internal and external Delphi survey panels (see Appendix Table A1). These projections included four focus areas: customer expectations, collaboration, technology, and innovations in the logistics service sector until 2020. The participants were invited to evaluate each projection regarding its expected probability of occurrence (on a scale from 0 to 100%), impact on industry in case of occurrence, and desirability to occur (both rated on a 5-point Likert scale). Assessments of 1 represented a low impact (undesirability) whereas assessments of 5 equal a high impact (desirability). For each evaluated dimension, the participants had the opportunity to provide reasons for their decisions in writing. In a tutorial, the participants could learn about the usage of the Delphi platform. The questionnaire design and the Delphi tools were introduced and examples on the surveyed dimensions and their evaluation were provided.

3.4. The Delphi process

The Delphi method is an interactive, iterative group communication process, in which participants evaluate projections about the future inter alia regarding their probability of occurrence, impact, feasibility, or rank a factor list according to the factor's potential [41,63,64]. The survey process is anonymous and participants receive feedback about their evaluations in relation to the other participants' evaluations. Furthermore, they have the opportunity to reassess their initial evaluations based on the group knowledge. Since its development by the RAND Corporation in the middle of the 20th century [65] the Delphi technique and related methods on the "wisdom of crowds" have become prevalent across all disciplines in research and practice [66], particularly in foresight and decision-making support [36,37]. Not only foresight in general [67], but also the Delphi method in particular, has been

Table 2
Comparison of internal and external panel information.

	External panel	Internal panel
Projections	Identical set of 18 projections concerning 4 strategic topics of the logistics service sector[a]	
Panelists	88 managers from 15 European countries in companies/institutes related to specific logistics service sector: 44 (potential) customers (50%), 16 logistics service providers (18%), 9 suppliers (10%), 19 academics and association representatives (22%)	57 managers from reference company (logistics service provider) with headquarters in Germany
Response rate	19.6%	87.7%
Sample period	March to April (7 weeks)	May (2 weeks)
Estimated probability [total/per person]	212/2.4	164/2.9
Estimated impact [total/per person]	108/1.2	92/1.6
Estimated desirability [total/per person]	130/1.5	131/2.3
Average number of logins	3.19; SD = 0.91	3.25; SD = 0.61
Number of written arguments [total/per person]	746/8.5	859/15.1
Average IQR of probability ratings (SD)	21.4 (9.6)	18.3 (6.9)

SD = standard deviation.
IQR = interquartile range.
[a] Customer expectations, collaboration, technology and innovation, sustainability.

criticized over the past decades [68,69]. Nevertheless, during decades of application, research supports the reliability and validity of the Delphi technique if it is conducted under rigorous standards [12,13]. The broad usage of the Delphi technique in research today demonstrates its popularity [70]. Literature reviews of different Delphi applications have been provided by Gupta and Clarke [71] or Nowack and his colleagues [72].

During the past decades, modifications to the classical Delphi survey have been made [15,73,74]. For our research, we applied an online, real-time Delphi survey format [41] that follows the rationale introduced by Gordon and Pease [75]. After each projection's first round assessment, our panelists received feedback almost immediately, or in real-time, about other panelists' evaluations. This method overcomes the main short-comings of the paper-and-pencil Delphi survey method: long time spans for the feedback loop, leading to response fatigue and high drop-out rates. The real-time Delphi method provides the same favorable characteristics of a classical Delphi survey, namely anonymity, controlled feedback, and summarized group response [36,76]. The Delphi portal was open for several weeks so that the participants could visit it as often as they liked to reassess their evaluations.

The two Delphi surveys were conducted sequentially: Directly after the external panel survey concluded, the internal company survey started. Information of the external panel was not provided to the internal panel; the surveys were fully independent of each other. In the following Table 2, we illustrate the key characteristics of the two panels.

The survey period for the internal Delphi study could be reduced to two weeks because the managers from the company responded faster and logged into the Delphi survey more often within a shorter period of time. The commitment of the company's top management and its interest in the survey results were very high. To give the external participants sufficient opportunity to re-enter the platform, the survey period was prolonged. Finally, the average number of logins of the two panels and therefore the number of survey rounds in which the panelists participated were almost equal for both panels.

Experience has shown that the majority of arguments are provided after the first round. The internal panel provided more written arguments despite the shorter survey period. As an indicator for consensus among the panels for expected

probability, we provide the average interquartile ranges and the respective standard deviation. For both panels, we split the panels into two halves randomly and compared the results to measure the reliability of results. No differences were found in this comparison, indicating reliability. An indicator for the stability of the data over the Delphi rounds [77] cannot be applied, as the Delphi survey was conducted in a real-time format and had no clearly defined rounds.

4. Analysis and results

4.1. Analysis and results of quantitative data

Previous researchers comparing two or more Delphi panels displayed the means of the panels [78] either via applied t-tests [79], t-tests that assume normality and Mann–Whitney U tests [24], or Chi-square tests [80], depending on the type of questions and the obtained data.

To identify the appropriate test for our data set, we started by analyzing the shapes of our data distributions. First a Kolmogorov–Smirnov test is conducted which reveals that the data at hand is not normally distributed. Second, we conducted Levene's test to test for equality of the variances of the two panels. The test finds three projections, for which the variances are not equal (projection 1: $F_{(1, 143)} = 11.375$, $\rho = 0.001$; projection 3: $F_{(1, 143)} = 8.115$, $\rho = 0.005$, projection 10: $F_{(1, 143)} = 4.968$, $\rho = 0.027$). Likewise, a subsequent Chi-square test for comparing the similarity of distribution identifies projections, for which the distribution is not equal (projection 1: $X^2(17) = 27.722$, $\rho = 0.048$; projection 2: $X^2(15) = 27.351$, $\rho = 0.026$; projection 3: $X^2(18) = 30.904$, $\rho = 0.030$; projection 12: $X^2(12) = 23.142$, $\rho = 0.027$). No particular pattern, such as bipolar distributions, could be found to provide insightful explanation for the differences in distribution. Levene's test and the Chi-square test support our Proposition 1a that the data obtained from the two different panels does not have the same distribution.

Since the Kolmogorov–Smirnov test showed that the data at hand is not normally distributed, we applied a Mann–Whitney U test for comparison of the means because it does not assume normally distributed values. As the previously mentioned tests revealed that the data sets did not have the

same shape, the Mann–Whitney U test compares the ranks of the means [81]. Therefore, a closer investigation of the distributional differences becomes necessary to obtain more information on the differences of the two panel data sets [82]. The comparison of the internal and external panels was conducted for initial and final evaluations of expected probability of occurrence, impact, and desirability.

First, the mean values are compared by investigating the initial assessments for expected probability of occurrence. For the initial assessments, differences between the panels are based on the varying levels of information that the participants have at the outset. Feedback from co-participants is not yet provided and therefore, cannot influence the participants' evaluations. The Mann–Whitney U test finds significant differences in seven of 18 projections on a 0.05 level of significance (see Table 3).

Second, we investigated the final assessments for expected probability of occurrence. The Mann–Whitney U test likewise reveals significant differences between our internal and external panel in seven of the 18 projections on a 0.05 level of significance for the final results. However, differences were not found for the same projections in the initial and final assessments. In the final assessments, we find a significant difference for projection 18 that is not present for the initial assessments. The initial assessments differ significantly for projection 7, where no significance is found for final results.

