Article

S|S|S

Social Studies of Science
1–23

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0306312718798497
journals.sagepub.com/home/sss
SAGE

How to train your oracle: The Delphi method and its turbulent youth in operations research and the policy sciences

Christian Dayé
Department of Sociology, Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

Abstract

Delphi is a procedure that produces forecasts on technological and social developments. This article traces the history of Delphi's development to the early 1950s, where a group of logicians and mathematicians working at the RAND Corporation carried out experiments to assess the predictive capacities of groups of experts. While Delphi now has a rather stable methodological shape, this was not so in its early years. The vision that Delphi's creators had for their brainchild changed considerably. While they had initially seen it as a technique, a few years later they reconfigured it as a scientific method. After some more years, however, they conceived of Delphi as a tool. This turbulent youth of Delphi can be explained by parallel changes in the fields that were deemed relevant audiences for the technique, operations research and the policy sciences. While changing the shape of Delphi led to some success, it had severe, yet unrecognized methodological consequences. The core assumption of Delphi that the convergence of expert opinions observed over the iterative stages of the procedure can be interpreted as consensus, appears not to be justified for the third shape of Delphi as a tool that continues to be the most prominent one.

Keywords

Cold War social science, operations research, RAND Corporation, social life of methods

Introduction

In March 2010, the Japanese National Institute of Science and Technology Policy (NISTEP) published the results of a survey on the contribution of science and technology

Correspondence to:
Christian Dayé, Department of Sociology, Alpen-Adria-Universität Klagenfurt, Universitätsstrasse 65-76, 9020 Klagenfurt, Austria.
Email: christian.daye@aau.at

to future society. Carried out at regular intervals since 1971 and for the ninth time in 2009, NISTEP's surveys attempt to assess when specific technological innovations will be made, when they will be implemented in Japanese society, how important their impact will be for Japanese society and which sectors of the knowledge economy will be the driving forces behind each innovation. NISTEP's surveys are highly regarded by many analysts and professionals around the globe. They apply the Delphi method: Their samples comprise people acknowledged as experts in the fields of interest, and each survey

> iterates two or more rounds of the same questionnaire to the same respondents, until the answers converge to some specific way of thinking. In the second and subsequent questionnaire, the respondents are allowed to change their answers based on the summarized information (i.e. general trend of thinking) of the previous round. Some of the respondents change their opinions, allowing the overall opinions to converge. (NISTEP, 2010: 3)

Participating in NISTEP's 9th Delphi Survey were 2900 technology experts. Based on their input, NISTEP (2010) forecasted, for instance, that solar photoelectric power generation plants in space that transmit electricity to the ground via microwaves or lasers will be 'technologically feasible' in 2027 and 'socially realized' in Japan ten years later (p. 12).

Like its precursors, NISTEP's 9th Delphi Survey attempted to provide information that can be used in decision processes in government, industry and education. Smaller, but methodologically comparable regular Delphi studies on science and technology development and policy are conducted in other countries, such as in Germany by the Fraunhofer Institute. Further, Delphi is used in market research (Deutsche Post AG, 2009), public opinion and media research (Ferguson, 2000), or intelligence analysis (Heuer and Pherson, 2011). Delphi studies use several consecutive rounds of questionnaires to assess the opinions of experts on statements describing potential future developments. With each new round, information on the distribution of the opinions from the previous round is fed back (Häder, 2009; cf. Linstone and Turoff, 1975). This iterative procedure is designed to result in a convergence of opinions, which in turn can be interpreted as consensus.

While Delphi now has a fairly stable methodological shape, this was not so in its early years. As this article shows, its creators repeatedly modified the procedures, the methodological premises and the epistemological character of their brainchild. This article contributes to the literature on the 'social life of methods' by focusing on a specific period, the first years of Delphi's life. In this period, the creators of a method usually face the task of finding a social community where their brainchild might fit in and lead a decent life once it has left the parental home. Delphi is a valuable case for exploring this phase, because it did not leave its home for a surprisingly long time and thus underwent a long phase of primary socialization. Starting in 1948, the ideas behind Delphi were developed at the RAND Corporation, a research organization headquartered in Santa Monica, California, that entertained strong ties to the US Air Force (on the history of RAND, see Abella, 2008; Collins, 2002; Ghamari-Tabrizi, 2000; Hounshell, 1997; Kaplan, 1983; Mirowski, 1999; Smith, 1966). However, Delphi was only introduced to the world outside RAND through publications in the 1960s.

Things are made more interesting by the fact that, metaphorically speaking, Delphi's parents were erratic. Delphi's social life began with a phase of ambivalent, confusing and in part self-contradictory attempts by its creators to shape it according to their ideas, with the prime issue being that these ideas were continuously changing. The various designs tested over the first two decades of Delphi's existence bear witness to a multitude of parental ideas, and the challenge arises how to explain these continuous changes.

This article describes the varying methodological designs developed and used in the years of Delphi's youth and proposes a social explanation for these instabilities and variations in Delphi's history. Focusing on the first two decades of the method's existence, i.e. from 1948 to 1968, the article is based on an intensive study of the original research reports and publications by the inventors of Delphi. It also draws from archival research at the RAND Corporation headquarters in Santa Monica, CA, and on a series of interviews with former RAND researchers. One approach to explaining the changes in Delphi's shape would be to claim there was some form of cognitive progress over time; the changes would then reflect successive stages of methodological reflection and sophistication. Yet I argue that this was not the case. Rather than trying to refine the procedure along a given set of objectives, Delphi's parents repeatedly changed objectives on a fundamental level. As I describe, they first conceived of Delphi as a new *technique*. A few years later, they developed an epistemological foundation for it, transforming Delphi into a scientific *method*. And finally, after a few more years, they again re-conceptualized it as a *tool*. Any attempt to explain these changes by reference to the notion of cognitive progress would thus have to claim a progressive order between these three forms.

A more plausible approach to explanation can be found in Delphi's social life. If there had been crucial personnel changes in the primary group that engaged in Delphi's education, this might explain modifications in its design. But such personnel changes did not take place. Yet while the group of educators remained rather stable over the course of Delphi's youth, the social community into which they sought to socialize their method – operations research, roughly speaking – was not. In line with Rocco's (2011: 303) observation that virtually all ideas and analytic approaches developed at RAND can be understood as attempts to create the image of a trustworthy organization, I argue that the multiple shapes that Delphi took in its childhood and youth reflect changes in the social community into which its parents wanted to socialize it. Delphi was perceived and framed by its creators as a new technique at a time when operations research was a field of applied research close to the military. It was recast as a method when operations researchers increasingly questioned the scientific character of their field and called for more sophisticated theorization. And it re-appeared as a tool at a time when future studies emerged as an interstitial field between operations research, policy sciences, and business management.

It was the third shape of Delphi, as a tool, that became widely disseminated and used without much reference to the earlier shapes – NISTEP is an example of this third shape. However, it violated principles earlier established as crucial, with fundamental methodological consequences. Most importantly, it undermined the interpretation of the convergence of opinions as expert consensus. As a tool, I argue, Delphi could not claim to establish consensus anymore, but rather produced annoyance and fatigue.

Averaging group predictions: The first Delphi study, 1951

Before the first Delphi study was carried out in 1951, a group of RAND researchers had already experimented with the idea of pooling experts' opinions as a means for sketching out possible futures.[1] In the first half of 1948, Abraham Kaplan et al. (1950), had organized a series of prediction experiments. In the introduction of their article, 'The prediction of social and technological events', they argue that

> [m]any policy decisions require foreknowledge of events which cannot be forecast either by strict causal chains (as can eclipses) or by stable statistical regularities (as can the number of traffic deaths in a given period). For prediction of such events, the policy maker has no recourse but reliance on the judgment of experts. (Kaplan et al., 1950: 93)

However, public opinion polls provided a means to counteract idiosyncrasies and thus to stabilize the basis for decision-making. Polls were an established business at this point. George Gallup, Archibald Crossley and Elmo Roper had successfully established the polling business and, as a prerequisite, the legitimacy of polling methods (especially of sampling techniques) in the late 1930s (cf. Igo, 2006, 2007: 103–149; Lusinchi, 2017). There was a well-established body of methods to assess opinions on specific issues, on which the methodology proposed by Kaplan, Skogstad and Girshick could be based. It did not matter that polls addressed primarily public opinion, rather than the opinions of experts. The modification proposed by Kaplan et al. concerned the epistemic status of opinion. Standard polls usually treated opinion, even expert opinion, 'as an expression of a point of view, rather than as verifiable prediction' (Kaplan et al., 1950: 95).[2] This could and, in the view of the authors, should be improved. Why not conceive of the interviewees as 'predictors'? Why not use experts as persons able to deliver reliable estimations on future developments?

