SOCIO-HISTORICAL STUDIES OF THE SOCIAL AND HUMAN SCIENCES

Experts, Social Scientists, and Techniques of Prognosis in Cold War America

Christian Dayé

palgrave macmillan

Socio-Historical Studies of the Social and Human Sciences

This series is the first to focus on the historical development and current practices of the social and human sciences. Rather than simply privileging the internal analysis of ideas or external accounts of institutional structures, it publishes high quality studies that use the tools of the social sciences themselves to analyse the production, circulation and uses of knowledge in these disciplines. In doing so, it aims to establish Socio-Historical Studies of the Social and Human Sciences as a scholarly field in its own right, and to contribute to a more reflexive practice of these disciplines.

More information about this series at
http://www.palgrave.com/gp/series/15409

Christian Dayé

Experts, Social Scientists, and Techniques of Prognosis in Cold War America

palgrave
macmillan

Christian Dayé
Graz, Austria

Socio-Historical Studies of the Social and Human Sciences
ISBN 978-3-030-32780-4 ISBN 978-3-030-32781-1 (eBook)
https://doi.org/10.1007/978-3-030-32781-1

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgments

I started writing this book on 6 August 2015. On this particular day, the radio station I usually listen to while preparing breakfast played the sound of the Peace Bell in the Peace Memorial Park in Hiroshima. It did so to mark the 70th anniversary of the dropping of the atomic bomb "Little Boy" over this town in South Japan. I had visited Hiroshima the year before. There, despite having been aware of much of the historical facts and circumstances of this attack, and the devastating suffering it brought to the city's inhabitants, the cultural rupture that the new superweapon brought with it became tangible to me. Little Boy and his fellow Fat Man, the bomb that was dropped over Nagasaki some days later, marked the rise of a new culture of insecurity, a culture that was, for the first time in history, truly of global scope.

I would like to start by thanking those who were available for face-to-face interviews. These were Daniel Ellsberg, Joan D. Goldhamer, Theodore J. Gordon, Nicholas Rescher, and Martin Shubik. My ambition and, indeed, hope was that they find resemblance between the events described in the book and their own recollections. I am grateful also for the permissions to use parts of the interviews in the text. I also relied on a series of interviews carried out by Martin J. Collins as part of Smithsonian Institution's RAND Oral History Project as well as on an interview with Olaf Helmer carried out by Kaya Tolon, which is included in the annex of Tolon's PhD thesis "The American Future Studies Movement

(1965–1975): Its Roots, Motivations, and Influences". Ames, Iowa: Iowa State University, 2011. All these sources provide a sense of the RAND culture that cannot be found in research reports and articles.

I visited RAND twice, in 2011 and 2013, and received considerable support at the place as well as via digital channels over the years. I am grateful to Vivian Artebery who, although her primary responsibilities were with other issues, found the time to support me prior to and during my first visit. Vivian also showed me the Olaf Helmer papers that had just been given to RAND upon his death in spring 2011. Roberta Shanman and Ann Horn provided additional support at that time. My second visit to RAND was prepared by Roberta and Susan Scheiberg, and Susan and Cara McCormick were instrumental in clarifying issues that popped up in the final stages of the research and manuscript preparation process. I specifically want to thank Cara for being of such great support in securing the access to historical photographs, some of which are included, with the generous permission of the RAND Corporation Archives, in the following pages.

Mary Osielski and Brian Keough from the M. E. Grenander Department of Special Collections and Archives, State University of New York at Albany, were kind hosts when I inspected the papers from Hans Speier in October 2007. Further, I am grateful to the staff at the archives of the Hoover Institution and at Hoover Library, both located at Stanford University, Stanford (CA); at the Library of Congress, Washington (DC); at the New York Public Library in New York City; and at the Massachusetts Institute of Technology Libraries, Department of Distinctive Collections, Cambridge (MA). These archives are paradises for historians (of social science), and this is in large parts due to the continued effort of their staff. Permission to reproduce materials was granted by the RAND Corporation Archives and the Massachusetts Institute of Technology.

People do not like insecurities, and over the centuries of human civilization, countless numbers of ways have been devised to reduce insecurity. Among these ways, the council of others deemed wise and experienced ranks high. A lot of wise and experienced people offered support over the years, among them Katelin Albert, Peter Baehr, Daniel Bessner, Martin J. Collins, Matthias Duller, Christian Fleck, Peter Gasser-Steiner, Sharon

Ghamari-Tabrizi, Robert Jackall, Philipp Korom, Stefan Laube, Barbara Louis, E. Stina Lyon, Jennifer Platt, Andrea Ploder, Werner Reichmann, Matthias Revers, Erik Schneiderhan, Peter D. Simonson, Mark Solovey, Richard Swedberg, Stephen Turner, Hans Vabitsch, Angelika Wetterer, and Mario Wimmer. Further, the book followed me through times of repeated professional changes, and I was offered support—in one way or the other—from people and institutions at the University of Graz, the University of Klagenfurt, and Graz University of Technology, where I am now based. If I did not accept all offers, this had nothing to do with their inherent qualities, but instead was dictated by the work progress; often, to receive the offer was support enough.

Christian Fleck and Karen Meehan took the burden to read the whole manuscript and provided valuable suggestions. Also, I am indebted to the series editors and two anonymous reviewers for providing helpful comments and constructive critique. The remaining errors are proud results of my own stubbornness.

I have used parts of the materials presented in this book in earlier articles that approached them with different research interests than the one that guides the argument here. In these cases, I made a reference to the respective article. The cover of this book was made using a poster that the artist Hristo Penev created for the Bulgarian National Peace Committee in 1983. Its title, set in orange capital letters on the top of the poster, was bilingual and read "We допуснем ли?" (Bulgarian, translit. We dopusnem li?) and below, in English: "Should we let it happen?".

Research for this book was supported by a fellowship from the School of Social and Economic Sciences, University of Graz, Austria (2011), a project funded by the Austrian Science Foundation (Fonds zur Förderung der wissenschaftlichen Forschung—FWF, project no. P 24693-G16, 2012–2015), and a grant awarded by the Wissenstransferzentrum Süd (2015). Further support came from various awards, among them the *Theodor Körner Preis* (2011); the *Young Scholar Prize* of the Research Committee on the History of Sociology, International Sociological Association (2012); and the *SOWI im Dialog 2013* prize for best dissertation, awarded by the School of Social and Economic Sciences, University of Graz, Austria.

Praise for *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*

"Much of modern social science has its origin in the famous RAND Corporation, and this includes studies of the future, or futurology. In this well-documented and exciting study, the author explores this part of the work at RAND with the help of little-known archival material and interviews. Extremely interesting!"

—Richard Swedberg, Professor Emeritus, *Cornell University, USA*

"Paying close attention to the historical context and practical details of social science methods in the Cold War, Dayé's book speaks to questions about truth and democracy, which are both timeless and urgent."

—Monika Krause, Assistant Professor, *LSE, UK*

Contents

1 Introduction: A Culture of Insecurity and Its Experts 1

2 Experts, Think Tanks, and the Delicate Balance of Public Trust 15

3 The Wisdom of the Group: RAND's First Experiments with Expert Prediction, 1947–1951 41

4 Negotiating Rules for the Game: Political Games at RAND, 1954–1956 77

5 The Oracle's Epistemology: Expert Opinions as Scientific Material, 1955–1960 129

6 The Boredom of the Crowd: The Long-Range Forecasting Delphi, 1963–1964 157

7 Conclusion: The Strength of Epistemic Hopes 205

Index 239

List of Figures

Fig. 2.1 Aerial view of the second RAND building, Santa Monica, CA. Photograph was taken in 1960 (RAND Corporation Archives, Santa Monica, CA) 25

Fig. 3.1 The convergence of estimates (Source: Dalkey and Helmer 1962, 15; reproduced with permission) 64

Fig. 4.1 Debate during the STRAW Strategic Air War Game, 1953; the man in the center without a tie is John D. Williams, longtime director of the Mathematics Division (RAND Corporation Archives, Santa Monica, CA) 84

Fig. 4.2 RAND experts discussing air strategy during a SAFE war planning game, January 1963; the man in the center wearing a black suit is Milton G. Weiner, a major proponent of war gaming at RAND (RAND Corporation Archives, Santa Monica, CA) 87

Fig. 4.3 A RAND analyst throwing dices to introduce "chance" into a SAFE strategic air-war planning game, January 1963 (RAND Corporation Archives, Santa Monica, CA) 88

Fig. 4.4 The SAFE game's umpire team observing the players, January 1963; Olaf Helmer, in the checked jacket on the left, his back turned to the camera, functioned as the game director (RAND Corporation Archives, Santa Monica, CA) 89

Fig. 4.5 Postmortem evaluation discussion between Olaf Helmer and Milton G. Weiner, SAFE strategic air-war planning game, January 1963 (RAND Corporation Archives, Santa Monica, CA) 90
Fig. 5.1 The degree of confirmation as a bridge concept 139
Fig. 6.1 Results of panel 1 on scientific breakthroughs. (Source: Gordon and Helmer 1964, 12; reproduced with permission) 172
Fig. 7.1 Work session of CONEX I at Endicott House, photograph, ca. 1968. (Photograph with seven men sitting around table. The man on the left is smoking a pipe and the third from the left adjusts his eyeglasses, undated, Lincoln P. Bloomfield Papers, MC 326, Box 11, CONEX I, folder 1/2. Massachusetts Institute of Technology Libraries, Department of Distinctive Collections, Cambridge, MA) 215
Fig. 7.2 "What do you mean you are going on strike?," photograph, 1968. (Lincoln P. Bloomfield is standing in the middle, face to the camera (Lincoln P. Bloomfield Papers, MC 326, Box 11, CONEX I, folder 1/2. Massachusetts Institute of Technology Libraries, Department of Distinctive Collections, Cambridge, MA)) 218

List of Tables

Table 3.1 Sample items of the study on exert prediction by Kaplan et al. (1950, 109–110) 46
Table 3.2 Predictive success per group (Adopted from Kaplan et al. 1950, 104) 51
Table 3.3 Success by basis statements (Adopted from Kaplan et al. 1950, 107) 52
Table 3.4 Bomb estimates in the first round (Adopted from Dalkey and Helmer 1962, 7) 60
Table 3.5 Bomb estimates in the third round (Adopted from Dalkey and Helmer 1962, 7, 10; own calculations) 62
Table 3.6 Bomb estimates (totals) in the fourth round (Adopted from Dalkey and Helmer 1962, 10, 12; own calculations) 63
Table 3.7 Final and corrected final bomb estimates (Adopted from Dalkey and Helmer 1962, 7, 10; own calculations) 63
Table 4.1 General and specific factors influencing Soviet policy. (Reproduced from Goldhamer 1955a, 7–8, with permission of the RAND Corporation) 105
Table 4.2 Matrix of the likelihood of Soviet aggression. (Reproduced from Goldhamer 1955a, 12, with permission of the RAND Corporation) 106
Table 6.1 Exemplary prediction items, results of round two. (Adopted from Gordon and Helmer 1964, 8; own calculations) 166

Table 6.2 Collected results for items nos. 19 and 10. (Adopted from Gordon and Helmer 1964, 8, 9, 10; own calculations) 168
Table 6.3 The probability of another major war. (Adopted from Gordon and Helmer 1964, 27) 173
Table 6.4 List of war prevention measures. (Adopted from Gordon and Helmer 1964, 28–31) 177
Table 6.5 Exemplary questions and results from Brown and Helmer's 1964 study. (Adopted from Brown and Helmer 1964, 3, 4, 6; own calculations) 187
Table 7.1 Types of experts according to Turner (2001) 228

1

Introduction: A Culture of Insecurity and Its Experts

A Culture of Insecurity

Like almost any technology, the atomic bomb had effects that extended far beyond the field of its immediate use into the wider sphere of culture. When media reports of their detonation over Hiroshima and Nagasaki circulated worldwide, "Little Boy" and "Fat Man" became the symbols of the emerging global culture of insecurity. To an extent, unseen before in a weapon, the cultural effect of the weapon became its primary asset. Writing in retrospect, US Secretary of War Henry Stimson made clear that "the atomic bomb was more than a weapon of terrible destruction; it was a psychological weapon" (Stimson 1947, 66). The two bombs had killed hundreds of thousands of people and left many more injured. Yet precisely because of its cruelty, the use of the atomic bomb as a means of deterrence became more effective than its actual detonation. The bomb's primary objective was political and cultural: to create an atmosphere of existential fear and insecurity among those threatening the values of the West. And its outreach was global: while the bombs had been dropped over Japan, the Soviet Union and its potential allies around

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*,
Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_1

the globe became the main addressee of the psychological and cultural effects of the bomb.

Yet, as US strategists soon were to realize, a strategy of deterrence always has repercussions on all parties involved. The culture of insecurity that was emblematized and initiated by the launching of the bombs over Hiroshima and Nagasaki was not restricted to the "East"—quite to the contrary, it pervaded US American culture, especially after the first successful test detonation by the Soviets in 1949. In contrast to its physical radiation, the bomb's psychological radiation could not be restricted in terms of space. Its cultural effects were global. As a "psychological weapon," it paradoxically also affected those who used it.

Two factors fostered the diffusion of the culture of insecurity in the United States. Chief among them was *technical ignorance*: were the Soviets capable of producing a bomb? How well developed was their knowledge of nuclear physics? Could they secure the service of experienced German scientists, as the United States did? Or, being one of the occupying forces in postwar Germany, did they get access to crucial data and results unknown to US scientists? In large parts, the Cold War game of deterrence was played along lines of technical ignorance, with each side attempting to occlude its capabilities and to deceive the other side into assuming the worst. However, a *theoretical* or *philosophical void* accompanied this technical ignorance. While those involved in the game of deterrence tried to apply their means to the most desired outcome, they had to do so without knowing the rules of the game. The atomic bomb profoundly changed how people thought about war. Upon reading about the dropping of the Hiroshima atomic bomb in the newspaper, accomplished strategist and Yale professor Bernard Brodie reportedly "turned to his wife and said, 'Everything that I have written is obsolete'" (Kaplan 1983, 10). To Brodie, and to a majority of his fellow strategists, the bomb had destined the entire body of military knowledge accumulated over the past centuries to be moved to the deep caves of archival oblivion. The world was confronted with a weapon of disastrous force but had not developed theories to understand, let alone handle it. "The whole conception of modern warfare, the nature of international relations, the question of world order, the function of weaponry, had to be thought through again. Nobody knew the answers; initially, not many had even the right questions" (Kaplan 1983, 10).

One common human reaction to ignorance and insecurity is to endow large and potentially unjustified amounts of trust in selected social or cultural positions and their proponents. When the world set out to return to a peacetime organization of life in the aftermath of World War II, there was a window of opportunity for a new social figure to climb up the ladder of cultural relevance. The age of the expert in foreign and military politics was about to dawn; and by creating a culture of ignorance and insecurity, the atomic bomb acted as leverage for the expert's success in entering the court of power. The bomb had completely shaken up the structures in this field. Almost every claim to authority had to be newly established and negotiated. Scientists entered the struggle by arguing that a scientific procedure was the most reasonable way to cope with the overwhelming task of restructuring US military defense and foreign policy. They claimed the opportunity to participate in political and military decision processes. In the same breath, they emphasized that earlier experience was no convincing guide in the realignment of the field.

Conceived of as a mediator between knowledge and power, the expert occupied an important position in US Cold War culture. To describe the epistemological characteristics of this position is the objective of this book. It focuses on the capacities ascribed to this social figure and the hopes that were related to it in this culture of insecurity. To anticipate the conclusion, and quite unsurprisingly, both the ascribed capacities and hopes were grand. In the early years of the age of the expert, mass media treated this new figure as a source of the general reason (Brint 1994; Herman 1995). With regard to experts in foreign policy, a widespread hope was that they were able to level out the warmongering impulses from military officials as well as the shortsightedness of political leaders. Since using the atomic bomb was so obviously irrational and inhuman, the expectation toward the civilian experts in foreign policy was that they would ensure a level of reason and rationality in the decision processes.

This expectation was not confined to mass media but was an essential part of the self-image of those scientists who came to be addressed as experts. As such, it influenced their doings. The analytic approach developed in this book makes use of this relation by examining a specific line of methodological thinking within the social sciences. The main idea of this line of thinking was that the expert could be used as a source of knowledge about the future. This idea was developed at the RAND Corporation, a research organization that emerged from a collaboration between the US Air Force and Douglas Aircraft Company shortly after the end of World War II. Here, scientists created a series of techniques that aimed at producing knowledge about the future by systematically collecting expert opinions and allowing for a certain degree of interaction among these experts.

The book explores a series of studies done in the 1950s and 1960s by two groups of RAND researchers. One group, consisting of members of the Mathematics Division, designed the Delphi technique; the other, consisting of members of RAND's Social Science Division, proposed and developed a technique they called political gaming. Delphi distributed questionnaires to a pool of experts, asking them to estimate when specific future events would take place. These estimations were then averaged and fed back to the participating experts with the intent to have them think about their initial answers again. The expectation behind this repeated procedure was that the estimates would converge over time to a range which could then be called expert consensus. The political games carried out at RAND invited experts to participate in various groups, each of them representing a national government, or a block of national governments (e.g., Western Europe). The groups were then asked to discuss how the government they represented would react toward specific actions of the other governments, thereby simulating a political and military crisis. After each step, game leaders would collect the decision of the groups and use them to synthesize a new game state. Both techniques are still in use today, mainly in the areas of applied policy, market, and trend research.

As the ensuing chapters show, there are considerable differences between the two techniques. Above all, they embody different epistemologies and philosophies of science. They bear the marks of the academic tribes from which their inventors came—the program of logical empiricism in the philosophy of science in the case of Delphi, the sociology of knowledge developed by German sociologist Karl Mannheim in the case of political gaming (cf. Dayé 2014). On the one hand, the logicians involved in creating Delphi, chief among them Olaf Helmer, Norman Dalkey, and Nicholas Rescher, all studied with important representatives of logical empiricism—Rudolf Carnap, Carl Gustav Hempel, Hans Reichenbach, to name a few (cf. Dayé 2016). On the other hand, the leading social scientists at RAND—Hans Speier, Herbert Goldhamer, and Paul Kecskemeti—had all known Karl Mannheim personally and had been influenced by his understanding of a social determination of ideas. RAND's political gaming incorporated ideas formulated in Mannheim's classic text, *Ideology and Utopia* (Mannheim 1997; orig. 1929; cf. Bessner 2014).

Thus, while there are interesting differences between the two techniques, some of which I explored in earlier articles, the main interest of this book is with the similarities of the two techniques of prognosis. Since both techniques rely on expert opinions or expert knowledge to produce statements about the future, these techniques can be understood as manifestations of the expectations and hopes related to the alleged capacities of the expert. Thus, an analysis of these expectations and hopes might help us understand how in a culture of insecurity, trust in a social figure was created, justified, maintained, and corroborated.

Techniques of Prognosis

Many in the military saw the advent of the expert as an attempt to oust military officers and other proponents of the armed forces from their positions of authority. However, this was more than just a struggle over organizational power. It concerned the question of whom to entrust with decision-making in the new culture of insecurity. The stakes were unprecedentedly high and nobody knew the rules of the game. This corroborated the experts' claim that what the world required in order to confront the challenges of the new culture was the production of new knowledge by the sciences, not the outdated wisdom passed on by one generation of military artisans to the next (Connelly et al. 2012). As a matter of fact, many military officials acknowledged, if somewhat grudgingly and hesitantly, that times had changed. As US Air Force General Curtis E. LeMay, who had been involved in planning and executing the strategic bombing campaign against Japan during World War II and after the war became deputy chief of Air Staff for Research & Development at the Pentagon, claimed in 1946: "Warfare is no longer a military problem" (cited in Jardini 2000, 314).

In the attempt to cope with the culture of insecurity, the newly appointed experts on foreign policy and military strategy perceived it as their task to develop techniques of prognosis, instruments, and procedures informed by (social) scientific methodology with the objective of "envisioning an unknown future" (Mallard and Lakoff 2011, 339). Bestowed with the expectation to deliver to the nuclear age what the augurs delivered to the people of ancient Rome, they searched for innovative ways of social scientific prognosis. The most established form of a scientific prognosis, statistical extrapolation, was deemed inadequate both with regard to scientific-technological advances and to social and cultural processes. In both cases, data were rare. Yet, more importantly, non-schematic actions on the micro level could lead to leap-like changes or revolutions that completely transformed the social, cultural, and societal scenery on the macro level. Statistical data of the past might help, the experts were convinced, but it would not suffice to allow for solid prognoses. "Thoughtful observers had recognized that the existential fact of the bomb altered time significantly and permanently" (Ghamari-Tabrizi 2012, 269). This had a huge impact on all those involved in decision-making in the nuclear age; for them, "the present, future, and conditional worlds ran together" (Ghamari-Tabrizi 2012, 269; see also Byrne 2010).

One solution to this methodological problem of social prognosis that the foreign policy experts explored is in the focus of this book. This solution was to conceive of experts as persons with a privileged amount of—explicit as well as tacit—knowledge and to devise techniques and tools that would make systematic use of this knowledge in producing prognoses. The methodological solution pursued by the RAND researchers, however, implies a telling irony. In search of ways to cope with the culture of insecurity, decision-makers asked experts to deliver prognoses. They trusted them to find methods and ways to deliver stable knowledge of the future of the social, technological, and political sphere. And the solution proposed by the experts was: ask scientific experts. The circular character of this argument astonishes, especially when one realizes that it went virtually unnoticed by the experts themselves. From the historical distance, however, we can use this circularity. To explore how experts feature in the two techniques informs us very broadly of the expert's role in US American Cold War culture. The epistemic role attributed to the expert within the relatively narrow frames of the techniques can be interpreted as a manifestation of the hopes and expectations attached very generally to the social figure of the expert during this period. And these expectations and hopes, in turn, formed the basis of trust.

Two concepts, thus, are at the core of this study: epistemic roles and epistemic hopes. Building on the traditional sociological concept of the social role as a bundle of expectations attributed to a specific social position, the epistemic role of the expert consists of the expectations related to the participating experts in the methodologies of the various techniques of prognosis. What knowledge can we expect from experts? What, and how, does she/he know? And what is the epistemic character of expert statements? Epistemic hopes, on the other hand, refer more generally to the cultural functions of expert knowledge. In a culture of insecurity, experts were trusted to bring clarity, certainty, and guidance into an increasingly Byzantine world. That these hopes were sometimes largely exaggerated, has been repeatedly observed (recently for instance by Collins 2014, 1–11); that coeval experts nonetheless thought them plausible is evinced by the abovementioned circularity. However, both the epistemic roles assigned to experts in the techniques of prognosis and the epistemic hopes attributed to them in contemporary culture are key to understanding the rise of the age of experts. It is their framing as sources for understanding a social figure so crucial for Cold War culture, the scientific expert, that motivates and at the same time justifies the in-depth study of the two techniques.

In focusing on techniques, this book takes a path only rarely followed in the historiography of the (social) sciences. Most publications in this field are concerned with either scholars or theories (Fleck 2015; Fleck and Dayé 2015). However, in concurrence with other scholars sharing this focus (e.g., Platt 1996), I argue that there is a lot to learn from analyzing in some depth the history of methods or techniques. While analyzing techniques sheds light on the actual practices of social researchers, it also allows for exploring the more intrinsic sediments of coeval ideas, conceivabilities, imaginations, and worldviews. To focus on these sediments means to write the history of the social sciences in a way that recognizes the core tenet of historical epistemology, as proposed by Gaston Bachelard (1984, 2002): scientific reasoning shows implicit and nonrational traces of human existence: fears, myths, beliefs, dreams, and so on; and sometimes, these traces can explain both the success of a scientific idea and its demise (on Bachelard, see Lecourt 1975; Tiles 1984; Chimisso 2013; Dayé 2019).

The history of social science—or, for that matter, of any science—can be written in various ways and for various reasons (Dayé and Moebius 2015). Yet, chief among the reasons is the wish to provide something of value to current problems tackled in the discipline. In order to do so, histories can attempt to strengthen the disciplinary identity, for example, by the construction of a disciplinary canon and the continued critical assessment of its value. They can serve as resources in the teaching of students about the core ideas of the discipline. They can reflect on the current shape of the discipline and its position in the social, cultural, and societal whole. And finally, they can attempt to inform current research and theorizing, for example, by historically contextualizing the development of important notions in order to ensure a more sophisticated, history-conscious handling of the notions, or by understanding the history of a science as a strategic research material (Merton 1987) that allows for addressing contemporary scientific problems (cf. Dayé 2018b).

This last position motivated the writing of this book, and it is the one that comes the closest to Bachelardian historical epistemology. The history of the two techniques of prognosis is of interest because it delivers knowledge that might lead to more sophisticated use of the two techniques. By systematically addressing the epistemic role of the expert, we question the tacit assumptions on which the two techniques—and those many that are similar in this regard—rely. Yet beyond that, the history of the two techniques also allows us to address contemporary (or perhaps eternal) questions regarding the role of hope in science—how it informed the invention, development, and reception of scientific techniques.

There is another aspect of Bachelardian historical epistemology that I had to confront in writing this book. To seek for the sediments of implicit ideas requires one to be very attentive to details. Apart from the risk that the reader might perceive an overwhelming precision with regard to outmoded and seemingly peripheral ideas to be exaggerated at best, unjustified at worst, there is also the fact that one is tempted to enter a methodological discourse about the method itself and its (current or optimal) use. While I attempted to keep reflections in this direction to a minimum in order to not endanger the readability of the text, such reflections certainly relate to what the word "epistemology" meant in this linguistic compound to Bachelard: investigate the history of a line of thinking in order to improve its current use by sensitizing current practitioners for the historical, psychological, cultural—brief: nonscientific—ideas it entails as sediments.

Some final words on the use of specific terms. To begin with the easier one, I use *technique* as a very generic term, comprising both procedures deemed to produce *true* knowledge (as implied, e.g., in the term scientific method) and procedures deemed to produce *useful* knowledge (as implied, e.g., in the term tool). As anyone with a basic understanding of the history of science will realize, the debate on whether there is a fundamental link between these two realms of knowledge is centuries-long and ongoing, and it is not the task of this book to explore the underlying understandings of science and its counterparts (cf. Dayé 2018a). The term technique is used in this book to cover all sorts of reasoned and explicit procedures in the realm of prognosis.

Further, there is considerable confusion in the literatures involved with regard to the use and the precise meanings of terms like prognosis, prognostication, prediction, prospection, forecast, and foresight. While I have not tried to iron out this confusion, the following definitions in my view still match the understanding currently shared by the majority of commentators. In this book, *prognosis* is used as a generic term, encompassing a variety of types of foreknowledge production, that is, knowledge about the future. If all the techniques discussed in this book are concerned, I will thus speak of techniques of prognosis.

Three specific types of foreknowledge are treated in this book: prediction, forecast, and prospection. Each of these types results in statements of different epistemic status. A *prediction* is defined as a statement made by a person (or an organization) without reference to evidence corroborating the statement. The credibility of the statement fully depends on the authority of the person (or organization) uttering the statement.[1] In contrast, a *forecast* always relies on a set of evidence (or data). Further, this evidence is assessed "using tools not easily employed by the general public" (Friedman 2014, x), but requiring some specialist (scientific) training. The relation, however, between evidence and forecast is not strictly logical or mathematical; a forecast is not simply a statistical projection but involves judgment. Finally, a *prospection* singles out factors that are relevant in shaping the future and explores their interdependence. A prospection refrains from making firm statements about the future, but instead indicates potentialities and tries to determine how current developments might play out in the future. Thus, as Mallard and Lakoff (2011, 339–340) note, techniques of prospection are sometimes used to understand the present rather than the future. For instance, in national security, the use of these techniques helps to understand whether or not ambiguous events in the present can pose security threats in the future.

None of these types of foreknowledge had its origin at RAND; and as the current historical literature on prognosis shows, scientists and entrepreneurs of various stripes developed techniques across the three types (Andersson 2018; Andersson and Rindzevičiūtė 2015; Friedman 2014; Harper 2012; McCray 2013; Pietruska 2018). Also, many of these earlier techniques had involved expert opinions. What the RAND researchers added to this already extant body of knowledge, however, was interaction. The techniques developed at RAND were diverse, but all provided for some form of (controlled) interaction between the experts. The idea that emerged at RAND was that interaction would lead the experts contributing to these prognoses to produce results of higher stability and, thus, credibility.

Each of the subsequent chapters introduces a specific phase of RAND prognostic studies. Since the focus is on the development of the core idea—systematically using expert opinions to produce prognoses—the chapters are ordered chronologically. After Chap. 2 introduces the place of origin of the various techniques of prognosis, the RAND corporation, Chap. 3 describes the first attempts at RAND to produce predictions based on expert knowledge. In 1951, these attempts led to the first Delphi study, yet there had been an earlier RAND study from 1948 that the Delphi developers conceived of as “precursor.” The analysis shows that the two studies confronted the expert with different epistemic expectations. While in the precursor study, the expert had the epistemic role to produce a prediction, the first Delphi study expected the expert to deliver a forecast. Chapter 4 then turns to a series of political games conducted by RAND’s Social Science Division between 1954 and 1959. This technique, in contrast, attributed to the expert the epistemic role of prospection. Chapter 5 describes an attempt to develop a philosophical foundation for the various techniques of prognosis developed at RAND. In line with the first Delphi study described in Chap. 3, the proposed “epistemology of the inexact sciences” was built around the idea that under certain conditions experts were able to deliver forecasts. The ensuing Chap. 6 assesses the long-range Delphi study carried out by members of the Mathematics Division in 1963. This study became the paradigmatic example for the use of the technique. The curious finding is that in spite of the earlier effort spent on philosophically corroborating the use of experts to produce *forecasts* (described in Chap. 5), the long-range Delphi study was again based on the idea that experts could come up with *predictions*. This applied also to a parallel study that addressed methodological issues of Delphi. Prior to a summary of the book’s main argument, Chap. 7 sketches the further trajectories of techniques of prospection.

Note

1. Although not treated in this book, a *prophecy* thus is a subtype of a prediction where the credibility fully depends on the ascribed transcendental abilities of the person making the prediction.

References

Andersson, Jenny. 2018. *The Future of the World: Futurology, Futurists, and the Struggle for the Post Cold War Imagination*. Oxford and New York: Oxford University Press.

Andersson, Jenny, and Eglė Rindzevičiūtė, eds. 2015. *The Struggle for the Long-Term in Transnational Science and Politics: Forging the Future*, Routledge Approaches to History 11. New York and London: Routledge.

Bachelard, Gaston. 1984. *The New Scientific Spirit*. Trans. Arthur Goldhammer. Boston: Beacon Press.

———. 2002. *The Formation of the Scientific Mind. A Contribution to a Psychoanalysis of Objective Knowledge*. Trans. Mary McAllester Jones. Manchester: Clinamen.

Bessner, Daniel. 2014. Weimar Social Science in Cold War America: The Case of the Political Game. In *More Atlantic Crossings? European Voices in the Postwar Atlantic Community*, Bulletin of the German Historical Institute Washington DC, Supplement 10, ed. Jan Logemann and Mary Nolan, 91–109. Washington, DC: German Historical Institute.

Brint, Steven G. 1994. *In an Age of Experts: The Changing Role of Professionals in Politics and Public Life*. Princeton, NJ: Princeton University Press.

Byrne, Peter. 2010. *The Many Worlds of Hugh Everett III: Multiple Universes, Mutual Assured Destruction, and the Meltdown of a Nuclear Family*. Oxford and New York: Oxford University Press.

Chimisso, Cristina. 2013. *Gaston Bachelard: Critic of Science and the Imagination*. London and New York: Routledge.

Collins, Harry. 2014. *Are We All Scientific Experts Now?* Cambridge, UK and Malden, MA: Polity Press.

Connelly, Matthew, Matt Fay, Giulia Ferrini, Micki Kaufman, Will Leonard, Harrison Monsky, Ryan Musto, Taunton Paine, Nicholas Standish, and Lydia Walker. 2012. 'General, I Have Fought Just as Many Nuclear Wars as You Have': Forecasts, Future Scenarios, and the Politics of Armageddon. *American Historical Review* 117 (5): 1431–1460.

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

———. 2016. 'A Fiction of Long Standing:' Techniques of Prospection and the Role of Positivism in US Cold War Social Science, 1950–1965. *History of the Human Sciences* 29 (4–5): 35–58. https://doi.org/10.1177/0952695116664838.

———. 2018a. How to Train Your Oracle: The Delphi Method and Its Turbulent Youth in Operations Research and the Policy Sciences. *Social Studies of Science* 48 (6): 846–868. https://doi.org/10.1177/0306312718798497.

———. 2018b. A Systematic View on the Use of History for Current Debates in Sociology, and on the Potential and Problems of a Historical Epistemology of Sociology. *The American Sociologist* 49 (4): 520–547. https://doi.org/10.1007/s12108-018-9385-1.

———. 2019. Historische Epistemologie der Soziologie? Probleme eines Theorietransfers. In *Zyklos 5: Jahrbuch für Theorie und Geschichte der Soziologie*, ed. Martin Endreß and Stephan Moebius, 17–40. Wiesbaden: Springer VS.

Dayé, Christian, and Stephan Moebius, eds. 2015. *Soziologiegeschichte. Wege und Ziele*. Berlin: Suhrkamp stw 2144.

Fleck, Christian. 2015. Skizze einer Methodologie der Geschichte der Soziologie. In *Soziologiegeschichte: Wege und Ziele*, ed. Christian Dayé and Stephan Moebius, 34–111. Berlin: Suhrkamp.

Fleck, Christian, and Christian Dayé. 2015. Methodology of the History of the Social and Behavioral Sciences. In *The International Encyclopedia of Social and Behavioral Sciences*, ed. James D. Wright, vol. 15, 2nd ed., 319–325. Oxford: Elsevier.

Friedman, Walter A. 2014. *Fortune Tellers: The Story of America's First Economic Forecasters*. Princeton and Oxford: Princeton University Press.

Ghamari-Tabrizi, Sharon. 2012. Cognitive and Perceptual Training in the Cold War Man-Machine System. In *Uncertain Empire: American History and the Idea of the Cold War*, ed. Joel Isaac and Duncan Bell, 267–293. Oxford and New York: Oxford University Press.

Harper, Kristine C. 2012. *Weather by the Numbers: The Genesis of Modern Meteorology*. Cambridge, MA and London: MIT Press.