For the investigation of the initial and final evaluations of impact and desirability provided by the internal and external Delphi panels, we again applied the Mann–Whitney U test. For each initial and final estimations, the difference is significant on a 0.05 level for only one projection (see Table 4).

The difference between the panels regarding initial assessments of desirability of occurrence is significant on a 0.05 level of significance for six projections. For the final estimations, a significant difference on at least a 0.05 level of significance is identified for eight projections, resulting in 44% of projections with differences in the final results.

The series of Mann–Whitney U tests (Tables 3 to 5) demonstrates that there were numerous projections for which the evaluations of internal and external participants differ significantly on a 0.05 level of significance. For the final results, which would be the basis for decision makers, we receive participants' evaluations regarding three dimensions that were judged for 18 projections, resulting in 54 outcomes. From these 54 evaluations, 16 deviate significantly from the internal to the external panel, corresponding to 30% deviation. The greatest differences are revealed for the evaluation of desirability of occurrence. Our findings thus support Proposition 1b. In fact, there were differences for all topics of the projections, which concerned customers, competition, innovation, and sustainability. Therefore, additional recommendations cannot be made whether there is a thematic area for which the differences are often noticeable.

In a more detailed analysis of the final estimations of probability of occurrence, we also observe differences in the distribution of the values (Table 6).

Quantifying the degree of consensus among Delphi panelists is an important component of Delphi data analysis and interpretation [83,84], but reaching consensus (as measured by a certain pre-defined threshold value) is not the general aim of a Delphi survey [85,86]. Among others, the IQR is often used as a measure for consensus in Delphi literature due to its robustness as a statistical measure [86–88]. The IQR is the range in which the middle 50% of the provided evaluations are located. Therefore, a smaller range is an indicator for a lower dispersion of the data [89]. Depending on the utilized scales, different thresholds for the IQR were defined to indicate that consensus among the experts was achieved [84]. After reviewing these previous measures and utilizing a scale from 0 to 100%, we defined an IQR equal to or smaller than 20 as an indication of consensus for the research at hand. Since such pre-defined measures have to be treated with caution, we also added summary statistics of the IQRs to Table 2 additionally strengthening the conclusions drawn

Table 3

Mann–Whitney U test for the comparison of internal and external Delphi panel assessments of probability of occurrence.

Projection number	Initial evaluations of expected probability of occurrence				Final evaluations of expected probability of occurrence			
	U	Mean [%]	Min [%]	Max [%]	U	Mean [%]	Min [%]	Max [%]
1	1,761.0**	35.5	0	95	1,561.0***	31.1	0	90
2	2,669.5	44.9	0	100	2,950.5	42.8	10	90
3	3,069.0*	58.8	0	100	3,346.0***	61.4	0	100
4	2,520.5	44.0	0	90	2,584.5	44.6	0	90
5	2,764.5	71.6	10	100	2,724.0	74.1	10	100
6	2,640.5	70.9	10	100	2,693.5	72.3	0	100
7	3,041.5*	61.2	0	100	2,749.0	65.8	5	100
8	2,294.5	48.1	0	100	2,257.0	47.8	0	100
9	2,561.0	70.4	10	100	2,422.5	73.1	20	100
10	3,009.0*	65.3	5	95	3,093.5*	66.8	5	95
11	3,156.0**	77.9	20	100	3,138.0**	78.6	20	100
12	1,879.5**	79.9	10	100	1,780.0**	80.7	10	100
13	2,362.5	76.7	5	100	2,397.0	77.6	5	100
14	2,637.0	80.4	10	100	2,699.5	80.8	10	100
15	3,355.5***	62.6	0	100	3,292.0***	62.0	0	100
16	2,439.5	63.5	0	100	2,413.5	64.3	10	100
17	2,245.0	66.3	0	100	2,186.0	67.0	0	100
18	2,067.5	64.3	5	100	1,860.5**	63.4	5	100

* $\rho \leq 0.05$.
** $\rho \leq 0.01$.
*** $\rho \leq 0.001$.

Table 4
Mann–Whitney U test for the comparison of internal and external Delphi panel assessments of impact.

Projection number	Initial evaluations of impact				Final evaluations of impact			
	U	Mean	Min	Max	U	Mean	Min	Max
1	2,498.0	3.3	1	5	2,446.5	3.2	1	5
2	2,314.0	3.3	1	5	2,148.0	3.4	1	5
3	2,329.0	3.7	1	5	2,296.0	3.8	1	5
4	2,521.0	3.2	1	5	2,447.5	3.2	1	5
5	2,282.0	3.9	2	5	2,282.0	4.0	2	5
6	2,294.0	3.8	1	5	2,318.0	3.8	1	5
7	2,601.0	3.8	1	5	2,577.0	3.8	1	5
8	2,124.0	3.3	1	5	2,340.0	3.3	1	5
9	2,694.0	3.6	2	5	2,781.0	3.7	2	5
10	2,600.0	3.9	2	5	2,619.0	3.9	2	5
11	2,885.5	4.0	2	5	2,835.0	4.1	2	5
12	2,170.5	3.9	1	5	2,187.5	3.9	1	5
13	2,273.0	3.9	2	5	2,304.5	3.9	2	5
14	2,669.5	3.9	2	5	2,685.5	3.9	2	5
15	2,331.0	3.8	2	5	2,158.0	3.9	2	5
16	2,435.5	3.3	1	5	2,468.0	3.3	1	5
17	1,944.5 [**,*,***]	3.9	2	5	1,974.5 [**]	3.9	2	5
18	2,519.0	3.5	1	5	2,683.0	3.5	1	5

* $\rho \leq 0.05$.
** $\rho \leq 0.01$.
*** $\rho \leq 0.001$.

regarding consensus. The IQR of the internal panel did not show consensus for four of 18 projections, while consensus was not achieved for 9 projections by the external panel. When adjusting the threshold to 25, another frequently used threshold, the difference becomes even more obvious with consensus for two additional projections in the internal survey and just one additional projection for the external panel. Moreover, the average of the IQRs and the respective standard deviations (Table 2) were higher for the external panel. Therefore, we conclude that consensus was reached more often for the internal than for external panel.

To shed further light on the dissent results, we conducted an outlier analysis for Delphi data [86]. It is expected that 1% of the

data, that is one evaluation, is an outlier. For the external panel, we have four projections for which we observe one outlier, three projections with two outliers, and five projections with three outliers. Since none of the eight projections with more than one outlier has an IQR above 20, this result does not function as an explanation for the observed dissent. For comparison, the outlier analysis for the internal panel results in six projections with one outlier, seven projections with two outliers, and only one projection with three outliers. As with the previous analysis, only one of these projections is a projection for which no consensus was reached — outliers do not function as an explanation for observed dissent. However, in comparing the two panels, we determined that the external panel had

Table 5
Mann–Whitney U test for the comparison of internal and external Delphi panel assessments of desirability.