The idea that expert opinion was – from an epistemological standpoint – firmer than lay opinion provided the first of two links to the later development of Delphi. However, Kaplan and colleagues went one step further. Not only was expert opinion seen as relatively epistemically firm, it also appeared possible to attribute to it a predictive quality. Because of their broad and, for the most part, implicit knowledge of a field or specialty, experts were seen to be able to put forth reasonably stable predictions. This understanding of the expert as predictor permeated the rhetoric of the study and, with the restriction that not all study participants qualified as experts, also determined its design. Kaplan and his colleagues assembled a group of 26 persons. The vast majority of them, fifteen, were mathematicians and statisticians, four were engineers, another four economists or business administrators. The remaining three predictors were 'one office manager, one secretary, and one writer' (Kaplan et al., 1950: 96). All but two group members had college educations; eight held doctorates. It is plausible to assume that most members of the group – if not all – were working at RAND.

After all prearrangements were settled, the study began in the first weeks of 1948. Predictions were gathered via questionnaires, each with about a dozen questions on technological and societal events. Each item described a status quo, gave a date in the near future (about twenty weeks ahead), and asked the participant to judge the 'likelihood of occurrence' of four alternative states (using percentages that summed to 100%). For

Table 1. Predictive success per group.

Groups	Success rate
All Predictors	53%
Best Informed Predictors (top half)	56%
Worst Informed Predictors (bottom half)	50%
Independent Group	52%
Cooperative Group	62%
Joint Group	67%
Mean Prediction	66%

Adapted from Kaplan et al. (1950: 104).

instance, in a questionnaire distributed on March 29, 1948, one item read: 'In view of the danger of war, there is a possibility that (1) production of automobiles for civilian use will be legally restricted to save steel and (2) one or more auto factories will be converted to military production.' (Kaplan et al., 1950: 110) It then asked the participants to estimate the likelihood that by August 17, 1948, option (1), option (2), both options, or none would have had occurred.

New questionnaires were distributed weekly for thirteen weeks in a row. The participants had three hours to answer the questionnaire, a measure intended to avoid in-depth research on the questions. Apart from these general restrictions, the social situation in which the participants filled out the questionnaires was deliberately varied. One half of the participants (Kaplan et al., 1950: 97; emphasis in original)

> constituted a special set for the study of group predictions. They were divided each week into three quartets (provision being made for one absentee). One quartet, known as the independent group, answered the questionnaire individually, as usual; this was the control group. One quartet, the cooperative group, discussed the questions together, and then answered them individually. The joint group discussed the questions, and then came to some collective decisions, giving one answer for the entire group. This phase was so designed that every individual participated four times in each of the three types of quartets, working together once with each of the other twelve predictors in this phase.[3]

As mentioned above, the idea of attributing to expert opinions a higher epistemological status is one link of this study to the later Delphi efforts. A second, more hidden link is a procedural finding. After the twenty weeks had passed, the authors compared the predictions with actual states. It turned out that the cooperative and joint groups had achieved a higher 'predictive success' than the independent group. The independent group had achieved a success rate of 52%, but the cooperative group scored 10% higher and the joint group 15% higher (62% resp. 67%; see Table 1). 'The group effort is thus significantly better than that of the individuals composing the group working independently', concluded the study (Kaplan et al., 1950: 103).

These results might have made an argument for fostering group discussions in the frame of such methods, if it had been interpreted as being caused by a higher level of predictive capacity. But that was not the case. The authors argued that the higher predictive success was not caused by any emergent group property, but by a kind of averaging of opinions that takes place within groups and discourages extreme positions. This interpretation was corroborated by an interesting finding: The averaging effect could be mathematically reconstructed. When the researchers took the estimates of the independent group and, instead of assessing the individual predictor's success, used the mean value of the group to calculate the success rate, it increased from 52% to 66%. It thus approached the rate of the joint group (67%), the highest success rate achieved in the experiment. The authors concluded that 'in this study the success of collective psychological effort was duplicated by statistical methods' (Kaplan et al., 1950: 104). If so, a high success rate could thus also be achieved without interaction amongst the group members. There was no need to organize group discussions, because the averaging effect that had led to the higher predictive success could also be achieved by mathematical means.

Taking the mathematical average as a substitute for group interaction averaging thus provided the second link between the precursor study and the later Delphi design. In the months following the publication of that first study, Olaf Helmer and Norman C. Dalkey picked up on the ideas there developed.[4] Like the earlier study, their procedure attempted to establish reliable estimates via questionnaires. However, unlike the precursor study and most probably inspired by contemporary discourses in cybernetics, they set up an iterative methodological structure and provided for a feedback loop: The participants received information about the results from the previous round and were instructed to consider them when restating their opinion. The name Delphi was proposed by Kaplan, as Dalkey remembered. He commented: 'In some ways, it [the name] is unfortunate – it connotes something oracular, something smacking a little of the occult – whereas as a matter of fact, precisely the opposite is involved' (Dalkey, 1968: 8). Its parents thus hoped that Delphi would induce transparency in the opaque sceneries of expert policy advice.

Helmer and Dalkey conducted the first Delphi study in the first half of 1951. The report, 'The use of experts for the estimation of bombing requirements', was issued internally on November 14, 1951, but remained classified for ten years.[5] In this report, the Delphi technique is introduced by claiming that it attempts to 'obtain the most reliable consensus of opinion of a group of experts … by a series of intensive questionnaires interspersed with controlled opinion feedback' (Dalkey and Helmer, 1962: 1). The experts were polled individually via questionnaires that were designed: (1) to assess their answers to so-called 'primary questions', (2) to allow for sketching the reasoning behind the answer, (3) to list the relevant causal factors, (4) to estimate these factors, and finally, (5) to provide 'information as to the kind of data that he feels would enable him to arrive at a better appraisal of these factors and, thereby, at a more confident answer to the primary question' (Dalkey and Helmer, 1962: 1f). The aggregated results from each questionnaire had been sent to the participating experts, together with a new questionnaire and additional information. The experts were then asked whether they wanted to revise their earlier answers. The expectations, which were confirmed, behind this procedure were that the estimates would converge and the range of estimates would diminish (see Figure 1).

Regarding the content of the study, the participating experts were invited to change their usual perspective: Their task was to select, 'from the viewpoint of a Soviet strategic planner', a list of 'optimal' US industrial targets, and to estimate the number of atomic bombs such that the odds would be even that dropping them would ruin those branches of the US economic system required in munition production.