Herman, Ellen. 1995. *The Romance of American Psychology: Political Culture in the Age of Experts*. Berkeley; Los Angeles; and London: University of California Press.

Jardini, David. 2000. Out of the Blue Yonder: The Transfer of Systems Thinking from the Pentagon to the Great Society, 1961–1965. In *Systems, Experts, and Computers: The Systems Approach in Management and Engineering, World War II and After*, ed. Agatha C. Hughes and Thomas P. Hughes, 311–358. Cambridge, MA and London: The MIT Press.

Kaplan, Fred. 1983. *The Wizards of Armageddon*. New York: Simon & Schuster.

Lecourt, Dominique. 1975. *Marxism and Epistemology. Bachelard, Canguilhem and Foucault*. Trans. Ben Brewster. London: NLB.

Mallard, Grégoire, and Andrew Lakoff. 2011. How Claims to Know the Future Are Used to Understand the Present: Techniques of Prospection in the Field of National Security. In *Social Knowledge in the Making*, ed. Charles Camic,

Neil Gross, and Michèle Lamont, 339–377. Cambridge, MA and London: Harvard University Press.
Mannheim, Karl. 1997. *Ideology and Utopia*. London: Routledge.
McCray, Patrick. 2013. *The Visioneers: How a Group of Elite Scientists Pursued Space Colonies, Nanotechnologies, and a Limitless Future*. Princeton and Oxford: Princeton University Press.
Merton, Robert K. 1987. Three Fragments from a Sociologist's Notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials. *Annual Review of Sociology* 13: 1–28.
Pietruska, Jamie L. 2018. *Looking Forward: Prediction & Uncertainty in Modern America*. Chicago: University of Chicago Press.
Platt, Jennifer. 1996. *A History of Sociological Research Methods in America, 1920–1960*. Cambridge, UK: Cambridge University Press.
Stimson, Henry L. 1947. The Decision to Use the Bomb. *Bulletin of the Atomic Scientists* 3 (2): 37–41; 66–67.
Tiles, Mary. 1984. *Bachelard: Science and Objectivity*. Cambridge; London; and New York: Cambridge University Press.

2

Experts, Think Tanks, and the Delicate Balance of Public Trust

The Organization of Scientific Policy Advice

The very idea to conceive of "experts" as potential sources of foreknowledge drew upon the relative prominence of the term and the idea of "experts" in US-American public discourse. When in a speech during his 1912 campaign, presidential candidate Woodrow Wilson (cited in J. A. Smith 1991, 1) exclaimed: "God forbid that in a democratic country we should resign the task and give the government over to experts," the term "expert" was already well known, albeit with a pejorative connotation.[1] To be sure, the role knowledge had to play in the organization of polity had been an issue of debate since ancient times. In Western political philosophy, this debate has centered on concepts like freedom, order, democracy, and the public good. Famously, in his programmatic utopia *New Atlantis* first published postmortem in 1627, Francis Bacon (1521–1626) envisioned an order called Salomon's House, "the noblest foundation (as we think) that ever was upon the earth." Dedicated "to the study of the Works and Creatures of God" (Bacon 1989, 58), the order applied scientific methods to inform policy-making and steer the development of the visionary society. Certainly, Bacon was not the first to

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*, Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_2

bestow scientists with political influence. There are many earlier examples of visionary societies steered by scholars or scientists, among them the reign of philosopher kings developed in Platon's *Politeia*. All of these visions and programs devise institutions that relate knowledge to political power.

However, in these classic accounts, the separation between knowledge and power was mainly one of spatial or social geography. In order to bring these two together, one simply had to devise a meeting place. The idea that science and politics operate according to specific logics and thus are in large parts incompatible with each other appears to be of later origin. Most probably, the rise of this idea was tied to the rise of an educated bourgeoisie seeking to increase their autonomy. Be it as it may, during the eighteenth and nineteenth centuries the classic project of creating meeting places found its manifestation in the establishment of national academies of science. However, as crucial as national academies were for those scientific disciplines already well established at that time, they were not the most fertile soil for the social sciences that were yet barely formed. In their struggle for recognition as scientific endeavors, the social sciences did not find considerable support from the academies and instead turned toward the universities.

> The need of the modern state for more exact knowledge on which to base its decisions had led to the emergence of new categories of knowledge already in the eighteenth century, but these categories still had uncertain definitions and frontiers. […] It was in this context that the university (which had been in many ways a moribund institution since the sixteenth century, the result of having previously been linked too closely with the Church) was revived in the late eighteenth and early nineteenth centuries as the principal institutional locus for the creation of [policy relevant] knowledge. (Wallerstein et al. 1996, 6)

Similar things can be said about the early history of the social sciences in the United States. It has been suggested that the institutionalization of the social sciences at the American universities toward the end of the nineteenth century was essentially driven by the need for trained personnel in dealing with the social problems caused by rapid urbanization and impoverishment of certain social strata (J. A. Smith 1991, 24ff). New graduate programs at Johns Hopkins, Columbia, Chicago, and Wisconsin produced the first generation of trained social scientists—most notably Richard T. Ely, John R. Commons, and Frank Lester Ward—who “created a pattern in which teaching and public service were combined” and, as the first and exemplary experts, “helped build even closer ties between experts and the government” (J. A. Smith 1991, 28). In France, to bring another example, although there existed an academy specifically for the social sciences, the protagonists of the institutionalization of the social sciences nonetheless concentrated on the universities. “The formation of economics, psychology, and sociology as university disciplines arose primarily out of a struggle against the doctrines and practices of the Academy [of Moral and Political Sciences], and the very success of these university pioneers is the main reason why the Academy has virtually disappeared from collective memory” (Heilbron 2015, 15). Similarly, the efforts of early proponents of culture-oriented anthropology in the German-speaking countries to gain recognition for their scientific project surmounted to an authority struggle against representatives of the National Academy and their rigid understanding of the natural sciences as models of science *in toto* (Weiler 2018). The formation of social scientific disciplines occurred against the backdrop of the differentiation of two fields, the political and the scientific, and academies of science were often seen as institutions of a time when these two fields had been too strongly intermingled for science to follow its own ethos.

Consequently, neither the organizational innovations of the nineteenth century nor the utopian visions of the centuries before referred to the idea of the expert (or used the noun). The term used was scientist. In contrast to the scientist or the philosopher, the expert is an interstitial figure. The expert bridges the realms of knowledge and power, of science and policy. He or she fills a void that opens only after the two fields have sufficiently differentiated from each other and follow their own logic. As interstitial figures, experts had to find ways to unite the divergent expectations emanating from four fields: academic science, politics, businesses, and the media public (cf. Medvetz 2012). The resulting ambivalence is obvious in the various labels these experts have received over the decades, some inventive, some pithy, some apparently neutral.

In US foreign policy discourses, they have been called *defense intellectuals* by some, a label that alludes to a breadth of knowledge and a concern over the developments of culture and society. Others called them *civilian strategists*, a term apparently neutral and still in use today. Sociologist Irving L. Horowitz (1963) dubbed them *civilian militarists*, a term living on the irony that originates from the intermingling of two presumably separate spheres of society. The originality award certainly goes to *thermonuclear Jesuits*. With reference to an article in the London *Times Literary Supplement*, Fred Kaplan used the term to describe how these men (and few women) moved as freely through the corridors of the Pentagon and the State Department as the Jesuits had moved through the courts of Madrid and Vienna three centuries earlier (cf. Kaplan 1983, 11). In her study on the history of rational choice liberalism, S. M. Amadae (2003) called them *defense rationalists* (see also Szalai 2014). And with a hint of sarcasm, historian Bruce Kuklick (2006) proposed the term *cerebral strategists* to emphasize the intellectual limits of the ways of thinking that in his view dominated the field throughout much of the Cold War era.

Immediately upon entering the scene, the new breed of experts had to confront the necessity of foreknowledge to come to reasoned decisions. Strategic decisions have their consequences; once a strategy is chosen, it might be virtually impossible or at least very costly to change it. There are path dependencies meaning that the initial decision predetermines which further decisions one may take and which not. In order to select the best path, knowledge of what one will encounter during the journey is crucial. The necessity of foreknowledge in strategic decision-making was, of course, not a new phenomenon *per se*. Yet again, the bomb had changed its urgency at least in two respects. First, it convinced even the utmost skeptic that the impact of science and technology on society had increased tremendously. To anticipate the future of society meant, to a large degree, to anticipate correctly the technological and scientific advances and their transformative force. Second, the presence of the bomb tied the fate of any nation to those of the others. A nation's future depended not only on its own decisions but also on those of all (or most) other nations. Isolationism was no longer an option.

Think tanks, modern organizations offering research and expertise to public and private agencies, became central players in this emergent scenery of expertise. Think tanks provided knowledge to whoever was willing to have it, and in most cases, this knowledge concerned social phenomena. They became far more important for the social sciences than for the natural sciences because for the latter, a void between knowledge and (economic and political) power never really opened up (Leslie 1993). Initially, think tanks were a unique US-American phenomenon, and it is safe to say that for the most part, they still are, since the majority of think tanks worldwide still is located on the US soil and their influence on policy processes undisputed (Grossman 2014; McGann 2015; Medvetz 2012; Ricci 1993; Rich 2004; J. A. Smith 1991). In a sense, the blossoming of this organizational form in the United States can be treated as a consequence of the early establishment of the expert within US politics. For European countries, political scientists have not been able to identify many comparable organizations but instead had to contend themselves with what they assumed to be "functional equivalents" (e.g., Gellner 1995; Braml 2004). Though some historians follow the traces of think tanks back to the seventeenth century (e.g., Soll 2009), it is reasonable to claim that the modern think tank appeared toward the end of World War II.

A New Ethos of Social Science?

Despite the ambivalence characterizing the expert as a participant in two social spheres, and ironically partly due to it, scientists successfully occupied this new social position in the first decades after the end of World War II. The experts in foreign policy and military strategy were a rather diverse bunch of people. They came from the social as well as from the natural sciences and the humanities. Quite a few of them had managed to flee from Europe immediately before the Nazi takeover or shortly afterward and had succeeded in finding a place for themselves in the US American scene. Others came from more or less well-established American families. Many of them had served in the wartime effort, some as soldiers, but most as members of US-based agencies like the Applied Mathematics

Panel (AMP), the Office of Strategic Services (OSS), or the Office of War Information (OWI), that produced a variety of studies commissioned by the military and the government. The wartime effort consolidated a new form of social science that had emerged earlier but only arrived at full blossom in the density of events and the resulting urgency of quick results brought about by the war. Proponents thought this new form engendered a complete rupture from earlier traditions of social science. Theoretically, the main thrust of the new social sciences consisted of a blend of positivism-inspired rigor with what one may be called practicality thinking. Organizationally, this thinking was well anchored in agencies like those named above during the war effort. After the war ended, the proponents of the new form of social science felt the danger that with society moving back into a peacetime mode of organization, this new form would lose its infrastructure. Supported by large philanthropic foundations and organizations like the Social Science Research Council (SSRC) or the National Science Foundation (NSF), they promoted the spread of this new form of social science (among other sources, see Hauptmann 2012, 2016; Heyck 2015; Solovey 2004, 2013). Part of their strategy was to establish a new label—"behavioral sciences"—that sought to demarcate the new form from the old (Pooley 2016). The emergent class of strategy experts appeared to be among the torchbearers of this new form of social science.

The closeness of the nexus between social sciences, philanthropic foundations, and political and military decision-makers appeared historically new to some critics. "For the first time in the history of their disciplines," C. Wright Mills wrote in 1959,

> "social scientists have come into professional relations with private and public powers well above the level of the welfare agency and the county agent. Their positions change—from the academic to the bureaucratic; their publics change—from movements of reformers to circles of decision-makers; and their problems change—from those of their own choice to those of their new clients" (Mills 2000, 95f).

The liberal practicality that had characterized US-American social science since the nineteenth century transformed into illiberal practicality: instead of the public, elites had become the target audience of the new social sciences. A consequence of this situation, one heavily criticized by Mills and other intellectuals, was a withdrawal of large parts of social scientific research from the public discourse. "Of central significance to the scientific identity that sociology was forging for itself was the corollary that the discipline's scientific endeavors could only flourish if they took place in isolation from public discourse and insulated from publics" (Haney 2008, 9).

Not discussed by Mills, but obvious to today's observers, was the gender dimension involved in this development. During the early decades of the twentieth century, male academic social scientists had spent considerable effort in differentiating themselves from allegedly "female" activities in social betterment and reform, an effort shaping the lives of women like Jane Addams (Deegan 1988; Schneiderhan 2015) as it shaped the face of the discipline. Yet, while successful in achieving and securing status, the men in academic social science still sought for a field to apply their knowledge. Foreign policy became a prime field for them—and the battle for the hearts and minds also became a struggle over the definition of (social) science (Cohen-Cole 2014; Wolfe 2018).

The rise to relevance of the experts in foreign policy and military strategy happened in the slipstream of yet another fundamental transformation in the social position of science, with which however it is not identical. This transformation concerned the ethos of science, and again, the atomic bomb, while not having initiated it, provided the final and crucial impulse. Some branches of the natural sciences had gained considerable recognition and public acclaim during the two World Wars. This was true especially for the physicists, epitomized—in the view of the public—by the team led by J. Robert Oppenheimer that had successfully developed the atomic bomb. The launching of this bomb and its devastating "success" in killing civilians led to fierce debates about the responsibility of the scientist for the results of his or her research. In his famous account of the ethos of science, written in the early 1940s, Robert K. Merton (1996) had not included responsibility as core norm. His argument was that the system of science functioned best if it adhered to the norms nowadays known by the abbreviation, CUDOS. The first norm of *communism* meant that the results of scientific research were not the property of the scientist, but of society. The second norm, *universalism*, referred to the expectation that the social system of science should provide structures to ensure that the validity of results was not related to the person of the researcher. Scientific standards should apply to all scientific work, regardless of its authors' ethnicity, religion, gender, or other subjective characteristics and preferences. Third, *disinterestedness* should steer the scientist. In the pursuit and development of ideas, the scientist shall not seek for their marketability, but only for the truth. Finally, fourth, science must institutionalize a system of *organized skepticism*: all necessary provisions must be taken to ensure that results can be assessed and evaluated by peers.

While not on Merton's list, *responsibility* became a most problematic issue for scientists in the war effort, both for those involved with the development of the new superweapon and for those evaluating operational plans or propaganda schemes. However, it did so in different ways. As mentioned above, the experts on foreign policy and military strategy saw it as their responsibility to enter the arenas—public and secret—of strategy and foreign policy deliberation, because without a proper scientific, "neutral," and matter-of-fact analysis, the culture of insecurity could not be managed or controlled and global nuclear devastation was imminent. They felt the responsibility to alert the public to the dangers of lacking scientific knowledge. However, other scientists, J. Robert Oppenheimer chief among them, felt responsible to warn the decision-makers and the public about the potential harm resulting from already available scientific knowledge. They realized that their scientific success had led them to a position where they could no longer keep detached from the practical uses of their results. Adopting the terminology by Hans Reichenbach, the atomic bomb forced many natural scientists to consider the context of using their findings more profoundly than they had deemed it necessary prior to Hiroshima and Nagasaki (on Oppenheimer, see Bird and Sherwin 2005).

Experts on foreign policy and military strategy thus occupied a different argumentative position than the natural scientists involved in weapons development. The trust in their capacities was fostered by successes that were not their own. Their rise depended on the fact that in the wake of breakthroughs in specific scientific fields, the public generalized from these specific instances of success to all branches of science. To borrow a term coined by Norbert Elias (Elias and Scotson 1994, xix), it was a *pars-pro-toto* distortion: the parts of science that were most visible to the public were taken to represent the whole of it—but as parts of "heroic science," they were "quite unrepresentative for most of the science that goes on" (H. Collins 2014, 9). Nonetheless, the experts on foreign policy and military strategy saw themselves confronted with the comfortable prejudices that they were competent and skilled to deliver solutions to the cultural and ethical dilemmas raised by the bomb. Hope and trust rested upon them. And although some experts retained their skepticism toward their own capability to fully live up to these expectations, they were nonetheless convinced that since there was no better alternative at hand, it was their historical duty to accept the task conferred upon them. The closeness of the nexus between scientists, philanthropic foundations, and political and military decision-makers promoting the new form of social science further nurtured this conviction.

A Brief History of RAND

The RAND Corporation, the place where the techniques of prognosis treated in this book originated, has been repeatedly called a paradigmatic case of a think tank, something like an organizational prototype. Initially installed as a collaborative project between high-ranking Air Force officers and representatives of Douglas Aircraft Company, it has maintained close links to the armed services and the Pentagon ever since. Even though it has steadily broadened its research spectrum, half of its annual revenue of $293.3 million in 2015 came from the Pentagon, the US Air Force (USAF), and the US Army. Today, it entertains four principal locations, three in the United States and one in Cambridge, United Kingdom. The organization's headquarters are still located close to the Santa Monica pier. In 2015, RAND employed 1875 persons.[2]

Since the history of the RAND Corporation has been in the focus of several books and articles (among them B. L. R. Smith 1966; Digby 1990; Hounshell 1997; M. J. Collins 2002; Abella 2008; Rocco 2008, 2011; Augier and March 2011, chap. 5), some brief remarks might be sufficient. The corporation was an offshoot of Project RAND (an acronym of Research ANd Development), a project initiated by US Air Force Officers together with Douglas engineers in March 1946. It was only one of a set of comparable institutions inaugurated at that time. Also in 1946, the US Navy established the Office of Naval Research (ONR; Allison 1985; Sapolsky 1990). Further, the US government founded the Research and Development Board (RDB) within the Department of Defense "to oversee the research activities of the military services" as well as the National Research Foundation (NRF) "with the purpose of invigorating academic research in the sciences and providing the military with a continuing flow of information and technique to improve weapons and operations" (M. J. Collins 2002, viii).

Project RAND began work on the premises of Douglas in Santa Monica, California, in early 1947 before moving into a building of its own. Since the end of World War I, Californian towns had already battled over the allocation of military organizations in their confines and had been quite successful (cf. Lotchin 1992). The Los Angeles area was home to several military bases as well as an impressive number of military-related enterprises when RAND opened its doors. By the end of the 1960s, half a million people reportedly worked in aerospace in Southern California, and this had been the result of efforts by local politicians to attract military bases already 40 years earlier (cf. Lotchin 1992, 4). Thus, RAND quickly became a part of a well-established network of military bases as well as firms working in the ship, aircraft, and weapon industries. Partly because of this already extant "metropolitan-military complex" and the chances and challenges it offered, Air Force representatives, Douglas Aircraft directors, and RAND researchers concluded that an excessively close relationship might lead to charges of both scientific intransparency and economic protectionism. Thus, the registration of RAND as a non-profit research organization in California was initiated. In parallel, mathematician John D. Williams, one of the first employees at RAND, was commissioned to recruit social scientists. With the help of Olaf Helmer, the head of the group that later invented Delphi, Williams organized a conference in New York that assembled a range of established social scientists (RAND 1948). Their task was to discuss directions for further research, but implicit was the objective to identify persons for some new to be founded divisions at RAND, the Economics Division and the Social Science Division. After some negotiations, Charles Hitch and Hans Speier agreed

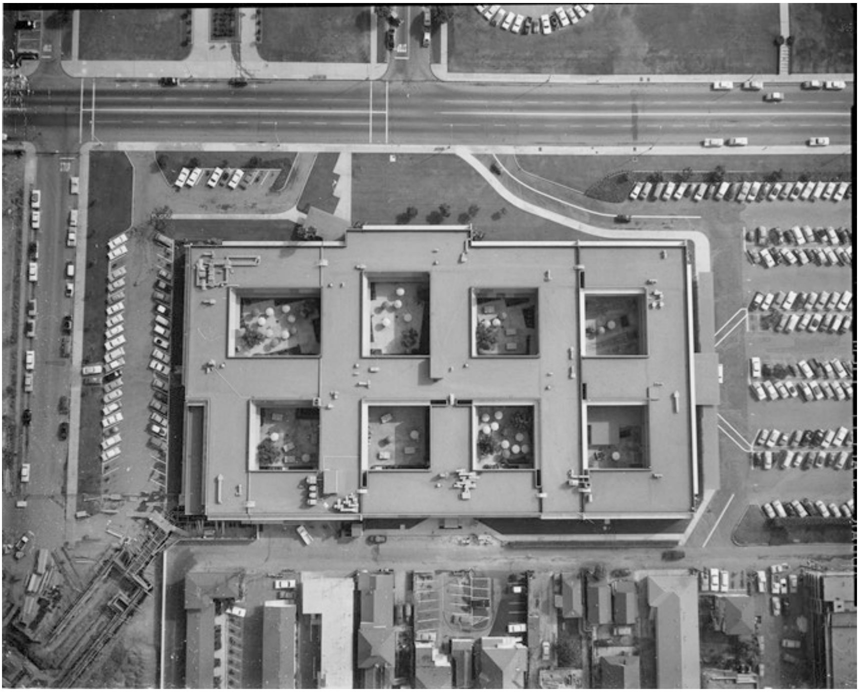

Fig. 2.1 Aerial view of the second RAND building, Santa Monica, CA. Photograph was taken in 1960 (RAND Corporation Archives, Santa Monica, CA)

to become the directors of the Economics Division and the Social Science Division, respectively. One condition for both of them was that the organizational split from Douglas Aircraft was completed, which finally happened in 1948 when the RAND Corporation was registered as a separate non-profit organization according to Californian law (Fig. 2.1).

Throughout the 1950s and for the most part of the 1960s, RAND engaged in research on strategic and technical analyses of (nuclear) warfare and its social, societal, and economic consequences (Kaplan 1983; Edwards 1996; Hounshell 1997, 2000; Mirowski 1999, 2002; Amadae 2003; Ghamari-Tabrizi 2005). During these years, RAND was involved in policy research on the Korean War, on the Berlin Crisis, and on the Cuban Missile Crisis. The Vietnam conflict escalated, and RAND researchers were busy with analyzing and developing strategies on which they briefed decision-makers in the US Air Force, the Pentagon, and other government agencies. Some RAND sponsors did not leave any doubt about their understanding of the role of military-funded science. In 1953, Don K. Price, then deputy chairman of the Department of Defense's Research and Development Board, expressed his bewilderment that scientists were "still struggling to reconcile their eighteenth-century devotion to science as a system of objective and dispassionate search for knowledge and as a means for furthering the welfare of mankind in general, with the twentieth-century necessity of using science as a means for strengthening the military power of the United States." The military unswervingly followed "its cardinal principle: it does not make research contracts for the purpose of supporting science, but only 'in order to get results that will strengthen the national defense'" (cited in Herman 1995, 132).

In retrospect, however, researchers working at RAND tend to emphasize the relative autonomy of RAND toward its sponsors and attribute this to the authority of director Franklin Collbohm. And indeed, in the first three decades of its existence, RAND was an important place for basic research in various disciplines. As regards the social sciences, RAND was among the small number of institutions to develop game theory after its initial conception by John von Neumann and Oskar Morgenstern (Leonard 1992, 2010; Erickson 2011). Game theory as practiced at RAND, that is, not only in a strict mathematical form but also in a more open form of thinking about decisions, led to a variety of influential publications. The vast majority of those economists who received a Nobel Memorial Prize in Economic Science for their contributions to game theory or decision theory—Kenneth Arrow (1972), Herbert Simon (1978), John F. Nash (1994), and Thomas Schelling (2005)—had spent considerable parts of their career at RAND.

Whereas this work was carried out predominantly in the Economics and Mathematics Divisions, also the Social Science Division comprised impressive number of staff: as of 1 January 1956, for example, the Social Science Department consisted of 39 persons.[3] Ten of them performed administrative tasks, whereas all the others in RAND's Santa Monica and Washington offices were engaged in research. Further, 23 persons provided "professional services"; they probably had short-term task-related employment contracts with RAND. Irving Janis, Otto Kirchheimer, Harold D. Lasswell, Paul F. Lazarsfeld, and Edward Shils belonged to this category. Six years later, on 1 January 1962, the Social Science Division had grown to a staff number of 46, 13 of them responsible for administrative and organizational tasks. This staff number was completed by a further 57 persons in the category professional services, where we find, in addition to some of those already mentioned above, scholars such as James S. Coleman, Morris Janowitz, Ithiel de Sola Pool, Philip Selznick, and Immanuel Wallerstein.

The researchers working for RAND during the first two decades of the Cold War were mainly engaged in foreign policy and strategy research. However, institutional mechanisms existed that clearly limited the Air Force's participation in determining research topics or questions. Thus, while the Air Force was the main financial source of RAND, the organization managed to convince USAF officials that sound research required certain freedom and independence. But still, this did not mean public accessibility. The money for research came from public sources, but for the most part, the results were not communicated in accessible outlets. Internal reports and briefings with Air Force and Pentagon officials were the main channels of communication. The cognitive authority of RAND experts was acknowledged among their peers, that is, among those persons with an academic background that worked for the government or the armed forces. It was less acknowledged by military officers since they felt the new civilian strategists threatened their position of power. And with some notable exceptions, academic social science also did not take notice of the work of RAND researchers, and only a few RAND social scientists participated in the publication and recognition game required for university careers.

Experts and the Media, Part 1: Strategic Surrender

Eventually, however, as RAND grew over the course of the years, the public became more and more interested in its doings. It was roughly around 1960 when the public began to question the political implications of the emergence of RAND and similar think tanks. In 1961, folk singer Malvina Reynolds recorded “The RAND Hymn” in which she described the work of RAND scientists with the following words:

> Oh, the Rand Corporation’s the boon of the world,
> They think all day long for a fee.
> They sit and play games about going up in flames;
> For counters they use you and me, honey bee,
> For counters they use you and me.[4]

The same year, Pete Seeger included a take of this song on his album *Gazette*, Vol. 2, and thus RAND had entered the public culture. A few years later, the movie by Stanley Kubrick, *Dr Strangelove or: How I Learned to Stop Worrying and Love the Bomb* (1964), modeled parts of its script to mimic RAND style research and thought. These and other products of popular culture constructed for RAND an image of cold-blooded ruthlessness, of rationalized unreasonableness.[5] Counter to its self-understanding as securing peace, RAND appeared to these creatives to be a stronghold of secret and elitist warmongers, and the threat that this posed to liberal democracy was assessed mostly by the stylistic device of ironic exaggeration.

Interestingly, for RAND, the chain of events leading to its public image as a stronghold of warmongers was triggered by an expert who acted like a scholar addressing its academic peers. In the following, I describe briefly how the relationship between RAND, its funders, and the public evolved over the first two decades of its existence. Analytically, the following interpretation is based on two premises. The first premise is that the public image of think tanks depends to a large degree on the acting of a handful of publicly visible experts. To take up an earlier reference to the *pars-pro-toto* distortion (Elias and Scotson 1994), there is a flip side to this mechanism of status distribution. While the public may perceive the best of those scientists visible in media to represent good science, it might also attribute the worst characteristics of visible experts to all think tank researchers. And indeed, it did. This, of course, was momentous. The second premise, thus, is that the public image of think tanks decisively influences the organizational behavior of think tanks and the substantial orientation of the work carried out under its auspices. This is why media work and PR became an important field of think tank activity (cf. Rich 2004; Medvetz 2012). By determining the character of the relations between think tanks and their principals, their public image shapes the agency of think tanks. At least twice in its history, RAND's agency was restricted to a degree that necessitated a comprehensive restructuration and transformation of the organization.

Throughout the first years of its existence, the media public took hardly any notice of RAND. Those newspaper articles that mentioned RAND reported research results and did not describe, neither normatively nor objectively, the place where these results were produced. Things began to change when in early 1958 RAND social scientist Paul Kecskemeti published a lengthy historical study on instances of *Strategic Surrender* (Kecskemeti 1958) with Stanford University Press. Kecskemeti compared the French surrender of 1940, the Italian surrender of 1943, and the surrenders of Germany and Japan in 1945. Based on these historical comparisons, he argued that the enforcement of the US American policy of unconditional surrender actually prolonged the phases of war and suffering and that conditional surrenders were more likely to be achieved sooner, and with less loss of human lives. First book reviews appeared in May 1958, most of them neutral or mildly sympathetic to the argumentation. Writing for the *Washington Post*, Forrest C. Pogue (1958) points out that Kecskemeti was "convinced that the atomic bomb was not necessary to force the Japanese surrender." *Hartford Courant's* Thomas E. J. Keena (1958) emphasized the moral implications of Kecskemeti's study: "[N]ations may have to limit drastically what they hope to gain [from the defeated opponent] if they are not to force all-out conflict. […] In today's world, total conflict could be self-defeating." So far, no tumult.

The outcry came a few months later. On August 5, the *St. Louis Post-Dispatch* published a review of the study written by retired Brigadier General Thomas R. Phillips. However, Phillips discussed *Strategic Surrender* within a context it did not address itself. While apart from the case studies, Kecskemeti used the neutral terms "winner" and "loser" and focused his analysis on the structural, theoretical aspects of conflict, Phillips saw Kecskemeti's considerations as speaking to the crisis caused by the launch of Sputnik 1 the year before and the widespread fear of a technological supremacy of the Soviet Union. Phillips, as one commentator noted (Sokolsky 1958a, b), aimed to use Kecskemeti's study to increase the pressure on politicians to invest more money in military-technological development. To the uncritical reader of the review article, however, it appeared that *Strategic Surrender* was, at least hypothetically, concerned with scenarios of future surrender by the United States. This, in turn, troubled senator Stuart Symington (D-Mo.), who filed the *Post-Dispatch* article in the Congressional Record with the explanation that "the article has prompted considerable correspondence from disturbed constituents who asked whether the article has any basis of truth" (cited in Sokolsky 1958a). The turmoil took off. Angered by the supposed fact that somebody, funded by public money, may have studied scenarios of possible US surrender, a group of Republican senators asked President Eisenhower about the truth of this. The president's attempt to calm things down was obviously not sufficient, as a few days later, Senator Richard B. Russell (D-Ga.) introduced an amendment to the debate on a pending appropriation bill, stipulating "to bar funds for any research or planning on 'when and how or in what circumstances the government of the United States should surrender this country and its people to any foreign power.' […] [T]he same amendment […] also would cut off the salary of any official who ordered such a study" (Donovan 1958a).

A heated and partly tumultuous debate over the amendment ensued on August 14 and 15, after the Senate's Democratic leader, Lyndon B. Johnson (D-Tx.), had suggested a recess until the next day. In the hours leading to the senate's decision, Eisenhower again tried to reassure senators of the pointlessness of their anger and authorized a spokesperson to issue a statement: "As far as the President is concerned, all this talk about surrender is nonsense. … There has been no public money spent to study how or when the United States will surrender. There will be no such money spent for such purpose. The whole matter is too ridiculous for any further comment" (Donovan 1958b). Regardless of this intervention, and despite the very fact that nobody in the senate was informed of any study, extant or planned, of how the United States might surrender in a nuclear war, the amendment passed with 88:2 votes (cf. B. L. R. Smith 1971, 56ff; Solovey 2013, 89–90; Rohde 2013, 29–30). In the culture of insecurity that characterized the Cold War, the fear of what the mere mentioning of such a study might signal to the Soviet enemy weighed heavier than any attempt to question the reasonability of such action.

Paul Kecskemeti was conspicuously absent from these debates. While a repeated contributor to various high-ranking newspapers before and after the incident, he did not use this channel to correct the misunderstandings surrounding his book. The reasons for this are unclear. RAND policies may have required him to remain silent, but I found nothing that would corroborate this assumption. His wife, Elisabeth Lang Kecskemeti, died on 23 November 1959, at a relatively young age, but this was more than a year after the debates. Since Paul did not resign from his duties at RAND, where he contributed substantially to the Social Science Division's efforts in organizing political games and began to write a book on the Hungarian Revolution of 1956 (Kecskemeti 1961), we can conclude that even if Elisabeth was severely ill, he still would have found the time to issue a statement. A bit harder to reject is the hypothesis that he was too preoccupied with other tasks; he was to participate in the first MIT-sponsored Endicott House game that took place in September 1958, and since he served as advisor to the organizer, Lincoln P. Bloomfield, he might have found no time to write a newspaper commentary.

Experts and the Media, Part 2: Thinking the Unthinkable

In the debate on Kecskemeti's *Strategic Surrender*, RAND figured as a more or less unknown actor. Obviously copying from each other, some newspapers introduced the organization incorrectly as a "non-profit scientific agency operated for the Air Force by a group of universities" (cf. Drury 1958; Lawrence 1958). Other articles were more precise and had obviously received information materials from RAND. *The Evening Independent*, published in St. Petersburg, Florida, offered the following description:

> Rand is an independent, non-profit organization which does research in the interest of national security and welfare. Its subjects range over virtually the whole field of warfare, mass behavior and war-making capabilities—of

> both foreign nations and the United States. Many of the corporation's studies are made at the request of the Air Force, for which Rand is a contractor. But some studies are made on the corporation's own initiative. (*The Evening Independent* 1958)

Most probably as a reaction toward the political uproar triggered by the *Strategic Surrender* study and the ensuing media inquiries, RAND increased its funds for public relations. One of the first initiatives of the respective office after the Kecskemeti incident was an invitation extended to selected reporters to visit RAND. Their efforts proved successful. On 11 May 1959, a large report on the RAND Corporation appeared in *Life Magazine*. Entitled "Valuable Batch of Brains—An Odd Little Company Called RAND Plays Big Role in U.S. Defense," Leonard McCombe's article described RAND as an intriguing mixture of unconventional, but highly intelligent humans who, with high concentration and without the distractions induced by academic teaching, were able to follow their scientific—and at the same time politically relevant—research interests. Although there were high security measures to keep unauthorized persons from entering the building or rooms with restricted access, the atmosphere was casual and uncomplicated once inside. RAND, in short, was presented as a place where highly talented scientists were given the freedom they needed to unfold their creativity.

However, the connotations were to change soon, and the image of the RAND Corporation quickly metamorphosed from being a Mount Olympus of intellectual creativity in the name of national security to being a Mordor of warmongers, a place where civilians create and assess plans for global nuclear disaster. This metamorphosis was triggered by another book-length study published by a RAND researcher, *On Thermonuclear War* by Herman Kahn (1961), the first edition of which was published in late 1960. In this book, Kahn put forth the argument that a nuclear war between the United States and the Soviet Union was not only possible at any moment, but indeed was even probable. In such a case, he continued, the United States should conduct the first strike, because this increased the likelihood of destroying such extensive parts of the Soviet nuclear arsenal that the counter-strike would not fully eradicate civilized life in America.