Projection number	Initial evaluations of desirability				Final evaluations of desirability			
	U	Mean	Min	Max	U	Mean	Min	Max
1	1,900.5 [*]	2.5	1	5	1,612.5 [***]	2.4	1	5
2	1,922.0 [*]	2.5	1	5	1,728.0 [***]	2.5	1	5
3	3,198.0 [**]	3.5	1	5	3,320.0 [***]	3.7	1	5
4	2,075.0	3.2	1	5	1,925.0 [*]	3.2	1	5
5	2,134.5	3.7	1	5	2,220.5	3.8	2	5
6	2,361.5	3.7	1	5	2,332.0	3.7	1	5
7	3,078.5 [*]	3.6	1	5	3,003.0 [*]	3.6	1	5
8	2,538.0	3.5	1	5	2,714.0	3.6	1	5
9	2,834.5	3.6	1	5	2,785.0	3.6	1	5
10	3,094.5 [**]	3.6	1	5	3,115.5 [**]	3.7	1	5
11	2,234.5	3.6	1	5	2,309.5	3.7	1	5
12	1,791.0 [**]	3.9	1	5	1,800.0 [***]	4.0	2	5
13	2,192.5	3.9	1	5	2,195.0	3.9	1	5
14	2,302.0	3.7	1	5	2,395.5	3.7	1	5
15	2,920.5	3.4	1	5	3,044.0 [*]	3.4	1	5
16	2,263.5	3.4	1	5	2,086.0	3.4	1	5
17	2,126.0	3.6	1	5	2,250.0	3.7	1	5
18	2,545.5	3.2	1	5	2,432.5	3.2	1	5

* $\rho \leq 0.05$.
** $\rho \leq 0.01$.
*** $\rho \leq 0.001$.

Table 6
Distribution of evaluations for expected probability of occurrence.

Projection number	External Delphi panel					Internal Delphi panel				
	IQR	Min [%]	Max [%]	SD	CV	IQR	Min [%]	Max [%]	SD	CV
1	30	0	90	21.8	− 18.4%	20	0	70	15.1	− 33.6%
2	30	10	90	23.0	− 10.5%	20	10	90	19.8	− 13.2 %
3	26.25	5	90	20.7	− 15.1%	20	0	90	16.5	− 19.1%
4	30	5	100	20.5	− 5.9%	30	0	90	20.9	− 8.3%
5	10	10	90	17.0	− 12.6%	20	30	100	16.8	− 6.9%
6	10	15	100	18.4	− 5.8%	15	0	100	22.0	− 4.9%
7	20	10	100	19.8	− 7.7%	20	5	95	20.0	− 7.0%
8	40	0	100	25.9	− 2.0%	30	0	90	21.5	− 12.7%
9	10	30	100	14.5	− 13.6%	20	20	100	16.7	− 7.2%
10	26.25	5	90	17.6	− 5.2%	10	20	95	14.3	− 10.7%
11	10	20	100	15.7	− 3.9%	10	25	100	13.0	3.3%
12	10	20	100	14.5	− 1.1%	5	10	100	15.7	− 2.0%
13	20	20	100	16.4	− 3.0%	15	5	100	15.1	− 5.2%
14	10	30	100	12.4	− 2.9%	10	10	100	15.7	− 2.1%
15	22.5	0	100	20.8	− 4.8%	25	15	100	18.7	− 2.3%
16	30	10	100	17.6	− 10.1%	20	20	90	15.0	− 8.1%
17	20	0	100	19.0	− 3.7%	15	10	100	14.9	− 6.6 %
18	30	20	100	21.0	− 2.5%	25	5	100	21.9	− 7.2%

IQR = interquartile range.
SD = standard deviation.
CV = convergence rate (% change in SD of initial and final mean estimate).

more outliers in total. One explanation might be the greater diversity in panel composition, but it could also be influenced by the larger panel size in the internal survey. To provide possible reasoning for the expected dissent for some of the projections, we conducted outlier and bipolarity analyses [90]. However, we could not detect a bipolar distribution of the data.

The total range of the evaluations is only higher for nine projections of the external panel; for seven projections the range is even larger for the internal results. This result is surprising since we expected the range to be broader for the external panel — the professional background of the panelists is more heterogeneous and, therefore, the range of provided evaluations is expected to be broader.

Although the average convergence rate (percent change in SD of initial and final mean estimate) is higher for the internal panel than for the external panel (−8.5% vs. −7.2% respectively), there are eight projections for which the evaluations of the external panel converge more than the internal panel. For one projection, the opinions of the internal panel even diverge (projection 11). The majority of projections for which the external panel evaluations converge more than those of the internal panel, are those for which the external interquartile range is smaller than the respective internal range. The low convergence of the evaluations therefore is not explained by a smaller range of initial evaluations. In conclusion, an unambiguous answer to Proposition 1c cannot be made.

Finally, we compare the frequency of changes in the evaluations of the two panels as displayed in Table 7. Here, we

Table 7
Total frequency of changes in estimations.

Total changes for	EP	I	D
Internal	177 (17.3%)	97 (9.5%)	140 (13.6%)
External	221 (14.0%)	108 (6.8%)	135 (8.5%)

EP = Estimated probability of occurrence.
I = Impact.
D = Desirability.

display the total number and the percentage of the changes over all eighteen projections.

We expected that the external panelists would change their evaluations more often, because they receive information from a more diverse group with different perspectives. Nevertheless, changes for the internal panel were more frequent: therefore, we reject Proposition 1d.

4.2. Analysis and results of qualitative Delphi data

For the analysis of the qualitative arguments, we followed the general and established coding procedure proposed by Miles and Huberman [91] with a-priori codes, which are changed and developed by additional codes arising during the coding procedure. In order to avoid bias, the coding was conducted by two different coders of the facilitator team. All discrepancies were resolved by mutual agreement between the coders; in case of differing opinions, a common code was developed after discussion. For each panel, each projection, and each dimension (high/low expected probability of occurrence, high/low impact, high/low desirability), we analyzed which aspects were provided by the participants in their arguments. One argument can include multiple aspects since these often comprised full sentences. Therefore, each aspect received a different code. Since different aspects can be mentioned by more than one individual, we aggregated the codes for each projection's dimension. We took into consideration that the same aspect might be new for different projections. This results in the possibility to have two different codes for the same aspect. This enables to differentiate whether one code was mentioned for more than one projection.

In a first step of analysis, we compared the codes for each projection in detail to determine to what degree the participants from one panel provided reasoning that was not mentioned by participants in the other panel. In this stage, we also compared: a) whether the panelists tended to provide reasoning only for one aspect of a dimension, or b) whether the panel reasoned for both directions equally, and c) whether the participants of

one panel provided the same codes as their co-panelists or d) whether some codes were mentioned only once.

For the analysis of the qualitative data, we coded the 859 arguments from the internal Delphi study as well as the 746 arguments from the external survey. While the internal participants provided a higher number of arguments, coding of the arguments resulted in only 99 different codes. On the other hand, the arguments provided by the external participants could be coded into 142 different codes. Therefore, we conclude that the external participants provided a greater variety of aspects in their arguments.