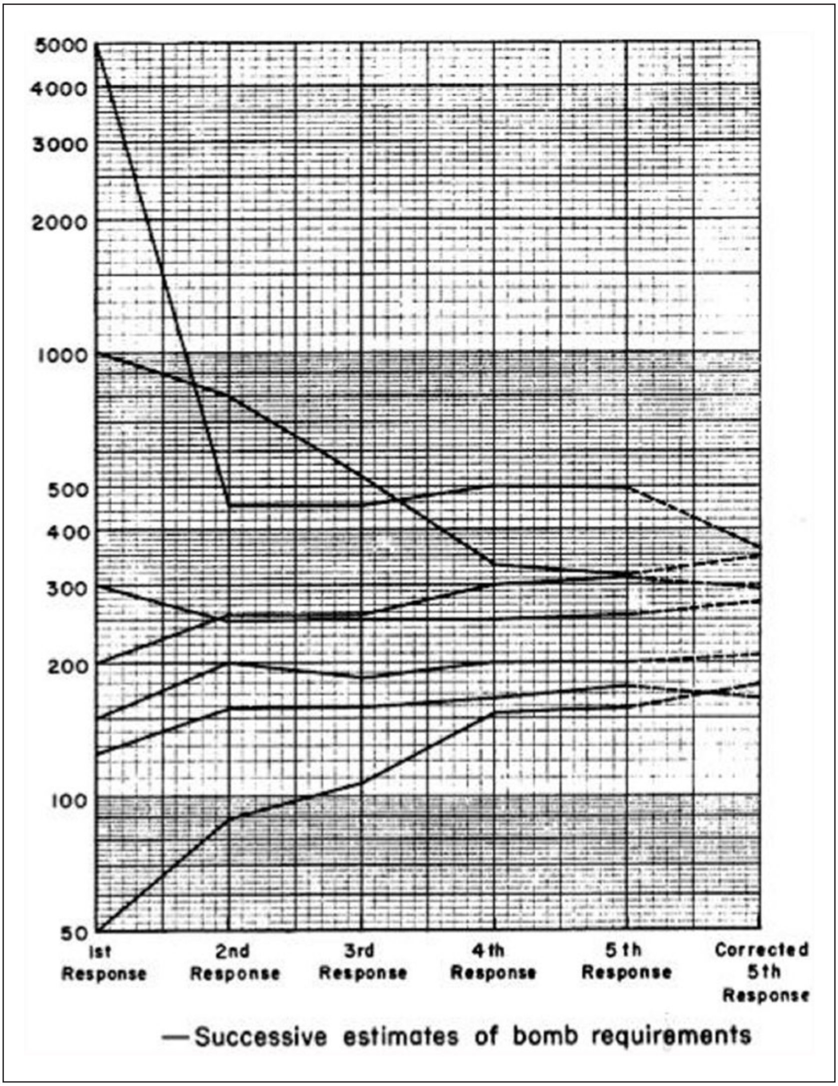

Figure 1. The convergence of estimates.
From Dalkey and Helmer (1962: 15), reprinted with the permission of the RAND Corporation.

The panel of experts for this study had seven people: four economists, a physical-vulnerability specialist, a systems analyst and an electronics engineer. The participants were strictly advised not to discuss these matters with colleagues and other scientists.

> This mode of controlled interaction among the respondents represents a deliberate attempt to avoid the disadvantages associated with more conventional uses of experts, such as round-table discussions or other milder forms of confrontation with opposing views. … Direct confrontation … all too often induces the hasty formulation of preconceived notions, an inclination to close

> one's mind to novel ideas, a tendency to defend a stand once taken or … a predisposition to be swayed by persuasively stated opinions of others. (Dalkey and Helmer, 1962: 2)

In addition to informing the participants about the median and the percentile distribution of the answers to the primary questions (1) given in the previous round, each new round of the Delphi study provided the experts with both (3) information on the factors considered relevant to the primary questions by other participants – 'e.g., the extent to which power transmission facilities permit reallocation of electric power' – and (5) data requested in the previous round – 'e.g., output statistics for steel mills' (Dalkey and Helmer, 1962: 2). The experts were asked whether – given the new data, selected justifications by other anonymous experts and the aggregated estimates – they wanted to revise earlier answers or whether they needed any additional information. The rationale for this broad approach to expert interrogation was that it allowed for assessing the factors informing the expert's judgment. This was intended to even out potential misconceptions and to bring into sight some factors that the expert might previously have ignored.

Altogether, five questionnaires were distributed to the participants in roughly weekly intervals. In addition to the questionnaires, the researchers decided to carry out interviews that were intended to capture aspects of the reasoning behind the estimates. These interviews were carried out after the first and the third questionnaire. Unlike many current Delphi studies, this first Delphi thus combined both quantitative and qualitative research techniques.

In contrast to its precursor, the first Delphi allocated a different task to the expert. Kaplan and colleagues had asked the participants to make predictions based on their implicit knowledge of the field, without any attempt to make this knowledge accessible. In contrast, Delphi sought to make explicit the experts' reasoning behind their estimates and thus to elicit some of the implicit knowledge involved in this form of prediction. A set of evidential material was collaboratively set up and made accessible to all participants, and the experts were asked to evaluate the extent to which this set altered their estimates. The iterative character of the method was used not only to feed back the aggregated estimations of the previous round, but also to distribute other kinds of information and data that might lead to a change in opinion or adjustment of estimates.

Despite these differences, the precursor and the first Delphi study were both conceived as *techniques* that should enlarge the arsenal of operations research, the most dominant branch of policy science at that time. While earlier traces can be found, operations research was a product of World War II (Thomas, 2015). The application of mathematics to the planning of both military and civil affairs during the war was deemed a great success by decision-makers in the government and the armed forces, and the very founding of RAND was an attempt to establish a place where the competence and skills developed during the war years could be held together in peacetime. The primary audience for the first version of Delphi were operations researchers and decision-makers. By systematizing expert opinions, Delphi promised to remedy an age-old problem, the conflict between the apparent need for expert knowledge and the dangers of idiosyncratic counseling in decision-making.

In its first life phase, Delphi consisted of a series of systematized steps that followed a specific logic. At the same time, it was flexible enough to allow for reacting to the specificities of the case under scrutiny. It thus reflected a philosophy quite characteristic of much of contemporary thinking in operations research: that what counted most was the usefulness of the results, that one could – and given the constraints of war should – start with testing and implementing even imperfect solutions, and that these imperfect solutions could then be gradually improved in a sort of trial-and-error scheme. Helmer (1963) described the craft of the operation researcher in a similar tune: '[W]e are pragmatists, our primary concern is with more effective manipulation of the real world, even if this may have to be accomplished without the desirable degree of understanding of all the underlying phenomena' (p. 1). In its first version, thus, Delphi was a technique that aimed at a more effective control of the factors deemed relevant for policy decisions. Yet, Delphi's parents were soon to change the 'identity' of their brainchild.

A flexible positivism: On the epistemology of the inexact sciences, 1958

Both Helmer and Dalkey had studied philosophy and logic before they joined RAND in the late 1940s. And although he was now working in a different research environment, Helmer still entertained an interest in problems of epistemology. In the early 1950s, he began to think about how to establish an epistemological foundation for the systematic use of experts. Discussions on such topics intensified when Nicholas Rescher joined the group in 1954.[6] Rescher recalled: 'When I came [to RAND], there was a kind of transition, a transition from people worrying specifically about one particular issue and people worrying about a more general methodology, and its rationale and how it might work' (Interview Rescher, September 1, 2011, 00:13:29). One outcome of these discussions was a longer paper entitled 'On the epistemology of the inexact sciences' (OEIS), co-authored by Helmer and Rescher.[7]

The dimension of the problem that Helmer and Rescher would address soon became apparent. It was to solve what Helmer (1963), a few years later, called the 'long-standing controversy as to whether operations research may be regarded as a scientific activity in the full sense of that phrase' (p. 1). In the view of both Helmer and Rescher, this question had to be answered yes, albeit with qualifications: Operations research was indeed a science, provided that one had a correct definition of what a science is. Such a definition, they thought, could not include exactness, nor precision as defining characteristics. Rather, '[f]or an enterprise to be characterized as scientific it must have as its purpose the explanation and prediction of phenomena within its subject-matter domain and it must provide such explanation and prediction in a reasoned, and therefore intersubjective, fashion' (Helmer and Rescher, 1959: 25).

With these criteria proposed, Helmer and Rescher sought an epistemological demonstration that operations research met them. A fundamentally new approach was required, they argued, because established epistemology had focused almost exclusively on some selected branches of the exact sciences. It had thus failed to consider a source of knowledge that was crucial and highly valuable in many other branches of the sciences – natural and social – that did not show a similar exactness but still were sciences in the sense explicated above. This source of knowledge were expert opinions. The task was to show how explanation and, even more so, prediction in the inexact sciences could rely on (expert) opinions in a systematical and epistemologically justified manner.