Virtually overnight, the book became a hot topic for popular media. The reactions it caused ranged from furious outrage and mild appreciation—also by pacifists like Bertrand Russell, who lauded the book's ruthlessness in thinking about nuclear war—to the conjecture that the whole book was a hoax invented by a mad brain. "Is there really a Herman Kahn?," asked a reviewer in *Scientific American*, just to answer "It is hard to believe. Doubts cross one's mind almost from the first page of this deplorable book: no one could write like this; no one could think like this" (cited in Ghamari-Tabrizi 2005, 19). The media turmoil that arouse around Herman Kahn had several consequences. The first consequence was that the RAND management did not object to Kahn's decision to leave RAND some months after the publication of *On Thermonuclear War*. He became the director of the Hudson Institute, a think tank he had previously co-founded. Kahn increasingly conceived of himself as public person and started systematically to raise media attention, thereby preforming a new social type of media expert. Understandably, RAND would have preferred to have, if any, noncontroversial media coverage of its endeavors. This, however, was at odds with the interests of Kahn to become a media figure. Ironically, while *On Thermonuclear War* had been cleared—albeit with reluctance—by both the Air Force and RAND, this happened just one month prior to the installation of a new review process at RAND, which stipulated "that manuscripts had to be cleared by every relevant air force office and other federal agency" before publication (Ghamari-Tabrizi 2005, 207).

Of more sustained impact, however, were the repercussions of this media event for the public image of RAND and of think tanks in general. Similar to the Kecskemeti incident, RAND found itself put in a semipolitical role it never sought to acquire (cf. B. L. R. Smith 1971, 56). While RAND's position and self-understanding were to contribute neutral and factual analyses, journalists and politicians were quick to represent RAND as the mouthpiece of whatever elites they wished to counter. Artists-activists like Malvina Reynolds and Pete Seeger and politicians from across the ideological spectrum seemed to agree that RAND, and think tanks in general, were potentially promulgating political interests. From the perspective of the organization, a new actor had entered the field: the media public. Managing this new actor was set to become a major source of concern and tensions for RAND in the following decades.

These tensions incidentally culminated in the Pentagon Papers affair. The Pentagon Papers were a comprehensive, 9000 pages long series of studies on the Vietnam War that testified to systematic misinformation of the American people by its government. They proved that the American government, and especially Minister of Defense Robert McNamara, had continued to assure the public that the war would be won although they were demonstrably informed about the devastating and irredeemable situation in Vietnam and the obvious failure of the US strategy. Understandably, the study received the highest security restrictions. However, it was available at RAND, and there it fell into the hands of Daniel Ellsberg. After he had worked for RAND as a young man, Ellsberg had moved to the Pentagon where he contributed to the strategic planning of the Vietnam War. Following his own wish, he served as a State Department civilian in Vietnam, where he arrived in mid-1965. While there, he saw the Pentagon calculations fail. Gravely disappointed, he returned to the United States and, though returning to RAND in the hope of changing war strategy from within, he finally ended up copying the Pentagon Papers and leaking them to the press (Ellsberg 2002). Several national newspapers published extracts of the Pentagon Papers in 1971.

This affair had immediate consequences for RAND. The US Air Force and other organizations of the military-government complex abolished the rules that granted most RAND analysts access to classified information. The process of opening RAND to new fields, like public health and social issues, which had been begun in the late 1960s (Light 2003), gained economic importance, because financial contributions from the armed forces decreased. More intensively than before, RAND now had to act as a participant struggling for income and orders under market conditions of scarcity. And as it is the rule for think tanks today, RAND had to learn to operate in a crossing of various fields that included academic science, political power, business interests, and public media. All these fields had their own interests, and all applied their own rules to understand the new organization. It is in this context that the two techniques of prospection treated in this book have been invented, used, and developed.

Notes

1. The *Oxford English Dictionary* gives the year 1825 as the earliest mentioning of the noun "expert" (cf. http://www.oed.com/view/Entry/66551, visited August 1, 2017) and 1868 for "expertise" (http://www.oed.com/view/Entry/66556, visited August 1, 2017).
2. Figures for 2015 are: $60.6 million (20.7%) came from the office of the US Secretary of Defense and other national security agencies, making it the second biggest source of income for RAND; $44.7 million (15.2%) came from the US Air Force, and $42.4 million (14.5%) from the US Army. Source: http://www.rand.org/about/clients_grantors.html, last visited September 7, 2016.
3. The following figures have been calculated from organization charts which were sent to me by Vivian Artebery (RAND Corporation). See also the table summarizing staffing 1946–1959 in Collins (2002, 141).
4. Lyrics are available at http://people.wku.edu/charles.smith/MALVINA/mr140.htm. Accessed September 13, 2016.
5. In the MARVEL universe, for instance, a "RAND Corporation" made its first appearance in "The Fury of Iron Fist," Marvel Premiere vol. 1 # 15 (May 1974).

References

Abella, Alex. 2008. *Soldiers of Reason*. Orlando; Austin; and New York: Harcourt.

Allison, David K. 1985. U.S. Navy Research and Development since World War II. In *Military Enterprise and Technological Change: Perspectives on the American Experience*, ed. Merritt Roe Smith, 289–328. Cambridge, MA and London: The MIT Press.

Amadae, S.M. 2003. *Rationalizing Capitalist Democracy: The Cold War Origins of Rational Choice Liberalism*. Chicago: The University of Chicago Press.

Augier, Mie, and James G. March. 2011. *The Roots, Rituals, and Rhetorics of Change: North American Business Schools after the Second World War*. Stanford, CA: Stanford University Press.

Bacon, Francis. 1989. *New Atlantis and The Great Instauration*, Croft Classics. Rev. ed. Edited by Jerry Weinberger Arlington. Heights, IL: Harlan Davidson.

Bird, Kai, and Martin J. Sherwin. 2005. *American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer*. New York: A.A. Knopf.

Braml, Josef. 2004. *Think Tanks versus "Denkfabriken"?* Baden-Baden: Nomos.
Cohen-Cole, Jamie. 2014. *The Open Mind: Cold War Politics and the Sciences of Human Nature*. Chicago and London: The University of Chicago Press.
Collins, Martin J. 2002. *Cold War Laboratory: RAND, the Air Force, and the American State, 1945–1950*. Washington, DC and London: Smithsonian University Press.
Collins, Harry. 2014. *Are We All Scientific Experts Now?* Cambridge, UK and Malden, MA: Polity Press.
Deegan, Mary Jo. 1988. *Jane Addams and the Men of the Chicago School, 1892–1918*. Transaction Publishers.
Digby, James. 1990. Strategic Thought at RAND, 1948–1963: The Ideas, Their Origins, Their Fates. N-3096RC. The RAND Corporation, Santa Monica, CA.
Donovan, Robert J. 1958a. Surrender Study Stirs Row. *New York Herald Tribune*, August 15.
———. 1958b. Ike, Senators Emphatic About No 'Surrender'. *The Boston Globe*, August 16.
Drury, Allen. 1958. Report of U. S. Surrender Study Arouses Angry Debate in Senate. *The New York Times*, August 15.
Edwards, Paul N. 1996. *The Closed World. Computers and the Politics of Discourse in Cold War America*. Cambridge, MA and London: The MIT Press.
Elias, Norbert, and John L. Scotson. 1994. *The Established and the Outsiders*. London; Thousand Oaks; and New Delhi: SAGE Publications.
Ellsberg, Daniel. 2002. *Secrets: A Memoir of Vietnam and the Pentagon Papers*. London: Penguin Books.
Erickson, Paul. 2011. Eine Neubewertung der Spieltheorie. In *Macht und Geist im Kalten Krieg*, ed. Bernd Greiner, Tim B. Müller, and Claudia Weber, 258–275. Hamburg: Hamburger Edition.
Gellner, Winand. 1995. *Ideenagenturen für Politik und Öffentlichkeit. Think Tanks in den USA und in Deutschland*. Opladen: Westdeutscher Verlag.
Ghamari-Tabrizi, Sharon. 2005. *The Worlds of Herman Kahn. The Intuitive Science of Thermonuclear War*. Cambridge, MA and London: Harvard University Press.
Grossman, Matt. 2014. *Artists of the Possible: Governing Networks and American Policy Change since 1945*. Oxford and New York: Oxford University Press.
Haney, David Paul. 2008. *The Americanization of Social Science: Intellectuals and Public Responsibility in the Postwar United States*. Philadelphia, PA: Temple University Press.

Hauptmann, Emily. 2012. The Ford Foundation and the Rise of Behavioralism in Political Science. *Journal of the History of the Behavioral Sciences* 48 (2): 154–173. https://doi.org/10.1002/jhbs.21515.

———. 2016. 'Propagandists for the Behavioral Sciences': The Overlooked Partnership Between the Carnegie Corporation and SSRC in the Mid-Twentieth Century. *Journal of the History of the Behavioral Sciences* 52 (2): 167–187. https://doi.org/10.1002/jhbs.21786.

Heilbron, Johan. 2015. *French Sociology*. Ithaca and London: Cornell University Press.

Herman, Ellen. 1995. *The Romance of American Psychology: Political Culture in the Age of Experts*. Berkeley; Los Angeles; and London: University of California Press.

Heyck, Hunter. 2015. *Age of System: Understanding the Development of Modern Social Science*. Baltimore, MD: Johns Hopkins University Press.

Horowitz, Irving L. 1963. *The War Game: Studies of the New Civilian Militarists*. New York: Ballantine.

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

———. 2000. The Medium Is the Message, or How Context Matters. In *Systems, Experts, and Computers: The Systems Approach in Management and Engineering, World War II and After*, ed. Agatha C. Hughes and Thomas P. Hughes, 255–310. Cambridge MA and London: The MIT Press.

Kahn, Herman. 1961. *On Thermonuclear War. Three Lectures and Several Suggestions. Second Edition with Index*. Princeton, NJ: Princeton University Press.

Kaplan, Fred. 1983. *The Wizards of Armageddon*. New York: Simon & Schuster.

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

———. 1961. *The Unexpected Revolution: Social Forces in the Hungarian Uprising*. Stanford, CA: Stanford University Press.

Keena, Thomas E. J. 1958. Terms of Surrender. *The Hartford Courant*, April 5.

Kuklick, Bruce. 2006. *Blind Oracles: Intellectuals and War from Kennan to Kissinger*. Princeton: Princeton University Press.

Lawrence, David. 1958. 'Surrender' Concepts Seen in Forms of Appeasement. *New York Herald Tribune*, August 15.

Leonard, Robert J. 1992. Creating a Context for Game Theory. In *Toward a History of Game Theory*, Annual Supplement to History of Political Economy

Volume 24, ed. E. Roy Weintraub, 29–75. Durham and London: Duke University Press.
———. 2010. *Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900–1960*. Cambridge and New York: Cambridge University Press.
Leslie, Stuart W. 1993. *The Cold War and American Science: The Military-Industrial-Academic Complex at MIT and Stanford*. New York: Columbia University.
Light, Jennifer S. 2003. *From Warfare to Welfare. Defense Intellectuals and Urban Problems in Cold War America*. Baltimore and London: The Johns Hopkins University Press.
Lotchin, Roger W. 1992. *Fortress California 1910–1961: From Warfare to Welfare*. New York and Oxford: Oxford University Press.
McGann, James. 2015. 2014 Global Go To Think Tank Index Report. Think Tanks and Civil Societies Program, University of Pennsylvania, Philadelphia, PA. http://repository.upenn.edu/think_tanks/8.
Medvetz, Thomas. 2012. *Think Tanks in America*. Chicago and London: The University of Chicago Press.
Merton, Robert K. 1996. The Ethos of Science. In *On Social Structure and Science*. Edited and with an Introduction by Piotr Sztompka, 267–276. Chicago: The University of Chicago Press.
Mills, C. Wright. 2000. *The Sociological Imagination*. Oxford and New York: Oxford University Press.
Mirowski, Philip. 1999. Cyborg Agonistes. *Social Studies of Science* 29 (5): 685–718.
———. 2002. *Machine Dreams: Economics Becomes a Cyborg Science*. Cambridge, UK and New York: Cambridge University Press.
Pogue, Forrest C. 1958. War's 'New Moment of Truth'. *The Washington Post*, April 5.
Pooley, Jefferson D. 2016. A 'Not Particularly Felicitous' Phrase: A History of the 'Behavioral Sciences' Label. *Serendipities—Journal for the Sociology and History of the Social Sciences* 1 (1): 38–81.
RAND. 1948. *Conference of Social Scientists. September 14 to 19, 1947—New York. R-106*. Santa Monica, CA: The RAND Corporation. https://www.rand.org/pubs/reports/R106.html.
Ricci, David M. 1993. *The Transformation of American Politics. The New Washington and the Rise of Think Tanks*. New Haven and London: Yale University Press.

Rich, Andrew. 2004. *Think Tanks, Public Policy, and the Politics of Expertise*. Cambridge; New York; and Melbourne: Cambridge University Press.

Rocco, Philip. 2008. Managing a Matrix: The Development of Interdisciplinary Social Science at the RAND Corporation, 1946–1957. *The International Journal of Interdisciplinary Social Science* 3 (3): 151–162.

———. 2011. Wissensproduktion in der RAND Corporation. In *Macht und Geist im Kalten Krieg*, ed. Bernd Greiner, Tim B. Müller, and Claudia Weber, 301–320. Hamburg: Hamburger Edition.

Rohde, Joy. 2013. *Armed with Expertise: The Militarization of American Social Research During the Cold War*. Ithaca and London: Cornell University Press.

Sapolsky, Harvey M. 1990. *Science and the Navy: The History of the Office of Naval Research*. Princeton, NJ: Princeton University Press.

Schneiderhan, Erik. 2015. *The Size of Others' Burdens: Barack Obama, Jane Addams, and the Politics of Helping Others*. Stanford, CA: Stanford University Press.

Smith, Bruce L.R. 1966. *The RAND Corporation. Case Study of a Nonprofit Advisory Corporation*. Cambridge, MA: Harvard University Press.

———. 1971. *The RAND Corporation. Wissenschaftliche Politik-Beratung in den USA*. Düsseldorf: Bertelsmann Universitätsverlag.

Smith, James Allen. 1991. *The Idea Brokers: Think Tanks and the Rise of the New Policy Elite*. New York: The Free Press.

Sokolsky, George E. 1958a. Surrender Psychology. *The Washington Post*, May 9.

———. 1958b. We Ought to Have a More Mature Way to Get Action. *Times Daily*, November 9.

Soll, Jacob. 2009. Think Tanks um 1640. Von der Akademie der Brüder Dupuy zu Colberts staatspolitischer Bibliothek. *Zeitschrift für Ideengeschichte* III (3): 44–60.

Solovey, Mark. 2004. Riding Natural Scientists' Coattails onto the Endless Frontier: The SSRC and the Quest for Scientific Legitimacy. *Journal of the History of the Behavioral Sciences* 40 (4): 393–422. https://doi.org/10.1002/jhbs.20045.

———. 2013. *Shaky Foundations: The Politics-Patronage-Social Science Nexus in Cold War America*. New Brunswick, NJ: Rutgers University Press.

Szalai, András. 2014. Wor(l)Ds of Dr. Strangelove: The Persuasiveness and Institutionalization of Defense Rationalist Ideas on Nuclear Strategy, 1948–1963. PhD dissertation, Central European University, Budapest. https://politicalscience.ceu.edu/sites/politicalscience.ceu.hu/files/attachment/basicpage/764/andrasszalaidissertation2015_0.pdf.

The Evening Independent. 1958. Senator Would Outlaw Surrender Study Fund, August 14.
Wallerstein, Immanuel, Calestous Juma, Evelyn Fox Keller, Jürgen Kocka, Lecourt Dominique, V.Y. Mudkimbe, Kinhide Miushakoji, Ilya Prigogine, Peter J. Taylor, and Michel-Rolph Trouillot. 1996. *Open the Social Sciences: Report of the Gulbenkian Commission on the Restructuring of the Social Sciences*, Mestizo Spaces–Espaces Métisses. Stanford, CA: Stanford University Press.
Weiler, Bernd. 2018. *The Order of Progress: The Rise and Fall of the Idea of Progress in 'Early' Anthropology*. Oxford: The Bardwell Press.
Wolfe, Audra J. 2018. *Freedom's Laboratory: The Cold War Struggle for the Soul of Science*. Baltimore: Johns Hopkins University Press.

3

The Wisdom of the Group: RAND's First Experiments with Expert Prediction, 1947–1951

Experts as Predictors

With the culture of insecurity on the rise after the bombings of Hiroshima and Nagasaki, policy analysts began to search for new ways of social prognosis. At RAND, researchers followed a specific approach to prognosis that was based on expert knowledge. The basic idea was that pooling the opinions of knowledgeable people, or experts, allowed for sketching possible futures. This, to be sure, was not new. Asking "experts" to provide long-term predictions has been a journalistic genre already in the late nineteenth century (cf. Pietruska 2018, 23), and experts had been sources of information and judgment in forecasting the economy or the weather (Friedman 2014; Harper 2012; Pietruska 2018). Yet, the techniques developed at RAND in the 1950s and 1960s that implemented this idea added another dimension, which hitherto was not used: controlled interaction. The forecasters of the nineteenth century had well listened to the opinions of other experts when devising their predictions. Yet, they did not systematically bring these opinions and their holders in interaction with each other. Notwithstanding the differences among them, the techniques developed at RAND were all informed by the notion that

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*, Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_3

systematic interaction among experts and/or expert opinions increased the stability of the prognosis. Exactly this differentiated the RAND approach from earlier ones. This chapter describes the earliest techniques of social prognosis developed at RAND. It starts by discussing RAND's first study on expert prognosis in the context of earlier approaches to group prediction and then turns to the first Delphi study.

In early January 1948, Abraham Kaplan, A. L. Skogstad, and Meyer Abraham Girshick carried out a series of experiments to test the predictive capabilities of a group of people. Not yet called a Delphi, and indeed showing procedural differences, the study by Kaplan et al. was a RAND-internal starting point for using groups to predict social and technological events. The study report first circulated as RAND Paper P-93 and was eventually published in *The Public Opinion Quarterly* (A. Kaplan et al. 1950).[1] In the introduction to their article on "The Prediction of Social and Technological Events," Kaplan, Skogstad, and Girshick explain their decision to seriously consider expert opinions in policy-making.

> Many 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. (A. Kaplan et al. 1950, 93)

However, the problem was how to improve the procedures in gathering expert judgments. And the solution proposed was to install more systematized procedures. The method deemed most appropriate for this was polling. In the United States, public opinion polls were a well-running business in the late 1940s. George Gallup, Archibald Crossley, and Elmo Roper had successfully established the legitimacy of polling methods (and especially of sampling techniques) in the late 1930s and in parallel created a market for their products (cf. Igo 2007, 103–149; Fleck 2011). The year 1936 had marked a watershed in the career of opinion polling. The "scientific pollsters" had challenged the results of a *Literary Digest* straw poll and correctly predicted the outcome of the presidential race between the Democrat Franklin D. Roosevelt and Republican Alfred Landon. While they relied on a sample size that was much smaller than that of the *Literary Digest*, it was constructed properly along the lines of statistical theory (cf. Igo 2007, 103–104).

However, Kaplan and his colleagues argued, the predictive capacities of polling techniques had not yet been systematically explored, let alone realized. Most contemporary polls understood their results "as an expression of a point of view rather than as a verifiable prediction." Moreover, polls would "rarely distinguish between verifiable predictions about matters of fact and unverifiable judgments of value" (A. Kaplan et al. 1950, 95). The approach proposed by Kaplan and his colleagues was to use the established techniques of polling but also to conceive of the results of these polls not as opinions, but as projections. Unlike election polls, where the analyst takes public opinions to construct a prediction, their approach was to invite the study participants to predict. The predictors, then, were not the poll takers, but the interviewees.

Only in passing do Kaplan, Skogstad, and Girshick mention that there have been earlier studies taking a similar approach. As a matter of fact, two papers had been published a decade earlier in two consecutive issues of the *Journal of Abnormal and Social Psychology*, "The Major Determinants of the Prediction of Social Events" by Douglas McGregor (1938) and "The Prediction of Social Events" by Hadley Cantril (1938).[2] Both papers offered very sophisticated analyses of the psychology of predicting and would have provided a sound basis for further development. However, despite the fact that at the time of writing, both McGregor and Cantril held professorships at two of the most prestigious American universities, MIT and Princeton, Kaplan and colleagues did not engage with their arguments. Whether this was caused by the wish to not compromise their own claims of novelty, by the lack of understanding on their behalf about the intricacies of the psychological arguments or by the trial-and-error approach to scientific study characteristics taken by some parts of RAND, has to remain open. However, there are indications that the last factor had at least some impact. RAND mathematician Olaf Helmer, who was of supportive help to Kaplan and his colleagues, remembered that the idea to pool expert opinions for predictions came from observing the forecasts on horse races.

> RAND to some extent was interested in future problems [with] which the Air Force might be confronted. So, we asked ourselves "how can I make some reasonable forecast about what's likely to happen?" And one of us—I cannot remember it was me or one of my colleagues, I don't want to claim it for myself, I really can't remember it as it was such a collaborative effort—came to the idea that we are to look at the forecast that are traditionally made about the outcome of horse races. So called horse race handicappers put out forecasts of who's likely to win. So, we persuaded the fellow in charge of procurement in RAND—who were used to quaint requests—to subscribe to these horseracing journals. So we collected this information, and what we found, which wasn't very surprising. Took any particular forecaster and followed his forecasts systematically, he lost a lot of money. Because he was more often wrong than right. So then we had the idea, of not using one, but combining a number of these forecasters and somehow seeing if we could improve these forecasts by using the recommendations of several forecasters. (Interview with Olaf Helmer by Kaya Tolon, 3 June 2009, p. 5)

Against this background, Kaplan et al.'s study appears to have been an attempt to explore this line of everyday thinking more systematically. They did not pursue a problem left open in earlier scientific literature but instead started with the problems of the decision-maker in the culture of insecurity. Later RAND authors interested in similar social epistemological approaches to prediction took the Kaplan et al. paper as their starting point and ignored the earlier literature.[3]

Sparing a direct discussion of the earlier work of the two eminent psychologists, Kaplan and colleagues proceed with explicating their research design. They gathered a group of 26 participants—in a side note, we should mention that McGregor (1938, 181) had 400 and Cantril (1938, 368) 500 study participants. A vast majority of them, 15, were mathematicians and statisticians, 4 were engineers, and another 4 were economists or business administrators. The other three participants were "one office manager, one secretary, and one writer" (A. Kaplan et al. 1950, 96). Except for two persons, group members had a college education; eight members held a doctorate. In their paper, there is no direct evidence, but it seems plausible to assume that some of the group members—if not all—were working at RAND.

The study authors decided to use questionnaires to gather the predictions. In line with the principles of experiment design, Kaplan, Skogstad, and Girshick decided to vary the setting in which the participants filled out the questionnaires. Thus, they separated the participants into groups. One half of the participants received the questionnaires with the instruction to answer them on their own, without any research, and within the following three hours (cf. A. Kaplan et al. 1950, 97)—the control group, in the parlance of experimental research. The other 13 participants—the treatment group—went through various settings. Each week, they were divided into three quartets. One quartet, dubbed the "independent group," filled in the questionnaire in the same setting as did the first half of the participants: alone, without research, and within a three-hour timeframe. The second quartet discussed the questions within the group of four and then went on to fill in the questionnaire individually; this was called the "cooperative group." The third quartet, the "joint group," was instructed to discuss each question and decide upon a single answer for the whole group. Each week, subjects participated in a different group with different co-members, so that in the end each had worked together once with each of the other 12 subjects.[4] The 13th person served as a form of backup in case some of the other participants were absent.

After a few months of preparation, the study participants received the first questionnaire in January 1948. New questionnaires were distributed weekly for 13 weeks in a row. Each of these questionnaires comprised about a dozen items.

> In each question the predictor was offered four exhaustive and exclusive alternative outcomes with a time limit for occurrence of the predicted event set at 20 weeks (or less) from the date of the questionnaire. The predictor was required to give for each alternative his judgment of "the likelihood of its occurrence," expressed as a value from 0 to 100, inclusive. (A. Kaplan et al. 1950, 96–97)

With regard to content, the questionnaire items concerned both technological and social or political events (see Table 3.1). Among others,

Table 3.1 Sample items of the study on exert prediction by Kaplan et al. (1950, 109–110)[5]

1. (Asked on January 12, 1948) The monthly average of the BLS Cost of Living Index for September was 164 (1935-39=100). The Cost of Living Index monthly average for April 1948 will be to the nearest integer: a. Less than 160. a. .... b. 161-165. b. .... c. 166-170. c. .... d. More than 170. d. .... 100% Basis of your judgment:	4. (Asked on February 17, 1948) On June 10, 1946, Italians elected by proportional representation a Constituent Assembly of 556 members, comprising 104 Communists, 115 Socialists, and 207 Christian Democrats. On April 18, 1948, there will be an election for 557 Deputies. Of these, the combined number of Communists and Socialists will be: a. Less than 150. a. .... b. 150-219. b. .... c. 220-290. c. .... d. Over 290. d. .... 100% Basis of your judgment:
2. (Asked on January 12, 1948) Charged particles have now been accelerated up to energies of less than 300 Mev. (million electron volts). Plans are now under ways to construct more powerful accelerators. By J-Day energies will be announced of: a. Up to 300 Mev. a. .... b. 301-500 Mev. b. .... c. 501-1000 Mev. c. .... d. Over 1000 Mev. d. .... 100% Basis of your judgment:	5. (Asked on March 16, 1948) The Radio Manufacturers Association estimated that the television output reached a new monthly peak in February with a jump of nearly 170 per cent in 6 months. There are now 19 stations operating in 22 cities, 82 have construction permits in 51 cities, and 93 applications are being investigated with 64 more pending. By August 3, 1948 the total number of television stations operating in the U.S. will be: a. 19 or fewer. a. .... b. 20-25. b. .... c. 26-30. c. .... d. More than 30. d. .... 100% Basis of your judgment:
3. (Asked on February 3, 1948) The Republican party at its convention June 20th will select as its presidential candidate: a. Dewey. a. .... b. Stassen. b. .... c. Taft. c. .... d. Other. d. .... 100% Basis of your judgment:	6. (Asked on March 29, 1948) 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. By August 17, 1948, there will occur: a. (1) only. a. .... b. (2) only. b. .... c. (1) and (2). c. .... d. Neither (1) nor (2). d. .... 100% Basis of your judgment:

participants were asked to estimate the development of a cost of living index, the number of TV stations operating in the United States, and results in both domestic and foreign elections.

The study by Kaplan, Skogstad, and Girshick addressed three problems: the problem of evaluation, the problem of improvement, and the problem of appraisal. The problem of evaluation, to begin with, is concerned with determining how successful people were in predicting events. Clearly, as regards their epistemic character, the statements collected by the questionnaires were to be taken as opinions. But was it justified to ascribe to opinions a predictive capacity? And, if yes, were there differences in predictive success over different topical areas? More precisely, was prediction easier with technological matters than with social issues?

Kaplan and his colleagues defined a prediction as successful when the alternative to which the predictor assigned the highest likelihood became true. They first discussed the success of individual predictors: study participants' success ranged from 28% to 71% with a mean of 53%. In addition, a measure for definiteness was developed based on the values individuals attributed to each of the four alternative answers. Definiteness was defined as the extent to which the attributed value differed from 25, the value that would result from an even distribution of all the available 100 points to the four alternative answers.

Quite against intuition, Kaplan, Skogstad, and Girshick found no significant correlation between success and definiteness. The rank correlation coefficient was not higher than $r = 0.2$. "Predictors who were often right were, on the whole, scarcely more definite in their predictions than those who were often in error" (A. Kaplan et al. 1950, 98). Definiteness thus cannot be taken as a simple measure of the quality of the expert predictor, but can also be caused by other factors: "A predictor might be indefinite, that is, not because he cannot appraise his prediction, but because he is correctly estimating the four alternatives to have approximately equal probabilities" (A. Kaplan et al. 1950, 98).

Kaplan et al. then turned to the question of whether the temporal proximity or distance was a factor influencing predictive success. The presumption was that the success of any prediction would be higher the closer the event was to the date of prediction delivery. This presumption was supported by the data. Predictions of events that occurred within up to six weeks from the date of the questionnaire showed 49% of success, whereas this success slightly decreased to 45% and 35% for the following two six-week periods. After that, somewhat counterintuitively, predictive success increased again, turning out to be 55% for the full 20 weeks horizon of the study. This increase is partly an artifact, because in many cases, the predictions were verified without the occurrence of a specific event: “Twenty weeks could be recognized by most of the predictors too short a time for the occurrence of events of the type asked about, and so a fair measure of success was achieved by selecting the ‘No Change’ alternative” (A. Kaplan et al. 1950, 100). They conclude that although predictive success decreased with time during the 20-weeks period, it was not justified to assume that this trend continued beyond that period. Instead, it was considered possible that on a larger time scale, predictive success increased again with time and distance.

In addition, Kaplan and his colleagues also turned to the question of whether predictive success depended on the subject matters. They were interested in predicting both social and technological events. Interestingly, and in spite of the fact that the majority of the study participants had received natural science training, results indicated that developments in political and economic affairs could be predicted with slightly more success than developments in science and technology (53% resp. 51%). Several success rates were calculated for different kinds of questions, and though not all differences passed the significance threshold, all pointed in the same direction. After the expectable caveat that these were only results of an exploratory study and therefore limited in generalizability, the authors offered an interesting interpretation of these differences in predictive success. They argue that

> [i]t must also be recognized that the differences revealed in this study may be due in part, and perhaps altogether, to the fact that the predictors were not specialists on precisely the matters to be predicted. While it may be true that, for the non-specialist, social science subject matters are somewhat more predictable than those of the natural sciences, this need not hold for the specialist. (A. Kaplan et al. 1950, 102)

This echoed a finding that had emerged from the study of McGregor (1938). McGregor, who was interested in predictions as instances of psychological inferences, had argued that social phenomena were easier to predict because our interpretations of daily life constantly involve predicting possible future developments. “Perusal of the front page of any newspaper, or observation of a discussion of current events at the dinner-table will reveal that we do not ordinarily interpret social events as momentary or isolated. We see them, instead, as meaningful links in a chain of happenings extending from the past into the future" (McGregor 1938, 180). Polling approaches to prediction might thus be more successful with regard to social issues.

The second problem that Kaplan, Skogstad, and Girshick considered, the problem of improvement, concerned the improvement of the reliability of experts' predictions by statistical and mathematical means. Questions concerned the methodological procedure itself: What options in designing the procedure did the researchers have if they wished to increase the predictive success? The study approached this question via two measurable subtasks. The first was to find out whether the predictive success of the group was improved if the answers of the participants were assigned weight in accordance with the degree of expertise or, if available, to their predictive success in earlier studies. Since they had no performance record of their study participants, the authors decided to administer a knowledge test. In order to assess how "informed" a predictor was, the study participant had to fill in parts of the "Cooperative General Culture Test" (Revised Series Form X) issued by the American Council on Education in 1947 (Blair et al. 1947). The two parts selected were "Current Social Problems," and "Science."[6] The scores obtained in these two tests by the participants were then related to their predictive success in the two knowledge fields of interest. For both fields, a significant positive correlation (0.6) resulted which meant that the greater the knowledge in a given field, the higher the likelihood of successful prediction. However, the effect was below the expectations of the authors. "The success of the best informed predictors was not vastly greater than that of the worst informed" (A. Kaplan et al. 1950, 102).

In an attempt to interpret this minor correlation, Kaplan and his colleagues list several potential causes. First, one possibility would be that "the tests used are an inadequate measure of the knowledge actually brought to bear in the predictions themselves" (A. Kaplan et al. 1950, 103). They might measure general knowledge, but general knowledge might not be the kind of knowledge required for prediction. It might also be that, apart from knowledge and also independent of it, a factor of "judgment" might be operative. "As to whether the first, second, or indeed some other explanation is preferable, our meager data on the psychological side of the matter provide no indication" (A. Kaplan et al. 1950, 103).

Again, Kaplan, Skogstad, and Girshick might have been able to come up with a better interpretation if they had paid attention to the work by the two psychologists a decade earlier. McGregor, for instance, had evidence that corroborated the hypothesis that the decisive factor was not the expertise of the individual, but the clarity—or ambivalence—of the problem. When the current state of events—the "stimulus situation," in McGregor's terms—was complex, diffuse, or ambivalent, it did not provide a stable and consistent basis for prediction. This effect, McGregor's data showed, had more or less the same strength across all levels of expertise (1938, 195). "It is the *nature* of one's information that is determinative, not the *amount*. And the nature of the information will depend upon both the ambiguity of the stimulus situation and the subjective factor of importance. If ambiguity is zero, more facts can only strengthen an already firm conviction. If the ambiguity of the stimulus situation is maximal, more information may but add to the confusion" (McGregor 1938, 194; emphasis in original).

Although one can debate their interpretation, Kaplan et al. maintained that the effect of prior knowledge on the success of the prediction is negligible. Therefore, knowledge assessments were no sensible means to improve predictive success, and Kaplan and his colleagues turned their attention to potential alternatives. The first alternative they discussed was to seek not for the predictive wisdom of an individual, but of a group. Could it be, they asked, that predictive success increased if a group, not an individual, produced the prediction? To answer this question, the authors referred to the subgroups described above—the independent group, the cooperative group, and the joint group. Hence, they compared the success rates of the three groups and found that while the independent group achieved a success rate of 52%, the cooperative group scored 10% higher, and the joint group even 15% (62% resp. 67%; see Table 3.2). They concluded that "[t]he group effort is thus significantly better than that of the individuals composing the group working independently" (A. Kaplan et al. 1950, 103).

Seeking explanations for this improved success, the authors asked whether it could be attributed to the social-psychological effects of the

Table 3.2 Predictive success per group (Adopted from Kaplan et al. 1950, 104)

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

group or whether it was the result of a process of averaging that could also be obtained by statistical means. Answering this question, they stated that their "data strongly indicates that the latter is the case" (A. Kaplan et al. 1950, 103). The collaborative effort of the group did not add anything—there was no specific emerging group property that allowed for higher predictive success. Rather, some kind of averaging took place in the group that countervailed or discouraged extreme positions. If there was wisdom in the group, it lay in the leveling-out of extreme opinions.