We investigated the distribution of arguments as provided by the participants along the three dimensions probability of occurrence, impact, and desirability. Participants provided the most amounts of arguments for probability of occurrence. Therefore, we decided to examine whether the number of arguments provided in the two panels for low and high probability of occurrence was almost equal. This could function as an indicator of balanced argumentation for or against the occurrence of an event. We can observe that for the internal panel, the number of projections for which we find dissimilar amounts of arguments for high or low ratings is higher than in the external panel. For the internal panel, 16 of the 18 projections show a difference in the number of arguments (delta) that is higher than five arguments. Two projections even had 23 arguments more for the one opinion direction, e.g. low, than for the other, e.g. high. For the external panel, the arguments were provided more equally for low and high probability of occurrence. Only in 9 of the 18 projections the difference (delta) in the number of arguments given is higher than five. The maximum deviation in the external panel is 19 arguments. Therefore, the results support Proposition 2a.

We summarized the frequency of occurrence of one code for each projection and each dimension. We can observe that in the external panel more codes appeared only once than in the internal panel, indicating that these codes were novel and unique. The codes might be outlier statements or specific information by knowledgeable individuals. In the internal panel, the codes were repeated more often. For instance the "*influence on the costs*" was stated most often, namely 66 times. These results support Proposition 2b.

In a second step, we aimed to identify differences in the answering behavior of the individuals. First, we counted the average number of words the participants used for their argumentation. However, it is important to note that an argument with more words is not necessarily better than a short one. The quality depends on the content. We assumed that the length of the argument is an indicator for the dedication and interest of the participant because he or she took the time to support his or her answers. The external panelists used 15.2 words on average while the internal panelists used 13.1 words. However, the results indicate that the difference in the length of the reasoning is very small.

Afterwards, we considered the type of argument. Based on previous research experience of the facilitator team, including 32 past Delphi studies with more than 20,000 qualitative arguments, we derived generic categories in which the arguments were sorted: first with regard to their content, second with regard to their syntax. The categories are shown in Table 8. The label "[company]" replaces a specific name provided by an expert.

One argument may contain variations of the mentioned types and, therefore, can be assigned to more than one type. The results of the comparison of the types of arguments the members of the two panels provided are summarized in Table 9.

Considering the large amount of different types of arguments in relation to the total number of arguments provided, the internal panel specified their answers more often, for example to make a distinction between kinds of products and talked more about their personal experiences than the external panel. The internal panel also used fragments and catchwords more frequently, while the external panel used whole sentences. This is an indicator that the external panelists took their time to provide reasoning for their evaluation — the usage of catchwords by the internal panelists might be due to a common internal understanding of certain terms. The external panelists discussed trends more often, topics where information is easily accessible to the public, and were not as specific in describing past experiences than the internal panelists. Furthermore, only one case was found, where the expert seemed to have misunderstood the projection at hand. However, in total the observed differences were very small. No real differentiation between the panels can be made on behalf of the kinds of arguments used. We therefore cannot support Proposition 2c.

In a last step, we considered the content of the arguments, especially focusing on the question whether the internal panelists discussed company-specific information. The participants in the internal Delphi survey took advantage of the fact that the discussion was only accessible by fellow colleagues and kept confidential. They provided statements regarding the reference company's core competencies, the demands or suggestions made by customers, the desirability of certain developments considering the business strategy of the company, etc. Such sensitive information and protection of specific knowledge are crucial to company success [92]. Therefore, these arguments would not have been stated in a survey where competitors could read them. On the other hand, the external participants provided more arguments concerning the macro economy than the internal participants. They considered the development of the employment rate, provided insights from customers' or manufacturers' perspectives more often, and were more skeptical about the highly integrated logistics service concept that is offered by our reference company. Proposition 2d is therefore supported.

5. Discussion

Our comparison of an external and internal company Delphi panel yields support for Proposition 1. One possible reason for the differences in the numerical evaluations could be the composition of the panels: the external Delphi panel has higher heterogeneity with regard to the participants' professional background and knowledge background than the internal Delphi panel [1,29]. Our research confirms that when conducting an internal Delphi survey, the participants are very likely to have a similar information base and similar biases, since they have a comparable desirability for certain events or have similar mindsets [21]. On the contrary, the external panel encompasses a wider range of perspectives. An indicator hereof is the high IQR in our research, indicating higher dispersion, which can have a high value for foresight projects as well, and might even be the purpose of research such as Policy Delphi studies [93].

Table 8
Examples of types of arguments.

Type of argument	Example from Delphi discussion
Particular cases	New retail concepts are gaining momentum in many countries (see e.g. [company] in Spain).
Figures	If they are giving 1–2% discount on the product price, that can be compensated by 10–20% decrease in logistics costs.
Trends	The outsourcing trend will continue.
(Historical) analogies	The automotive industry has already done so, this sector should try to do so as well.
Cause–effect relationships	If the customer totally relies on the expertise of the food logistic provider, the customer may decide to assign their logistics partner with the design and operation.
Developments	Competitors will be challenged in the future.
Beliefs	It is fundamental!
Experience	Customers already require this.
Differentiation	It depends on the type of firm.
Lack of information	Compare with my first evaluations…
Misunderstandings	Do not understand what the projection means…
Syntax	
Whole sentences	The focus on customers does not allow focus on other jobs.
Phrases	Limited time of consumer
Catchwords	Environment

It was not possible to ascertain whether differences occurred more often for a certain topical area. The purpose of long-term foresight activities is not to find true statements about the future, but to identify possible developments that are not expected, evaluations that differ from the opinion of the majority are welcome to find a basis for robust strategy development [94]. Differing ranges of opinions can be observed for the external panel, where we found dissent in the evaluation of the expected probability of occurrence in 50% of the projections.

In the analyses of the qualitative data, internal panelists provided more arguments than the external panelists, which might be due to higher commitment. The kinds of arguments are similar, but the external panelists provided more statements that were unique in reasoning. More codes were derived only once within the external panel than in the internal panel. While the internal panelists took advantage of the opportunity to explain their past experiences and to talk about company-specific information, external panelists often described cause–effect relationships followed by expressions of their beliefs. This is in line with our expectations.

The rather identical knowledge base of employees from one company causes a limited amount of novel opinions, as indicated by the lower number of codes for our internal Delphi survey. Members from different companies in a panel contribute innovative and uncommon ideas, thereby adding to the richness of information, and sometimes shared this indirectly in their estimations or directly in their written arguments. However, in sum the differences between the arguments of the external panel and those of the internal panel are limited in number. Thus, we could not find full support for Proposition 2, as Proposition 2c is not supported.

In addition to analysis of results, Preble [24] pointed out the difference in costs related to intra-company and inter-company panels and suggested choosing intra-company since the costs related to conducting an internal Delphi survey are lower. He argued that the costs for recruiting external participants and the subsequent procedural costs are much higher than for an internal company Delphi study. However, the research at hand uses an online Delphi format, which has proven to be more cost efficient than the paper-and-pencil Delphi surveys: recruiting participants via Internet and telephone has become cheaper. Moreover, the online format is more convenient, the amount of time required to participate in online surveys is less, and the participant drop-out rates are lower. The panelists receive direct feedback and the survey finishes after a preset time span; not when consensus or stagnation have been reached [41,95].

Table 9
Types of arguments and panel comparison.