Central to their argument is the degree of confirmation. Originally a concept in the philosophy of science, the degree of confirmation was a value that described the extent to which a hypothesis *H* is confirmed by a given set of evidence *E*. The literature on this concept consisted in large parts of formulas describing its correct computation (e.g. Hempel and Oppenheim, 1945; Hosiasson-Lindenbaum, 1940), and Helmer himself had contributed to it before joining RAND (Helmer and Oppenheim, 1945). OEIS took up this linguistic frame, yet had the degree of confirmation functioning as a bridge between the credibility of individual predictions and the use of expert estimations in the inexact sciences on the other. The authors claimed that it justified the assumption that persons with large background knowledge in the field of study attributed personal probabilities to hypotheses that are reasonably close to the degree of confirmation a more formal calculation would produce. In the case of predictive hypotheses, where a formal calculation is not possible, it was therefore justified to take the expert's judgment of the probability of a given hypothesis as a reasonable approximation of the actual degree of confirmation (*dc*) of the hypothesis:

> *dc(H,E)* is intended to be a conceptual reconstruction of the personal probability which an entirely rational person would assign to *H*, given that his entire relevant information is *E*. In practice this relation can be applied in both directions: In simple cases where we have a generally acceptable definition of '*dc*' we may judge a person's rationality by the conformity of his personal probabilities – or of his betting behavior – with computable (or, if his information *E* is uncertain, estimable) *dc*-values. Conversely, once a person has been established as rational and possibly even an expert in a field, we may *use his personal probabilities as estimates, on our part, of the degrees of confirmation* which should be assigned given hypotheses. (Helmer and Rescher, 1959; emphasis added)

The epistemology of the inexact sciences proposed by Helmer and Rescher did not simply declare expert opinions to be evidence. Rather, experts were to evaluate the likelihood of a specific hypothesis in light of their implicit knowledge as well as a shared set of evidence. 'The predictive use of an expert', Helmer and Rescher write, 'can be characterized as follows: We wish to investigate the predictive hypothesis H; with the expert's assistance, we fix upon the major items of the body of explicit evidence E which is relevant to this hypothesis' (Helmer and Rescher, 1958: 48). Only under this condition did it seem justified to use the expert's personal probability as estimate of the degree of confirmation.

By establishing the degree of confirmation as a bridge between the credibility of hypotheses and experts' estimations, Helmer and Rescher attempted to philosophically consolidate a thought that emerged as a pragmatic solution to a policy problem: that experts have something useful to say about future or otherwise unknown phenomena and that it is reasonable to listen to them. However, they did more than that. They claimed that expert opinions, if properly used, were a justified source of scientific knowledge and thus had a place in science. Given the background of Helmer, Dalkey, and Rescher in neo-positivist philosophy, the importance of this shift can hardly be exaggerated (Dayé, 2016).

To conceive of experts as evaluators in the sense introduced above is also a non-trivial semantic departure from the initial conceptualization, in the study by Kaplan et al., of experts as predictors. However, this reflected a transformation that affected the entire social community into which Delphi was to be socialized. During the 1950s, operations research was transforming from a field of research into a scientific discipline. Increasingly, theoretical treatises were published and discussed, a development that 'does not seem to have resulted from any deliberate campaign, nor was it seriously resisted, sporadic grumbling notwithstanding' (Thomas, 2015: 196). Thus, to refashion Delphi not only as a technique of analysis, but as a scientific *method* required a fundamentally new epistemology, which Helmer and Rescher hoped to have established. As a scientific method, Delphi would claim to deliver useful results, but since science was the search for truth, it would also claim that under the conditions of inexactness and with a lack of precision relative to some other parts of science, its results would come close to the truth. Yet in the years to come, Delphi's parents decided to again modify the identity of their brainchild. In line with changes in the structure and orientation of the fields addressed, they re-designed the procedure in the early 1960s to become a tool. While continuing to toy with the picture of science as the search for truth, they neglected the epistemological principles formulated in OEIS.

Crafting the paradigm: The long-range study and the dissemination of Delphi, 1964

As mentioned above, RAND researchers and administrators initially hesitated to present Delphi publicly. Though originally written in 1951, the report on the first Delphi study was made available via the RAND bookstore only in 1962, three years after the publication of OEIS (Helmer and Rescher, 1959). The decision of the RAND leaders to declassify and release the report was one step in a broader attempt to develop and disseminate the Delphi procedure. When the text was published in *Management Science* in October 1963, Theodore J. Gordon and Olaf Helmer had already started a new Delphi study concerned with predictions into the further future.[8] After twelve months of study, the 'Report on a long-range forecasting study. P-2982' (Gordon and Helmer, 1964) was issued and made publicly available. This study led to the breakthrough of Delphi, especially after Helmer (1966) included it, with only minor adaptations, in his widely read book *Social Technology*: 'Not until after the 1966 publication of Helmer's Social Technology', Rescher (1997) notes, 'did Delphi effectively penetrate beyond the RAND Corporation orbit' (p. 353).[9]

The approach taken by Gordon and Helmer modified several features of the earlier Delphi. Most importantly, the number of participating experts was increased, and the collaborative compiling of a set of empirical evidence and relevant information was omitted. These modifications, however, were neither documented nor explained. Rather, Gordon and Helmer emphasized the continuities of their study with the previous efforts within RAND. The interested public thus conceived of Gordon and Helmer's Delphi study as defining the methodology in general, and the 1964 study became its paradigmatic example.[10] As a consequence, Delphi experienced a methodological consolidation. The phase of invention and initial development of Delphi was terminated. Taking up the metaphor of the social life of methods, its parents had decided that Delphi was sufficiently trained to leave the family home. After the publication of the study, there was not only a bigger pool of researchers contributing to the further development of the method, but there was also a ‘groundbreaking’ model to guide this development. Thus, this consolidation also meant that there were fewer degrees of freedom to think about the procedure.

At the outset of their report, Gordon and Helmer (1964) explained that ‘[p]rediction-making is a fundamental part of technological, military, commercial, social, and political planning in the modern world’ (p. v). But whereas it was relatively common to predict whether an event would occur in the following twenty-four hours, and such short-term predictions were usually accurate and trustworthy, many decision required the estimation of longer periods of time. Extending the time period would lead to specific problems:

> [A]s the period of concern is moved further and further into the future, uncertainties multiply, confidence in prediction is degraded, and the scientific theories and techniques of forecasting increasingly give way to intuitive judgment. The fact remains, however, that for better or for worse, trend predictions – implicit or explicit, ‘scientific’ or intuitive – about periods as far as twenty or even fifty years in the future do affect current planning decisions (or lack of same) in such areas as national defense, urban renewal, resource development, etc. (Gordon and Helmer, 1964: v)

Thus, while one cannot claim that all predictions about how the world will look in fifty years will eventually be verified, there still is some orientation value to such long-range predictions. ‘For the more distant future, as the uncertainties grow, increased reliance on intuitive (as opposed to theory-supported) contingency forecasts becomes inevitable. Yet this does not deter us from planning ten to fifty years ahead’ (Gordon and Helmer, 1964: 3). And

> [u]ntil a satisfactory predictive theory of the phenomena in question becomes available, it would seem that any improvement in reliability, however slight, that could be achieved by replacing casual guess with the controlled use of intuitive expertise would be desirable because of the benefits that long-range public policies might derive from it.’ (Gordon and Helmer, 1964: 4)

The difference to the position taken in OEIS is obvious. While OEIS had gone a long way to establish that procedures like Delphi in fact were scientific, the 1964 study again introduced a differentiation between scientific prediction and ‘the controlled use of intuitive expertise’. Delphi’s parents had again changed their pedagogical thrust. They had forgotten or neglected their earlier decisions, and took measures for a different future of their brainchild.

Before I discuss this change and its consequences, some more remarks on the study design are appropriate. The longer temporal outlook of this study marked a first difference from the predictive studies discussed above. To indicate this difference, the authors preferred to speak of forecasts instead of predictions. Whereas the latter proposed a coherent picture of the future, they argued that forecasts mapped out several possible futures (cf. Gordon and Helmer, 1964: 1).