But did the data support this interpretation? In order to assess that, the authors argued that the averaging effect they assumed to be operating in groups is similar to mathematical averaging. Thus, if instead of taking the individual estimates, one took the mean value of all estimations to calculate the success rate, it should increase to a level close to the rates of the cooperative and joint groups. And indeed, by so doing, the success rate of the whole group increased from 52% to 66%, which meant that its predictive success was comparable to those of the joint group (67%) The high success rate of the joint group, in other words, could also be achieved without any kind of social interaction between the group members. Avoiding further reflection on this finding, the authors concluded that "in this study the success of collective psychological effort was duplicated by statistical methods" (A. Kaplan et al. 1950, 104).

Finally, Kaplan, Skogstad, and Girshick addressed what they call the problem of appraisal. The core question here was whether predictions made with high confidence were more likely to be successful than predictions made with lower confidence. The authors argued that the percentages provided by the participants were a reasonable indicator of their confidence. Assuming this, they showed that the degree to which their participants were convinced about the veracity of their predictions correlated with the success of their predictions: the higher the percentage attributed to an option, they found, the higher the success rate. Predictions that were given a likelihood of 90–100% came true in 73% of the cases; predictions thought to come true with a chance of 51–89% realized in 61% of the cases. Neglecting the differences between the self-assessed conviction and the success rate, which showed that human predictors tend to be overly certain about their predictions, Kaplan et al. maintained that the positive correlation between the degree of conviction of the predictor and predictive success justified using the self-assessed conviction as a means of appraising estimates.

Finally, they turned their attention to the field in the questionnaire where the study participants were asked to explain the basis of their estimation. In the first step, the answers were classified into four categories. Statements were ordered into the category "justification" if they showed "some degree of logical warrant for the prediction" (A. Kaplan et al. 1950, 107). Factual elaborations of details or appeals to evidence were included, as was the formulation of hypotheses and so on. The category "rationalization" comprises statements that on the level of rhetoric pretend to be logical arguments but in fact were not. "Rationalizations consist in … references to completely unspecified 'evidence of past experience'; in appeals to what is 'reasonable,' 'obvious,' etc.; and in mere statements of belief" (A. Kaplan et al. 1950, 107). The third category, "guess," refers to statements were participants frankly disclosed their ignorance of the topic. And, finally, the category "special" comprised statements that commented on the question rather than provided information on the basis of the prediction.

Related to predictive performance, the results are the following (see Table 3.3):

Table 3.3 Success by basis statements (Adopted from Kaplan et al. 1950, 107)

Basis category	Number of predictions	Frequency (%)	Success (%)
"Guess"	290	11	40
"Rationalization"	181	7	48
"Special"	67	2	55
"Justification"	499	19	62
No comment	1161	61	51
Total	2653	100	52

Again, these final results from the Kaplan et al. study implicitly corroborate the finding of McGregor, who claimed the nature of knowledge, and not its amount, to be of crucial importance for successful prediction. In situations where people have a clear and unambiguous knowledge and thus can come up with a sound justification, they are more likely to predict correctly than in other situations. However, even when they guess, they are more likely to choose the correct answer than a random procedure running over the four alternatives (i.e., even guesses have a success rate considerably higher than 25%).

If one were to assess the novelty of the approach taken by Kaplan and his colleagues, it should be clear that the two major methodological ideas—first, the idea that a structured interrogation of experts can be used to produce predictions in fields where there is nobody of adequately formalized theory to do so, and second, the idea that questionnaires might be a valuable research tool for this purpose—had already been put forth a decade earlier. While there is not much reason to doubt that those who developed the Delphi design perceived the Kaplan et al. study as a precursor to their own approach, it is clear in historical hindsight that it was less innovative than it appeared to the actors. Nonetheless, the study was influential at RAND. It paved the way for how RAND researchers conceived of the predictive capabilities of experts and possible ways to use them. Apart from the organizational proximity, the continuation of this line of research at RAND was also fostered by the involvement of Olaf Helmer in the Kaplan et al. study, as a footnote acknowledged. Shortly after the completion of the report, and based on its results, Helmer, together with Norman C. Dalkey, began with the development of an improved procedure, a project which the two decided to call Delphi.

The First Delphi Study

The very name Delphi was another sign of the closeness of the two research groups. It had been, as Dalkey remembered, suggested by Abraham Kaplan. In retrospect, however, Dalkey thought that this baptism was "unfortunate—it connotes something oracular, something smacking a little of the occult—whereas as a matter of fact, precisely the opposite is involved. It primarily is concerned with making the best you can of a less than perfect fund of information" (Dalkey 1968, 8). At its core, Delphi established estimations, mostly in the form of numeric estimates. In contrast to the study by Kaplan et al., however, it introduced a multilevel assessment of the opinions of experts. Participants were interrogated repeatedly, and their answers were fed back into the next phase of data collection.

One particular result of the Kaplan et al. study had caught the attention of Helmer. While the study had shown that some kind of cooperation or interaction resulted in a higher predictive success, it had also shown that face-to-face interaction was not required to achieve this positive effect. Instead, the beneficial averaging of the experts' opinions could also be achieved by way of mathematics. There was thus no reason to physically assemble a group of experts. It would be sufficient, Delphi inventors thought, to make them interact only indirectly by feeding back aggregated questionnaire results. Moreover, the possibility to avoid interpersonal face-to-face situations was cherished as a methodological advantage, because one could rule out the social-psychological processes otherwise unavoidable in human groups, for example the desire to establish and maintain a positive self-image within a group, or the fear of social exposure.

When studying the forecasts of horse race handicappers, the group around Helmer realized that using a pool of several forecasters produced forecasts that realized more often than those of the single handicapper. Still, they were not highly successful.

> What we found was, if we did that [pooling of single forecasts of horse race handicappers], the result was that we would have lost much less money, but we still would've lost money. At least you didn't lose as much money, so the purses would improve by combining the forecasts. So that was the intellectual basis for Delphi. […] We asked ourselves, if we have a group of people who are asked to make a forecasted [!] about a particular development, whatever it might be. How can you make the best use of this group of forecasters in order to get the most reliable forecast? (Interview with Olaf Helmer by Kaya Tolon, 3 June 2009, p. 5)

The answer, inspired partly by coeval lines of reasoning from cybernetics and general systems theory, was to introduce a feedback loop.

Delphi thus became a technique designed 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 expectation was that this iterative procedure would lead to more stable results and a higher level of predictive success.

Olaf Helmer (1910–2011) had been born in Berlin. He completed his studies of mathematics and logic at the University of Berlin with a dissertation that he had begun under the direction of Hans Reichenbach. When Reichenbach decided to leave Germany due to the power seizure by the Nazis in 1933, Wolfgang Köhler became Helmer's *Doktorvater*. Helmer finished his studies in the following year; he submitted the dissertation as "Olaf Helmer-Hirschberg," a name he hardly ever used afterward. Rescher assumes that by so doing, Helmer was following a request to make explicit his Jewish ancestry. While I found no evidence of any official regulation in this regard, Hirschberg in fact had at some point been the name of Olaf's family. His father, born Fritz Hirschberg, was an actor, who struggled to make ends meet in the years after World War I (see the interview with Olaf Helmer included in Tolon 2011, 109–210). Partly to increase the likelihood of further engagements, Fritz's agent one day suggested he should consider adopting a stage name. This was in 1914–1915, at a time when Hirschberg was playing Torvald Helmer in Henrik Ibsen's *A Doll's House*, and in allusion to this part, he decided to pick Helmer.[7] Most probably, however, he did so without initiating an official change of name, neither for himself nor for the rest of the family; still, the doorbell might have read "Helmer-Hirschberg," to accommodate both private and professional needs. The student record at the University of London, to where Olaf emigrated shortly after receiving his doctorate, shows the double name "Helmer-Hirschberg" (The University of London n.d., 662), as does the entry in the *Sonderfahndungsliste G. B.*, the Nazi Black Book for emigrants in Great Britain (Forces War Records n.d.). The reasons leading Olaf to prefer Helmer alone upon his arrival in the United States remain obscure.

Be it as it may, when Helmer left Germany for Britain, he enrolled at the University of London and completed a second doctorate in philosophy in 1936. In 1937, Helmer was appointed a research assistant to Rudolf Carnap at the University of Chicago. There, he met Carl Gustav Hempel again, his close friend from Berlin student times, who also was a research associate to Carnap with a Rockefeller fellowship. Helmer went on to teach mathematics first in Urbana, Ill., and then in New York City. When the United States entered World War II, he took a position as mathematician for the National Defense Research Council, where he met John Williams (cf. Rescher 2006, 288). When Williams became one of the first scientists to sign a work contract with RAND in 1946, Helmer almost immediately followed him. Helmer's RAND ID card bore the number 5, and he was obviously proud of that fact (cf. Rescher 2005, 184).

Norman Crolee Dalkey (1915–2003) was born in Santa Clara, California. He took graduate courses with Carnap in philosophy at the University of Chicago from 1939 to 1940, where he met both Helmer and Hempel. He continued his studies at UCLA, where he received a PhD two years later with a dissertation on "The Plurality of Language Structures," written under the supervision of Hans Reichenbach. Probably upon the initiative of Helmer, he joined RAND's Mathematics Division in 1948. He remained at RAND for the rest of his career.

Clearly, thus, the inventors of Delphi had received their academic education not in the social sciences, but in philosophy, mathematics, and logic. They shared this background with both Abraham Kaplan and Meyer A. Girshick; of these early RAND prognosticators, only Skogstad, an economist, had a background in the social sciences. Furthermore, it is also clear that the Delphi inventors shared the same academic pedigree. They all were trained by philosophers who were committed to the program of logical empiricism or neopositivism blossoming in Europe before the Nazis came into power, with Hans Reichenbach, Rudolf Carnap, Carl Hempel, and Paul Oppenheim being among the most important contributors to this program. Today's observers sometimes assemble them as members of the so-called Berlin School of Logical Empiricism (Rescher 2006). It is important to keep in mind their common epistemological socialization since it helps to understand some of the peculiarities of their approach as well as some differences to other approaches to expert prognosis followed at RAND.

Helmer and Dalkey conducted the first Delphi study in the first half of 1951. Their initial report, published 14 November 1951 under the title "The Use of Experts for the Estimation of Bombing Requirements," remained classified for about ten years. In summer 1962, an abridged version was declassified, now entitled "An Experimental Application of the Delphi Method to the Use of Experts. RM-727/1-ABRIDGED" (Dalkey and Helmer 1962). This text was eventually published under the same title in *Management Science* one year later (Dalkey and Helmer 1963). Since the original report is still secret, I have relied on the abridged version of the RAND report. Apparently, the abridgments concerned mainly the results, not the design of the methodological procedure.

The study's objective was twofold. First, the authors wanted to determine a reliable estimate of a factor relevant to military decision-making. 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 important US industrial targets—"an optimal U. S. industrial target system" (Dalkey and Helmer 1962, 1). Based on that, participants had to estimate the number of atomic bombs required to ruin those branches of the US economic system that were required for munition production. The questionnaire distributed to the participating experts started by sketching a scenario:

> Let us assume that a war between the U.S. and the S.U. breaks out on 1 July 1953. Assume also that the rate of our total military production (defined as munitions output plus investments) at that time is 100 billion dollars and that, on the assumption of no damage to our industry, under mobilization it would rise to 150 billion dollars by 1 July 1954 and to 200 billion dollars by 1 July 1955, resulting in a cumulative production over that two-year period of 300 billion dollars. Now assume further that the enemy during the first month of the war (and only during that period) carries out a strategic A-bombing campaign against U. S. industrial targets, employing 20-KT bombs. Within each industry selected by the enemy for bombardment, assume that the bombs delivered <u>on target</u> succeed in hitting always the most important targets in that industry. What is the least number of bombs that will have to be delivered on target for which you would estimate the chances to be even that the cumulative munitions output (exclusive of investment) during the two year period under consideration would be held to no more than one quarter of what it otherwise would have been? (Dalkey and Helmer 1962, 6; emphasis in original)

The question at the end of this paragraph was called the primary question of the study. When we paraphrase it and leave aside some of the technical details, the study thus was concerned with finding out how many atomic bombs were needed to damage the relevant US industries (steel, petroleum, aluminum, etc.) to such a degree that as a consequence, only a fourth of the expected munition could be produced. To keep things in perspective: A 20 kt bomb would have a higher blast yield than *Little Boy* that detonated over Hiroshima with approximately 15 kt, and only a little less than Nagasaki's *Fat Man* (ca. 21 kt). In the culture of insecurity, apparently, the exploration of disastrous futures was part of the responsibility of the social scientist.

The second objective of the study was methodological. Dalkey and Helmer wanted to test whether the proposed innovation—to ask experts to give their opinion repeatedly while feeding back the results of previous rounds as well as additional materials—resulted in a convergence over time of the individual estimates, so that in the end, the range of estimates was smaller than in the beginning. In line with Kaplan et al.'s finding that the leveling of opinions could be reproduced by mathematical means, measures were taken to avoid direct contact between the experts. The experts were interrogated individually via questionnaires that were designed to (1) assess the answers to so-called primary question; (2) allow for sketching the expert's reasoning that led to her answer to the primary question; (3) list the factors considered relevant for the primary question, thus informing the answer; (4) provide estimates of these factors; and finally, (5) ask for "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 questionnaires were sent to the panel of participating experts, which comprised seven persons: 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. In the words of Dalkey and Helmer (1962, 2),

> [t]his 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 incli-

> nation 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.

Estimates were collected in a first round of questionnaire interrogation and used to calculate established measures of central tendency. 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 experts, and the aggregated estimates—they wanted to revise their first answers or whether they needed any additional information. The rationale for so doing was that it allowed for exploring the factors informing the individual estimates. This made it "possible to correct any misconceptions that he [the individual expert] may have harbored regarding empirical facts or theoretical assumptions underlying those factors, and to draw his attention to other factors which he may have overlooked in his first analysis of the situation" (Dalkey and Helmer 1962, 3).

Altogether, five questionnaires were distributed to the participants in roughly weekly intervals. In addition to the questionnaires, Dalkey and Helmer decided to interview the experts in order to explore the reasoning behind their estimates. This first Delphi thus combined both quantitative and qualitative research techniques. The interviews were carried out after the experts had completed the first and third questionnaires. Further, in addition to the 50% likelihood of successful destruction mentioned in the primary question, they also asked the participants to give estimates for 10% and 90% likelihoods.

This procedure resulted in the first distribution of estimates (see Table 3.4). Based on these results and the rationales given both in the questionnaires and in the follow-up interviews, Dalkey and Helmer drafted and distributed a second questionnaire. This questionnaire was concerned with establishing within the pool of experts a common understanding of the issues under scrutiny. At the outset, it identified four

Table 3.4 Bomb estimates in the first round (Adopted from Dalkey and Helmer 1962, 7)

	Respondent						
Response	1	2	3	4	5	6	7
Primary response (50% confidence)	125	50	150	300	200	1000	5000
10% confidence	75	25	100	250	70	–	2500
90% confidence	200	150	175	800	500	–	10,000

major factors that appeared to be relevant for answering the primary question: (A) the vulnerability of industrial plants and infrastructures, (B) the recuperability of industries, (C) the initial stockpiles available to the industries, and (D) complementarities among the different industries. The questionnaire of the second phase did not ask the experts to estimate, but instead invited them to expose their lines of reasoning on the subject. It was distributed together with a list of those ideas and statements related to the aspects of vulnerability (A) and recuperability (B) that had been mentioned in the follow-up interviews.

The second questionnaire had a completely different design than the first. It comprised six questions, which all invited the participants to lay down in full sentences their reasoning and the basis of their analytical approach (Dalkey and Helmer 1962, 8):

Question 1. Does the preceding breakdown of the problem [into A, B, C, and D] agree with your intuitive approach to a solution? If not, explain in detail; in particular, are there major items in addition to A, B, C, D which should be taken into consideration?

Question 2. [With regard to the items concerning A on the distributed list:] What additional factors, if any, do you consider relevant to the problem of vulnerability? Which of the factors listed do you consider irrelevant?

Question 3. [With regard to the items concerning B on the distributed list:] What additional factors, if any, do you consider relevant to the problem of recuperability? Which of the factors listed do you consider irrelevant?

Question 4. [With regard to C:] What factors should be taken into account for our problem in assessing the size and role of initial stockpiles?

Question 5. [With regard to D:] What factors should be taken into account for our problem as regards determining complementarities among industries?
Question 6. Are there any general comments which you wish to make?

The data collected by this second questionnaire was rich, and Dalkey and Helmer report that only selected parts of it could be considered in the design of the following phases of the study. However, the breadth and depth of the information provided by the participants allowed for a more sophisticated analysis of the study's primary question. It helped the analysts to appreciate the intricacies of the phenomenon under scrutiny in more detail—a cognitive progress that they set out to share with their participants. The openness and flexibility of their research design allowed for exactly that.

The third questionnaire focused again on the primary question. It provided data on both the US economy—informing for instance about the share of the national output of specific metals that went into munitions production—and on the vulnerability of industrial (infra-) structures. For the latter, they used "[e]xamples of damage with 20-KT bomb obtained from Japanese bombings" (Dalkey and Helmer 1962, 9).[8] With this shared knowledge basis established, the researchers went on to ask five questions, the most important of which were the first and the fifth (Dalkey and Helmer 1962, 9f):

Question 1. What is your revised answer to the primary question of Questionnaire 1?
[…]
Question 5. For the following industries, how would you allot the minimum number of bombs on target called for in the primary question?

Steel	Heavy steel fabrication
Petroleum refining	Machine tools
Aluminum	Electron tubes
Copper	Aviation fuel
Power	Anti-friction bearings
A-bombs	Other industries
Aircraft engines	

Table 3.5 Bomb estimates in the third round (Adopted from Dalkey and Helmer 1962, 7, 10; own calculations)[9]

	Respondent						
Response	1	2	3	4	5	6	7
Questionnaire 3	158	89	200	250	256	800	450
Interview	158	106	184	250	256	525	450
Difference	0	+17	−16	0	0	−275	0
Difference to Questionnaire 1 (50% confidence)	+33	+56	+34	−50	+56	−475	−4550

After this third questionnaire, Dalkey and Helmer again carried out a series of follow-up interviews to clarify some of the issues involved, which for some experts resulted in minor revisions of their estimate. The results thus produced in round three are given in Table 3.5.

At this point it was already conceivable that the procedure would result in a convergence of estimates. Those experts who had initially given rather low estimates tended to increase their figures, while those with initially high estimates decreased them.

The fourth questionnaire, in addition to providing even more data and information on specific issues, employed two innovative forms of questioning. First, it contained sheets with empty coordinate systems and asked the participants to draw graphs which indicated "the estimated progress of steel and of munitions output recuperation after bombing" (Dalkey and Helmer 1962, 11). Second, and more important to our context, it made use of the estimates obtained in the above-stated Question 5 of the preceding questionnaire. Question 5 had resulted in 7 different allocations of bombs to 13 industries. These "bombing schedules" were then used in Questionnaire 4 to construct a highly sophisticated item. The schedules "were roughly ordered cyclically in such manner that each was as similar as possible to its two neighboring schedules" (Dalkey and Helmer 1962, 10). Together with Questionnaire 4, each participant received those two schedules that were most similar to his own schedule. The participants were first asked to revise their figures again, if they wished to do so, and then provide reasons for assessing their bombing schedule as superior to the other two schedules. The revised bombing

Table 3.6 Bomb estimates (totals) in the fourth round (Adopted from Dalkey and Helmer 1962, 10, 12; own calculations)[10]

	Respondent						
Response	1	2	3	4	5	6	7
Questionnaire 4	166	153	200	250	300	332	500
Difference to Questionnaire 3	+8	+47	+16	0	+44	−193	+50

Table 3.7 Final and corrected final bomb estimates (Adopted from Dalkey and Helmer 1962, 7, 10; own calculations)[12]

	Respondent						
Estimates	1	2	3	4	5	6	7
Final (Questionnaire 5)	177	159	200	255	312	314	494
Corrected final	167	179	206	276	292	349	360
Difference	−10	+20	+6	+21	−20	+35	−134
Difference to Round 4	+1	+27	+6	+26	−8	+17	−140
Difference to Round 3	+9	+73	+22	+26	+36	−176	−90
Difference to Round 1	+42	+129	+56	−24	+92	−651	−4740

schedules that thus resulted from Questionnaire 4 were then used to calculate an answer to the primary question, that is, a revised total estimate of bombs to be delivered on target. Thus, instead of directly restating the primary question, it was calculated by summing up the estimates given for the various industries. Table 3.6 shows the results.

Finally, Dalkey and Helmer circulated the fifth questionnaire. Again, they added some new information and data, and some of the drawings of the participants were inconclusive to the extent that the study authors wished to offer the participants the opportunity to correct them. After this was done, the participants were asked to revise their bombing schedules one last time. Dalkey and Helmer decided to apply a series of mathematical operations to these final estimates. Selected individual estimates on single industries were replaced by the median of the final estimates, producing what they called, in a striking move, "a consensus of estimates" (Dalkey and Helmer 1962, 13f).[11] The final and corrected final estimates are given in Table 3.7.

The figures are summed up by Dalkey and Helmer in a graph (see Fig. 3.1) which, in their words, "brings out very clearly the gradual convergence of the answers" (Dalkey and Helmer 1962, 14).

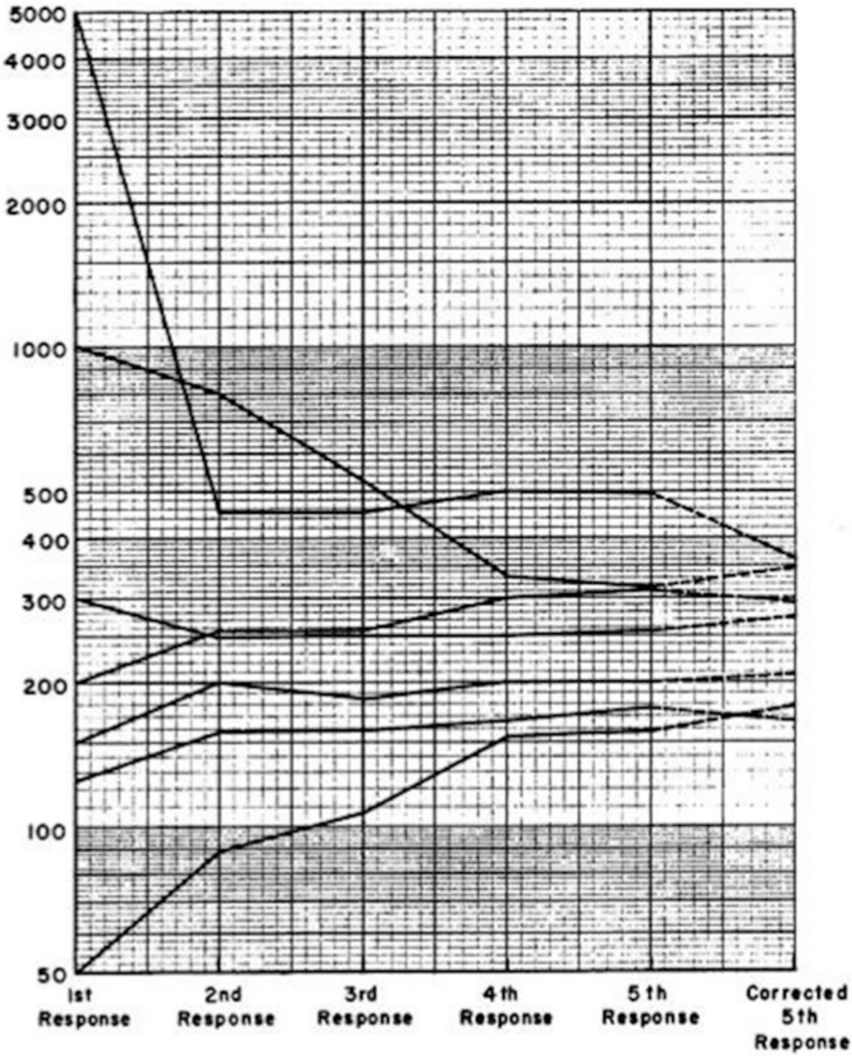

Fig. 3.1 The convergence of estimates (Source: Dalkey and Helmer 1962, 15; reproduced with permission)[13]

Thus, while the initial answers ranged from 50 to 5000, the iterative interrogation together with the additional materials and the specification of the primary question led the study participants to revise their answers in such a way that they finally ranged from 167 and 360. The range was thus reduced from 4950 to 193. Dalkey and Helmer interpreted this convergence as the emergence of a consensus. It was understood that the study participants had agreed that "the least number of bombs that will have to be delivered on target," that is, on facilities of the US munition industry, to decisively damage production ranged between 167 and 360.

Whether this figure was perceived as posing a threat to US security can only be guessed. There is no interpretation in the study. The study authors as well as its readers were certainly aware of the fact that the Soviet Union had the capacity to produce atomic bombs. The first Soviet atomic bomb, RDS-1 (nicknamed "Joe-1" in America), was successfully detonated on a testing site in the Kazakh plains near Semipalatinsk in late August 1949, two years prior to the publication of the Delphi study report. Still, there were different opinions within the US government and military forces about whether the Soviet Union also had developed the capacities to produce large numbers of atomic bombs. Further, since President Truman had authorized the research required to produce hydrogen bombs in late 1949, four months after the detonation of Joe-1, the destructive capacity of a warhead was likely to increase tremendously within the next decade. Scientists expected this increase to amount to multiplication with factor 1000 (cf. Ghamari-Tabrizi 2005, 105). The first US tests of the new superweapon, the hydrogen bomb applying the Teller-Ulam configuration, took place in 1952. It delivered a yield of 10.4 megatons—in other words, 10,400 kilotons or 520 times the bomb dropped over Nagasaki. Contemporary defense intellectuals expected the Soviet nuclear capacities to increase approximately at the same pace as those of the United States. Moreover, with a president hesitant to increase federal military expenditures as rapidly as military officials would have hoped for, the fear that the Soviet Union would soon outweigh the United States in nuclear capacities became more and more widespread. Most prominently, the 1957 report of the Security Resources Panel of the President's Science Advisory Committee, better known as the Gaither Report after the panel's chair (and longtime RAND trustee) Rowan Horace Gaither (1909–1961), claimed without any hint of doubt that the Soviets had "produced fissionable material sufficient for at least 1500 weapons" (Sprague and Foster 1957, 4). Origin of the infamous Missile Gap thesis, this report is an utterly valid indicator of the widespread fear of US inferiority in the nuclear arms race (Amadae 2003, 47–57; F. Kaplan 1983, 123–173). Its hope was that by transforming this fear into unwavering, seemingly factual statements, it would finally be taken seriously.

By 1951, as we know today, the Soviet Union had installed six of at least ten Atomgrads, closed cities in which scientists carried out research on weapon design and production. However, from what we know today, it is improbable that the factories in these closed cities were already able to produce such numbers of bombs. The current estimations are that by 1964—more than ten years after the scenario of the first Delphi study—the Soviet Union had assembled 500 warheads, whereas the United States had a stock of 6800 (M. Bowker 2002, 95). This, however, is knowledge from hindsight. Even though we know today that the Soviet Union had by far no sufficient number of bombs—and that signs indicating this were generally ignored by decision-makers—, the estimates from this first Delphi were probably assessed as worrisome by those who produced and read them.

The Wisdom of the Expert Group

What is the epistemic role of the expert in the two or, including the studies from the 1930s, three approaches outlined above? What is the expert expected to know, how is she expected to know it, and what follows from this for designing a prospective methodology based on expert estimation? In comparing these studies, we find a series of continuities. The first and, since it functions as the axis of comparison, the most obvious similarity is of course the idea of *interrogating experts via questionnaires* to produce valid predictions. Also, one can note a very peculiar continuity in style. The words chosen to describe the studies come from the repertoire of the experimental design. Even Dalkey and Helmer, in their own text, repeatedly refer to themselves as “experimenters.” This is as astonishing—their study lacks the constitutive features of experimental design—as it is telling. Their understanding of the nature of an experiment was the understanding of philosophers, not of psychologists (or, for that matter, of natural scientists). When using the term experiment, they did not think about the detection of causal relations. Rather, they were thinking of a trial-and-error approach to testing the feasibility of the research technique they were developing (cf. Dayé 2018).

Another striking feature of RAND's first studies in expert prediction is the obviously unquestioned belief in the *predictive superiority of experts* over laypeople. The earlier studies by McGregor (1938) and Cantril (1938) had shown no difference in predictive success between experts and laypeople, and Kaplan, Skogstad, and Girshick also found evidence supporting the view that prior (general) knowledge had no influence on predictive success. Nonetheless, Dalkey and Helmer continued to focus on experts.

In the context of their study, however, this appeared justified. In contrast to the earlier studies, their questionnaire items were not concerned with everyday predictions, but with a primary question that assumed an extremely high level of expert knowledge and intellectual capacities. One only has to repeat the primary question to see that it is appropriate for addressing a general audience: "What is the least number of bombs that will have to be delivered on target for which you would estimate the chances to be even that the cumulative munitions output (exclusive of investment) during the two year period under consideration would be held to no more than one quarter of what it otherwise would have been?" (Dalkey and Helmer 1962, 6). For the study participant, yet to understand what the task already involved required a considerable amount of acquaintance with logical complexity and mathematical thinking.

In line with this difference in research interest—various general predictions in Kaplan et al., a very concrete and specific prediction in Dalkey and Helmer—we can find another crucial difference between the two studies. This concerns how the study designs treated the participants. Kaplan, Skogstad, and Girshick asked the participants to make predictions based on their knowledge of the field. The interest in assessing the nature and content of the knowledge behind the participant's predictions was small. The items provided some space for providing the "basis of judgment," but the analysis was superficial. More importantly, however, the participants' reasoning behind their answers was not a decisive element in the methodological design. Since the study by Kaplan et al. did not apply an iterative design, there was no option to feed back the participants' reasoning into the prediction process. And while there certainly was an exchange of opinions and arguments in those groups who were given the chance to discuss (the collaborative group and the joint group), the inherent intellectual quality of these arguments never was a concern of the researchers. Their interpretation of the differences in predictive success across the groups further corroborates the impression that Kaplan et al. did not deem the quality of arguments important. What they saw happening in the groups with discussion was not the victory of the best argument, but a procedure of mutual adjustment toward the center, a leveling-out of the most deviant opinions—in fact, an averaging that could also be mathematically modeled.

In contrast, the reasoning behind the estimates was a crucial element of the first Delphi study. Most of the effort by the study leaders concentrated on elucidating and communicating these ways to think about the problem, always with the intention to improve among the participants the mutual understanding of the approaches of others. In other words, Dalkey and Helmer asked experts to evaluate the extent to which the iteratively improved data and informational base accessible to all of them altered their estimates. As they elaborated in a later article described in Chap. 4, the basic idea was to collaboratively construct a set of evidence (explicit knowledge) which the experts, based on their tacit or implicit knowledge, would then be asked to evaluate with regard to the question or hypothesis at hand. And the iteration implemented by the feedback loop gave this collaboration a structure.

This is one of the two major differences in how the first Delphi study and the precursor study conceived of the epistemic role of experts. While Kaplan et al. had expected their experts to predict *ex nihilo*, Helmer and Dalkey expected them to evaluate a primary question in the light of an evolving set of evidential materials that was made accessible to all. Their task was not simply to give their opinion on a question regarding the future, but to evaluate how a body of information related to such a question. While, in a word, the research designs by Kaplan et al. (and Cantril and McGregor) had conceived of the participant as *predictors*, Helmer and Dalkey, by emphasizing the relevance of factual information, conceived of them as *forecasters*. While *predictors* deliver statements (predictions) without reference to corroborating evidence, *forecasters* use a set of available evidence to judge the likelihood of realization of a specified development or hypothesis.

The second difference between the two RAND studies described in this chapter is hidden behind what at first appears to be a continuity. As already mentioned, Dalkey and Helmer started from the precursor study's finding that the highest rate of predictive success was shown in a collaborative setting. This success, the authors had argued, resulted from some kind of averaging effect that took place within the group. But they had also shown that this averaging effect could be modeled by means of mathematics without leading to a decline in predictive success.

In the argumentation of the first Delphi study, however, the point established by Kaplan et al. was put in a subtly different context. For Dalkey and Helmer, the fact that the averaging effect could be modeled was reason enough to completely dismiss any form of direct interaction between the study participants in favor of a methodological structure in which interaction was fully controlled by the researchers. In the eyes of Dalkey and Helmer, mathematics fully substituted for direct interaction among the experts: the predictive success being the same, there was no need to have direct interaction. Indeed, since it increased the control on behalf of the study leaders, avoiding such interaction was perceived as an advantage.

In a concluding section titled "Critique of the experimental procedure," Dalkey and Helmer mentioned several shortcomings of their study. The list opened with three points that are worth investigating more thoroughly[14]:

> (i) The experts' responses were not strictly independent. Although the respondents on the whole complied with the initial cautioning not to discuss the experiment with one another while it was in progress, their other working assignments on related subjects required some contact among several of them. (ii) At least one of the respondents was also used by the experimenters as a consultant on one aspect of the subject matter of the experiment. (iii) Some "leading" by the experimenters inevitably resulted from the selection of the information supplied by the experts. (Dalkey and Helmer 1962, 16)

As these points clearly suggest, the authors' initial intention was to create an experimental method that was as "clean" as possible. They wished to completely control what information was communicated to the participating experts; they wished to avoid direct interaction between the participants; and they wished that their operating as experimenters would not influence the process of opinion formation throughout the multiple stages of the study. Certainly, these wishes were only partly fulfilled in practice. But more important is the fact that they mirror an unquestioned understanding of group interaction as being not only valueless but potentially distorting. *Face-to-face interaction was a threat* to the quality of the results. "[I]f a group of people are brought together," a later commentator remarked, "a great many socio-psychological interactions occur that detract from the development of a good forecast or a good decision. The Delphi technique is a way of allowing only those interactions to occur that are likely to improve the quality of the forecast or decision" (Cornish 1977, 119). Which social-psychological effects exactly the authors had in mind, remained obscure.