Type of argument	External panel [total number/percentage share]	Internal panel [total number/percentage share]	Percentage difference
Particular cases	5/0.7%	6/0.7%	0%
Figures	0/0%	4/0.5%	0.5%
Trend	116/15.5%	78/9.1%	6.4%
(Historical) analogy	6/0.8%	5/0.6%	0.2%
Cause–effect relationship	368/49.3%	422/49.1%	0.2%
Developments	60/8.0%	70/8.1%	0.1%
Beliefs	279/37.4%	332/38.6%	1.2%
Experience	127/17.0%	199/23.2%	6.2%
Differentiation	67/9.0%	92/10.7%	1.7%
Lack of information	5/0.7%	6/0.7%	0%
Misunderstandings	1/0.1%	0/0%	0.1%
Whole Sentence	568/76.1%	588/68.5%	7.6%
Phrases	163/21.8%	209/24.3%	2.5%
Catchwords	21/2.8%	71/8.3%	5.5%

Therefore, online surveys can result in better reliability and validity [41]. However, the effort required to identify external participants might be worth the benefit of having external independent advisors support in improving decision accuracy. A Delphi survey conducted with external participants might be a good solution although recruiting costs are still high and the response rate is considerably lower for the external panel than for the internal panel, where managers are motivated to participate by company interests.

An external Delphi panel does have additional disadvantages in addition to participant identification and recruiting efforts. An important precondition for advice taking is that the advisors are trustworthy and not intentionally misleading [3,44]. Since participants for an external Delphi panel could include individuals from competing companies, the question of whether these participants would be truthful cannot be answered definitely. Moreover, customers might want to encourage a company to endorse risks, so that they do not have to do it themselves. In order to counteract possible deceptive responses, outliers' impacts can be qualified in the analysis of Delphi results. Furthermore, for a heterogeneous panel, such as the external panel, a larger sample size is required in order to cover all relevant parties [70].

Based on the aforementioned information, the decision of selecting an internal or an external Delphi panel depends on the survey topic and the purpose for which the data will be used. Internal panels should be consulted when the Delphi survey strives for consensus-building. The employees of one company are more likely to arrive at a shared opinion about the future. Our analysis proved that it is useful to discuss a topic in depth which is related to the company's strengths and weaknesses since it provides the opportunity to share knowledge and to involve the employees in the strategic decision making process. Therefore, it is applicable for conventional Delphi surveys, where in addition to structuring the group communication process, the focus is on finding consensus and the most reliable forecast for certain topics [68]. The same applies to decision Delphi applications, where decision making is to be prepared and supported [96]. Due to the shorter preparation time and the lower effort required to recruit experts, internal panels are a useful source of advice when the time frame of the survey period is short or the resources that are available for the foresight project are limited, which might especially be an argument for small and midsize enterprises (SME) [97]. Finally, we can conclude that the knowledge obtained from own management sources might be a sufficient source of information, if companies are rather independent of external influences and act in a stable environment [98].

Our research supports that external Delphi panels are beneficial when a variety of opinions is desired, such as in policy Delphi surveys [85,93] or also decision Delphi applications. For the development of scenarios, it is also desirable to have divergent data as a basis to develop multiple scenarios, which should challenge mental models of the future [99]. This is the target application of the data generated by the Delphi survey at hand. Furthermore, the usage of external panels is recommendable when a broad set of projections is to be evaluated. When the projections cover different topics, as in our survey, it is even more likely that the external experts have knowledge beyond that of internal panelists and higher levels of expertise for some of the topics. Finally, involving external panelists in a Delphi

survey is a form of proactive stakeholder management. The company conducting the external Delphi survey strives to anticipate the stakeholders' expectations of the future in order to be able to proactively act in case their expectations do not match with the company's expectations [35,100]. The interests of the stakeholders in the company should be enhanced. Knowing the stakeholders' objectives not only affect firm performance positively, it is also a way of legitimating decisions [34,101]. Stakeholders, such as (potential) customers, can rate the topics regarding their expected probability of occurrence, impact and desirability: dimensions which should be considered when deciding about actions to adhere to stakeholder interests [102]. The reference company benefited from fruitful discussions with its partners even after the study.

Conducting not only one internal or external Delphi survey, but to conduct both simultaneously as separate panels might have additional advantages. In addition to being able to determine, prepare for, and adapt to stakeholders expectations of the future, another advantage from the company's perspective is that the employee evaluations remain confidential. Of course, the costs of conducting two separate Delphi surveys need to be evaluated against the expected benefit for every application individually.

6. Limitations and future research

We are aware of limitations to our research, which lead to future research directions. We only presented a limited number of projections to the panels. The logistics service sector will face more than these issues in the future. However, in order to prevent response fatigue and to ensure a vivid text-based discussion in the survey, we limited the number of topics presented.

Furthermore, one might assume that the internal panelists were obliged to participate. Their 'real' commitment might be weaker (artificial motivation) than that of the voluntary external participants. However, the amount of changes to evaluations per person and the number of Delphi platform logins were even higher in our internal panel, which can be perceived as an indicator of their honest commitment. Moreover, the number of logins can refute the possible doubts regarding comparability of the two panels because of the differing survey periods. A shorter surveying period was possible due to the motivated internal panel which re-logged-in several times.

In our analysis of the quantitative data, we utilize the Mann–Whitney U test even though the panel data does not meet the often mentioned requirement of same shape. According to literature, the Mann–Whitney U test may be utilized for data that does not fulfill this requirement, but then the test compares the ranks of the means and not of the medians [82].

In the literature review, we mention rivalry as one aspect because of which internal advice might be neglected. The degree of perceived rivalry might be reduced due to the anonymous survey design of the Delphi method and by the selection of panelists, for instance by selecting only one specialist from each department.

Finally, we cannot state that the evaluations made by the external Delphi panel are more accurate than those made by the internal panel. Since we consider events in the future, we can only indicate reasons why we expect their evaluations to be more accurate: due to the higher level of diversity within the

panel and the resulting bracketing effect. However, with our research we do not claim to make accurate forecasts for the long-range, but we aim to structure a systematic group discussion on the future, to facilitate a strategic conversation among experts, and to compare different panels during that process.

Future research might particularly focus on the different dimensions of heterogeneity. In our research, we focused on diversity in professional background and a diverse range of companies for whom the external participants work for (operational focus). Additionally, the biases individuals can have should be further investigated since they might have influenced the participants' evaluations.

7. Conclusion and managerial implications

In our research, we compared the results of two independent Delphi panels on an identical set of projections. The first panel consisted of managers from a reference company, a global logistics service provider. The participants from the second panel came from the company environment of the reference company and included not only competitors that offer similar logistics services, but also (potential) customers, suppliers, and academics. An analysis of the results revealed that the expectations of both Delphi panels differ significantly. We find significant differences for 30% of the criteria.

From the discussion of our results in the context of the reviewed literature we derive the following conclusions:

Internal panels should be consulted when the Delphi survey is designed to promote consensus-building for certain decisions. Internal topics can be discussed; information is kept confidential. Moreover, less time is needed to recruit participants and their response time is shorter. Therefore, internal Delphi surveys are applicable when the time frame is limited or the resources are restricted. External Delphi panels should be consulted when numerous perspectives on different subjects are desired. Applications include policy Delphi surveys, decision Delphi surveys, and, depending on their purpose and scope, Delphi-based scenario projects. Indeed, conducting Delphi surveys with both internal and external panels offers the most diverse and plentiful opportunities for decision makers. The differing expectations can be analyzed to find group patterns and to more aptly match strategies.