Gordon and Helmer (1964) identified six areas of interest: (1) scientific breakthroughs, (2) population control, (3) automation, (4) space progress, (5) war prevention, and (6) weapon systems (cf. p. 2). Next, six panels of experts were set up, one for each area. The study leaders invited roughly 150 persons to participate, and 82 responded to one or more questionnaires. Slightly more than half of the respondents were RAND employees (35 persons) or consultants (7 persons), and the remaining forty persons had no official connection to RAND. Six of those with no connection to RAND were from Europe.[11] Each panel of experts was addressed with four consecutive questionnaires which were sent by mail approximately every two months. The first questionnaire was issued in June 1963, the fourth and last in January 1964 (cf. Gordon and Helmer, 1964: 27).

Gordon and Helmer (1964: 6ff) illustrated the procedure by describing the panel on scientific breakthroughs (Panel 1). The first questionnaire distributed to this panel asked the participants 'to list below major inventions and scientific breakthroughs in areas of special concern to you which you regard as both urgently needed and feasible within the next 50 years' (Gordon and Helmer, 1964: 7). The answers were compiled; multiple nominations were listed only once. In total, the list contained 49 possible inventions or breakthroughs. The list was then, in a second step, distributed to the participants, together with a set of nine time intervals (Gordon and Helmer, 1964: 7):

> 1963–65, 1965–68, 1968–72, 1972–78, 1978–86, 1986–97, 1997–2013, Later than 2013, Never

The participants were asked to indicate for each of the intervals the estimated probability of actual implementation of the given invention or breakthrough. This allowed for approximately assessing for each item the year to which the respondent attributed a 50% probability of actual implementation. In a next step, median and quartiles were calculated for these 50% values. It must be emphasized that all the values reported in the following by Gordon and Helmer referred to this 50% value, meaning that the resulting figures did not represent the year by which the experts are certain that an invention will be implemented, but rather a year in which the experts see the chances of its implementation evenly distributed. This had also been their approach in the first Delphi study.

The procedure of the second round was repeated another three times. Some items were added to the list; some were eliminated; and some were rephrased in order to avoid misunderstandings. As with the first Delphi study by Dalkey and Helmer (1962), the procedure of repeatedly asking approximately the same questions – together with a request to give reasons for deviating views – led to a convergence of estimates which was then be interpreted as consensus of experts. The graphical displays that presented the final results used a pentagonal shape to inform about median and the two quartiles for each item. For instance, the reader could see that the estimates for the 'creation of a primitive form of artificial life (at least in the form of self-replicating molecules)', had a median of 1989 and the two quartiles of 1979 and 2000. For two items, 'long duration coma to permit a form of time travel' and 'use of telepathy and ESP [Extra-sensory perception] in communications', half of the experts estimated the chances to be even that they would never happen.

Intensive dissemination of the Delphi method had already begun in parallel with the study itself. In addition to publishing *Social Technology* (Helmer, 1966), Helmer (1963) addressed audiences at the Third International Conference on Operational Research in Oslo, Norway and at the 10th Annual Meeting of the Western Section of the Operations

Research Society of America ORSA at Honolulu, Hawaii (Helmer, 1964); he also addressed the board of the Air Force Advisory Group AFAG (Helmer, 1967b) and published two papers reporting results of the long-range forecasting study (Helmer, 1967a, 1967c)

With Helmer preparing to leave RAND for the Institute for the Future, Norman C. Dalkey took over the dissemination task. In October 1967, he presented the method to the participants of the Second Symposium on Long-Range Forecasting and Planning, Almagordo, New Mexico (Dalkey, 1967). He introduced Delphi at the National Meeting of the American Chemical Society in San Francisco, California (Dalkey, 1968), and included a brief description of Delphi in his speech at the National Conference on Fluid Power in Chicago, Illinois, in October that year (Dalkey, 1968). Also, an article by Bernice Brown (1968), entitled 'Delphi process: A methodology used for the elicitation of opinions of experts' was published in *ATSME Vectors* in early 1968 and distributed as RAND report P-3925.

In their descriptions of the Delphi procedure, all these papers are quite similar to each other. The following description, taken from Dalkey (1967), can be considered representative:

> A typical [Delphi] exercise is initiated by a questionnaire which requests estimates of a set of numerical quantities, e.g., dates at which technological possibilities will be realized, or probabilities of realization by given dates, levels of performance, and the like. The results of the first round will be summarized, e.g., as the median and inter-quartile range of the responses, and fed back with a request to revise the first estimates where appropriate. On succeeding rounds, those individuals whose answers deviate markedly from the median (e.g., outside the inter-quartile range) are requested to justify their estimates. These justifications are summarized, fed back, and counter-arguments elicited. The counter-arguments are in turn fed back and additional reappraisals collected. … One additional feature of present Delphi procedures [is that respondents] are requested to make some form of self-rating with respect to the questions. (p. 4)

This quotation describes neatly the specific form of Delphi that was disseminated in the 1960s. The Delphi paradigm that emerged around this definition relied exclusively on the 1964 study (Gordon and Helmer, 1964), thereby negating both the earlier study from 1951 (Dalkey and Helmer, 1962) and – implicitly, albeit not overtly – the methodological position elaborated in OEIS (Helmer and Rescher, 1958). In this form, the Delphi paradigm also became codified in the first textbook-like publications on the method, namely in the series on 'The DELPHI Method,' begun and supervised by Helmer and issued by RAND for the purpose of public relations.[12]

The loss of a juvenile depth of thought, and its consequences

The 1964 Delphi differed fundamentally from both the 1951 study and the epistemological position developed in the 1958 OEIS. It sacrificed the flexibility that had characterized the first Delphi study in order to increase the number of participants. The 1951 study had assembled a small group of participants, which allowed for considerable flexibility in the whole procedure, for a combination of a variety of methods, including many open items in the questionnaire and qualitative interviews, and for setting up a shared base of information material. In the long-range forecasting study from 1964, the participants gave mainly numerical estimates, and were in turn informed of the distribution of the answers to the previous round by means of three figures: the median and the two quartile values. Virtually no additional information or data was communicated, and no common set of evidential material established. At most, the formulation of some questionnaire items was clarified. Furthermore, only those researchers whose estimates deviated most strongly from the median were asked to provide a justification for their opinions. Thus, instead of allowing for an exchange of the reasoning behind the individual estimates, the procedure was more likely to put pressure on the outliers.

Unlike in the original Delphi study, the participants in the long-range forecasting study did not collaboratively set up a set of evidence with the aim of evaluating its meaning for a specific prospective question or hypothesis. Rather, the participants were asked to predict the date of future events based on their implicit knowledge, without any attempt to make explicit the substance of this knowledge. All of this counteracted the ideas formulated in OEIS, and quite explicitly the postulate that established that expert opinions can be used as a surrogate (or estimate) for the degree of confirmation of a predictive hypothesis. This postulate assumed that the opinions of the experts related to a set of evidence that they had themselves helped to set up. And since it did not establish such a set of evidence, the long-range Delphi study did not meet the criteria established by Helmer and Rescher.

But if it was neither a technique nor a scientific method, what, then, was this new Delphi? Again, this question can be answered best by considering the social community into which its creators wanted to release it after socialization. And this was no longer operations research. Rather, it had become the broader field of policy science, which experienced the beginning of a period of growth in the 1960s and which sustained the idea to apply operations research techniques developed in military contexts to civilian fields (cf. Jardini, 2000; Light, 2003; Sapolsky, 2004). More precisely, the intended audience of the new Delphi was a subfield of policy science that itself was only in its infancy, future studies (Andersson, 2012; Tolon, 2011, 2012) Over the 1960s, Helmer had grown increasingly annoyed with the decisions taken by RAND management in its attempts to develop the organization. There was an increasing influx of private research contracts, which in Helmer's perspective reduced the degree of freedom of research.[13]

> At the time, you see, in fact, my feeling was – this was supported by Williams as well – that RAND ought to make an effort to expand some of its research methods to fields other than the military – for instance to social problems. That was the reason why, since we couldn't persuade the RAND management at the time, I decided together with some of my colleagues that maybe we should set up our own organization and pursue that idea to pursue some of the rand techniques in areas applicable to social problems. (Interview with Helmer by Kaya Tolon, in Tolon, 2011: 198)

Together with Gordon and Paul Baran, Helmer founded the Institute for the Future in 1968, which, after securing an affiliation with Wesleyan University in Connecticut (Gordon, 2011: 1099). Future studies had become Helmer's core interest. 'There is little question that Helmer's main academic contribution was to the field of Futuristics. Once he became engrossed in matters of prediction and futurology this replaced all other concerns' (Rescher, 1997: 349).