The epistemic role of the expert in these first predictive studies at RAND, thus, was the following: Against stable evidence, the belief was upheld that experts were more capable than laypeople of coming up with good predictions. To gather their predictions, questionnaires were a sensible instrument, not least because they allowed for research designs that avoided direct interaction amongst the experts—a factor suspected of having distorting effects. As regards the presumed capacities of the experts and, as a consequence, the epistemic status of the results of the techniques, there was ambivalence. While the study by Kaplan et al. was an exercise in prediction, the experts in Helmer and Dalkey's study had produced a forecast.

Notes

1. The affiliations stated in the article—UCLA for Kaplan, Stanford for Girshick, RAND for Skogstad only—might obscure that it reported, in fact, a study carried out at RAND. Kaplan joined RAND initially as a consultant in June 1947, but quickly became an employee in September 1947 and only resigned from this (part-time) position when he advanced from associate to full professor at UCLA's Department of Philosophy in

1952. Nonetheless, Kaplan continued to work for RAND as consultant until 1965. Girshick had joined RAND in summer 1947 and stayed for a little more than a year before getting hired by Stanford University in September 1948 (cf. Blackwell and Bowker 1955; Daly 1955; A. H. Bowker et al. n.d.).

Before working on the prediction study, Kaplan wrote a research memorandum on "The Concept of Military Worth. RM-37" (A. Kaplan 1948) that became an influential point of reference for RAND's analyses in these early years. Robert Leonard (2010, 281 fn. 46) suggests that the concept of military worth had been introduced by mathematician Merrill Flood, then at Princeton's Fire Control Research Office, in a 1944 report that applied game theory to the study of World War II bombing campaigns. This might match with, but add to the finding of William Thomas (2015, 127f), who identifies a report written in 1945 by Warren Weaver, then the director of the Applied Mathematics Panel (AMP) as an early instance of a text applying this notion. The Fire Control Research Office was working under contract for the AMP, and despite working at Princeton, Flood was the secretary of AMP's steering committee (cf. Erickson 2015, chap. 3). The main achievement of this concept was to establish a direct link between military strategy and tactics and the established tools of economic analysis. At RAND, military worth became a core concept for roughly a decade. John D. Williams, a former AMP researcher who became the first mathematician employed by Project RAND in 1946, immediately built up a "Military Worth Evaluation Section" which only later was renamed into Mathematics Division (cf. Collins 2002, 119ff). Flood joined RAND in the late 1940s and continued to work on game theory-based strategic analyses. Despite the change of the name of the section, then, military worth remained a central concept of RAND analyses.

2. McGregor (1938) develops a methodology to discern the effects of an individual's attitudes, wishes, and prior knowledge on his or her prediction. He shows that the degree of ambiguity in an individual's knowledge as well as the importance he or she attributes to the question at hand are much more significant determinants of the prediction than his or her expertness. Building on that, Cantril (1938) emphasizes the relevance of external frames of reference for the objectivity of a prediction.
3. Repeatedly, for instance, Nicholas Rescher referred to the Kaplan et al. study as the "precursor" of the Delphi technique (e.g., Rescher 1997, 353; 2007, 104).

4. 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).
5. The meaning of Question 2 remains unclear. J-Day usually refers to the day an assault occurred. Apart from corroborating the assumption that the study participants were familiar with military parlance, the use of this term is puzzling, since J-Day is variable as concerns the factual date of occurrence. The term denotes the day an operation will be carried out. One possible interpretation is that the authors identified a specific J-Day for the questionnaire and forgot to mention it in the article. Another interpretation is that the participants, closer to the event than we are today, understood by J-Day what Americans nowadays call V-J Day, the Victory over Japan Day or, more precisely, the day of the Japanese surrender. V-J Day is celebrated in the United Kingdom on August 15, the day of the actual surrender, and in the United States on September 2, the day of the formal surrender ceremony on the USS Missouri. However, assuming that the date given in the questionnaire (January 12) is correct, both dates transcend the proclaimed maximum of 20 weeks (22 resp. 24 weeks).
6. A few examples might be appropriate: Item 23 of the "Current Social Problems" part reads: "Which of the following has most recently become an important feature of dispute between labor and industry? (1) Sit-down strikes (2) Lockouts (3) The yellow-dog contract (4) Government intervention (5) Sabotage" (Blair et al. 1947, 3). Item 38 of the "Science" part, in turn, reads: "Osmosis is a process of (1) oxidation (2) diffusion (3) adsorption (4) reduction (5) magnetic attraction" (Blair et al. 1947, 22).
7. With this name, he starred in "Im Bewusstsein der Schuld" (1916), a movie directed by William Wauer that received wide acclaim from contemporary critics (on Wauer, see Wedel 2014, 90–103). Since Torvald Helmer is by far no heroic character, Hirschberg's choice of name is interesting.
8. Both the wording "Japanese bombings" and the fact that figures from the devastating attacks on Hiroshima and Nagasaki are used as examples to estimate the destructive force of A-bombs on US territory might appear odious to the reader in our times. They should however not be misunderstood as cynical or bloody-minded. In the culture of insecurity that characterized the Cold War era, many perceived using all available means to counteract the potential disaster as their prime moral duty toward humanity.

9. The difference to the estimates from the first round was calculated with the figures from the interviews.
10. Again, the difference was calculated with the figures from the interviews.
11. The procedure applied was intricate: "Our procedure was, first of all, to tabulate for each of the industries considered the medians of (i) the expected numbers of plants respectively producing 50%, 75%, and 100% of the total output in mid-1953, and (ii) the number of plants requiring two rather than one bomb on target for destruction. We then listed (iii) the percent of damage to each industry that each expert intended as indicated from the figure he gave for the numbers of plants in mid-1953, the number of bombs needed to destroy 75% and 100%, and of bombs to be allocated to each industry, and (iv) the corresponding numbers of bombs as computed with the aid of the tabulation obtained under (iii)" (Dalkey and Helmer 1962, 14).
12. The differences to the estimates from the previous rounds were calculated with the corrected final figures. The figures from round 3 are those of the follow-up interview.
13. To be clear, the 2nd and 3rd responses had been given in Questionnaire 3 and the follow-up interview, respectively.
14. The other items on the list concern the time schedule ("[t]he experiment was terminated prematurely"), the comparative task in Questionnaire 4, the vague wording of some questions in Questionnaire 2, and the missing theoretical foundation for the correction of the final estimates (Dalkey and Helmer 1962, 16f).

References

Amadae, S.M. 2003. *Rationalizing Capitalist Democracy: The Cold War Origins of Rational Choice Liberalism*. Chicago: The University of Chicago Press.

Blackwell, David, and Albert H. Bowker. 1955. Meyer Abraham Girshick 1908–1955. *The Annals of Mathematical Statistics* 26 (3): 365–367.

Blair, Norman J., Jeanne M. Bradford, Miriam M. Bryan, Paul J. Burke, and Herbert Danzer. 1947. *Cooperative General Culture Test. Revised Series Form X*. Cooperative Test Services.

Bowker, Mike. 2002. Brezhnev and Superpower Relations. In *Brezhnev Reconsidered*, ed. Edwin Bacon and Mark Sandle, 90–109. Basingstoke and New York: Palgrave Macmillan.

Bowker, Albert H., Kenneth J. Arrow, and Herman Chernoff. n.d. Memorial Resolution: Meyer Abraham Girshick (1908–1955). Stanford University Faculty Memorials, Stanford Historical Society. histsoc.stanford.edu/pdfmem/GirshickM.pdf.
Cantril, Hadley. 1938. The Prediction of Social Events. *Journal of Abnormal and Social Psychology* 33 (3): 364–389.
Collins, Martin J. 2002. *Cold War Laboratory: RAND, the Air Force, and the American State, 1945–1950*. Washington, DC and London: Smithsonian University Press.
Cornell, John A. 1993. W.J. Youden—The Man and His Methodology. *ASQC (=American Society for Quality Control) Statistics Division Newsletter* 13 (2): 9–18.
Cornish, Edward. 1977. *The Study of the Future. An Introduction to the Art and Science of Understanding and Shaping Tomorrow's World*. Washington, DC: World Future Society.
Dalkey, Norman C. 1968. Predicting the Future. P-3948. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3948.html.
Dalkey, Norman C., and Olaf Helmer. 1962. An Experimental Application of the Delphi Method to the Use of Experts. RM-727/1-ABRIDGED. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM727z1.html.
———. 1963. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science* 9 (3): 458–467.
Daly, Joseph F. 1955. Meyer Abraham Girshick, 1908–1955. *The American Statistician* 9 (3): 6.
Dayé, Christian. 2018. How to Train Your Oracle: The Delphi Method and Its Turbulent Youth in Operations Research and the Policy Sciences. *Social Studies of Science* 48 (6): 846–868.
Erickson, Paul. 2015. *The World the Game Theorists Made*. Chicago and London: The University of Chicago Press.
Fleck, Christian. 2011. *A Transatlantic History of the Social Sciences: Robber Barons, the Third Reich and the Invention of Empirical Social Research*. Trans. Hella Beister. London and New York: Bloomsbury Academic.
Forces War Records. n.d. Hitler's Black Book—Information for Doctor Olaf Helmer-Hirschberg. Accessed March 28, 2017. https://www.forces-war-records.co.uk/hitlers-black-book/person/197/doctor-olaf-helmer-hirschberg/.
Friedman, Walter A. 2014. *Fortune Tellers: The Story of America's First Economic Forecasters*. Princeton and Oxford: Princeton University Press.

Ghamari-Tabrizi, Sharon. 2005. *The Worlds of Herman Kahn. The Intuitive Science of Thermonuclear War*. Cambridge, MA and London: Harvard University Press.

Harper, Kristine C. 2012. *Weather by the Numbers: The Genesis of Modern Meteorology*. Cambridge, MA and London: MIT Press.

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

Kaplan, Abraham. 1948. The Concept of Military Worth. RM-37. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM37.html.

Kaplan, Fred. 1983. *The Wizards of Armageddon*. New York: Simon & Schuster.

Kaplan, Abraham, A.L. Skogstad, and Meyer A. Girshick. 1950. The Prediction of Social and Technological Events. *Public Opinion Quarterly* 14 (1): 93–110.

Leonard, Robert J. 2010. *Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900–1960*. Cambridge and New York: Cambridge University Press.

McGregor, Douglas. 1938. The Major Determinants of the Prediction of Social Events. *Journal of Abnormal and Social Psychology* 33 (2): 179–204.

Pietruska, Jamie L. 2018. *Looking Forward: Prediction & Uncertainty in Modern America*. Chicago: University of Chicago Press.

Rescher, Nicholas. 1997. H2O: Hempel-Helmer-Oppenheim, An Episode in the History of Scientific Philosophy in the 20th Century. *Philosophy of Science* 64 (2): 334–360.

———. 2005. *Studies in 20th Century Philosophy*. Walter de Gruyter.

———. 2006. The Berlin School of Logical Empiricism and Its Legacy. *Erkenntnis* 64 (3): 281–304.

———. 2007. *Autobiography*. Walter de Gruyter.

Sprague, Robert C., and William C. Foster. 1957. Deterrence & Survival in the Nuclear Age. Security Resources Panel of the Science Advisory Committee, Washington, DC.

The University of London. n.d. Graduates List 1935–1936. Accessed March 28, 2017. http://www.senatehouselibrary.ac.uk/our-collections/special-collections/archives-manuscripts/university-of-london-students-1836-1934.

Thomas, William. 2015. *Rational Action: The Sciences of Policy in Britain and America, 1940–1960*. Cambridge, MA: MIT Press.

Tolon, Kaya. 2011. The American Future Studies Movement (1965–1975): Its Roots, Motivations, and Influences. Iowa State University, Ames, Iowa.

Wedel, Michael. 2014. *Filmgeschichte als Krisengeschichte: Schnitte und Spuren durch den deutschen Film*. Transcript Verlag.

4

Negotiating Rules for the Game: Political Games at RAND, 1954–1956

War Simulations, Hot and Cold

Unlike Delphi, which was in essence a social science method developed by philosophers of science, political gaming had been developed by social scientists. They were members of RAND's Social Science Division (SSD) headed by Hans Speier (1905–1990), a native German who escaped to the United States in the early 1930s. The development of political gaming must be seen in the context of, as well as in contrast to, other approaches to policy analysis then dominant at RAND, most prominently game theory. In brief, while the main thrust of game theory consisted in applying mathematical and logical reasoning to strategic problems, gaming approaches allowed for exploring the potential trajectories of a conflict or any other social process by simulating them in a step-by-step manner, similar to the moves of a game. The relations between game theory and political gaming are manifold and complex, both as regards discussions about their scientific and methodological complementarity or antagonism and their historical development. Insofar as they have been important for the development of political gaming, some of these relations are treated in the following.[1] The most important similarity, of

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*, Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_4

course, is that in both approaches, the guiding idea is that the results of a process are an emergent property of the interaction between various parties. What results from a specific move depends on the move of the other players. In this perspective, strategic action always implies a moment of uncertainty, a moment that is beyond the capacities of one party to control.

Political gaming was neither the only type of gaming used at RAND nor the first. RAND analysts had already been involved in war gaming before the SSD developed its proposal.[2] And of course, the use of game boards and pawns to represent battles was not a modern invention. Earlier versions of today's probably most famous war game, chess, have been played since the sixth century. Whereas the objective of war games was entertainment in the majority of cases, boards and pawns were also used as devices in probing and evaluating factual military tactics. In his analysis of war games in the European Middle Ages and the Baroque period, Philipp von Hilgers found that "mathematical and military semiotics could initially coincide entirely with the concept of the game and only gradually underwent a differentiation. Only in this way can it become clear that the divided mathematical and military professions of the twentieth century ultimately remain, at a subterranean level, in thrall to the game as a medium" (Hilgers 2012, x). The heuristic function of war games from a military perspective was systematically used, if not before, in the nineteenth century. The Prussian Army had conducted a *Kriegsspiel* around 1865, and an American counterpart was devised by Major William Roscoe Livermore in 1879 and played by the Volunteer Militia of Rhode Island and other states in the years following the civil war (cf. Specht 1957, 7). In short, at the time RAND was created, war gaming had been a military art for more than a century at the very least.

The most important change that occurred in this tradition in the twentieth century was the transformation of this military art into a scientific technique. On the social level, this transformation had been prepared by the increasing involvement of scientists in foreign policy research during World War II. The event that finally triggered the transformation, however, took place on the cognitive level. It was the publication in 1944 of *Theory of Games and Economic Behavior* by John von Neumann and Oskar Morgenstern (1953) that operated as the metaphorical catalyst of this transformation. To become a field of activity to the civilian militarists at RAND, gaming had to gain scientific character. The mathematical and the military semiotics first had to become related again. This is what *Theory of Games and Economic Behavior*, widely known by the abbreviation *TGEB*, accomplished.

In the places where the two authors received their scientific training—von Neumann's Göttingen, Morgenstern's Vienna—parlor games were very popular. Not only was chess a game commonly played in Central-European coffee houses but as a mode of thinking, it had also entered into the scientific discourses in which the two scientists participated, which was mathematics in the case of von Neumann and economics in the case of Morgenstern. Both were trained, thus, in cultural contexts where the idea to conceive of a parlor game as scientific heuristic was not completely beyond the scope of reasonable consideration, even if, as Leonard (2010, 77) argues, the two initially used chess in different intellectual contexts. While the mathematician in Göttingen conceived of chess as a logical structure allowing for the mathematical formulation, the Viennese economist rather took it as a metaphor for the infinite struggle of competing interests and the high degree of complexity and psychological factors involved in such struggle.

That the two overcame their differences in intellectual interest and outlook to write *TGEB* is, in Leonard's perspective, a result of the historical events. From the 1920s on, the political situation worsened in both Germany and Austria, and anti-Semitism was on the rise. In these times of high insecurity, writings and talks by Karl Menger led to Morgenstern's conviction that societal and economic problems—group coherence, social integration, and societal stability—could be addressed with the instruments of mathematical logic. Sticking to this conviction increasingly alienated Morgenstern from his Viennese fellow economists. As regards von Neumann, he began to concern himself with political problems as a reaction to the political events in Europe. "Unlike at Göttingen a decade previously, where he [von Neumann] had been concerned with the behaviour of the chess- or poker-player, his concern was now with the rationality of the social actor or unit. […] His concern became that of understanding social coalitions" (Leonard 2010, 222–223), and he attempted to model them by mathematical means. This is the point where von Neumann's and Morgenstern's interests converged.

In spite of the claim to demonstrate the applicability of mathematical procedures on decision-making in economics or other fields of strategic action, game theory as formulated in *TGEB* yielded hardly any immediate response in economics nor in the social sciences. It was, in Leonard's (1992, 59) words, "without a 'natural' audience:" neither mathematicians nor economists initially felt attracted to game theory. So when RAND, under the aegis of the director of the Mathematics Division, John D. Williams, took up game theory already at the time of the establishment of Project RAND in 1947, it became virtually *the* place to be for those researchers who pursued von Neumann and Morgenstern's program.[3] Although admittedly not the only habitat for game theorists in the field's first decade, RAND nonetheless was one with considerable resources and a genuine organizational interest in developing game theory. However, because of its thematic and organizational setting, RAND promoted peculiar lines of development with more emphasis than and to the disadvantage of others. Williams' initial objective for the Mathematics Division was the development of a general, mathematical theory of war, which he wanted to base on game theory. A trained astronomer, Williams had worked at the Applied Mathematics Panel (AMP) in New York City during World War II (cf. Collins 2002, 80). The AMP was a subunit of the National Defense Research Council, itself a subsidiary organization of the Office of Scientific Research and Development. Established in 1942, its creator and first director Warren Weaver took measures to focus the AMP's research activities on operations research (Thomas 2015, 102–109). Williams' inspiration and enthusiasm for applying mathematical analysis to problems of combat was certainly a result of his time at the AMP. He also met a series of people who later joined him—as employees or consultants—at the RAND Corporation, among them John von Neumann, Olaf Helmer, or Ed Paxson, the creator of systems analysis. After the war had ended and on the recommendation of Weaver, Williams moved to Santa Monica to become one of the first employees of Project RAND.

In the negotiations over his first assignments, Williams had agreed to take over the task of recruiting social scientists and economists for the envisaged new Social Science and Economics Divisions. The reason for Williams' willingness to do so related to his larger project of a general theory of war. Though the theory he projected was basically mathematical, algorithms still needed figures. And it appeared logical to him that social scientists were in charge of delivering these figures, most preferably in the form of preference matrices or utility functions for each of the involved war opponents. If that was too much to ask, Williams thought, they should at least produce studies in a form that made possible an empirically funded attribution of figures (cf. Kaplan 1983; Leonard 1992; Abella 2008).

Members of the Social Sciences Division never complied with that expectation and it can be doubted whether the RAND management shared Williams' views of what kind of research should be done at RAND. A general science of war implied "that Williams' mathematical group ought to oversee and coordinate the work of RAND's various departments and sub-units—which […] did not happen. RAND's emerging research departments […] were more or less equals in the life of the organization" (Collins 2002, 136f). Although Williams himself soon veered away from this endeavor, the work on game theory at RAND continued to follow the line of thinking that had inspired Williams' initial plans. For instance, while von Neumann and Morgenstern had focused preponderantly on games with *n* participants, the RAND researchers concentrated on games with only two opponents. This made sense in the specific organizational context of RAND, but of course could be seen as less challenging in terms of mathematical theory. Nonetheless, "[t]his postwar work was done with von Neumann's sanction and encouragement, even though he alone had developed the mathematics of cooperative games, had devoted the bulk of the Theory of Games to that topic, and saw the analysis of n-person games as the theory's crowning achievement" (Leonard 1992, 30). In the 1940s and 1950s, RAND was the regular meeting point of a group of mathematicians, young and old, most of them having received their degrees from Ivy League universities like Harvard, Columbia, or Princeton, and many of them are still remembered for their contributions to game and decision theory: RAND had contracts with, in alphabetical order and with no claim of completeness, Kenneth Arrow, George Dantzig, Melvin Dresher, Merrill Flood,

R. Duncan Luce, J. C. C. McKinsey, John F. Nash, Lloyd Shapley, and Martin Shubik.[4] Von Neumann and Morgenstern both held consultant contracts with RAND;[5] the former also supported the construction of RAND's first computer, which the engineers, to honor this contribution, named JOHNNIAC.[6]

Perhaps as a result of the currency of game theory in RAND's first years, RAND proved to be fruitful soil for game thinking more generally. Apart from developing the mathematics of game theory, several approaches to games and gaming were devised, ranging from the simulation of the man-machine interaction involved in military operations logistics (Edwards 1996, 122–124) to the simulation of the political intricacies of conflicts and to economic bargaining theory (Schelling 1968).[7] Considerable effort was spent on experiments testing game theoretical propositions and derivations. One early set of experiments was conducted by Merrill Flood, the co-inventor, with Melvin Dresher, of the infamous prisoner's dilemma. Flood was concerned about the lack of empirical or experimental testing of game theory. As he put it "[i]t has too often been forgotten that a theory is simply something to be tested experimentally, and to be rejected for any application where it fails to fit the condition" (1952, 1). To remedy this, Flood had conducted a "preference experiment" with three Harvard undergraduates in 1951. He had used the facilities of the Harvard Laboratory of Social Relations, then directed by Robert F. Bales, and acknowledged the valuable suggestions on the experiment's design that he received from Bales, C. Frederick Mosteller, and Samuel A. Stouffer (cf. Flood 1951, 3). The students were sitting together in the laboratory and were asked to select one of a range of alternative options. "The experiment is arranged so that each individual has a strong incentive to guide the group selection toward the alternative that he prefers" (Flood 1951, 1). The experimenters urged the group to discuss the problem but informed them that the first decision announced by anyone of them was final.[8] Flood complemented such formal experiments by informal ones concerned with ordering dinner at a restaurant (cf. Flood 1951, 11–13), or RAND colleagues selling their cars (cf. Flood 1952, 5–14). In each of these cases, it was attempted to relate the experiment to game theory and inform or modify the latter according to the experimental results.

A later instance of RAND's experimental game theory endeavors was the *Public Opinion Game* construed by W. Phillips Davison (1960). It attempted to simulate opinion-forming processes within groups. The participants were assigned "roles" arbitrarily. These roles were defined by the following information: sex, age, marital status (including number of children), socioeconomic status, occupation (in the case of a "housewife," the occupation of the husband was given), religious affiliation, memberships, and the affiliation to a primary group (or better subgroup) among the participants. Additionally, a scenario was described in the instruction sheets that were distributed at the beginning of the experiment. One such scenario read (Davison 1960, 6):

> You live in a medium-sized city in a mid-Western state, with a large farming hinterland. Recently a number of developments of some importance have taken place in your area, although these have not been publicly announced or generally discussed. First, geologists from the local university have discovered that one of the local parks in the city almost certainly has a large supply of oil beneath it. Second, the county clerk's office has noted that the percentage of unwed mothers has doubled during the past year. Third, several members of the city council are considering the introduction of a law to require all motorists to install an exhaust-purification mechanism, costing about 25 dollars, on their cars and trucks.

Using such scenarios as points of departure, experimenters expected to observe the processes by which group opinions emerged and consolidated.

However, the dominant approach to gaming in RAND's early years—and as a consequence, the second reference point for political gaming apart from game theory—was war gaming, that is, the simulation of battles with pawns on a stylized map (see Fig. 4.1). A report by Robert D. Specht (1957), published by RAND in 1957 as "War Games. P-1041," provides an overview of various approaches to war gaming. It opens with some remarks on the history of war gaming. Beside the already mentioned Prussian and American versions of *Kriegsspiel*, attention was drawn to the fact that Japan had established a *Total War Research Institute* in October 1940 (Sōryokusen Kenkyūjo). Throughout Summer 1941, the institute played games with

Fig. 4.1 Debate during the STRAW Strategic Air War Game, 1953; the man in the center without a tie is John D. Williams, longtime director of the Mathematics Division (RAND Corporation Archives, Santa Monica, CA)

players representing "the Italo-German axis, Russia, United States, England, Thailand, Netherlands, East Indies, China, Korea, Manchuria, and French Indochina" (Specht 1957, 2). An increase in complexity and realism was reached by representing Japan not as a single group of decision-makers but "as an uneasy coalition of Army, Navy, and Cabinet" (Specht 1957, 2), in other words by instituting multiple national organizations with the intent to simulate the conflicts *within* the Japanese governing forces. These games took place in immediate relation to the contemporary diplomatic exchange between Japan and the United States. As of 17 August 1941, the status of this exchange was that President Franklin D. Roosevelt had welcomed the Japanese proposal to restart informal discussions, provided that Japan committed to complete openness and avoided strategic maneuvers. Based on the war games as well as on other studies, the institute's suggestions was: "To the proposal of America, we shall neither give our word clearly concerning the position of Japan, but adopt a delaying policy by diplomatic negotiations, repleting war preparations in the meantime" (Boister and Cryer 2008, 1:501).

A few weeks later, the Japanese Navy conducted a war planning game at the Naval War College in Tokyo, which was concerned with two general problems, "first, the details of a surprise raid on Pearl Harbor and, second, a carefully worked-out schedule for occupying Malaya, Burma, the Dutch East Indies, the Philippines, the Solomons, and the Central Pacific Islands, including (ultimately) Hawaii" (Specht 1957, 4; cf. Boister and Cryer 2008, 1:505). Even the infamous code fixed to initiate the attack, "The cherry blossoms are in all their glory," was part of the series of games. Specht also states that some of the plans that resulted from these games, for example, with regard to the control of consumer goods, were eventually put into force in the aftermath of the attack on Pearl Harbor.

Presumably, war gaming had been practiced at RAND very early in its history. So, when in 1954 mathematician Alexander McFarlane Mood (1954) wrote a paper on "War Gaming as a Technique of Analysis. P-899," he was able to critically assess some of the methodological developments of the previous years. Whereas the traditional understanding of the scope of war games has been that it represented the final step in the preparation of a war plan, the procedure had recently been "modified to make it a method for solving problems previously thought to be beyond analysis and answerable only by appeal to the judgment of experts" (Mood 1954, 1). This modification, Mood explained, concerned the nature of the problems one conceived as being solvable by such games. In principle, one could discern problems that can be treated detached from their context, so-called factorable problems, and problems where such detachment cannot reasonably be achieved. A factorable problem, for instance, might be to determine "the best size for a bomber cell [a small group of bombers which flies together over enemy territory and attacks one or a few targets] given various reasonable assumptions regarding enemy defenses, target lists, electronic countermeasures, and so on. The cell-size question can be studied intelligently" (Mood 1954, 1) and the solution of such factorable problems along explicit criteria is called "sub-optimization."

On the other hand, non-factorable problems are withdrawn from any easily calculable solution because of the complexities of the situation. In a war, for instance, decisive elements of the situation "are in the control of the enemy" (Mood 1954, 3). Thus, unless one assumed full understanding and explicitness of the opponent's strategy, it could not be calculated. This was not a small problem for the way in which games had been discussed in the previous decade. Inspired by the publication of *TGEB*, Mood said, this period saw much theoretical investigation of games and rational modes of behavior in conflict situations. This had led to "a considerable body of clarifying ideas and a technique which can analyze quite simple economic and tactical problems. These techniques are not even remotely capable, however, of dealing with complex military problems" (Mood 1954, 3f).

This is where gaming, the actual playing of a game in contrast to calculating the matrices of strategies and outcomes, entered the stage (Fig. 4.2). At RAND, war gaming had players forming two or three teams who sat in separate rooms. All teams had the same map of the war region on which, in most cases, a grid of pentagons was protracted. War games at RAND followed one of two basic designs. One could either institute an umpire who, based on his or her military experience, assessed the value and effectiveness of any move, "the link between the tactics chosen by Red and by Blue and the results of the engagement as measured by movement and attrition of forces" (Specht 1957, 9). The structure of such a "conventional" war game was simple, at least from Mood's perspective.

> Three teams (Red, Blue, and the Umpire Team) play through a war in detail on maps. Blue's strategy is fixed by the plan. Red devises a strategy particularly intended to expose weaknesses in the plan. The results of the players' moves are adjudicated (after a certain amount of debate) by the Umpire team. It is not a game in the ordinary sense in which players freely choose their strategies and the outcomes of their moves are determined by written rules. (Mood 1954, 4)

Alternatively, the game could be played without an umpire but rather be based on a comprehensive body of rules and planning factors: "Some of these planning factors were fixed numbers, others were in the form of probability distributions—the toss of a coin or roll of dice being used to choose between alternatives" (Specht 1957, 10). This second alternative design would not forgo the expertise of experienced militarists; instead, "[e]xperience and judgment and intuition are here but they have been used to set up the rules and factors of the game" (Specht 1957, 10). RAND researchers used both designs (Fig. 4.3).

Clearly, the two alternatives, for which RAND researchers found various names but which we will, for their rigidity toward formalized rules, call the non-rigid and the rigid variety, provided for different functions to be fulfilled by the umpire team. The ubiquitous Olaf Helmer, who during the 1950s together with Lloyd Shapley had developed the *Strategic War Planning* game (SWAP) which formed the basis for the *Strategy and*

Fig. 4.2 RAND experts discussing air strategy during a SAFE war planning game, January 1963; the man in the center wearing a black suit is Milton G. Weiner, a major proponent of war gaming at RAND (RAND Corporation Archives, Santa Monica, CA)

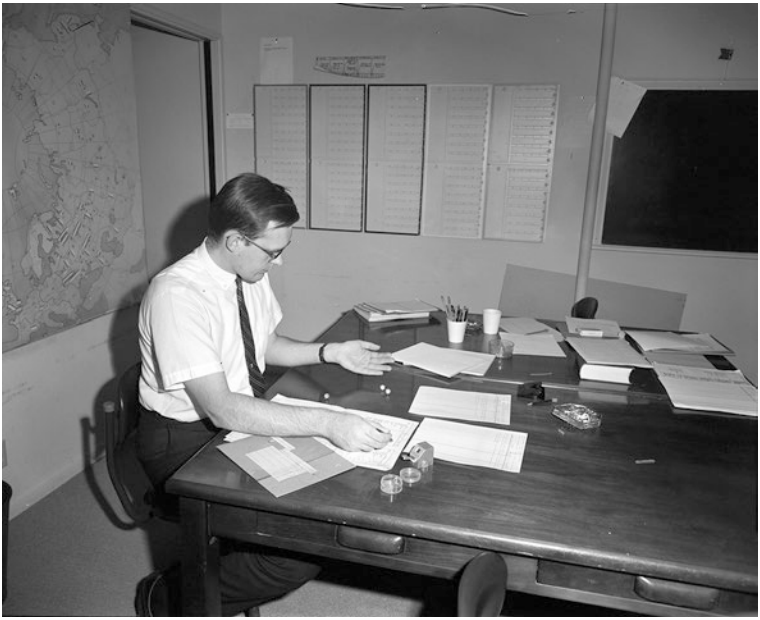

Fig. 4.3 A RAND analyst throwing dices to introduce "chance" into a SAFE strategic air-war planning game, January 1963 (RAND Corporation Archives, Santa Monica, CA)

Force Evaluation game (SAFE) introduced in the early 1960s (Helmer and Bickner 1961), described these functions as follows:

> If the game is non-rigid, additional expertise, through the person of the umpire, must come into play. In this case, if realistic simulation is intended, his discretionary intervention may be necessary in order to eliminate by fiat such player behavior as, in his predictive opinion, is incompatible with the decision making behavior of the player's real-life counterpart. If the game is rigid, on the other hand, since the umpire then acts merely as a procedural technician rather than a substantive expert, the burden of realism falls either on the game constructor, who through rule constraints may be able to enforce realistic player behavior, or on the good sense of the players themselves. (Helmer 1960, 18)[9]

However, divergent opinions existed as to the epistemic potential of such games (Figs. 4.4 and 4.5). Specht, to begin with, thought that the

Fig. 4.4 The SAFE game's umpire team observing the players, January 1963; Olaf Helmer, in the checked jacket on the left, his back turned to the camera, functioned as the game director (RAND Corporation Archives, Santa Monica, CA)

main value of such a war game "is not that it predicts the future nor that it allows the testing of a war plan" (Specht 1957, 12). The crucial tension between replicability and realism prohibited serious testing of strategies. On the one hand, attempts to increase realism quickly led to overly complex simulations which could not be replicated within a reasonable period of time. Easy replicability, on the other hand, per force implied simplification, which in turn undermined its degree of realism. Thus, "[t]he

Fig. 4.5 Postmortem evaluation discussion between Olaf Helmer and Milton G. Weiner, SAFE strategic air-war planning game, January 1963 (RAND Corporation Archives, Santa Monica, CA)

value of the game—and this is a value to be treasured—is that the players are taught to consider carefully all their resources—ground, air, atomic" (Specht 1957, 12; emphasis in original).

Alexander Mood, a close collaborator of Specht, had a quite different view of that issue. He claimed that war games could be used to test war plans. A game forced the analyst to evaluate each step in the order of the (simulated) events "with a view to uncovering flaws in that plan" (Mood 1954, 4). Of course, since the very notion of testing implied replicability, the installing of umpires to adjudicate moves and outcomes was not suitable. Testing was not possible in an open game since such a game "must be easily playable and must be played numerous times by the same players so that they can develop a knowledge of the structure of the game and a feel for good strategies" (Mood 1954, 4). In view of Mood, this could only be achieved by defining a fixed set of rules. From the perspective of the players, such a set of rules ensures that the "experience gained in one play is valid in other plays" (Mood 1954, 4). Further, "[t]he game should include whatever context is needed for a proper treatment of the problem at hand, but no more" (Mood 1954, 4). And the contextual factors included should be simplified and/or combined to make them easily manipulable, which, we might add, was achieved most probably by representing them in numerical form.

As Mood reported, RAND had developed three such games which were located at different levels of the military decision hierarchy, namely at the level of the divisional commander, the theater commander, and the commander-in-chief. They all shared the characteristics that they

> have two opposing teams which make a succession of moves. The outcomes of moves are determined by rules, not by umpires. Generally the teams are small, often just one or two persons. Because it is envisaged that a game must be played frequently if it is to be understood, every effort is made to keep the playing time short; most games can be played through once in a few hours or a few days. (Mood 1954, 7f)

None of these characteristics of the games Mood presented applied to political gaming as it was under development at the time when Mood wrote his report. And the political gamers also had a different take to offer on the problem of realism.