The reference company benefited in many ways from the external Delphi survey. Not only did the company receive knowledge about their stakeholders' expectations, which enables its management to act proactively to suit their requirements. The company also involved stakeholders in its decision processes and thereby increased their commitment and confidence in the company. Our reference company received very positive feedback from their customers and potential customers regarding their foresight activities. After the Delphi project concluded, the reference company invited customers, who participated in the external survey, to follow-up workshop sessions. The reference company took advantage of different media channels to communicate their activities and presented the study at conferences, in in-house magazines, and in popular logistics magazines. Furthermore, the project was the keystone for the foundation of two interdisciplinary working groups, who work on the upcoming topics from the foresight project, such as sustainability certification.

Appendix A

Table A1
Set of projections.

Projection number	Projection text
1	2020: The demand for one-stop shopping concepts has significantly decreased within the food service logistics industry.
2	2020: Long-term contracts between customers and food service logistics providers represent exceptional cases.
3	2020: Food service logistics providers are responsible for the design and operation of the customers' food supply chains.
4	2020: Small- and medium-sized logistics service providers have been able to increase their competitiveness against large logistics enterprises.
5	2020: The food service logistics industry is characterized by high pressures to innovate.
6	2020: Customers demand food service logistics companies to take care of reverse logistics processes (e.g. removal and recycling of all fractions of waste).
7	2020: Food service logistics providers take over in-store logistics processes (e.g. inventory management).
8	2020: Customers are willing to pay a premium for hygiene and safety in the food supply chain.
9	2020: Food service logistics providers have been able to extend their distribution networks to the supply of kiosks, petrol stations, automated shops etc.
10	2020: Customers increasingly ask their food service logistics providers to take over supply planning.
11	2020: Challenges in city supply demand for innovative logistics services.
12	2020: Customers force food service logistics providers to reduce their carbon footprint (without compensating for occurring additional cost).
13	2020: New technologies (e.g. smartphones, apps, RFID etc.) have helped food service logistics providers to increase customer satisfaction.
14	2020: Customers ask for high flexibility in service offer from their food service logistics providers.
15	2020: Food service logistics providers are forced to switch to alternative modes of transport (e.g. rail, sea, air, pipeline).
16	2020: Workplace design and ergonomics (e.g. senior oriented infrastructures) in the food service logistics industry are focused to meet the requirements/capabilities of elderly employees.
17	2020: Local sourcing for food products has become more important than global sourcing in the food service logistics industry.
18	2020: Customers demand full cost transparency from their food service logistics providers (e.g. disclosure of cost structures, cost breakdown, etc.).