However, the problems addressed by future studies were not operational, and they were not scientific. What decision-makers in business, industry and policy needed, so the futurists were convinced, were *tools* that allowed for quick and stable predictions (or forecasts). In its 1964 form, Delphi delivered exactly that. It was reconfigured as a tool that could be used without specific social scientific training – the items themselves came from the expert panels, no qualitative interviews were required etc. If the outline procedure was followed, virtually anyone could use it.

The 1964 Delphi promised exactness and stability of results, and caveats regarding the validity of their procedure were quickly discounted with a reference to the necessity of foreknowledge for good decision-making. The tool claimed to produce credible knowledge, and the core currency of credibility was a mixture of scientific lingo and quantification. The decision to reduce flexibility in order to be able to increase the number of participants can also be seen in the context of the contemporary trend of quantification in the social sciences (Platt, 1996; Steinmetz, 2005). Quantification was deemed to be the main avenue to objectivity not only within the sciences, but within virtually all cultural fields (Porter, 1995). This view also had important proponents at RAND, most notably in the Mathematics Division to which Helmer, Dalkey, and Rescher belonged. Through the 1960s, RAND experienced large changes in its institutional environment. It saw many of its formerly strong support channels break away or diminish. While other comparable research organizations reacted to these changes by retiring from the public (and academic) sphere (Rohde, 2013), RAND management decided to enter new markets. It reorganized its knowledge production and addressed new clienteles. The concentrated effort to publish and disseminate Delphi contributed to this strategy. It was expected that Delphi studies would augment their persuasiveness by not relying solely on in-house experts. Clearly, this required either inviting these experts to stay at RAND for the duration of the study, rendering Delphi a very expensive method, or relying on current communication technology (i.e. postal services) and accepting restrictions in the interactive structuring of a shared set of evidence.

However, more important than explaining why Delphi left RAND in the shape it did is to assess the methodological consequences of this final form. These are fundamental, because the final procedure was at odds with the hitherto unquestioned interpretation of convergence. Despite the unfounded expectation that opinions would always converge – Delphi researchers repeatedly, though not in the period at RAND covered here, observed the building of several opinion clusters – many proponents continued to interpret the convergence of estimates as *consensus*. If the procedure had followed the considerations explicated in OEIS, this could have been justifiable. However, lacking the collaborative composition of a set of evidence that evolves and alters from round to round, the 'consensus' produced by the long-range forecasting Delphi had no substantial backbone. Since they could not agree on the epistemic value of a set of evidential material, what were they expected to converge upon? Or, to put it the other way around, what is the likely psychological effect of being repeatedly asked to revise one's answers in the face not of new evidence, but solely of the opinion of the majority? Does such a procedure result in convergence? Or, rather, in a mixture of annoyance and fatigue? Without a substantial argument amongst the participating experts and no shared evidence to frame such an argument, the convergence that should have been a rational result became an artifact of the Delphi procedure. The exchange between experts offered a justification for the iterative procedure as a means of establishing consensus. When such exchange is lacking, there is no reason to interpret convergence as consensus, and other possible interpretations – among them the most likely candidate being fatigue – become plausible.

Conclusion: The fatigued expert

In hindsight, the label ‘Delphi method’ turns out to be an umbrella term for a heterogeneous group of approaches. The changing shape, as I argue here, can be understood as a consequence of various alliances that the authors of Delphi, and especially Helmer, wanted to forge. The socialization of Delphi was defined by its parent’s explicit hopes and tacit sorrows. In the years after World War II, Delphi was trained to be a technique used by expert operation analysts to produce figures that could be used in further planning. With the movement to establish operations research as a scientific field, the technique was given an epistemological foundation, effectively transforming it into a method with a claim to approach truth. Finally, however, the dawning emergence of future studies in the early 1960s led Delphi’s parents to devise yet another identity for their brainchild and Delphi became a tool that could be used without prior training in social research methodology.

In a sense, however, and despite its unstableness, this socialization led to success. Future studies attracted interest and resources, and Delphi as a tool became used in a variety of fields. In a paper written during his 1977 stay as a guest researcher at the International Institute of Applied Systems Analysis (IIASA) in Laxenburg, Austria, Helmer described Delphi as one step of a larger framework for long-term forecasting (on IIASA, see Duller, 2016; Rindzevičiūtė, 2016; Riska-Campbell, 2011). After having identified potential future developments, Delphi or other tools could be used to ‘[o]btain forecasts … regarding these developments’ (Helmer, 1977: 8). Then, the analyst had to estimate the connections among these developments and to use cross-impact analysis – another invention of Helmer’s (1977) – ‘to establish the relative sensitivity of the developments to one another’ (p. 8). Four further steps followed. Delphi had transformed into a step in a more comprehensive framework, into a suitable tool amongst others in the toolbox of future studies and policy science.

Although in terms of citations, OEIS appears to have been very influential in the history of Delphi, the path outlined in its pages was not taken. Most studies that used a Delphi design after it had left RAND relied on big samples of experts and on small, if any, explication of the evidential materials influencing the individual estimations. Surprisingly, however, Delphi’s training had engendered oblivion. Delphi’s parents did not realize that the final shape they had given to their brainchild violated the quality criteria that they themselves had formulated a few years earlier. This had two immediate consequences, one minor, the other major. The minor one was that, according to their own definition, Delphi was no scientific method anymore. It relinquished the claim to truth that had been so painstakingly established. More importantly, however, was that when Delphi sought higher participant numbers and less substantive interaction among participants, it destroyed the epistemological justification for interpreting the expected convergence of expert opinions as ‘consensus’. Even if a convergence can be observed, the lack of an identical set of evidence makes it hard, if not impossible, to interpret the convergence as indicating consensus and not simply as fatigue pertaining to the procedure on the part of the experts. Or, to put it differently, in the process of socialization, Delphi's initial aspirations had to be downsized to match the realities of the epistemic market.

Acknowledgements

I have discussed earlier versions of this argument at the 108th Annual Meeting of the American Sociological Association (ASA) in New York, August 2013, and the XIV ISA World Congress of Sociology in Yokohama, July 2014. The article is part of a larger project, and being unable to recollect all the scholars who have offered support and critique over the past decade, the author extends a general thanks to all who might feel addressed. A special thank you goes to Karen Meehan for her continuous support, to three anonymous reviewers for their comments and last not least to Sergio Sismondo for valuable suggestions, tireless commitment and patience.