The Realism of Political Gaming

In the section of his report on "War Gaming as a Technique of Analysis" in which he described the promise of war games, Alexander Mood (1954, 9ff) included some remarks on how realistic such games were. It should be clear that the term realism here was used with the meaning attributed to it in simulation and gaming ventures—as a measure of a simulation's resemblance with reality—and not in a philosophical sense. In stark contrast to Specht's (1957, 12) warning that part of the price for the "heightened effectiveness" in rigid-rule games was "that the game may persuade us equally convincingly of things that are not true in the real world," Mood (1954, 9f) contended that realism was not much of an issue for the designers of games.

> [T]he question of how correctly games can reflect reality does not trouble game makers for two reasons. In the first place, they feel that the number of significant factors in any given situation is not so large as to be out of the question in the game representation. In the second place, modern high-speed computers will enable the number of factors which can be included in a game to be increased tremendously, if necessary, without adding to the complexity of the game from the player's standpoint. The computer can be made to make a host of minor decisions on the basis of certain general instructions from the players.[10]

Amongst other things, it was this understanding of realism being a function of the amount of relevant quantitative factors that appalled the group of social scientists around Speier. Even if such an understanding might be appropriate on the operational level of a battle, this was truly not so for the political aspects of war. War could never be reduced to an operational level where all relevant factors are factorable. War, they argued, was always political. And it was impossible to capture the social and cultural levels of any conflict, armed or not, by means of numerical variables in a way that one could justifiably describe as a realistic model.[11]

A place where the social scientists presented this argument was a description of a political game carried out within the German army in the late 1920s. Citing from the then freshly published recollections of Erich von Manstein, Herbert Goldhamer, and Hans Speier wrote:

> Before Hitler assumed power in 1933, the leaders of the German Reichswehr were much concerned about Polish military strength and political designs. The German armed forces were then restricted to 100,000 men in strength. In 1929, a young staff officer, the later General Erich von Manstein,

> charged with the responsibility for the organization of a war game involving German defense against a Polish attack on East Prussia or Upper Silesia, realized that the outbreak of war would be preceded by mounting political conflict. In that conflict, he thought, Germany would have to avoid giving France and Czechoslovakia cause for entering the war as Poland's allies and the League of Nations a pretext for not declaring Poland the aggressor. Manstein proposed that the strictly military exercise be introduced by a political game in order to let political and military leaders learn from each other. High-ranking members of the Foreign Office played the roles of the president of the League of Nations Council and of the Polish and German Foreign Ministers. In his recently published memoirs, Manstein writes that the inventiveness of the player representing Poland in alleging German provocations left his German counterpart "completely speechless" and that the skillfully simulated procrastination of the League was grimly appreciated by all participants. "We had the impression also," Manstein reports, "that the gentlemen from the Foreign Office, to whom such a playing-through of possible conflicts seemed to be completely novel, were thoroughly convinced of the value of the game." (Goldhamer and Speier 1959a, 71f)

Thus, both game theory and war gaming formed a context that simultaneously offered points of critique and of connection for Speier's group of social scientists, with realism being at the center of critique. The key members of the Social Science Division in promoting the genuinely new approach to gaming were Hans Speier, Herbert Goldhamer, Paul Kecskemeti, and Victor Hunt. Hans Speier (1905–1990) was born in Berlin five years before Olaf Helmer.[12] He studied economics and sociology in Heidelberg with Emil Lederer and Karl Mannheim (Bessner 2018). Via Lederer, he met Alvin Johnson who had made plans together with Lederer about whom among the German social scientists to approach for the University in Exile Johnson planned to install at the New School for Social Research in New York City. After his emigration to the United States in 1933, Speier became a faculty member of the "University in Exile" at the New School. Shortly after the war had begun in Europe, he met the émigré psychoanalyst Ernst Kris, with whom he successfully carried out a project on German Radio Propaganda funded by the Rockefeller Foundation (Kris and Speier 1944).[13] With the United States preparing to enter World War

II, this made him a sought-after specialist in Washington. Speier held several positions in government service, among them with the propaganda analysis unit of the Foreign Broadcast Intelligence Service (FBIS) of the Federal Communications Commission (FCC) and with the United States Office of War Information (OWI) both located in Washington, DC. While there, he continued to analyze Nazi propaganda, and these analyses were distributed among various government agencies, among them the State Department and the Department of Defense. In 1947, he was invited to participate in the RAND-sponsored conference on social science organized by John D. Williams, after which he was asked to become the first director of RAND's Social Science Division. Speier accepted this offer after some of his concerns—most notably that RAND would establish an office in the political center of the United States, in Washington DC—had been resolved to his satisfaction. He worked for RAND from 1948 to 1963 (Speier 2007; Bessner 2013), then becoming a professor of sociology at the University of Massachusetts, Amherst.

Goldhamer (1907–1977) was born in Canada and after earning B.A. and M.A. degrees from the University of Toronto and spending a year of graduate study at the London School of Economics (1933–1934), enrolled at the University of Chicago where he received a Ph.D. in sociology in 1938. Upon completing his doctorate, he was appointed assistant, later associate professor at Stanford University (1938–1946) before he returned to Chicago for another two years. He joined RAND in 1948 and remained there virtually for the rest of his life, until 1972 as a senior staff member, and afterward as resident consultant. He received some fame for being the only civilian to participate in the Korean peace treaty negotiations (cf. Robin 2001, 126). Given the importance of gaming at RAND and, more generally, the influence of parlor games on the development of game theory (Leonard 2010), it is significant that Goldhamer was an outstanding chess player. For four consecutive years (1925–1929), he won the championship of the University of Toronto's Hart House Chess Club.[14] In 1956, this passion also had him play against the later grandmaster Bobby Fisher (1943–2008), then only 13 years old, at the Eastern States Open in Washington, DC. Fisher, already famed for having won the "Game of the Century" against Donald Byrne (1930–1976) some weeks earlier, beat Goldhamer in 25 moves.[15]

The biographical data on Paul Kecskemeti (1901–1980, born Pál Kecskeméti), who has already been mentioned in the discussion of the impact his study *Strategic Surrender* had on RAND, is scarce and not very reliable. With this caveat in mind, it is secured that Kecskemeti was born on 31 October 1901, in Makó, Hungary. He moved to Berlin in the late 1920s and, after a short stay in London, allegedly managed to get one of the last ships leaving from Casablanca to the United States, where he arrived on 2 August 1942. Some sources (e.g., Reisch 2005) describe him as a student of philosopher Charles W. Morris. Speier remembered that the two became friends in Washington during the war effort, with Speier working for the State Department and Kecskemeti for the War Department.[16] Kecskemeti worked for RAND from 1948 to 1966 and afterward held various visiting professorships at Brandeis University and MIT. After the death of Karl Mannheim, who had been the husband of his wife's sister, Kecskemeti edited and partly translated three volumes of Mannheim's essays on the sociology of knowledge (Mannheim 1952), on sociology and social psychology (Mannheim 1953), and on the sociology of culture (Mannheim 1956). Apart from the widely received publication of *Strategic Surrender* (Kecskemeti 1958), he wrote a book on the Hungarian Revolution entitled *The Unexpected Revolution* (Kecskemeti 1961). The title of the latter book is telling. The situation in Hungary had been a topic in at least one of the political games in which Kecskemeti participated; yet, the revolution was not foreseen.

Victor Myron Hunt (1908–1965) trained as a historian in Berkeley, California. After teaching for ten years at San Jose State College, he entered government service in 1943. He joined the Office of War Information and worked there together with Speier. Some months later, he transferred to the State Department where he "played a leading role in developing American information policy with respect to the Soviet Union" (DeWeerd 1966a, 161; cf. 1966b). He joined RAND in 1948.

Another important, yet not always acknowledged inspiration for the shape of political gaming was Project Troy, in which Speier had participated. Starting in October 1951, Project Troy was a "brainstorming council" on propaganda and psychological warfare that was carried out in a joint effort by the Massachusetts Institute of Technology MIT and Harvard University (Nedell 1998). Apart from Speier, participants included, in alphabetical order: radio expert Dana K. Bailey, psychologist

Alex Bavelas, radio engineer Lloyd Berkner, psychologist Jerome S. Bruner, law professor Burnham Kelly, anthropologist Clyde K. M. Kluckhohn, psychologist Donald Marquis, economist Max Millikan, historian Elting Morison, geographer John A. Morrison, electronics expert John R. Pierce, physicist Edward M. Purcell, and electronics professor Jerome Wiesner (cf. Nedell 1998, 11f).[17]

After introductory briefings in Washington, DC, the participants of Project Troy met in MIT's Lexington Field Station northwest of Boston. They stayed there for the following three months, "interrupted only by a group sojourn to Washington to meet with [Secretary of State] Dean Acheson" (Nedell 1998, 13). Clearly, Project Troy was neither a simulation nor a game. It was a summit, a work meeting of people outside the government who were invited to confront a specific problem of foreign policy. Donald Marquis remembered the meetings to be "very cloak-and-daggerish. We met in an underground, windowless building in Lexington and I'd go home [to Michigan] weekends" (Oral history interview with Donald Marquis, Ford Foundation Archives, cited in Schwoch 2009, 185).

The problem Project Troy assessed was given by the State Department, which financed the project with a contribution of $150,000 (Nedell 1998, 11), a sum approximating $1,507,400 in 2019.[18] It concerned the fact that the Soviet Union had found ways to jam the Voice of America broadcasts; Project Troy should counteract by proposing measures for a propaganda offensive to get "the truth behind the Iron Curtain" (broadcaster Justin Miller, cited in Nedell 1998, 3). "Modeled after the Manhattan Project, Project Troy undertook an unprecedented interdisciplinary and collaborative approach to work between natural scientists, engineers, and social scientists" (Gilman 2003, 157).

The participants were free to decide about the organization of their work. They were divided up into panels, attempting to staff them with experts from various fields. These panels were assigned problem areas and had the task to produce outlines of solution approaches. The outlines were discussed by the whole group and forwarded to working groups which comprised predominantly, but not exclusively, persons with expertise in the area of concern. The reports produced by these working groups were again discussed by the whole group. Based on these preparations, an editorial committee was put in charge of drafting the final report, which again, was submitted for review by the whole group before finalization. It is important to note that in spite of the blossoming behavioral sciences creed to which many of the project participants felt attracted, the final report concluded that "a psychological superweapon, one that could apply to all countries and all segments of the enemy population, conflicted with empirical data" (Robin 2001, 45). Despite being united in their communist conviction, cultures differed, and the strategies to address various populations would have to take these cultural differences into account. "Communism in China and Southeast Asia does not constitute a simple extension of Soviet power. Mao in China and Ho in Vietnam are not automatic tools of the Kremlin […]" (Project Troy report, cited in Robin 2001, 45–46).

In the end, a report of 80 pages was sent to the State Department on 15 February 1951, where it was received very positively (cf. Nedell 1998, 13, 19f; Gilman 2003, 157). As regards how to counter the Soviet jams, the report singled out no specific tactics. Rather, the innovative feature of Project Troy's report was the comprehensive perspective on political warfare. As the authors of the report noted, "the newness of our idea […] lies in the understanding of the strategic power of the several elements [of political warfare] when combined as a well rounded and coordinated whole" (Project Troy report, cited in Nedell 1998, 14). This, it was argued, was a direct result of the interdisciplinary collaboration realized in Project Troy.

Project Troy was important for the history of political gaming in three ways. First, Hans Speier's active participation had convinced him that the interaction of experts from a variety of disciplines can result in a sophisticated, comprehensive picture that acknowledges the complexities of reality to a higher degree than any single analyst might reasonably do. As Bessner (2015) has convincingly argued, the interdisciplinarity embodied in political gaming was a strategy of the SSD to further its integration within RAND. Second, the positive reception at the State Department of this joint interdisciplinary effort encouraged Speier. Finally, Project Troy had a third effect which incidentally would later become relevant to the history of political gaming. After the project's completion, the administrators of MIT and Harvard announced a follow-up project. MIT president James Killian led the effort, aiming at "the establishment of a permanent research facility aimed at continuing and extending the interdisciplinary collaborative research exemplified and advocated by the Project Troy report" (Gilman 2003, 157). In January 1952, the MIT Center for International Studies (CENIS) was established, and, apparently after Speier declined an initial offer (cf. Schwoch 2009, 64), another former participant of Project Troy, Max Millikan, became its first director. Initial funding came from the CIA, but the Ford Foundation, which had recently established a Behavioral Science Program (cf. Solovey 2013, chap. 3; Crowther-Heyck 2005, 152–156), decided to allocate $875,000 ($8,450,120 in 2017) to the newly founded center.[19] As discussed in more detail in Chap. 6, CENIS was the organization that continued the development and use of political gaming after RAND withdrew from such endeavors toward the end of the 1950s.

The idea of using a *gaming* approach—as opposed to a game theory approach with a complete set of formal rules—to simulate a likely chain of events in the arena of the Cold War was first discussed in early 1954 in RAND's Washington office (cf. Goldhamer 1955a, 1). Participants of this discussion were Goldhamer, Joseph M. Goldsen, Hunt, and Speier. It was concluded that this idea should be followed by members of RAND's Social Science Division as soon as "some of the Division's immediate commitments permitted" (Goldhamer 1955a, 1). Over the following summer, Goldhamer and Hunt continued discussions on devising a Cold War game in Santa Monica, and finally proposed their idea to RAND director Franklin Collbohm. Collbohm agreed that pursuing the idea would be worthwhile, and in the autumn months of 1954, Goldhamer began to prepare materials that could form the basis of deliberation in the envisaged Cold War game. As a first step, he tried to put together a comprehensive analysis "of the current cold-war situation, of the major threats to which the U.S. was currently and prospectively exposed, immediate and long-range U.S. objectives, major policy alternatives, and the effect of weapon developments on these" (Goldhamer 1955a, 2). However, this analysis turned out to be beyond his capacities, and only "[l]ittle progress was made" (Goldhamer 1955a, 2). Thus, the preliminary analysis

> served effectively to throw into relief the many areas of expert knowledge which a proper analysis would require. In pursuing this analysis, it was pos-

> sible to see more clearly the manner in which collaborative activity in the framework of a cold-war game might serve to overcome some of the difficulties faced by the current attempt at analysis. (Goldhamer 1955a, 2)

The experience with his attempt to put together a comprehensive analysis of the status quo of American foreign policy led Goldhamer to put forth some methodological considerations on why and how an open form political game can be expected to deliver valuable results. These are collected in the first RAND-internal text on political gaming, a piece entitled "Toward a Cold War Game. D(L)-2603" (Goldhamer 1954). To Goldhamer, although the conventional division of labor in the military and foreign policy analysis could deliver stable and useful descriptions and diagnoses, it systematically failed when it came to strategic and tactical planning. Here, the interrelatedness of relevant factors called for the collaborative effort of a wide range of specialists. Yet again, as with Delphi, the question was how to organize the interaction of experts. Games, he claimed, offered an interesting option for several reasons. Gaming structures were likely to raise the motivation of the participating scientists and to stimulate creativity and innovation. Also, with reference to the conventional division of labor, he thought that gaming procedures might achieve an exchange across the specialties, thus permitting "a more effective collaboration, one which does not divide up the problem in a manner which obscures the issue" (Goldhamer 1954, 1).

Finally, the proposed game would also result in a higher degree of realism:

> It is likely that a high degree of realism can be more readily achieved by a cold war game than by an analyst. This is a surprising statement. Most gaming procedures introduce a variety of simplifying assumptions and special restrictions which have quite the opposite effect. This is not the case with the game proposed here. The plan of the game encourages the introduction of many real life details that an analyst might not deal with. (Goldhamer 1954, 2)

Quite in contrast to other approaches to gaming and policy analysis in general, the main thrust of the approach proposed by the SSD was not to achieve simplification but to be as omnivorous as possible with regard to factors informing the game. Of course, this was not to say that all factors were assessed as relevant. Nor was this to mean that the games could be carried out without any reduction of complexity. However, their argument was that in order to carry out games of war and conflict with a high degree of realism, one had to be as open and non-rigid as possible.

Building on this methodological viewpoint, the SSD proposal was to assemble a group of experts and order them into three groups: the players, the referees (or Umpires), and the so-called Committee on Nature. The players were the largest group. Their task was to represent national governments, thereby taking into account possible frictions between and amongst political and military leaders. While, in line with their claim to principle openness, a game could include all the countries affected by a specific crisis, it was decided that to "economize manpower," the proposed Cold War game would provide for only three teams—the United States, the Soviet Union, and "other governments" (Goldhamer 1954, 2).[20] Two or three experts would be selected to each of these groups and were given the task of representing the respective governments. They would be allowed to make any move they could reasonably justify as a plausible action or reaction of the government they represented. Further, they could also specify moves "that are not directly government actions but can reasonably be supposed to be produced by government decision and influence" (Goldhamer 1954, 3).

Apart from the players, the proposal provided for a "Committee on Nature." This committee was conceived of as a subgroup of the referees; however, unlike the referees, Nature was entitled to make moves on its own. Its responsibility was to introduce events independent from the government action, but also to make moves that were plausible consequences of the decisions of the players but not taken by them, be it because they were beyond the scope of (governmental) action or because the players hoped to avoid them. "The Nature Committee thus performs a vital function since without it reality would be reduced to government initiated action" (Goldhamer 1954, 4). Finally, referees were in charge of organizing the communication process: they were entitled to "transmit appropriate information to the players, judge the feasibility of moves, and evaluate the outcome of actions and counteractions that are undertaken by the government players" (Goldhamer 1954, 5). Together with the Committee on Nature, the referees would be "responsible for defining the situation at any point of the game" (Goldhamer 1954, 5).

Goldhamer suggested not to attempt to devise too many rules before the game but instead let the referees and the members of the Committee on Nature develop a set of rules in the course of the game. Clearly, the lack of fixed rules in advance of the game countered any attempt to conceive it as an experimental situation. The comparison of the results of several successive games was rendered difficult. "But this disadvantage is more than offset by the gain in realism; elements of the game can become formalized without endangering the objective of the game" (Goldhamer 1954, 6). The issue of realism was also addressed in "War Games and Political Games. D-2849," a research memorandum by Paul Kecskemeti (1955). Kecskemeti reported that it had been repeatedly argued that in comparison to war games, political games lacked stability. To him, however, this lacuna did not result from the superiority of knowledge available to run good operational war games. Rather, it was caused by the ontological difference of the subject matters. To make his point, Kecskemeti claimed that the notion of games comprised two semantic levels that had to be discerned properly. There was the meaning of a game as "play," and there was the meaning of game as "contest for real stakes," in the sense of a serious game. Consequently, Kecskemeti used the abbreviations "P-game" and "S-game." In his view, all approaches to gaming were P-games. However, war games and political games differed with regard to the formal correspondence between them and the S-games they ought to represent. In war games, on the one hand, "there is a far-reaching formal correspondence between war as an S-game and war games as P-games" (Kecskemeti 1955, 3). This allowed for a fair amount of abstraction without endangering the realism of the game. On the other hand, political P-games were (Kecskemeti 1955, 9)

> played against an existing S-game constellation as a background […]. In this P-game activity, the formal structure of one actual play of an S-game (politics) is preserved; it is not broken down into 'abstract' elements. But precisely because of this, the political P-game as such does not create the image of the political S-game as a whole [as does the war P-game], and hence its playing does not give proficiency in playing the game itself.

Thus, if the results from political games differed from those obtained from rigid war games, it was "not because less political knowledge than military knowledge is now available or because our political science is just too imperfect, but because the political S-game universe is different from the military S-game universe and the conditions of P-gameability are different" (Kecskemeti 1955, 13). However, Kecskemeti's argument of the ontological differences of the realities which the games attempted to simulate was not taken up by later commentators, who continued to see the lack of "knowledge" or "theory" as responsible for the difficulties in achieving both realism and replicability in games simulating political processes.

From the available materials, it appears that the development of political gaming was a cautious process. The contributing scientists—Goldhamer, Speier, Hunt, Goldsen, and Kecskemeti—were apparently not determined to move forward at any rate. Rather, they appear to have groped their way through the forest of factual and organizational intricacies. Documents of repeated discussions with RAND director Collbohm and lengthy preparatory memoranda debating the potential of such an approach testify to the hesitation with which the social scientists proceeded. In reality, their approach was novel, and their game was "one of the earliest post-World War II games conducted in the United States" (Hermann 1968, 275). But quite in contrast to the typical way of approaching things at RAND, there is no sign of a light-hearted trial-and-error philosophy. They were aware that their proposal was unusual for RAND—and at the same time, they were determined to ensure that its status as a proper research procedure was acknowledged by relevant others. The stakes were high for the social scientists: they wanted to prove their division's value for RAND, and political gaming appeared to provide a means for that.

Players, Nature, Referees: Three Rounds of Political Gaming, 1955

Bolstered in repeated briefings and conversations, the proposal outlined by Goldhamer (1954) was accepted by RAND management and plans were made to further develop and finally test the proposed procedure during the first half of 1955 (cf. Goldhamer 1955a, 3). In February 1955, after some preparatory sessions, the first game was played in Santa Monica. Apart from the already mentioned Speier, Hunt, and Goldhamer, the participating scientists were Andrew Marshall, Paul Kecskemeti, Nathan Leites, Harvey DeWeerd, and Eleanor Sullivan Wainstein. Joseph Goldsen, who had contributed to earlier discussions in the Division's office in Washington, did not travel to Santa Monica. Apart from Marshall, who was a member of the Economics Division, all participants were members of the Social Science Division. "It was hoped that later an opportunity would arise for the participation of other members of the Social Science Division, and that at a still later period we would be in a position to arrange for the collaboration of additional persons in other divisions" (Goldhamer 1955a, 3).

In the first sessions in February, Goldhamer's proposal was presented and assessed. It was then decided not to enter the game at this stage, but instead to discuss openly some of the issues the participants expected would almost certainly arise in the course of such a game. The purpose of this delay was that, on the one hand, the participants would get an impression of how substantive issues could be integrated into the proposed game. On the other hand, such a phase of open discussion would allow the participants to develop some steps of the analysis required later in the game, with the effect that in the game, decisions could rely on their results. This phase of open discussion, however, consumed most of the time allocated for the game. Participants openly discussed a series of issues related to the playing of such a game—for example, the debate on limitation or non-limitation of nuclear weapons then of topical interest to international policy-makers (cf. Goldhamer 1955a, 4). On the request of the participants, Goldhamer prepared "a classification of the threats and dangers which U.S. foreign policy was required to meet" (Goldhamer 1955a, 4) which would serve in steering the discussion. With this classification at hand, the participants decided to focus on only one specific class of conflicts. Furthermore, it was decided that the game should be restricted to only one arena (or theater) and that this arena should be selected along the available knowledge of the participants. The decision was that this arena should be Western Europe. The aim was

> to arrive at a statement of those factors which were likely, on the one hand, to increase and, on the other hand, to decrease the likelihood of the S.U. making a decision to attack Western Europe […] In the light of the factors increasing or decreasing the likelihood of Soviet military attack on Western Europe, we then wished to specify what U.S. policies seemed appropriate in order to give greater force to the inhibiting factors and less force to the inciting factors. (Goldhamer 1955a, 6)

The final list of factors on which the participants agreed is given in Table 4.1. As such, this enumeration provides one of the two major outcomes of RAND's first political game. Each factor was intensively discussed, sometimes controversially, but an agreement—the authors did not use the term consensus—was reached on all of them as well as on their relative importance as regards decreasing or increasing the risk of a Soviet attack.

The Soviet willingness to attack was conceived of as an immediate product of the listed factors. As Goldhamer (1955a, 11; emphases in original) puts it, the game participants "found that the maintenance of a U.S. nuclear strategic threat and capability was a dominant factor inhibiting the Soviet Union and that this factor was capable of compensating for the combined effect of several other factors that might move the S.U. toward aggression." However, Goldhamer added, the participants acknowledged that the mere fact that the United States was in the possession of nuclear weapons was not sufficient—rather, the SU would also have to be convinced that this is the case. The question resulting from this analysis was how to foster the Soviet belief in American nuclear capabilities.

In a next step, the participants decided to define two options for each of these factors, a low-cost variant (L) and a high-cost variant (H), always speaking from the perspective of the Soviet Union. For instance, when the nuclear ability of the United States (factor 2) was low or inexistent, this factor would cause only low costs, which in turn increased the likelihood of a Soviet attack. This resulted in eight possible situations ("cases"). Factor 6c—harmony at the Kremlin—was omitted because participants perceived this to be a prerequisite to Soviet aggression. The resulting

Table 4.1 General and specific factors influencing Soviet policy. (Reproduced from Goldhamer 1955a, 7–8, with permission of the RAND Corporation)

I. General motivation	Power	Denial of Western Europe to the West Gain of Western Europe to the SU

II. Specific factors governing Soviet calculations

1. *Strategic use of nuclear weapons*
   US agreement to the proscription of strategic use of nuclear weapons would increase Soviet willingness to accept risk and cost of attack.
2. *Nuclear tactical ability*
   A limitation on nuclear tactical weapons or a low capability in the use of them would increase Soviet willingness to attack.
3. *Non-nuclear capability immediately and prospectively available in Europe*
   The higher this capability, the less the likelihood of an attack. We recognized here and in point 2 above, however, that an impending future large-scale increase in these capabilities can provide a motive to attack before they are attained […].
   [T]his point suggested that a distinction be made between offensive and defensive military capabilities (to the extent that some weapons can be distinguished in this manner): the political consequences of these capabilities differ. Thus, for example, a belt of nuclear mines is obviously more defensive, and therefore less provocative, than deliverable nuclear weapons.
4. *European political receptivity*
   The greater the likelihood, in the event of a Soviet attack, of sabotage or civil disorder in the Allied countries, the greater the likelihood of Soviet attack. Here again we recognize, however, that, if Western Europe were to move politically leftward, this would inhibit Soviet attack, since the likelihood of a cheaper acquisition of Europe would be in prospect.

5 (a) *US involvement elsewhere*
   The more committed and the more preoccupied the United States is elsewhere, politically and particularly militarily, the greater is the likelihood of Soviet attack.
   (b) *US indifference to Europe*
   The more the United States appears to have written off Europe, the greater is the likelihood of Soviet attack. Similarly, any breach between the United States and its European allies would operate in the same direction.

6 *The SU internal situation*
   (a) *The military domain*
   Included here are a variety of points, for example, military loyalty and morale, existing military capabilities, commitments elsewhere.
   (b) *Civil domain*
   Included here are, among others, civilian loyalty and morale, the state of the economy.
   (c) *Harmony in the Kremlin*
   It is held unlikely that the SU will engage in a war at a moment when a high degree of internal dissension exists at the top levels of the leadership. This, of course, does not preclude the possibility that a drive toward war may itself produce disharmony.

Table 4.2 Matrix of the likelihood of Soviet aggression. (Reproduced from Goldhamer 1955a, 12, with permission of the RAND Corporation)

	Factors						
Case	1	2	3	4	5	6a	6b
1	L	L	L	L	L	L	L
2	L	L	L	H	L	L	H
3	L	L	L	L	H	L	H
4	L	L	L	H	H	L	H
5	L	L	H	–	–	L	H
6	L	H	L	–	–	L	H
7	L	H	H	–	–	L	H
8	H	H	H	–	–	H	H

matrix (see Table 4.2) informed about what consequences the participants expected to arise in case one of these specific (future) situations would become real or, more precisely, how the risk of Soviet aggression was expected to change in case one or more of the enlisted factors were changed.

For instance, consider case 4 which denotes the situation that factors 1, 2, 3, and 6a put the only low cost on a Soviet attack (and are thus highly incentive) whereas factors 4, 5, and 6b require high costs. In such a situation, what spoke in favor of a Soviet attack was the United States' agreement to limit the use of nuclear weapons to strategic purposes (factor 1); the US military having shown inability to use the available nuclear weapons (factor 2); a lack of other military forces in the Western European theater (factor 3); and a high military potential and determinedness on the Soviet side (factor 6a). Countering these incentives for aggression, what spoke against a Soviet attack was the high probability of resistance from the population of Western Europe (factor 4); the minor involvement of the United States in other conflicts and a demonstrated high interest in the fate of Western Europe (factor 5); and the internal political situation of the Soviet Union which would make it hard for Soviet leaders to explain the aggression to their own population (factor 6b).[21]

Only after clarifying all these issues, the participants felt "that we had reached the point where our discussion had provided us with some useful themes for cold-war gaming and had clarified some of the gaming procedures themselves" (Goldhamer 1955a, 14). Starting in the second half of February 1955, the game "was played in a very informal and tentative fashion" (Goldhamer 1955a, 14). Goldhamer, Hunt, Marshall, and DeWeerd formed the US team; DeWeerd additionally played Great Britain. Leites played both the Soviet Union and France, Speier played Germany, and Kecskemeti the "rest of the World." Further, Speier and Kecskemeti took over the tasks of the referees and the nature committee. The role of the German, British, French, and SU desks in the State Department was played by Speier, DeWeerd, and Leites.

The game started from the actual date of the first session, which "meant that the state of the world as defined in the game was the real state of the world as we understood it to be from both classified and unclassified sources" (Goldhamer 1955a, 15). Unlike in later games, moves were not written on paper and handed to the referees for inspection. Rather, participants discussed openly. They did not attempt to have some form of secrecy, which meant that all moves and all justifications were known to all the participants all the time.[22] "In the first, very informal exercise almost all the work of the exercise was done in full sessions of the participants, the moves being transmitted orally and accompanied by considerable analysis as we went along" (Goldhamer 1955b, 7). As a consequence both of the informal and open character and of the decision to have an extended phase of discussion prior to the start of the game, there was little success in "mov[ing] much ahead of the time point at which the exercise began" (Goldhamer 1955b, 7).

Based on these first experiences with the new methodology, a second political game was played in May 1955 in Washington, and a third game again in Santa Monica, taking a full month, from 11 July to 11 August 1955. From the second game onward, participants were required to make their moves in written form and to justify them; open discussion and critique of moves as well as the plenary evaluation of the overall situation took place in occasional plenary sessions during the game and after its conclusion.

Again, the participants of the third game were mostly members of the Social Science Division (Schnitzer 1955, iii). The US team consisted of Joseph Goldsen, Victor Hunt, and Paul Kecskemeti. The Soviet government was represented by J. Gliksman, Léon Gouré, and Andrew Marshall (Economics Division). Harvey DeWeerd and Ewald Schnitzer made up the European team. Herbert Goldhamer, Abraham M. Halpern, and

Hans Speier were the referees. Additional consultancy came from Oleg Hoeffding (Economics Division) and Herman Kahn (Nuclear Energy Division). In a side play on the "Formosa question," Halpern played the Asian part.[23]

Furthermore, an innovation was introduced in the third game, namely the "international conference": the teams could arrange meetings with some other teams to discuss specific issues. However, it was not clear how the results of these conferences could effectively be integrated into the currently running game without losing too much time (cf. Goldhamer 1955b, 7). Game time started on the day when the game itself started and proceeded for eight months, so that the analysis covered the period from August 1955 to March 1956.

Again, the game focused on the European theater. The US team decided that it would want to put to test the current framework of US American policy for Europe, namely the maintenance and strengthening of North Atlantic Treaty Organization (NATO). The Soviet team continued the strategy it perceived as being in current operation. In the course of the game, a series of specific events took place of which Goldhamer (1955b, 17ff) perceives eight worth mentioning:

1. A four-minister conference on German reunification and a Western security system. Neither side was willing to concede the points of major interest to the other side, and the conference was a standoff, although both sides were anxious to issue an amicably worded joint communiqué. This conference had been preceded by a visit of Adenauer to Moscow, which had been equally indecisive.
2. A unilateral SU demobilization of seven divisions and the withdrawal of a corresponding number of troops from East Germany. At the end of the exercise the SU was preparing to announce further demobilizations and also the withdrawal of troops from the Satellites.
3. A unilateral SU declaration that it would not employ nuclear weapons unless they were first used by another power. The SU made strenuous but so far unsuccessful efforts to enforce a similar declaration from the West. At the same time, the SU tested three thermonuclear devices in the five- to twenty-megaton range, and demonstrated short-range missile prototypes in Red Square.

4. The SU announced the dissolution of the Cominform and moved toward creating a worldwide front of left groups, communist and non-communist. In secret, the Cominform continued.
5. The SU instigated a European economic conference as a prelude to further attempts to increase East-Western trade. During the conference the Western European powers requested free access to Western trade delegations to Iron Curtain countries.
6. The United States agreed to give Germany naval units within the limits imposed by the Paris Pact.
7. A four-power disarmament conference led to no positive results but also ended with an amicably worded joint communiqué. Neither party was willing to concede the minimum requirements of the other. Among the proposals advanced was one by the United States to establish internationally manned radar warning stations located along a sea frontier in suitable areas between the United States and SU.
8. The United States, through the mediation of the United Kingdom, arrived at an agreement with Communist China and the Nationalists that provided for: a five-year cessation of hostilities between the latter two; a plebiscite at the end of this period; admission of Communist China to the UN Assembly; annual election of an Asian power to Nationalist China's current seat on the Security Council.

The interpretation of these results, agreed upon by the participants, was that the Soviet Union had not considerably improved its position within the eight months covered by the game. "Its only important immediate success, however, was its progressive alienation of Yugoslavia from a Western or neutralist orientation" (1955b, 20). Furthermore, the Soviet Union (in the game) had declared publicly that it did not intend to use its nuclear capacity in an aggressive manner.

The Fourth Round, April 1956

After the third game, it was discussed whether or not to continue with organizing such games. Although the members of the Social Science Division saw that the potential of the method had, mostly for procedural and/or technical reasons, not yet been fully explored, they were also aware of the considerable efforts required to conduct a further gaming exercise, especially if it was to counter the shortcomings of earlier rounds. Yet, the decision was to give it a go. "The immediate objective for having a Fourth Round was to enable us to reach a more definitive judgment on the worth of the gaming technique" (Goldsen 1956, 2).