References

[1] G.S. Day, P.J.H. Schoemaker, Driving through the fog: managing at the edge, Long Range Plan. 37 (2004) 127–142.
[2] I. Yaniv, Receiving other people's advice: influence and benefit, Organ. Behav. Hum. Decis. Process. 93 (2004) 1–13.
[3] G.E. Schrah, R.S. Dalal, J.A. Sniezek, No decision-maker is an island: integrating expert advice with information acquisition, J. Behav. Decis. Mak. 19 (2006) 43–60.
[4] N. Harvey, I. Fischer, Taking advice: accepting help, improving judgment, and sharing responsibility, Organ. Behav. Hum. Decis. Process. 70 (1997) 117–133.
[5] E.M. Miller, Risk, uncertainty, and divergence of opinion, J. Finance 32 (1977) 1151–1168.
[6] J. Landeta, J. Barrutia, People consultation to construct the future: a Delphi application, Int. J. Forecast. 27 (1) (2011) 134–151.
[7] M. Will, Talking about the future within an SME?: corporate foresight and the potential contributions to sustainable development, Manag. Environ. Qual. Int. J. 19 (2008) 234–242.
[8] E. Hiltunen, Crowdsourcing the future: the foresight process at Finpro, J. Futur. Stud. 16 (2011) 189–196.
[9] C. Decker, I.M. Welpe, B.H. Ankenbrand, How to motivate people to put their money where their mouth is: what makes employees participate in electronic prediction markets? Technol. Forecast. Soc. Chang. 78 (2011) 1002–1015.
[10] V. Miemis, J. Smart, A. Brigis, Open foresight, J. Futur. Stud. 17 (2012) 91–97.
[11] S. Dalal, D. Khodyakov, R. Srinivasan, S. Straus, J. Adams, ExpertLens: a system for eliciting opinions from a large pool of non-collocated experts with diverse knowledge, Technol. Forecast. Soc. Chang. 78 (2011) 1426–1444.
[12] J. Landeta, Current validity of the Delphi method in social sciences, Technol. Forecast. Soc. Chang. 73 (2006) 467–482.
[13] R. Parenté, J. Anderson-Parenté, A case study of long-term Delphi accuracy, Technol. Forecast. Soc. Chang. 78 (2011) 1705–1711.
[14] C. Hussler, P. Muller, P. Rondé, Is diversity in Delphi panelist groups useful? Evidence from a French forecasting exercise on the future of nuclear energy, Technol. Forecast. Soc. Chang. 78 (2011) 1642–1653.
[15] G. Rowe, G. Wright, F. Bolger, Delphi: a reevaluation of research and theory, Technol. Forecast. Soc. Chang. 39 (1991) 235–251.
[16] T. Menon, L. Thompson, H.-S. Choi, Tainted knowledge vs. tempting knowledge: people avoid knowledge from internal rivals and seek knowledge from external rivals, Manag. Sci. 52 (2006) 1129–1144.
[17] G. Tichy, The over-optimism among experts in assessment and foresight, Technol. Forecast. Soc. Chang. 71 (2004) 341–363.
[18] P. Ecken, T. Gnatzy, H.A. von der Gracht, Desirability bias in foresight: consequences for decision quality based on Delphi results, Technol. Forecast. Soc. Chang. 78 (2011) 1654–1670.
[19] I. Yaniv, M. Milyavsky, Using advice from multiple sources to revise and improve judgments, Organ. Behav. Hum. Decis. Process. 103 (2007) 104–120.
[20] J.B. Soll, R.P. Larrick, Strategies for revising judgment: how (and how well) people use others' opinions, J. Exp. Psychol. Learn. Mem. Cogn. 35 (2009) 780–805.
[21] I. Yaniv, Group diversity and decision quality: amplification and attenuation of the framing effect, Int. J. Forecast. 27 (2011) 41–49.
[22] E. Mannix, M.A. Neale, What differences make a difference? Psychol. Sci. Public Interest 6 (2005) 31–55.
[23] M.F. Wiersema, K.A. Bantel, Top management team demography and corporate strategic change, Acad. Manag. J. 35 (1992) 91–121.
[24] J.F. Preble, The selection of Delphi panels for strategic planning purposes, Strateg. Manag. J. 5 (1984) 157–170.
[25] S.M. Campbell, T. Shield, A. Rogers, L. Gask, How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? Qual. Saf. Health Care 13 (2004) 428–434.
[26] G. Fraser, D. Pilpel, J. Kosecoff, R. Brook, Effect of panel composition on appropriateness ratings, Int. J. Qual. Health Care 6 (1994) 251–255.
[27] D. Kahneman, Maps of bounded rationality: psychology for behavioral economics, Am. Econ. Rev. 93 (2003) 1449–1475.
[28] A.S. Alexiev, J.J.P. Jansen, F.A.J. Van den Bosch, H.W. Volberda, Top management team advice seeking and exploratory innovation: the moderating role of TMT heterogeneity, J. Manag. Stud. 47 (2010) 1343–1364.
[29] T. Menon, J. Pfeffer, Valuing internal vs. external knowledge: explaining the preference for outsiders, Manag. Sci. 49 (2003) 497–513.
[30] D. Tourish, Critical upward communication: ten commandments for improving strategy and decision making, Long Range Plan. 38 (2005) 485–503.
[31] G. Wright, G. Cairns, P. Goodwin, Teaching scenario planning: lessons from practice in academe and business, Eur. J. Oper. Res. 194 (2009) 323–335.
[32] J.F. Preble, A comparison of intracompany and intercompany LEAP forecasts, Technol. Forecast. Soc. Chang. 26 (1984) 389–397.
[33] M. Javidan, Core competence: what does it mean in practice? Long Range Plan. 31 (1998) 60–71.
[34] A.O. Green, L. Hunton-Clarke, A typology of stakeholder participation for company environmental decision-making, Bus. Strateg. Environ. 12 (2003) 292–299.
[35] J.F. Preble, Toward a comprehensive model of stakeholder management, Bus. Soc. Rev. 110 (2005) 407–431.
[36] M.A. Jolson, G.L. Rossow, The Delphi process in marketing decision making, J. Mark. Res. 8 (1971) 443–448.
[37] G. Rowe, G. Wright, The Delphi technique as a forecasting tool: issues and analysis, Int. J. Forecast. 15 (1999) 353–375.
[38] P. Lamberson, S.E. Page, Optimal forecasting groups, Manag. Sci. 58 (2012) 805–810.
[39] N.L. Kerr, R.S. Tindale, Group-based forecasting?: a social psychological analysis, Int. J. Forecast. 27 (2011) 14–40.
[40] F. Bolger, G. Wright, Improving the Delphi process: lessons from social psychological research, Technol. Forecast. Soc. Chang. 78 (2011) 1500–1513.
[41] T. Gnatzy, J. Warth, H. von der Gracht, I.-L. Darkow, Validating an innovative real-time Delphi approach — a methodological comparison between real-time and conventional Delphi studies, Technol. Forecast. Soc. Chang. 78 (2011) 1681–1694.
[42] T.R. Johnson, D.V. Budescu, T.S. Wallsten, Averaging probability judgments: Monte Carlo analyses of asymptotic diagnostic value, J. Behav. Decis. Mak. 14 (2001) 123–140.
[43] S. Bonaccio, R.S. Dalal, Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences, Organ. Behav. Hum. Decis. Process. 101 (2006) 127–151.
[44] J.A. Sniezek, G.E. Schrah, R.S. Dalal, Improving judgement with prepaid expert advice, J. Behav. Decis. Mak. 17 (2004) 173–190.
[45] J.B. Soll, Intuitive theories of information: beliefs about the value of redundancy, Cogn. Psychol. 38 (1999) 317–346.
[46] A. Mojzisch, S. Schulz-Hardt, Information sampling in group decision making: sampling biases and their consequences, in: K. Fiedler, P. Juslin (Eds.), Information Sampling and Adaptive Cognition, Cambridge University Press, Cambridge; New York, 2006, pp. 299–325.
[47] G.M. Wittenbaum, E.S. Park, The collective preference for shared information, Curr. Dir. Psychol. Sci. 10 (2001) 70–73.
[48] M. Naim, G. Aryee, A. Potter, Determining a logistics provider's flexibility capability, Int. J. Prod. Econ. 127 (2010) 39–45.
[49] S. Wadhwa, A. Saxena, F.T.S. Chan, Framework for flexibility in dynamic supply chain management, Int. J. Prod. Res. 46 (2008) 1373–1404.
[50] T.J. Chermack, S.A. Lynham, L. van der Merwe, Exploring the relationship between scenario planning and perceptions of learning organization characteristics, Futures 38 (2006) 767–777.
[51] F.C. Vasconcelos, R. Ramirez, Complexity in business environments, J. Bus. Res. 64 (2011) 236–241.
[52] G. Wright, K. Van Der Heijden, G. Burt, R. Bradfield, G. Cairns, Scenario planning interventions in organizations: an analysis of the causes of success and failure, Futures 40 (2008) 218–236.
[53] H.A. von der Gracht, I.-L. Darkow, Scenarios for the logistics services industry: a Delphi-based analysis for 2025, Int. J. Prod. Econ. 127 (2010) 46–59.
[54] J.G. Covin, D.P. Slevin, The influence of organization structure on the utility of an entrepreneurial top management style, J. Manag. Stud. 25 (1988) 217–234.
[55] J. Barthélemy, B.V. Quélin, Complexity of outsourcing contracts and ex post transaction costs: an empirical investigation, J. Manag. Stud. 43 (2006) 1775–1797.
[56] R. Cavana, B. Delahaye, U. Sekaran, Applied Business Research: Qualitative and Quantitative Methods, Wiley, Singapore, 2000.
[57] M.I. Piecyk, A.C. McKinnon, Forecasting the carbon footprint of road freight transport in 2020, Int. J. Prod. Econ. 128 (2010) 31–42.
[58] J.S. Armstrong, T.S. Overton, Estimating nonresponse bias in mail surveys, J. Mark. Res. 14 (1977) 396–402.
[59] G. Rowe, G. Wright, Expert opinions in forecasting: the role of the Delphi technique, in: J.S. Armstrong (Ed.), Principles of Forecasting: A Handbook for Researchers and Practitioners, Kluwer Academic Publishers, Boston, 2001, pp. 125–144.
[60] J.R. Salancik, W. Wenger, E. Helfer, The construction of Delphi event statements, Technol. Forecast. Soc. Chang. 3 (1971) 65–73.
[61] M.E. Porter, The five competitive forces that shape strategy, Anglais 86 (2008) 78–93.

[62] F.J. Parenté, J.K. Anderson-Parenté, Delphi inquiry systems, in: G. Wright, P. Ayton (Eds.), Judgmental Forecasting, John Wiley & Sons, Chichester, 1987, pp. 129–156.

[63] R.D. Klassen, D.C. Whybark, Barriers to the management of international operations, J. Oper. Manag. 11 (1994) 385–396.

[64] M.R. Czinkota, I.A. Ronkainen, International business and trade in the next decade: report from a Delphi study, J. Int. Bus. Stud. 28 (1997) 827–844.

[65] N. Dalkey, O. Helmer, An experimental application of the Delphi method to the use of experts, Manag. Sci. 9 (1963) 458–467.

[66] R.S. Duboff, The wisdom of (expert) crowds, Anglais 85 (2007) 28.