Notes

1. There were earlier studies on the predictive capabilities of groups (Cantril, 1938; McGregor, 1938), but neither the application of such predictions to policy analysis nor the idea that experts were better predictors than lay people had been focuses of concern.
2. Election polls were an exception. In these cases, however,

   'the prediction studied is the poll-taker's own, not that of the persons polled. Though much is known about the prediction of public opinion itself, and perhaps something about the prediction of social events from a knowledge of opinion, the general problem of prediction … is largely unexplored'. (Kaplan et al., 1950: 96)

3. This part of the study design, the authors add in a footnote, was contributed by William J. 'Jack' Youden (1900-1971), a chemist and statistician known for his works in test design (cf. Cornell, 1993).
4. Helmer (1910–2011) was born in Berlin. At the age of 24, he completed his studies of mathematics and logic at the University of Berlin with a dissertation begun under the direction of Hans Reichenbach. Shortly afterwards, he emigrated to Britain, where he completed a second doctorate in philosophy at the University of London. In 1937, Helmer was appointed research assistant to Rudolf Carnap at the University of Chicago. During the war years, 'Helmer was drawn into mathematics-based work for the National Defense Research Committee under the direction of John Williams' (Rescher, 2006: 288), and in 1946, when Williams became one of the first scientists engaged by RAND, Helmer decided to join him. Norman Crolee Dalkey (1915–2003) was born in Santa Clara, California. From 1939 to 1940, he took graduate courses in philosophy at the University of Chicago, where he met Helmer. Dalkey completed his PhD two years later at the University of California in Los Angeles, having written a dissertation supervised by Hans Reichenbach. Probably upon the initiative of Helmer, he joined RAND's Mathematics division in 1948. For more biographical background on Helmer and Dalkey, see Dayé (2014, 2016).
5. In 1962, an abridged version was declassified, now entitled 'An experimental application of the Delphi method to the use of experts' (Dalkey and Helmer, 1962). This text was eventually published in *Management Science* one year later (Dalkey and Helmer, 1963). Since the original report is still secret, I rely on the abridged version. Apparently, the abridgements mainly

concerned the results, not the design of the methodological procedure. Two reasons might have led RAND management to agree to a declassification of the abridged report in 1962. Apart from being mentioned in one publication by Helmer and Rescher (1959) and in briefings and informal talks, the Delphi method had not yet been disseminated. There might have been a demand from outside RAND for more information about this method. More importantly, however, was that by then, the invention and perfection of Inter-Continental Ballistic Missiles (ICBMs) and the successful launch of Sputnik in October 1957 had rendered most of the study's results obsolete, because the scenario used in the study had departed from the then – in the early 1950s – valid assumption that A-bombs would have to be delivered by airplanes.

6. Born in Germany, Rescher (1928–) moved to the US with his family at the age of ten. He studied at Queens College, New York City, and took some courses led by Helmer's longtime friend Carl G. Hempel. He moved to RAND upon the invitation of Helmer in 1954 and stayed there for almost three years before continuing his career in academia.
7. *OEIS* was first issued as RAND paper P-1513, in October 1958 (Helmer and Rescher, 1958), and published in *Management Science* in 1959 (Helmer and Rescher, 1959). This published version was then re-issued in February 1960 as RAND report R-353 (Helmer and Rescher, 1960).
8. At that time, Theodore Jay Gordon (1930–), an engineer, was a consultant to RAND. Some years after the 1964 long-range forecasting study, he co-invented with Helmer another widely known method of future studies, the cross-impact analysis, which in the years following its publication was tested and then systematically applied in the CIA (cf. Heuer and Pherson, 2011: 107). In 1971, Gordon founded The Futures Group, a spinoff from the Institute for the Future (IFTF) founded by the internet pioneer Paul Baran, Helmer, and Gordon. (The IFTF itself was a spinoff from RAND.) The Futures Group established itself as an international consulting firm that in the first years predominantly 'contracted to perform Delphi studies for corporations on a proprietary basis' (Rescher, 1997: 354, fn. 32). Gordon retired as CEO of The Futures Group after twenty years but continued to work in future studies, amongst others as a senior fellow of the Millenium Project (http://www.millennium-project.org/about-us/planning-committee/ted-gordon/, accessed 31 January 2018).
9. Also included in *Social Technology* was the report of a smaller study carried out by Bernice L. Brown under the auspices of Helmer, which was originally published as 'Improving the reliability of estimates obtained from a consensus of experts. P-2986' (Brown and Helmer, 1964) in September 1964, practically at the same time as the report on the long-range forecasting experiment. However, for reasons that will become clear later on in the discussion, it was the report by Gordon and Helmer that had the most sustained impact on the then emergent future studies community.
10. The use of the term 'paradigmatic' here is oriented to the second meaning noted by Kuhn (1970) in his Postscript to the *Structure of Scientific Revolutions*, namely that concrete solutions of scientific problems become paradigmatic when they are understood as examples of best practice; such paradigmatic cases can also replace explicit rules (Isaac, 2012).
11. Two of the European participants, 'Professor Dennis Gabor and Monsieur Bertrand the Jouvenel' (Gordon and Helmer, 1964: ix) are thanked by name in the acknowledgements section of the report. De Jouvenel (1903–1987), author of the concept of futuribles and founder of an organization and a journal with the same name, is one of the best known 20th century futurologists. Dennis Gabor (1900 – 1979), a British physicist of Hungarian origin, is probably best known for inventing holography, which earned him the 1971 Nobel Prize in Physics.
12. The series included 'The DELPHI method: An experimental study of group opinion. RM-5888-PR', by Norman Dalkey (1969); 'The DELPHI method, II: Structure of experiments. RM-5957-PR', by Bernice Brown, psychologist and RAND consultant Samuel

Cochran, and Dalkey (1969); ‘The DELPHI method, III: Use of self-ratings to improve group estimates. RM-6115-PR’, by Dalkey et al. (1969) and, finally, albeit with some restrictions, ‘The DELPHI Method, IV: Effect of percentile feedback and feed-in of relevant facts. RM-6118-PR’, by Dalkey et al. (1970).

13. It must be added, though, that government and military agencies had reduced their funds after a series of fierce public debates about RAND and think tanks in general, among them two that were triggered by RAND publications: Paul Kecskemeti’s (1958) misinterpreted study on *Strategic Surrencer* and Herman Kahn’s (1961) *On Thermonuclear War*.

ORCID iD

Christian Dayé https://orcid.org/0000-0001-5530-371X

References

Abella A (2008) *Soldiers of Reason*. New York: Harcourt.

Andersson J (2012) The great future debate and the struggle for the world. *American Historical Review* 117(5): 1411–1430.

Brown BB (1968) *Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts* (P-3925). Santa Monica, CA: The RAND Corporation.

Brown BB and Helmer O (1964) *Improving the Reliability of Estimates Obtained from a Consensus of Experts* (P-2986). Santa Monica, CA: The RAND Corporation.

Brown BB, Cochran S and Dalkey NC (1969) *The DELPHI Method, II: Structure of Experiments* (RM-5957-PR). Santa Monica, CA: The RAND Corporation.

Cantril H (1938) The prediction of social events. *Journal of Abnormal and Social Psychology* 33(3): 364–389.

Collins MJ (2002) *Cold War Laboratory: RAND, the Air Force, and the American State, 1945-1950*. Washington, DC: Smithsonian University Press.

Cornell JA (1993) W.J. Youden – The man and his methodology. *American Society for Quality Control Statistics Division Newsletter* 13(2): 9–18.

Dalkey NC (1967) *Delphi* (P-3704). Santa Monica, CA: The RAND Corporation.

Dalkey NC (1968) *Predicting the Future* (P-3948). Santa Monica, CA: The RAND Corporation.

Dalkey NC (1969) *The Delphi Method: An Experimental Study of Group Opinion* (RM-5888-PR). Santa Monica, CA: The RAND Corporation.

Dalkey NC and Helmer O (1962) *An Experimental Application of the Delphi Method to the Use of Experts* (RM-727/1-ABRIDGED). Santa Monica, CA: The RAND Corporation.

Dalkey NC and Helmer O (1963) An experimental application of the Delphi method to the use of experts. *Management Science* 9(3): 458–467.

Dalkey NC, Brown BB and Cochran S (1969) *The Delphi Method, III: Use of Self-Ratings to Improve Group Estimates* (RM-6115-PR). Santa Monica, CA: The RAND Corporation.

Dalkey NC, Brown BB and Cochran S (1970) *The Delphi Method, IV: Effect of Percentile Feedback and Feed-In of Relevant Facts* (RM-6118-PR). Santa Monica, CA: The RAND Corporation.

Dayé C (2014) In fremden Territorien: Delphi, political gaming und die subkutane Bedeutung tribaler Wissenskulturen. *Österreichische Zeitschrift für Geschichtswissenschaften* 25(3): 83–115.