The game took place in April 1956 and was organized like a summer school. Participants stayed in Santa Monica for the whole time, so that unlike in previous games, "all team members devoted practically full time to the exercise" (Goldsen 1956, 6). It took more than three weeks of intensive playing and, in terms of man-months spent, was the largest of the series. As with the earlier games, it dealt with the "activities of the United States and the Soviet Union with respect to each other and to Western Europe" (Goldhamer and Speier 1959a, 75). Participants produced 150 papers in total (Goldsen 1956, 11). In addition to the RAND-internal report by Goldsen (1956), W. Phillips Davison compiled a "summary of RAND's experiments in 'political gaming' [...] prepared for Mr. Henry Heald, President of the Ford Foundation" (Davison 1958, ii) from which RAND had received financial support.[24] The results of RAND's experiences with political gaming were further summarized in what appears to be the best-known paper from this series, namely an article cowritten by Goldhamer and Speier, "Some Observations on Political Gaming" (Goldhamer and Speier 1959a).[25]

The fourth round differed from the earlier ones in various respects. Most importantly, it did not recruit its participants exclusively from RAND. Speier and his colleagues decided to invite external experts to join RAND experts. Prior to the start of the game, leading members of the Foreign Service Institute (FSI), the US Department of State's diplomatic training institution, had shown interest in learning about the game. Fascinated by its potential applicability as a training device, the FSI arranged for the participation of three officers (cf. Goldsen 1956, 4ff). Partly as a consequence of this decision, the teams were considerably larger than in the previous exercises. Together with Andrew Marshall and Hans Speier, the Dean of FSI's School of International Studies, Albert B. Franklin, formed the referee team. The United States was represented by Joseph Goldsen, Victor Hunt, and

Jeffrey C. Kitchen, an officer from the US Department of State. The Soviet Team comprised another Foreign Service Officer, Edward Page, Jr., as well as Raymond Garthoff and Robert Tucker from RAND's Social Science Division. Finally, the Western European Team consisted of Harvey DeWeerd, Paul Kecskemeti, and Nathan Leites.

In addition, the game provided for a set of RAND experts that were available to the teams to tackle questions that went beyond their specialties. The list of these Consultants on Special Problems comprised Lewis Bohn, Abraham Halpern (both from the Social Science Division), the economists Malcolm W. Hoag, Oleg Hoeffding and Charles Wolf, Jr., and the physicists Herman Kahn and Arnold Kramish. Ewald Schnitzer operated as a Special Assistant. In addition, "[t]he players were assisted by a sizable secretarial staff" (Goldhamer and Speier 1959a, 74).

One lesson drawn from the previous rounds was to disentangle game time from real time. Instead of starting game time on the date of the opening session of the game, the game organizers decided to use a scenario which defined as the start time of the game a time point in a not too distant future but sufficiently remote from the present as to inhibit actual events from interfering with the game. The scenario of the fourth game was written in March 1956 and described a possible state of the world of 1 January 1957. The scenario thus allowed for a higher amount of control, which engendered two improvements. It "rid them [the participants] of the intrusion of current news into the game and served to focus it on problems of special analytical interest" (Goldhamer and Speier 1959a, 74). While the scenario was some sort of extrapolation from the state of affairs at the time of writing, and thus was deemed to be plausible to the expert, the game organizers also used this increased control capacity with the intent to direct the game: "A certain degree of plausibility was worked into the scenario, but some events were introduced primarily in order to provide challenging problems for action and analysis" (Goldsen 1956, 7).

What did the scene look like? From a global perspective, no wars were to begin. In the United States, the favorite Republican candidate, President Dwight D. Eisenhower, was to be elected for a second period, and the scenario described the atmosphere in US foreign policy with the following words (cited in Goldsen 1956, 8):

> Serious U.S. observers note the passing of the year 1956 with considerable uneasiness. There is greater confidence that total war is rejected by all major powers as a calculated act. The fear of local war abates, partly because no U.S. forces are involved in war and partly because it is known among interested persons that the U.S. government is making a real effort to develop a force suitable for peripheral wars.
>
> Whereas the military strength of the U.S. makes some gains, commentators do not spot any significant political tide turning in favor of the West. France's stability leaves much to be desired. The German scene is of greater concern than it was in 1955. Nationalism and anti-colonialism in Africa and Asia offer the Soviets more opportunities for activity than are seen or seized by the U.S. government. The policies of deterrence and containment are seen as more on the defensive than before, while rollbacks of communism appear less possible. With 1956 being an election year, the U.S. Congress and Administration are careful to avoid risky initiatives in foreign affairs lest domestic political opposition be aroused. The uneasy peace continues as the new year dawns.

The Soviet Union, in the perspective of the scenario writers, made no overly aggressive move in the nine months covered by the projection. It continued to modernize its armed forces and to pursue a moderate foreign policy.[26] Further, it was assumed that throughout 1956, a war between Arabs and Israelis could be avoided, but that the whole region was characterized by high political instability.

The scenario was distributed prior to the beginning of the game. Thus, lengthy and retarding discussions about the state of the world at the beginning of the game, as had happened in the first game, were avoided. Similar to the earlier games, all teams except the US team were told to make only the most plausible moves. They "were expected to behave according to their best judgments of how in fact the governments they represented would behave in the context of the events developed in the game" (Goldsen 1956, 10). In contrast, the US team had more freedom to choose strategies that at first sight appeared unlikely to be those taken in real life. They were not required to follow what they perceived as being the most plausible alternative but instead could test several strategies. Although their moves were tested for plausibility by the referees, they were deliberately and openly less strict. A paper assembling these and further rules was distributed among the participants prior to the start of the game.

In principle, players could make two kinds of moves, open moves and "game classified" moves. Open moves were announced to all participants after inspection by the referees. Game-classified moves, on the other hand, were the (Goldsen 1956, 11)

> equivalents of private or secret diplomatic exchange or of intelligence. The Referees, in their capacity as controllers of information, could "leak" the contents of "game classified" papers, in whole or in part, accurately or in distorted form—thus serving as surrogates for the intelligence function in the political process or for actual "leaks" of classified information in the free press.

Moves had to be put down on paper, "dated in game time and numbered sequentially with reference both to the action of each team and to the game as a whole" (Goldhamer and Speier 1959a, 75). They were then checked by the referees for plausibility; the referees could also ask for further clarification or justification. In turn, the decisions of the referees were also open to criticism and revision by the game participants. As it turned out, the majority of the 150 papers produced by the participants were open moves; sometimes, game classified papers providing background information on the policy calculations behind the moves accompanied them. The events simulated in the three weeks of playing reached far into the summer 1957 game time. After the game had finished, all records were thrown open for inspection by the participants, who then assembled in "assessment sessions" to discuss the outcomes and procedural features of the game.

Initially, the social scientists at RAND had embarked on the political gaming exercises to find answers to a set of questions which the final report (Goldsen 1956, 1f; cf. Brewer and Shubik 1979, 61) summarizes as follows:

1. Would a political "game" be a useful technique for generating forecasts of political developments and for sharpening our estimates of the probable consequences of policies pursued by various governments in international affairs? Is gaming a useful way to test the comparative worth of political strategies and tactics?

2. Would the partial simulation of reality in the game, in which participants try to play the role of responsible government officials, provide a stimulus to political inventiveness? Does the game situation intrinsically foster a stronger motivation for imaginative political thinking than other forms of political stimulation [!]? Would a game approach add significantly to the quality of political analysis over and beyond that which might be achieved by the more conventional techniques of research and analysis?
3. Would the game device serve to call sharpened attention to those problems of international politics in need of special research and further study? Would these problems be elicited in a clearer and more useful way than might otherwise be the case if one looked for worthwhile research projects outside of the action context imposed by the game?
4. Would the game procedure be especially useful for educational and training purposes? Would it substantially help research people to acquire a heightened sensitivity to problems of political strategy and policy consequences? And would it be useful for giving policy-makers a means for analyzing more deeply the implications of events in the context of an unfolding political process?

The answers of RAND's social scientists to these questions were partly positive, partly dismissive (cf. Goldsen 1956; Goldhamer and Speier 1959a). As regards the first question, the social scientists were skeptic that the proposed procedure could be used for testing political strategies. Although the third and the fourth games took several weeks each, the time was not sufficient to thoroughly develop strategies or sensibly compare several strategies against each other. "Sets of policies were enunciated and several important problems were worked on, but the three weeks of play were far too brief a time for a 'test.'" (Goldsen 1956, 31) Goldsen concluded that while the SSD now felt safe to claim that the technique could theoretically be used to compare various strategies, practical matters made this more or less impossible. Referring to a huge collaborative effort in tactical and strategic simulation undertaken by RAND at that time, SIERRA, he stated that

> [f]rom a practical standpoint we have grave doubts about the wisdom of attempting a program of the requisite scale involving a relatively prohibitive commitment of manpower and expenditure of other resources. Since we must consider the marginal utility of our resources we have concluded that it is beyond our capacity to embark upon a program of the magnitude of SIERRA, or quite possibly larger. It is our judgment that the results which might be achieved would not be worth the sacrifice of the time, money, energy, and neglect of other pressing problems. (Goldsen 1956, 32; emphasis in original)

Nonetheless, several instances could be observed where the real developments did "'confirm' a number of the statements made or propositions taken in the game in the sense that reality has 'imitated' the play" (Goldsen 1956, 33). However, Goldsen warned against taking these "confirmations" too seriously. Quite a few of them, he said, were not a proof of the players' capacities, but rather of the rootedness of the future in what can be observed in the present.

When assessing the second question, the social scientists argued that the potential of political gaming with regard to inventiveness should be assessed in comparison to the situation where a single analyst—or a team of analysts collaboratively—produces an evaluation of a specified set of strategies. While being reluctant regarding the stability of results, the social scientists were nonetheless convinced that "the game does provide some testing of a strategy prior to the test made by history itself" (Goldhamer and Speier 1959a, 78). It did so by restricting the influence of the individual analyst. The game provided a structure of analysis that traded the freedom of the individual analyst for a higher degree of realism. In a game, the outcome emerged from the interaction of various ideas, thereby moving the control away from the single analyst. The analysis itself is thereby transformed into a collaborative endeavor, and it was this feature that justified the inclusion of political gaming in the analyst's toolkit.

Finally, questions three and four were answered positively. The involved social scientists saw strong evidence that gaming led to an increased political sensibility. They reported that this was a widespread impression among the participating experts. "Seeing new inter-connections of earlier insights or the acquisition of factual knowledge seems to have been considerably fostered by the game and by the psychological impact of the game context on the productivity of the players" (Goldsen 1956, 36). This was especially the case, Goldsen reported, with the participating officers from the US Department of State, who "were particularly emphatic in their belief that much had been learned" (Goldsen 1956, 36). Considering the use of political games in academic training, Goldsen cautioned that the same educational value would probably not emerge in academic training: "We are convinced that the educational value of the RAND Exercise is largely due to the fact that very considerable area knowledge and political sophistication are brought to bear" (Goldsen 1956, 37). Nonetheless, students could profit from both the

> lively setting in which students of politics, acting as observers or apprentice participants, can learn a good deal about the structure of the contemporary political world and about some of the reasons behind political decisions. Factual information takes on a new interest and importance when it is required for intelligent participation in the game, and political principles assume special significance when they are illustrated by political actions and situations with which the student is associated as a participant. (Goldhamer and Speier 1959a, 79)

A further "educational benefit" realizable by political gaming could be seen in its potential "to give players a new insight into the pressures, the uncertainties, and the moral and intellectual difficulties under which foreign policy decisions are made" (Goldhamer and Speier 1959a, 79).

Considering the massive financial and administrative effort required to carry out political games, the SSD finally suggested not to pursue this approach any further. RAND stopped with political gaming after this fourth round and left it to other organizations—most notably a group directed by Lincoln P. Bloomfield at the MIT—to further develop this approach (see Chap. 7).

The Surplus of Direct Interaction

The political gaming structure developed by RAND's Social Science Division in the 1950s received wide acclaim. In hindsight, Brewer and Shubik (1979, 61) contended that the game's "place in the development of manual or free-form gaming, both within Rand and in the larger gaming community, is central." True, political gaming was novel, and RAND—as a place where the concepts of game and gaming were already held in high currency—turned out to be the fertile soil required to turn an invention into an innovation. Yet what kind of innovation political gaming represented was not so clear. A constant cautioning against any premature labeling of political gaming as "scientific," "reliable," or even "predictive" research procedure permeated the papers produced by the inventors of political gaming. Summarizing RAND's experiments with political games, Goldhamer and Speier (1959a) maintained that the principal difficulties encountered by political gaming are those inherent in all applied social scientific work: To predict future developments in the social life on the basis of currently available empirical and theoretical knowledge was impossible.

But, granted this, the authors continue, "there still remains the problem of making the most effective use of any given level of empirical knowledge and theory" (Goldhamer and Speier 1959a, 77). And political gaming was a method which attempted to make use of available knowledge. It did so by setting in interaction the knowledge brought to the game by the various experts. Expert opinions crucially informed the discussions both within the smaller group and in the plenary sessions. But expert opinions were certainly not on the same epistemological level as empirical data or theoretical propositions. Especially in comparison to Delphi, there was a high agreement amongst the practitioners as to the purposes for which one could use political gaming, and which expectations the technique would probably never be able to meet. As a consequence, the shape of the technique and the methodological reasoning that accompanied its uses show a high degree of continuity: the paradigm, so to speak, was stable from the very beginning. The caveats put forth by Goldhamer in the first paper on political gaming, written even before the first game had been carried out, are essentially the same as the caveats mentioned later by Lincoln Bloomfield and his colleagues at MIT.

With the Delphi inventors, the developers of the political game agreed that *experts are valuable sources of information* in policy analysis. Both the proponents of Delphi and of political gaming argued that the implicit knowledge of experts should be used to supplement available knowledge and to point out implicit, non-formal relations and interdependencies between empirical and theoretical knowledge. Both techniques of prospection represented attempts to pool expert knowledge and expert opinions in a way that transgressed the limitations involved in listening to single experts. However, the expectations toward the expert, and consequently their epistemic roles, differed considerably (cf. Dayé 2014, 2016). Delphi implied that the success of the method increased with the amount of control the study leader had on the interaction amongst the participants. If possible, any direct interaction had to be avoided, because social-psychological processes of mutual exchange and adaptation potentially distorted the individual expert's opinion. Political gaming, on the other hand, deliberately used *face-to-face interaction as a means to achieve a more realistic outlook*. Thus, questionnaires were not deemed useful methods for gathering data, but group discussions were.

The first Delphi study had claimed that the expert was able if not to predict, then at least to forecast—that is, to evaluate how likely a statement on the future appeared in view of a set of evidence. Political gaming also produced perspectives on the future, but from an epistemological standpoint, these perspectives were different. Rather than producing a forecast, political gaming resulted in a collective version of what psychologists call foresight, that is, the mental capacity of humans to imagine a future as an aid to decision-making in the present (Suddendorf and Corballis 2007; Bulley et al. 2017). Following Mallard and Lakoff (2011), this collective version of the subjective foresight can be called *prospection*. The perspective emerging from political gaming was marked by higher uncertainty than the one resulting from Delphi; yet, it helped to orient decision-making in the present, and it did so within a collective, which in turn diminished the uncertainty.

Notes

1. A comprehensive and systematic treatment of the issue has been produced by Martin Shubik (1975).
2. In the popular use of the term, war gaming can refer to a rather diverse range of unrelated activities. The term is used to denote lay actors who,

dressed in historical uniforms, reenact famous battles (e.g., cf. Thompson 2010), or the activities of the owners of tin-soldier armies. In our case, war games are simulations that aim to provide insights into problems related to operations on a battlefield for testing tactics and thereby supporting military decision-making (Weiner 1959). Similar procedures can also be used for the scientific reconstruction of historical battles (e.g., see Sabin 2009) and for educational purposes (Sabin 2012).
3. The history of game theory has been the subject of a range of studies from scholars with diverse backgrounds. One important source is the collection of essays edited by Roy Weintraub (1992); more recent contributions to the intellectual history of Cold War game theory include Erickson (Erickson 2015) and Amadae (2016).
4. The range and reverberation of the work of RAND-affiliated game theorists can for instance be assessed by browsing the bibliography of the first textbook-like monograph publication in the field, Luce and Raiffa (1957), or the collection of essays edited some years later by Shubik (1964). A similar listing of prominent game theorists in William Poundstone's popular book on game theory, *Prisoner's Dilemma* (Poundstone 1993, 94), wrongly names Anatol Rapoport. I thank Andreas Diekmann, Zurich, for pointing out that Rapoport never had an affiliation with RAND.
5. There is an anecdote on John Williams offering a RAND consultant contract to von Neumann. In a letter to von Neumann, Williams described his expectations as follows: "In practice I would hope […] that members of the Project [RAND] with problems in your line (i.e., the wide world) could discuss them with you, by mail and in person. We would send you all working papers and reports of RAND which we think would interest you, expecting you to react (with frown, hint or suggestion) when you had a reaction. In this phase, the only part of your thinking time we'd like to bid for systematically is that which you spend shaving: we'd like you to pass on to us any ideas that come to you while so engaged" (Poundstone 1993, 94).
6. JOHNNIAC is said to be a tongue-in-cheek acronym for John v. Neumann Numerical Integrator and Automatic Computer. The machine is based on the architecture von Neumann developed for the computer of the Princeton Institute for Advanced Studies (Gruenberger 1968; Campbell 2004; Akera 2007; Gerovitch 2002).
7. Certainly the most sophisticated overview of RAND gaming activities is Brewer & Shubik (1979, esp. 59–66). The authors discern five types of

games and simulations: (1) political, diplomatic, military, and crisis games; (2) strategic games; (3) tactical games; (4) logistics games; and (5) applications of game theory.

8. On the next page, one finds a nice example for the state of the art in research ethics at that time. "A series of trials will be made, in some instances, in which some subjects will repeat. Such a series enables the experimenter to condition his subjects. It also provides information concerning the dynamics of the group process. 'Stooges' may be used to give the experimenter stronger control over the subjects. Drugs, hypnosis, and surgery could eventually be employed for similar purposes. It may be instructive to make some trials with rats, or pigeons, as well as with normal and abnormal human subjects" (Flood 1951, 2).
9. On the series of SAFE games played at RAND (or with RAND involvement), see also Helmer and Brown (1962), Brown and Paxson (1975), and Brewer and Shubik (1979, 103–106).
10. One of the games developed at RAND reportedly relied on the IBM 704 computer and some predefined algorithms (Mood 1954, 10).
11. In *The War Game*, sociologist Irving L. Horowitz made a similar point: "A major difficulty with the thinking of the new civilian militarists is that they study war while ignoring politics. […] [I]n the universe of cold war we are confronted precisely with politics. The difficulties of United States foreign policy in relation to those emerging nations undergoing rapid social and economic changes stem, in the main, from nonmilitary causes. The difficulties of Soviet foreign policy in relation to such ostensible socialist allies as China and Yugoslavia also arise from nonmilitary sources. Hence, the settlement of world problems cannot very well be made in terms of a 'delicate balance of terror' [a concept developed by RAND's Albert Wohlstetter] simply because this mode of analysis happens to nicely fit a two person, zero-sum framework supplied by game theory. Quite on the contrary, if von Clausewitz was correct in asserting that war implies politics by *other* means, then it must also be the case that peace implies policies by *political* means" (Horowitz 1963, 60–61; emphases in original).
12. Apart from Speier's autobiographical writings, an important source is the interview conducted with him by Martin J. Collins in Speier's home in Hartsdale, NY, on 5 April 1988, as part of the Smithsonian Institution's RAND Oral History Project, Box 10, Folder 14; see http://sova.si.edu/record/NASM.1999-0037 (last visited 25 April 2017).

13. Another collaborator on the project was political scientist Harold Lasswell who had written his dissertation on propaganda during World War I. When the US entered World War II, both Speier and Lasswell went to Washington, where Lasswell served as Chief of the Experimental Division for the Study of War Time Communications at the Library of Congress (Herman 1995, 24, 32, 160).
14. Cf. https://harthousechess.com/history/ (last visited 23 August 2019).
15. The game can be inspected at http://www.chessgames.com/perl/chessgame?gid=1044407 (last visited 23 August 2019) and https://www.youtube.com/watch?v=CyFma_qk-J0 (last visited 23 August 2019).
16. Transcript of interview with Hans Speier by Martin J. Collins, 1988, Smithsonian Institution RAND Oral History Project, Box 10, Folder 14, p. 60. One thus can conclude that even though they both lived in Berlin at the same time, they did not meet there.
17. Professionals who could not arrange their participation included McGeorge Bundy, George Kennan, and Robert K. Merton.
18. Here and below, the current equivalents were calculated using http://www.dollartimes.com/inflation/dollars.php.
19. The story behind the founding of Ford's Behavioral Science Program (BSP) is a wonderful example of the density of the personal networks that defined Cold War social science. The BSP was instigated by a report written for Ford by a committee led by H. Rowan Gaither, the San Francisco lawyer and long-term trustee of RAND, in 1949 (cf. Solovey 2013, 112–19). Internal discussions at Ford continued, and in early 1951, the foundation's new president, Paul Hoffman, asked Gaither to implement the report's suggestions and create a program for behavioral science. Gaither agreed, and asked three persons to join him: psychologist Donald Marquis, who had already been his co-author for the 1949 report and also was a Project Troy veteran; Bernard Berelson, a public opinion researcher who would later become the head of the BSP; and, finally, another Troy veteran, Hans Speier, for whom Gaither arranged a four-months leave from RAND with a monthly salary of $1416.67 ($14,240 in 2017) plus social expenses (letter from Goldstein to Gaither, 15 June 1951, HSP Box 9 Folder 17). It was this group who decided to support the founding of CENIS. Already prior to Project Troy, Speier reportedly had close ties to the Ford Foundation, fostered in part by the fact that a considerable proportion of the foundation's activities were steered from their offices in Pasadena, CA, and thus not far from RAND's

headquarters (cf. Schwoch 2009, 185). Speier was also involved in the foundation of Center for Advanced Study in the Behavioral Sciences (cf. Bessner 2018, chap. 7).

20. Goldhamer (1954, 3) emphasizes that the members of the latter two groups do not have to speak with one voice: "The fact that a group of players represents a multiplicity of 'other governments' or, in the case of the communist bloc, such countries as the Soviet Union, China, Poland, etc. does not, of course, imply that the players representing these groups operate as if all these governments were without interest of their own. Teams representing diverse governments are simply assuming multiple roles and are not operating as the representative of a unified interest."
21. A further "result" of this matrix and the accompanying discussion was that "relationships of dominance" between the various factors could be determined (Goldhamer 1955a, 12):

    1, 2, 3, 6a > 4, 5, 6b
    1, 2, 6a > 3
    1, 6a > 2
    6c > 1, 6a

    Apart from the harmony in the Kremlin, thus, the most dominant factors increasing the likelihood of a Soviet attack were, in the view of the experts, the limitation of nuclear weapons use on the American side and the morale and status of the Soviet armed forces. Put in terms of policy advice, in order to avoid a major nuclear war, one has to undermine morale in the Kremlin (6c) and the Soviet army (6a), and maintain and develop the potential to destroy the enemy by means of nuclear weapons (1, 2).
22. Nonetheless, since they had access to restricted information, participants were expected not to discuss the games with people outside RAND. Herbert Goldhamer, his wife recalled, never mentioned anything substantial from the games at home (cf. Interview with Joan D. Goldhamer by the author, 24 August 2011, pp. 8–9).
23. Formosa was the name the Portuguese sailors gave to the Island of Taiwan when they reached it in the mid-sixteenth century. In January 1955, some months before the Social Science Division started the first game, the Eisenhower Administration had decided on a commitment of the United States to shield specific territories in the West Pacific against an imminent invasion of the Chinese army. The Formosa Resolution was a

reaction to the Chinese bombing of several islands in the Taiwan Strait, and eventually led to a peace agreement between China and the United States.

24. In the documents I inspected, there were no details regarding the amount of support or the exact source. Such information can very likely be found in the Ford Foundation Records in the Rockefeller Archive Center in Sleepy Hollow, NY. However, one can assume that the money came from the BSP to which Speier still functioned as a consultant. And the fact that there was a separate report prepared for the FF's president fosters the conclusion that the financial support was not insignificant.
25. The report was distributed as P-1679-RC at RAND in the spring of 1959 (Goldhamer and Speier 1959b) and appeared in print in *World Politics* in October that year (Goldhamer and Speier 1959a). It was republished at least twice: in Martin Shubik's (1964) reader *Game Theory and Related Approaches to Social Behavior* and in Speier's (1969) book *Force and Folly*.
26. We might note here that the Hungarian revolution of October 1956 and the violent reaction by the Red army had not been anticipated in the scenario.

References

Abella, Alex. 2008. *Soldiers of Reason*. Orlando; Austin; and New York: Harcourt.

Akera, Atsushi. 2007. *Calculating a Natural World*. Cambridge, MA and London: The MIT Press.

Amadae, S.M. 2016. *Prisoners of Reason: Game Theory and Neoliberal Political Economy*. New York: Cambridge University Press.

Bessner, Daniel. 2013. The Night Watchman: Hans Speier and the Making of the American National Security State. PhD Dissertation, Duke University, Durham.

———. 2015. Organizing Complexity: The Hopeful Dreams and Harsh Realities of Interdisciplinary Collaboration at the Rand Corporation in the Early Cold War. *Journal of the History of the Behavioral Sciences* 51 (1): 31–53. https://doi.org/10.1002/jhbs.21699.

———. 2018. *Democracy in Exile: Hans Speier and the Rise of the Defense Intellectual*, The United States in the World. Ithaca, NY: Cornell University Press.

Boister, Neil, and Robert Cryer, eds. 2008. *Documents on the Tokyo International Military Tribunal: Charter, Indictment and Judgments*. Vol. 1. Oxford: Oxford University Press.

Brewer, Garry D., and Martin Shubik. 1979. *The War Game: A Critique of Military Problem Solving*. Cambridge, MA and London: Harvard University Press.

Brown, T.A., and E.W. Paxson. 1975. A Retrospective Look at Some Strategy and Force Evaluation Games. R-1619-PR. RAND Corporation Archives, Santa Monica, CA.

Bulley, Adam, Gillian Pepper, and Thomas Suddendorf. 2017. Using Foresight to Prioritise the Present. *Behavioral and Brain Sciences* 40 (January): e79. https://doi.org/10.1017/S0140525X16000996.

Campbell, Virginia. 2004. How RAND Invented the Postwar World. *Invention & Technology* 2004: 50–59.

Collins, Martin J. 2002. *Cold War Laboratory: RAND, the Air Force, and the American State, 1945–1950*. Washington, DC and London: Smithsonian University Press.

Crowther-Heyck, Hunter. 2005. *Herbert A. Simon. The Bounds of Reason in Modern America*. Baltimore and London: The Johns Hopkins University Press.

Davison, W.P. 1958. A Summary of Experimental Research on 'Political Gaming'. D-5695-RC. RAND Corporation Archives, Santa Monica, CA.

———. 1960. A Public Opinion Game. P-2042. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2042.html.

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

———. 2016. 'A Fiction of Long Standing': Techniques of Prospection and the Role of Positivism in US Cold War Social Science, 1950–1965. *History of the Human Sciences* 29 (4–5): 35–58.

DeWeerd, Harvey A. 1966a. In Memoriam: Victor Myron Hunt, 1908–1965. *The Public Opinion Quarterly* 30 (1): 160–161.

———. 1966b. Victor Myron Hunt, 1908–1965. *The Western Political Quarterly* 19 (1): 216.

Edwards, Paul N. 1996. *The Closed World. Computers and the Politics of Discourse in Cold War America*. Cambridge, MA and London: The MIT Press.

Erickson, Paul. 2015. *The World the Game Theorists Made*. Chicago and London: The University of Chicago Press.

Flood, Merrill M. 1951. A Preference Experiment. P-256. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P256.html.

———. 1952. Some Experimental Games. RM-789-1. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM789-1.html.

Gerovitch, Slava. 2002. *From Newspeak to Cyberspeak: A History of Soviet Cybernetics*. Cambridge, MA and London: The MIT Press.

Gilman, Nils. 2003. *Mandarins of the Future. Modernization Theory in Cold War America*. Baltimore and London: The Johns Hopkins University Press.

Goldhamer, Herbert. 1954. Toward a Cold War Game. D(L)-2603. RAND Corporation Archives, Santa Monica, CA.

———. 1955a. Summary of Cold-War Game Activities in the Social Science Division. D-2850. RAND Corporation Archives, Santa Monica, CA.

———. 1955b. The Political Exercise. A Summary of the Social Science Division's Work in Political Gaming, with Special Reference to the Third Exercise July–August 1955. D-3164-RC. RAND Corporation Archives, Santa Monica, CA.

Goldhamer, Herbert, and Hans Speier. 1959a. Some Observations on Political Gaming. *World Politics* 12 (1): 71–83.

———. 1959b. Some Observations on Political Gaming. P-1679-RC. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P1679.html.

Goldsen, Joseph M. 1956. The Political Exercise. An Assessment of the Fourth Round. D-3640-RC. RAND Corporation Archives, Santa Monica, CA.

Gruenberger, Fred J. 1968. The History of the JOHNNIAC. RM-5654-PR. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM5654.html.

Helmer, Olaf. 1960. Strategic Gaming. P-1902. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P1902.html.

Helmer, Olaf, and R.E. Bickner. 1961. How to Play SAFE—Book of Rules of the Strategy and Force Evaluation Game. RM-2865-PR. RAND Corporation Archives, Santa Monica, CA.

Helmer, Olaf, and T.A. Brown. 1962. SAFE: A Strategy-and-Force-Evaluation Game. RM-3287-PR. RAND Corporation Archives, Santa Monica, CA.

Herman, Ellen. 1995. *The Romance of American Psychology: Political Culture in the Age of Experts*. Berkeley; Los Angeles; and London: University of California Press.

Hermann, Charles F. 1968. Simulation: Political Processes. In *International Encyclopedia of the Social Sciences*, ed. David L. Sills, vol. 14, 274–281. New York: Macmillan and The Free Press.

Hilgers, Philipp von. 2012. *War Games: A History of War on Paper*. Trans. Ross Benjamin. Cambridge, MA and London: The MIT Press.
Horowitz, Irving L. 1963. *The War Game: Studies of the New Civilian Militarists*. New York: Ballantine.
Kaplan, Fred. 1983. *The Wizards of Armageddon*. New York: Simon & Schuster.
Kecskemeti, Paul. 1955. War Games and Political Games. D-2849. RAND Corporation Archives, Santa Monica, CA.
———. 1958. *Strategic Surrender: The Politics of Victory and Defeat*. Stanford, CA: Stanford University Press.
———. 1961. *The Unexpected Revolution: Social Forces in the Hungarian Uprising*. Stanford, CA: Stanford University Press.
Kris, Ernst, and Hans Speier. 1944. *German Radio Propaganda*. London and New York: Oxford University Press.
Leonard, Robert J. 1992. Creating a Context for Game Theory. In *Toward a History of Game Theory*, Annual Supplement to History of Political Economy Volume 24, ed. E. Roy Weintraub, 29–75. Durham and London: Duke University Press.
———. 2010. *Von Neumann, Morgenstern, and the Creation of Game Theory: From Chess to Social Science, 1900–1960*. Cambridge and New York: Cambridge University Press.
Luce, R. Duncan, and Howard Raiffa. 1957. *Games and Decisions. Introduction and Critical Survey*. New York: John Wiley & Sons.
Mallard, Grégoire, and Andrew Lakoff. 2011. How Claims to Know the Future Are Used to Understand the Present: Techniques of Prospection in the Field of National Security. In *Social Knowledge in the Making*, ed. Charles Camic, Neil Gross, and Michèle Lamont, 339–377. Cambridge, MA and London: Harvard University Press.
Mannheim, Karl. 1952. *Essays on the Sociology of Knowledge*. International Library of Sociology and Social Reconstruction. Edited by Paul Kecskemeti. London and New York: Routledge & Paul.
———. 1953. *Essays on Sociology and Social Psychology*. International Library of Sociology and Social Reconstruction. Edited by Paul Kecskemeti. London and New York: Routledge & Paul.
———. 1956. *Essays on the Sociology of Culture*. International Library of Sociology and Social Reconstruction. Edited by Ernest Manheim. London and New York: Routledge & Paul.
Mood, Alexander M. 1954. War Gaming as a Technique of Analysis. P-899. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P899.html.

Nedell, Allan A. 1998. Project Troy and the Cold War Annexation of the Social Sciences. In *Universities and Empires: Money and Politics in the Social Sciences During the Cold War*, ed. Christopher Simpson, 3–38. New York: The New Press.

von Neumann, John, and Oskar Morgenstern. 1953. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Poundstone, William. 1993. *Prisoner's Dilemma. John von Neumann, Game Theory, and the Puzzle of the Bomb*. New York: Anchor Books.

Reisch, George A. 2005. *How the Cold War Transformed Philosophy of Science: To the Icy Slopes of Logic*. Cambridge: Cambridge University Press.

Robin, Ron. 2001. *The Making of the Cold War Enemy. Culture and Politics in the Military-Intellectual Complex*. Princeton and Oxford: Princeton University Press.

Sabin, Philip. 2009. *Lost Battles: Reconstructing the Great Clashes of the Ancient World*. London and New York: Continuum.

———. 2012. *Simulating War. Studying Conflict Through Simulation Games*. London and New York: Continuum.

Schelling, Thomas C. 1968. *The Strategy of Conflict*. London; Oxford; and New York: Oxford University Press.

Schnitzer, E.W. 1955. Third Political Exercise. Summary and Documents. D-3163-RC. RAND Corporation Archives, Santa Monica, CA.

Schwoch, James. 2009. *Global TV: New Media and the Cold War, 1946–69*. Urbana: University of Illinois Press.

Shubik, Martin. 1964. *Game Theory and Related Approaches to Social Behavior. Selections*. New York; London; and Sydney: John Wiley & Sons.

———. 1975. *Games for Society, Business and War*. New York; Oxford; and Amsterdam: Elsevier.

Solovey, Mark. 2013. *Shaky Foundations: The Politics-Patronage-Social Science Nexus in Cold War America*. New Brunswick, NJ: Rutgers University Press.

Specht, Robert D. 1957. War Games. P-1041. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P1041.html.

Speier, Hans. 1969. *Force and Folly*. Cambridge, MA and London: The MIT Press.

———. 2007. Nicht die Auswanderung, sondern der Triumph Hitlers war die wichtige Erfahrung. Autobiografische Notizen eines Soziologen. In *Die Intellektuellen und die moderne Gesellschaft*, 353–375. Graz and Wien: Nausner & Nausner.