[67] W.B. McDermott, Foresight is an illusion, Long Range Plan. 29 (1996) 190–194.

[68] H. Sackman, Delphi Assessment: Expert Opinion, Forecasting, and Group Process, The RAND Corporation, Santa Monica, California, 1974.

[69] C.M. Goodman, The Delphi technique: a critique, J. Adv. Nurs. 12 (1987) 729–734.

[70] G. Skulmoski, F. Hartman, J. Krahn, The Delphi method for graduate research, J. Inf. Technol. Educ. 6 (2007) 1–21.

[71] U.G. Gupta, R.E. Clarke, Theory and applications of the Delphi technique: a bibliography (1975–1994), Technol. Forecast. Soc. Chang. 53 (1996) 185–211.

[72] M. Nowack, J. Endrikat, E. Guenther, Review of Delphi-based scenario studies: quality and design considerations, Technol. Forecast. Soc. Chang. 78 (2011) 1603–1615.

[73] T. Kuwahara, Technology forecasting activities in Japan, Technol. Forecast. Soc. Chang. 60 (1999) 5–14.

[74] S. Di Zio, A. Pacinelli, Opinion convergence in location: a spatial version of the Delphi method, Technol. Forecast. Soc. Chang. 78 (2011) 1565–1578.

[75] T. Gordon, A. Pease, RT Delphi: an efficient, "round-less" almost real time Delphi method, Technol. Forecast. Soc. Chang. 73 (2006) 321–333.

[76] J.C. Anderson, M. Rungtusanatham, A theory of quality management underlying the Deming management method, Acad. Manag. Rev. 19 (1994) 472–509.

[77] J.S. Dajani, M.Z. Sincoff, W.K. Talley, Stability and agreement criteria for the termination of Delphi studies, Technol. Forecast. Soc. Chang. 13 (1979) 83–90.

[78] R.T. Nakatsu, C.L. Iacovou, A comparative study of important risk factors involved in offshore and domestic outsourcing of software development projects: a two-panel Delphi study, Inf. Manag. 46 (2009) 57–68.

[79] M.R. Kastein, M. Jacobs, R.H. van der Hell, K. Luttik, F.W.M.M. Touw-Otten, Delphi, the issue of reliability: a qualitative Delphi study in primary health care in the Netherlands, Technol. Forecast. Soc. Chang. 44 (1993) 315–323.

[80] J.S. Martin, L.H. Chaney, Determination of content for a collegiate course in intercultural business communication by three Delphi panels, J. Bus. Commun. 29 (1992) 267–283.

[81] M.J. Campbell, Teaching non-parametric statistics to students in health sciences, ICOTS 7, IASE, ISI, Salvador, Bahia, Brazil, 2006.

[82] A. Hart, Mann–Whitney test is not just a test of medians: differences in spread can be important, Br. Med. J. 323 (2001) 391–393.

[83] J.V. Meijering, J.K. Kampen, H. Tobi, Quantifying the development of agreement among experts in Delphi studies, Technol. Forecast. Soc. Chang. (2013), http://dx.doi.org/10.1016/j.techfore.2013.01.003 (Available online 8 February 2013, ISSN 0040-1625).

[84] H.A. von der Gracht, Consensus measurement in Delphi studies: review and implications for future quality assurance, Technol. Forecast. Soc. Chang. 79 (2012) 1525–1536.

[85] H.A. Linstone, M. Turoff, Delphi: a brief look backward and forward, Technol. Forecast. Soc. Chang. 78 (2011) 1712–1719.

[86] J. Warth, H.A. von der Gracht, I.-L. Darkow, A dissent-based approach for multi-stakeholder scenario development — the future of electric drive vehicles, Technol. Forecast. Soc. Chang. 80 (2013) 566–583.

[87] M.K. Rayens, E.J. Hahn, Building consensus using the policy Delphi method, Policy Polit. Nurs. Pract. 1 (2000) 308–315.

[88] P.K. Ray, S. Sahu, Productivity management in India: a Delphi study, Int. J. Oper. Prod. Manag. 10 (1990) 25–51.

[89] E. De Vet, J. Brug, J. De Nooijer, A. Dijkstra, N.K. De Vries, Determinants of forward stage transitions: a Delphi study, Health Educ. Res. 20 (2005) 195–205.

[90] J. Warth, H.A. von der Gracht, I.-L. Darkow, A dissent-based approach for multi-stakeholder scenario development — the future of electric drive vehicles, Technol. Forecast. Soc. Chang. 80 (4) (2013) 566–583.

[91] M.B. Miles, A.M. Huberman, Qualitative Data Analysis: A Sourcebook of New Methods, 2nd ed. Sage, Thousand Oaks, 2005.

[92] D.J. Teece, Strategies for managing knowledge assets: the role of firm structure and industrial context, Long Range Plan. 33 (2000) 35–54.

[93] M. Turoff, The design of a policy Delphi, Technol. Forecast. Soc. Chang. 2 (1970) 149–171.

[94] R.J. Lempert, D.G. Groves, S.W. Popper, S.C. Bankes, A general, analytic method for generating robust strategies and narrative scenarios, Manag. Sci. 52 (2006) 514–528.

[95] M.R. Geist, Using the Delphi method to engage stakeholders: a comparison of two studies, Eval. Program Plann. 33 (2010) 147–154.

[96] W. Rauch, The decision Delphi, Technol. Forecast. Soc. Chang. 15 (1979) 159–169.

[97] K.D. Brouthers, F. Andriessen, I. Nicolaes, Driving blind: strategic decision-making in small companies, Long Range Plan. 31 (1998) 130–138.

[98] H. Catling, P. Rodgers, Forecasting the textile scene: an aid to the planning of a research programme, R&D Manag. 1 (1971) 141–146.

[99] P. Goodwin, G. Wright, Enhancing strategy evaluation in scenario planning: a role for decision analysis, J. Manag. Stud. 38 (2001) 1–16.

[100] I.M. Jawahar, G.L. McLaughlin, Toward a descriptive stakeholder theory: an organizational life cycle approach, Acad. Manag. Rev. 26 (2001) 397–414.

[101] S.L. Berman, A.C. Wicks, S. Kotha, T.M. Jones, Does stakeholder orientation matter? The relationship between stakeholder management models and firm financial performance, Acad. Manag. J. 42 (1999) 488–506.

[102] B.E. Perrott, Managing strategic issues in the public service, Long Range Plan. 29 (1996) 337–345.

Bernadette Förster is an external research assistant at the Institute for Futures Studies and Knowledge Management (IFK) of the EBS Business School in Wiesbaden, Germany. She holds a Diploma of Industrial Engineering from RWTH Aachen University (GER). Her research interests include corporate foresight, strategic planning, technology management and innovation management.

Dr. Heiko A. von der Gracht is the Founder and Managing Director of the Institute for Futures Studies and Knowledge Management (IFK) and post-doctoral researcher at EBS Business School in Wiesbaden, Germany. His research interests are corporate foresight, Delphi and scenario techniques, decision support, and quality in futures research. His works have been published in several books and in peer-reviewed journals, among them *Technological Forecasting & Social Change*, *Futures*, and the *International Journal of Production Economics*.