Dayé C (2016) ‘A fiction of long standing’: Techniques of prospection and the role of positivism in US Cold War social science, 1950–65. *History of the Human Sciences* 29(4–5): 35–58.

Deutsche Post AG (2009) *Delivering Tomorrow: Kundenerwartungen im Jahr 2020 und darüber hinaus. Eine globale Delphistudie*. Available at: https://www.competence-site.de/trends-der-zukunft-kundenerwartungen-im-jahr-2020/ (accessed 27 August 2018).

Duller M (2016) Internationalization of Cold War systems analysis: RAND, IIASA and the institutional reasons for methodological change. *History of the Human Sciences* 29(4–5): 172–190.

Ferguson SD (2000) *Researching the Public Opinion Environment: Theories and Methods*. Thousand Oaks, CA: SAGE.

Gordon TJ (2011) Obituary—Olaf Helmer, futures thinker. *Technological Forecasting and Social Change* 78: 1099–1100.

Gordon TJ and Helmer O (1964) *Report on a Long-Range Forecasting Study* (P-2982). Santa Monica, CA: The RAND Corporation.

Häder M (2009) *Delphi-Befragungen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Helmer O (1963) *The Systematic Use of Expert Judgment in Operations Research* (P-2795). Santa Monica, CA: The RAND Corporation.

Helmer O (1964) *Convergence of Expert Consensus through Feedback* (P-2973). Santa Monica, CA: The RAND Corporation.

Helmer O (1966) *Social Technology*. New York; London: Basic Books.

Helmer O (1967a) *Analysis of the Future: The Delphi Method* (P-3558). Santa Monica, CA: The RAND Corporation.

Helmer O (1967b) *Systematic Use of Expert Opinions* (P-3721). Santa Monica, CA: The RAND Corporation.

Helmer O (1967c) *The Future of Science* (P-3607). Santa Monica, CA: The RAND Corporation.

Helmer O (1977) *The Utility of Long-Term Forecasting* (IIASA PP-77-5, Olaf Helmer Papers). Santa Monica, CA: The RAND Corporation.

Helmer O and Oppenheim P (1945) A syntactical definition of probability and of degree of confirmation. *Journal of Symbolic Logic* 10(2): 25–60.

Helmer O and Rescher N (1958) *On the Epistemology of the Inexact Sciences* (P-1513). Santa Monica, CA: The RAND Corporation.

Helmer O and Rescher N (1959) On the epistemology of the inexact sciences. *Management Science* 6(1): 25–52.

Helmer O and Rescher N (1960) *On the Epistemology of the Inexact Sciences* (R-353). Santa Monica, CA: The RAND Corporation.

Hempel CG and Oppenheim P (1945) A definition of ‘degree of confirmation’. *Philosophy of Science* 12(2): 98–115.

Heuer R Jr and Pherson RH (2011) *Structured Analytic Techniques for Intelligence Analysis*. Washington, DC: CQ Press/SAGE.

Hosiasson-Lindenbaum J (1940) On confirmation. *Journal of Symbolic Logic* 5(4): 133–148.

Hounshell D (1997) The Cold War, RAND, and the generation of knowledge, 1946-1962. *Historical Studies in the Physical and Biological Sciences* 27(2): 237–267.

Igo SE (2006) ‘A gold mine and a tool for democracy’: George Gallup, Elmo Roper, and the business of scientific polling, 1935–1955. *Journal of the History of the Behavioral Sciences* 42(2): 109–134.

Igo SE (2007) *The Averaged American: Surveys, Citizens, and the Making of a Mass Public*. Cambridge, MA; London: Harvard University Press.

Isaac J (2012) *Working Knowledge: Making the Human Sciences from Parsons to Kuhn*. Cambridge, MA: Harvard University Press.

Jardini D (2000) Out of the blue yonder: The transfer of systems thinking from the Pentagon to the great society, 1961-1965. In: Hughes AC and Hughes TP (eds) *Systems, Experts, and Computers: The Systems Approach in Management and Engineering, World War II and After*. Cambridge, MA; London: The MIT Press, 311–358.

Kahn H (1961) *On Thermonuclear War: Three Lectures and Several Suggestions*. Princeton, NJ: Princeton University Press.

Kaplan A, Skogstad AL and Girshick MA (1950) The prediction of social and technological events. *Public Opinion Quarterly* 14(1): 93–110.

Kaplan F (1983) *The Wizards of Armageddon*. New York: Simon & Schuster.

Kecskemeti P (1958) *Strategic Surrender: The Politics of Victory and Defeat*. Stanford, CA: Stanford University Press.

Kuhn TS (1970) *The Structure of Scientific Revolutions*, 2nd Enlarged edn. Chicago, IL: The University of Chicago Press.

Light JS (2003) *From Warfare to Welfare: Defense Intellectuals and Urban Problems in Cold War America*. Baltimore, MD; London: Johns Hopkins University Press.

Linstone HA and Turoff M (1975) *The Delphi Method: Techniques and Applications*. London; Reading, MA: Addison-Wesley.

Lusinchi D (2017) The rhetorical use of random sampling: Crafting and communicating the public image of polls as a science (1935–1948). *Journal of the History of the Behavioral Sciences* 53(2): 113–132.

McGregor D (1938) The major determinants of the prediction of social events. *Journal of Abnormal and Social Psychology* 33(2): 179–204.

National Institute of Science and Technology Policy (NISTEP) (2010) The 9th Delphi Survey (Summary). Available at: http://data.nistep.go.jp/dspace/bitstream/11035/696/1/NISTEP-NR140-SummaryE.pdf (accessed 27 August 2018).

Platt J (1996) *A History of Sociological Research Methods in America, 1920-1960*. Cambridge: Cambridge University Press.

Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Rescher N (1997) H2O: Hempel-Helmer-Oppenheim, an episode in the history of scientific philosophy in the 20th century. *Philosophy of Science* 64(2): 334–360.

Rescher N (2006) The Berlin school of logical empiricism and its legacy. *Erkenntnis* 64(3): 281–304.

Rindzevičiūtė E (2016) *The Power of Systems: How Policy Sciences Opened Up the Cold War World*. Ithaca, NY: Cornell University Press.

Riska-Campbell L (2011) *Bridging East and West: The Establishment of the International Institute for Applied Systems Analysis (IIASA) in the United States Foreign Policy of Bridge Building, 1964-1972*. Helsinki: The Finnish Society of Sciences and Letters.

Rocco P (2011) Wissensproduktion in der RAND Corporation. In: Greiner B, Müller TB and Weber C (eds) *Macht und Geist im Kalten Krieg*. Hamburg: Hamburger Edition, 301–320.

Rohde J (2013) *Armed with Expertise: The Militarization of American Social Research during the Cold War*. Ithaca, NY: Cornell University Press.

Sapolsky HM (2004) The science and politics of defense analysis. In: Cravens H (ed.) *The Social Sciences Go to Washington: The Politics of Knowledge in the Postmodern Age*. New Brunswick, NJ: Rutgers University Press, 67–77.

Smith BLR (1966) *The RAND Corporation: Case Study of a Nonprofit Advisory Corporation*. Cambridge, MA: Harvard University Press.

Steinmetz G (2005) *The Politics of Method in the Human Sciences: Positivism and Its Epistemological Others*. Durham, NC: Duke University Press.

Thomas W (2015) *Rational Action: The Sciences of Policy in Britain and America, 1940-1960*. Cambridge, MA: The MIT Press.

Tolon K (2011) *The American Futures Studies Movement (1965-1975): Its Roots, Motivations and Influences*. Ames, IA: Iowa State University.

Tolon K (2012) Future studies: A new social science rooted in Cold War strategic thinking. In: Solovey M and Cravens H (eds) *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature*. New York: Palgrave Macmillan, 45–62.

Author biography

Christian Dayé is a Postdoctoral Assistant in the Department of Sociology at the Alpen-Adria-Universität Klagenfurt, Austria. His work is located at the intersections between sociology of science, sociology of knowledge, and the history of the social sciences. He has held a long-standing interest in the various uses of social scientific knowledge and their repercussions on the organizational and intellectual shape of these sciences.