Suddendorf, Thomas, and Michael C. Corballis. 2007. The Evolution of Foresight: What Is Mental Time Travel, and Is It Unique to Humans? *Behavioral and Brain Sciences* 30 (3): 299–313. https://doi.org/10.1017/S0140525X07001975.

Thomas, William. 2015. *Rational Action: The Sciences of Policy in Britain and America, 1940–1960*. Cambridge, MA: MIT Press.

Thompson, Jenny. 2010. *War Games: Inside the World of 20th-Century War Reenactors*. Washington, DC: Smithsonian Books.

Weiner, M.G. 1959. War Gaming Methodology. RM-2413. The RAND Corporation, Santa Monica, CA.

Weintraub, E. Roy, ed. 1992. *Toward a History of Game Theory*, Annual Supplement to History of Political Economy Volume 24. Durham and London: Duke University Press.

5

The Oracle's Epistemology: Expert Opinions as Scientific Material, 1955–1960

Pragmatic Positivism

In many ways, RAND was a special place in the first decades of its existence. An informal yet rigorous style of intellectual exchange distinguished it from contemporary academia. A famous episode has it that John von Neumann, "one of the greatest mathematicians of our time" (Interview with Hans Speier by Martin Collins, 1988, p. 39), was giving a talk at RAND. Hans Speier remembered the scene:

> So here is this great man, Johnny von Neumann, and writes something on the blackboard, and people listen. An unknown person from the mathematics division in his early twenties says, "No, no, that can be done much more simply." […] Now my heart stood still because I wasn't used to this sort of thing. Johnny von Neumann said, "Come up here, young man. Show me." He goes up, takes the piece of chalk, and writes down another derivation, and Johnny von Neumann interrupts and says, "Not so fast, young man. I can't follow." (Interview with Hans Speier by Martin Collins, 5 April 1988, p. 39)

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*, Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_5

As required for such stories, it later turned out that the young man was right. Moreover, Lloyd S. Shapley (1923–2016), as the young man was called, would go on to receive the 2012 Nobel Memorial Prize in Economic Sciences (with Alvin E. Roth).

Such events certainly made a place like RAND unique, at least for some observers. For others, the distinct character of RAND arose from the high value it placed on interdisciplinary collaboration. “The very nice thing about RAND,” philosopher Nicholas Rescher stated, was that it was “less given to tight disciplinary matrices and had more opportunities for interaction and interrelationships. They weren’t always seized, but they were at least there as opportunities” (Interview with Nicholas Rescher by the author, 1 September 2011, p. 5). RAND researcher Albert Wohlstetter emphasized that interdisciplinarity distinguished RAND from the universities where “there was really no genuine interdisciplinary work” (Wohlstetter, cited in Bessner 2015, 31).[1] To be sure, interdisciplinarity at RAND was not without its frictions. In his biographical recollections, Speier pointed out that some colleagues assumed an implicit hierarchy of the scientific disciplines, with the social sciences at the bottom and mathematics, physics, and the highly mathematized forms of economics practiced at RAND at the top. Speier complained:

> Only a few mathematicians, physicists, or engineers had an adequate understanding of the contribution social scientists could make in working on problems of national security. Either they believed in the omnipotence of social scientists, expecting that they could build castles in the air the way one builds bridges across a river, or they held the equally inappropriate view that socials scientists were not concerned with anything scientific at all and the solution of political problems could therefore just as well be left to astrologers, charlatans, or poets. (Speier 1989, 22)

Former RAND affiliate Martin Shubik concurred:

> There were a few of us, and RAND encouraged it, who did considerable interdisciplinary work. And certain groups, like the political group that I mentioned, Wohlstetter, Goldhamer, did talk to the mathematicians and to the war gamers fairly openly and often. But to a certain extent, disciplines still stayed apart. And it was only a handful of us who really wandered over

> all of the disciplines. So I say this because to some extent, I saw RAND from a heavy interdisciplinary area, but had I been an engineer, it would have looked very different. Or had I, you know, there were lots of other people, or for … except for Herman Kahn, the physicists were basically off in a corner. (Interview with Martin Shubik by the author, 2 September 2011, p. 4)

Thus, praising RAND as the (lost) paradise of true interdisciplinarity, as it is sometimes done (e.g., Campbell 2004), almost certainly conceals the underlying frictions, inequalities, struggles, and disagreements (cf. Bessner 2015).

Another aspect positively emphasized by RAND affiliates was the easiness with which the boundaries between RAND and the universities could be crossed. Researchers from both camps took sabbaticals to spend time on the other side. RAND invited university professors to spend the summer months in their Santa Monica buildings close to the Pacific Ocean. In turn, RAND researchers held visiting professorships at prestigious universities or were invited as fellows to places like Stanford's Center for Advanced Study in the Behavioral Sciences. In fact, as Joy Rohde (2013) argues, this ease of transgressing the boundaries between academic and non-academic sites of knowledge production was a more general phenomenon. Many campuses across the country housed organizations engaged in researching topics of relevance to military and political decision-makers, mostly in the form of Federal Contract Research Centers (FCRCs). To repeat the metaphor Rohde uses, a gray area of policy-relevant knowledge production had emerged, uniting FCRCs with universities, non-academic research organizations, the armed forces, government agencies, and other, especially philanthropic, patrons. Only in the late 1960s, the gray of this area began to separate more clearly into its two constituents black and white, when student protests and other developments led to an increasing separation of university and non-university research and forms of oppositions like basic versus applied became more important tropes in the self-description of academic social science. "By retreating from government research and advising, academic social scientists unwittingly ceded more territory to the Pentagon's increasingly insular network of contract research agencies" (Rohde 2013, 147). The walls had grown, and intellectual exchange faded. As a consequence, non-university policy research disappeared from the fora of academic quality control.

Nonetheless, during the 1950s, the situation was still characterized by a high permeability between academia and other research organizations. This concerned both individuals and their ideas. While the close relationship to the Air Force certainly fostered a technical and practical spirit, RAND researchers were also able to pursue more academic, or basic, projects. A case in point is Olaf Helmer. In parallel to his contributions to the development of techniques of prognosis, like Delphi and various war game designs, Helmer still entertained an interest in philosophical and epistemological issues. In the years following the first Delphi study, he became interested in establishing an epistemological foundation for the systematic use of experts. Discussions on such topics intensified when Nicholas Rescher (1928–) joined the Mathematics Division in 1954. Born in Germany, Rescher had moved to the United States with his family at the age of ten. Later, he enrolled at Queens College, New York, where he took some courses led by Carl G. Hempel. After graduate studies in Princeton, Rescher earned a Ph.D. in Philosophy in 1952 upon submitting a dissertation on Leibniz's philosophy of science. This earned him a teaching assignment for the following year, which he also used to collaborate with another German émigré philosopher Paul Oppenheim (cf. Rescher 2006, 292). After two years of military service spent mainly in Washington, DC, Rescher was happy to learn that RAND had accepted his job application and moved to California together with his wife. At RAND, Rescher remembers, things were changing when he arrived. "When I came, 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 with Nicholas Rescher by the author, 1 September 2011, p. 3). Helmer and Rescher realized that they shared an interest in epistemology and decided to pursue this interest together. In early 1955, they began meeting weekly in either Helmer's or Rescher's home—in Mandeville Canyon Road or Bestor Boulevard, respectively—and "spent many pleasant evenings […] talking epistemology" (Rescher 2002, 91).[2] The outcome of these discussions was a longer piece entitled "On the

Epistemology of the Inexact Sciences" (*OEIS*). *OEIS* was first issued as RAND paper P-1513, in October 1958, and published in *Management Science* in the following year (Helmer and Rescher 1959).[3]

The article provided the basis of several to follow in which Helmer addressed the question whether or not operations research, policy analysis, or however one wanted to call the research activities carried out at RAND and elsewhere, figured as "science," a question that puzzled many, so by far not all, contemporaries (cf. Thomas 2015, 193–198). It opened with an attack on the belief that among the academic sciences, some were exact and some inexact. This belief was, as the authors put it, "a fiction of long standing," because it "finds a difference in principle where there is only one of degree" (Helmer and Rescher 1959, 25). Exactness, they claimed, was not a defining characteristic for a science; neither was the use of formalization or mathematical notation. Rather, what rendered any intellectual effort scientific was, firstly, the attempt to explain and predict specific and specified phenomena and, secondly, to do so in an intersubjectively verifiable manner. This definition of science, they emphasized, potentially included all branches of science, the natural as well as the social sciences as well as the humanities. "[A] discipline which provides predictions of a less precise character, but makes them correctly and in a systematic and reasoned way, must be classified as a science" (Helmer and Rescher 1959, 25).

We might note in passing that this definition, by pushing the definitional weight to the concept of discipline as well as to terms like "reasoned" without any further explication, remained itself unclear. However, it was already a servant to the main thrust of their position: to rehabilitate those sciences that were pejoratively labeled as inexact. To the authors, exactness as a criterion for judging about sciences was a myth held as widely as—and related to—the belief in the existence of two classes of sciences. Instead, a closer look at the sciences would reveal that whole branches of the natural sciences were inexact. Indeed, they maintained, only a minority of fields in the natural sciences were exact in the sense that they predominantly consisted of formal and mathematical expressions. Referring to medicine, parts of aerodynamics and high temperature physics, Helmer and Rescher argued that many parts of the natural sciences were "still intermingled with unformalized expertise" and that "inexactness is not a prerogative of the social sciences" (Helmer and Rescher 1959, 26). Were exactness a sensible criterion, one would have to negate the scientific character of these branches of activity. On the other hand, the definition they submitted would allow calling these endeavors scientific, because obviously, they attempted to explain and predict phenomena in an objective, that is, intersubjectively verifiable manner.[4]

This misconception of the character of science as adherent to the myth of exactness has had disturbing consequences. "Indeed, the artificial discrimination between the physical sciences with their (at least in principle) precise terms, exact derivations and reliable predictions as opposed to the vague terms, intuitive insights and virtual unpredictability in the social sciences has retarded the development of the latter immeasurably" (Helmer and Rescher 1959, 26).[5] But if there was such thing as inexact science, what was specific about it? And taking into account that the exact sciences dominated much of the thinking in the philosophy of science, how might "the foundations for a uniform epistemology of all of the inexact sciences" (Helmer and Rescher 1959, 27) be laid?

In seeking to answer this question, Helmer and Rescher explored the role of selected epistemological concepts in the inexact sciences. They began with historical laws, which they define "as a well-confirmed statement concerning the actions of an organized group of men under certain restrictive conditions (such group actions intended to include those of systems composed conjointly of men and nonhuman instrumentalities under their physical control)" (Helmer and Rescher 1959, 27). In contrast to the idea that history is merely a descriptive undertaking, they claimed that any historical argument required such general statements.[6] Historical laws shared three features: they were (or could be) formulated in a law-like manner, they had clear references to (past) time and space, and they were loose. "Loose" in this context meant that the conditions implied in historical laws cannot be spelled out completely. To make their point, Helmer and Rescher discussed in more detail the example of marine battles in early modern history: "In the sea fights of sailing vessels in the period 1653–1803, large formations were too cumbersome for effectual control as single units" (Helmer and Rescher 1959, 27). The claim that historical laws were loose was supported by pointing out that the law-like statement on sailing fleet tactics depended on knowledge of sailing, weather, and ordnance; to take only the latter, knowledge about the ordnance ramified into ballistics, metallurgy, mining, and so on.

As the authors noted, this looseness obviously contradicted the universal character usually attributed to natural laws. Since the historical complexity required a selection of specific factors out of the pool of all relevant factors, historical laws must perforce allow for exceptions. It was thus more precise to understand historical laws as rules, Helmer and Rescher wrote, proposing to call such rule-like laws "quasi-laws." In order for a quasi-law to be valid,

> it is not necessary that no apparent exceptions occur; it is only necessary that, if an apparent exception should occur, an adequate explanation be forthcoming, an explanation demonstrating the exceptional characteristic of the case in hand by establishing the violation of an appropriate (if hitherto unformulated) condition of the law's applicability. (Helmer and Rescher 1959, 29)

In line with their earlier argument, Helmer and Rescher claimed that quasi-laws were found quite often even outside the historical or social sciences. They would appear frequently in the natural sciences, a fact that writers on the methodology of the physical sciences tended to overlook.[7] "Indeed some branches of the social sciences are in better shape as regards the generality of their laws than various departments of physics, such as the theory of turbulence phenomena, high-velocity aerodynamics, or the physics of extreme temperature" (Helmer and Rescher 1959, 30). Therefore, quasi-laws appeared not only to deserve a more sustained interest by philosophers of science; rather, they might also serve as a stable and fertile starting point for an epistemology of the inexact sciences. Philosophers of sciences, Helmer and Rescher (1959, 30) pleaded, "should realize that the seemingly thin line between vagueness and vacuity is solid enough to distinguish fact from fiction reasonably well in practical applications."

The first litmus test of quasi-laws as starting point of a new epistemology concerned their function in explanation and prediction. The authors noted that Hempel and Oppenheim's (1948) well-known definition of explanations as deductive-nomological inferences failed to include statistical laws and quasi-laws. As both statistical laws and quasi-laws were non-universal statements, a deductive inference was impossible. Nonetheless, the basic requirement remained the same: since a satisfactory explanation relying on non-universal laws could not logically entail the hypothesis, it must succeed "in making the statement to be explained highly *credible* in the sense of providing convincing evidence for it" (Helmer and Rescher 1959, 31; emphasis in original).

On the basis of the discussion so far, Helmer and Rescher turned against the parallel usually drawn by fellow philosophers of science between explanation and prediction. A standard proposition in contemporary philosophy of science—and certainly with Hempel and Oppenheim[8]—was that explanation and prediction were logically identical, the only difference being that the former is concerned with past events and the latter with future ones. In line with other coeval positions, Helmer and Rescher set out to challenge this claim.[9] From their point of view, the first difference between explanation and prediction was that "there are such things as *unreasoned* predictions—predictions made without any articulation of justifying argument" (Helmer and Rescher 1959, 32; emphasis in original). At first, this argument might appear naïve. One is tempted to rebut it by pointing to the obvious case that in science, even the best predictions probably would not get on without justifying arguments. However, there was more to this argument, and this related to credibility as a substitute for logical deduction. Predictions must be credible prior to the events that will potentially verify them. Whereas explanations achieved credibility by providing a justifying argumentation and "plausible arguments," predictions "may, for example, reside in proving sound *ex post facto* through a record of successes on the part of the predictor or predicting mechanism" (Helmer and Rescher 1959, 32)—hence being credible, but unreasoned predictions.

Further, Helmer and Rescher saw a second difference between explanation and prediction with regard to the "logical strength," or the degree of epistemological exclusivity that rendered an explanation or a prediction sufficiently credible. "By the very meaning of the term, an explanation must *establish* its conclusion […]. On the other hand, the conclusion of a (reasoned) prediction need not be well established in this sense; it suffices that it be rendered *more tenable than comparable alternatives*" (Helmer and Rescher 1959, 32; emphases in original). Although predictions might be as firmly based on empirical material and laws as explanations are, they need not be. In practical circumstances, even predictions that do not conform with the strict rules applied to explanations might be immensely useful.

The belief that explanation and prediction were identical in their logical structure had, so the authors concluded, too long detained philosophers of science from developing a methodology of prediction. To pursue such a project would probably lead to a reorientation within philosophy of science, because a better understanding of prediction would result in a reappraisal of "possibly unorthodox items of methodological equipment, such as quasi-laws" (Helmer and Rescher 1959, 33). Furthermore, such "unorthodox items" could be used to develop specific predictive methods.

Apart from quasi-laws, a major element of the class of unorthodox items was the opinion of experts. Since essential parts of expert opinion were tacit and non-explicit, the traditional philosophy of science had excluded cognizant, yet unreasoned assessments of phenomena as reliable sources of knowledge. *OEIS* undertook the attempt to remedy the consequences of this exclusion by constructing a methodological rationale for the use of expert opinions for prediction. It was meant to fill an epistemological gap, especially at RAND, where many approaches to strategy analysis had relied profoundly on the knowledge, estimation, and opinion of experts—a complex of cognitive elements and actions that Helmer and Rescher, in the following, subsume under the term "expertise." As we have seen in Chap. 3, Helmer had been a crucial member of the team that developed Delphi, and in fact, a passage in OEIS appears to be the first published mention of Delphi (Helmer and Rescher 1959, 47).[10] However, it is important to note that *OEIS* was intended to deliver a philosophical foundation not only for this technique, but also for many other approaches to predictive analysis developed and used at RAND, including war games and political gaming.

Probabilities and the Degree of Confirmation

Having thus established that the logical structures of explanation and prediction differed in the inexact sciences, the authors approached the problem of constructing a logical structure of prediction. Their efforts centered on the theory of probability, a theory that, as the authors noted, found itself in a peculiar position, drawing from fields like formal semantics and logics, but also from the empirical social sciences. Moreover, in the view of the authors, "[e]ven for the logical part of the theory, the foundations are not yet established very firmly, and only with regard to applications of the simplest forms of one-place predicate languages has real progress been made to date" (Helmer and Rescher 1959, 33). In addition, apparently several concepts of probability existed, which sometimes caused difficulties in understanding. Helmer and Rescher discussed three of these concepts: relative frequency, degree of confirmation, and personal probability. Relative frequency referred to a class of objects of which one or more characteristics are known. Examples are the frequency of males in the United States and the frequency of days with rain in the Los Angeles area. If we know that in 2010, there were 23 days of precipitation, we can form the statement that the probability of having a rainy day in Santa Monica, CA, is 0.063, or 6.3%.[11]

From this relative frequency concept of probability, the degree of confirmation must be discerned. While the mathematical procedure is identical, the crucial difference is its theoretical value. The degree of confirmation is the degree to which a hypothesis *H* is confirmed by the available evidence *E*. In the simplest case, there is a uni-dimensional set of evidence that can be described in the form of a statistical record. Formally, this would result in a set with *n* entries, all of them either positive or negative toward the hypothesis *H*. *H*, in this context, is understood as a predictive and singular statement and not, in usual parlance, as a statement aiming at generalizability. It is a statement about a further (and possibly future) event at a given spatiotemporal point. It would not have the form of "The sun always shines in Santa Monica," but instead read "Tomorrow will be a sunny day in Santa Monica." This statement *H* was then related to a record of past events. Let the data mentioned above be the content of our set of evidence *E*. Each day enters this set of evidence, informing us whether there was precipitation or not. To make things simple, we use the above mentioned 23 days with precipitation. Further, we apply the rule that in case there was precipitation, the sun did not shine. Each day when there was precipitation is thus an entry in *E* that contradicts *H*, whereas each day without precipitation contributes to the confirmation of H.[12] We thus can conclude that the degree of confirmation of H on the basis of E is

$$dc(H,E) = \left(\frac{365 - 23}{365}\right) = 0.937$$

Thus, the hypothesis that tomorrow will be a sunny day in Santa Monica, CA, is confirmed by the available (or reasonably selected) evidence with a degree of confirmation of 0.937, or 93.7%. This, of course, is exactly the converse probability to the one from the previous example. This is a result of the different statements, which focused on rain and sun, respectively. However, as mentioned above, the difference between the two kinds of probability is not mathematical but theoretical, since the second kind refers to a process of confirmation.

The concept of the degree of confirmation just described formed the core of the proposed epistemology of the inexact sciences. It functioned as a bridge between the credibility of predictions on the one hand and the use of expert estimations on the other (see Fig. 5.1). The link between the degree of confirmation and credibility was established by identifying the former with the amount of credibility one can rationally attribute to the statement under scrutiny. "The degree of confirmation of *H* on the basis of *E* is intended to be a measure of the credibility rationally imparted to the truth of *H* by the assumed truth of *E*" (Helmer and Rescher 1959, 34; emphases in original). The degree of confirmation thus related to the concept of credibility by assuming that the mathematical result of the formula for the degree of confirmation of a hypothesis corresponds to the amount of credibility that can rationally be attributed to it. If the degree

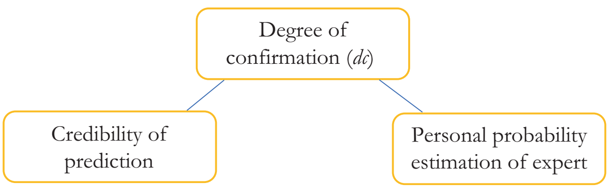

Fig. 5.1 The degree of confirmation as a bridge concept

of confirmation of the Santa Monica-sunshine hypothesis is 0.937, then it was 97.3% rational to believe in the hypothesis. In contrast to explanations à la Hempel-Oppenheim, where the explanandum was contained in the sentences of the explanans and then deduced by means of logical conclusion, the prediction à la Helmer-Rescher attempted to achieve credibility by providing evidence that corroborated the rationality of believing in the truth of a prediction.

The second link offered by the bridge concept of degree of confirmation, that is, its link to the use of expert estimations, was established in a comparable, though slightly different move. It referred to the third concept of probability, that is, the personal probability, a concept identical to the Bayesian concept of probability, although the authors did not acknowledge this. The personal probability was understood as indicating how convinced or confident a person was toward the truth of a given statement. One could measure such personal probability "behavioristically," as the authors put it, for instance by asking individuals to bet on a given situation. If a person was asked to bet which side of a coin would show up, she should be indifferent toward the bet as the likelihood of the two possible outcomes was the same (or $p1 = p2 = 0.5$; in other words, she will hold a bet of 1:1), provided that she was rational. If, like in the mentioned example, the objective likelihood of the outcome was known, a comparison between this objective likelihood and the personal probability was possible. This conception of rational behavior subsequently formed the basis of a definition of the predictive expert:

> We shall call a person "rational" if (1) his preferences (especially with regard to betting options) are mutually consistent or at least, when inconsistencies are brought to his attention, he is willing to correct them; (2) his personal probabilities are reasonably stable over time, provided he receives no new relevant evidence; (3) his personal probabilities are affected (in the right direction) by new relevant evidence; and (4) in simple cases where the evidence *E* at his disposal is known, and *E* and *H* are such that *dc(H,E)* is defined, his personal probability regarding *H* is in reasonable agreement with the latter […]. A (predictive) "expert" in some subject-matter is a

> person who is rational in the sense discussed, who has large background knowledge *E* in that field, and whose predictions (actual or implicit in his personal probabilities) with regard to hypotheses *H* in that field show a record of comparative successes in the long run. (Helmer and Rescher 1959, 36; emphases in original)

In behavioristic terms, then, one could identify a person as rational by comparing her personal probability as expressed in betting behavior with the objective odds, provided that they were known. On the basis of the same information, rational persons attributed personal probabilities to hypotheses that were identical or reasonably close to the computed degree of confirmation.

The idea of Helmer and Rescher was to turn this relation upside down. While at first, they had been talking about assessing the rationality (and partly the expertise) of a person by comparing her personal probabilities with (stable and known) odds, they proposed to assess the probability of a given event by using the estimations of experts. In cases where the "objective odds" were unknown, Helmer and Rescher argued, a person's personal probability toward a given hypothesis could reasonably be taken as approximation of the hypothesis' actual degree of confirmation, provided that the person is both rational and an expert in the field under question. The Bayesian "degree of belief" was taken to be a stable estimator of the degree of confirmation of a hypothesis concerning future events. In their words,

> *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, 36; emphases in original)

Of utmost importance, then, for the success of predictions, was the set of evidence *E* available and relevant for the hypothesis.

Evidence and Implicit Expertise in Prediction

While much attention so far had been dedicated to the link between the credibility of a prediction and the personal probability of an expert, the underlying notion of a set of evidential materials *E* had not yet received similar scrutiny. Acknowledging this, Helmer and Rescher pointed out that knowledge came in a variety of forms. Some forms would not easily integrate into a set of statistical sentences as required for *E* in the understanding suggested by the earlier examples. They described the example of a fare for riding a bus that, after amounting to 10 cents throughout the previous months, had increased to 15 cents. Despite the fact that the evidence *E*, if understood as a statistical set of data, confirmed to a high degree the hypothesis that the fare would decrease to 10 cents again, this hypothesis was not reasonable for those understanding how prizes vary in such markets. Because it was based on our background knowledge about how and why new fare schemes are adopted, the personal probability would differ considerably from the degree of confirmation calculated with the formalisms introduced above.

Thus, knowledge of past instances or statistical distributions in many cases does not suffice to set up convincing predictions.

> [T]he evidential use of such *prima facie* evidence must be tempered by reference to background information, which frequently may be intuitive in character and have the form of a vague recognition of underlying regularities, such as analogies, correlations, or other conformities whose formal rendering would require the use of predicates of a logical level higher than the first. (Helmer and Rescher 1959, 37–38; emphasis in original)

Such tempering was the primary function of expertise in prediction in the inexact sciences. Among them, Helmer and Rescher thought, the social sciences certainly stood out, because they constantly dealt “with situations in which statistical information matters less than knowledge of regularities in the behavior of people or in the character of institutions, such as traditions and customary practices, fashions and mores, national attitudes and climates of opinion, institutional rules and regulations, group aspirations, and so on" (Helmer and Rescher 1958, 30). The background knowledge, which on many occasions was a precondition of reasonable prediction, was often "non-explicit"; while being only intuitively understood, it was nonetheless "typical" of the inexact sciences. "Hence the great importance which must be attached to experts and to expertise in these fields" (Helmer and Rescher 1959, 38).

In formal notation, the definition of the degree of confirmation must be supplemented by the body of potentially relevant background knowledge K, so that $dc_{K(A)}(H,E)$—meaning the degree of confirmation of H on E in the view of expert A's background knowledge $K(A)$. This degree of confirmation is estimated via A's personal probability $pp_A(H,E)$. "Thus the device of using personal probabilities of experts, extracted by appropriately devised techniques of interrogation, can serve as a means of the dc-type even in cases where there is no hope of application of the formal degree-of-confirmation concepts" (Helmer and Rescher 1959, 39).

That $K(A)$ is non-explicit might be seen as threatening the scientific claim for objectivity. But this, Helmer and Rescher argued, was wrong. Of course, the selection of experts should follow objective criteria. The procedure should be based not on mere personal preference but on the predictive performance record of an expert. One could conceive of the expert "as an objective indicator" (Helmer and Rescher 1959, 43). Doing this would ensure that the expert's predictive statements became "an integral, intrinsic part of the subject matter" (Helmer and Rescher 1959, 43). They suggested no less than a complete incorporation of the expert into epistemology. His or her utterances were to be understood as elements of science comparable to other elements. "Our 'data' are supplemented by the expert's personal probability valuations and by his judgments of relevance […], and our 'theory' is supplemented by information regarding the performance of experts" (Helmer and Rescher 1959, 43). If such complete integration were ensured, the use of expertise was "no retreat from objectivity or reversion to a reliance on subjective taste" (Helmer and Rescher 1959, 43).

The Problem of Confirmation in the Philosophy of Science

Notwithstanding its crucial role in the argument, the concept of the degree of confirmation in *OEIS* remained curiously flexible and unsophisticated. One might be tempted to attribute that to ignorance regarding the state of the discussion on this topic in contemporary philosophy of science, were it not the case that Helmer himself had been a contributor to this discussion. In order to interpret this, it appears appropriate to briefly interrupt the discussion of *OEIS* and to look back to the earlier collaboration of Helmer with Carl G. Hempel and Paul Oppenheim. This discussion will also help interpret the role of the expert in *OEIS* and throw light on an aspect where *OEIS*, though almost hidden, featured an important methodological progress.

Prior to the publication of the "Studies in the Logic of Explanation" (Hempel and Oppenheim 1948), Helmer had worked closely with Hempel and Oppenheim on a formalization of the degree of confirmation and related topics. This line of work had been opened by Janina Hosiasson-Lindenbaum (1899–1942), a Polish philosopher of science and logician who figures as a member of the Lwów–Warsaw school of logic. In an article in *The Journal of Symbolic Logic*, Hosiasson-Lindenbaum (1940) proposed a formal solution to the problem of confirmation and described what came to be known as the Raven paradox, or Hempel's paradox. It stated that, if one accepted the basic laws of formal logic, the general proposition "All ravens are black" was confirmed both by observations of the kind "This raven here is black" and by observations about objects that were no ravens and not black, for example, "This shoe is white" or "This herring is red." This was because the general proposition about the black ravens was logically equivalent to the general proposition "Every non-black object is not a raven."[13]

In close exchange with Paul Oppenheim, Hempel pursued this line of thought after both Oppenheim and he had relocated to the East Coast of the United States. Several articles emanated from this collaboration, both single-authored by Hempel (1943, 1945a, b) and co-authored by Hempel and Oppenheim (1945). Around that time, Olaf Helmer joined the group, eventually leading to a further paper on the degree of confirmation (Helmer and Oppenheim 1945). The co-authored piece by Hempel and Oppenheim remarked at the outset that its content, a theory of confirmation, was developed "jointly with Dr. Olaf Helmer" (Hempel and Oppenheim 1945, 99). The group's work in this direction led to intense discussions in scholarly journals (e.g., a discussion initiated by a comment by C. H. Whiteley 1945, which led to answers by Hempel 1946, and Nelson Goodman 1946, who on this occasion formulated a second important paradox in the logical theory of confirmation, the Goodman paradox).

Hempel and Oppenheim's (1945) "Definition of 'Degree of Confirmation'" relied on the same formulaic notation that was later used in *OEIS*, expressing the degree of confirmation of a hypothesis relatively to a set of evidence sentences as $dc(H, E)$ and formulating as a basic condition that the degree of confirmation of a given hypothesis lies between 0 and 1.[14] Also with regard to its content, the continuity from these earlier writings to *OEIS* is obvious, for instance in the declaration by Hempel and Oppenheim (1945, 102) that "[o]ne of the guiding ideas in our attempt to construct a definition of confirmation will be to evaluate the soundness of a prediction in terms of the relative frequency of similar occurrences in the past." Moreover, far from identifying explanation with prediction as regards their logic, the authors stated that confirmation had to be discerned from verification. To interpret confirmation in the sense of verification necessarily resulted in making the concept inapplicable "to a hypothesis about an event which is temporally posterior to the data included in the evidence" (Hempel and Oppenheim 1945, 99, fn. 2). In this perspective, to talk about confirming an explanation made no sense.

Though identical in thrust, the text by Hempel and Oppenheim relied more heavily on formalization than the definition and introduction of the degree of confirmation in *OEIS*. It is, however, less heavily laden with mathematics than Helmer and Oppenheim's (1945) "A Syntactical Definition of Probability and of Degree of Confirmation," which appeared some weeks after the piece by Hempel and Oppenheim but was written in close collaboration. Whereas this is not the place to fully reconstruct their argument and formalisms, it appears appropriate to quickly discuss two aspects where the authors go beyond the description delivered more than a decade later in *OEIS*—although it must be emphasized that the description of the degree of confirmation was deliberately kept short in *OEIS*. Firstly, Hempel and Oppenheim referred to a set of evidence *E* that informs about two binary variables, implying that their argument can be applied even to multivariate sets of evidence. In their example, we have not only the variable sunny day with the two properties "yes" and "no," but we have objects that are "blue" or "not blue" and "round" or "not round." Consequently, they speak of four classes of objects: blue and round, blue and not round, not blue and round, and neither blue nor round.

A second notable aspect where the Hempel-Oppenheim paper was more sophisticated than the discussion in OEIS can be found in the discrimination between the set of evidence and the related frequency distributions, and with the latter between the real (and infinite) distribution and an "optimum distribution relatively to E" which must be construed as the real distribution is not fully observable. In other words, as *E* is only a sample of reality, it can only be used to infer on a hypothetical distribution which is optimal relatively to *E*, but not necessarily representing the real distribution. Invoking R. A. Fisher's maximum likelihood method as model procedure, Hempel and Oppenheim define the optimum distribution Δ_E as the distribution that assigns the highest probability to the available set of evidence *E*. In their definition, the degree of confirmation is calculated by inferring, "[o]n the basis of the given evidence E, […] the optimum distribution (or distributions) Δ_E and then assign to H, as its degree of confirmation, the probability which H possesses relatively to E according to Δ_E" (Hempel and Oppenheim 1945, 108). The formulaic expression of this definition reads:

$$dc(H,E) = pr(H,E,\Delta_E)$$

The degree of confirmation of a hypothesis *H* on the basis of the evidence *E* is thus defined as the probability *H* possesses with regard to a distribution that results from the set of evidence. In the written definition, Hempel and Oppenheim indicate that in some cases, several optimum distributions can be ascertained. This holds especially true when *E* contains only a small number of elements. In such cases where *dc* is not single-valued due to several optimum distributions, the authors suggest using the smallest value, though they also indicate some undesirable consequences of so doing (Hempel and Oppenheim 1945, 111 fn. 17).

The relative simplicity of the degree of confirmation concept introduced in *OEIS* is interesting especially with regard to the role Helmer had played in the attempts to achieve a convincing formalization about 15 years earlier. Reflecting on the role of Helmer in the period leading to the publication of the "Studies in the Logic of Explanation," Nicholas Rescher (1997, 350–351, emphasis in original) cites a personal reminiscence of Oppenheim's which is worthwhile to be reproduced here at length. (Note that Carl G. Hempel was called Peter by some of his friends.)

> One of the main problems [addressed by Hempel and Oppenheim in the early 1940s] still reflected our earlier interest in "ordering concepts." It concerned the development of a precise definition and a theory of a *quantitative concept of confirmation*. In this enterprise, we had the important help of Olaf Helmer who, at that time, was working with me in Princeton. We spent the early summer of 1944 with Olaf at Saranac Lake (where I introduced Peter to Einstein, who was then vacationing there.) By midsummer, we felt quite pleased to have formulated an explicit definition of "degree of confirmation" for certain simple formalized languages. But, in the meantime, we heard from Carnap that he was at work on much the same problem, and we were curious—to say the least—whether he had been proceeding along the same lines. Fortunately, we were soon to have a chance to find out, for the Carnaps had invited Peter to spend the latter part of the summer with them in their vacation house at Santa Fé. We had agreed Peter would wire Olaf and me "Stop working" if it should turn out that Carnap had gotten far ahead of us or if he should find a decisive flaw in our approach. A few days later, the fateful wire did come. As Peter explained in a subsequent letter, Carnap had been amazed to find that we had indeed been thinking along very similar lines defining the concept of degree of confirmation as the quotient of two range measures; but he had pointed out that we had chosen a measure-function, which he attributed to Wittgenstein, and which incorporated, as he was able to show, a non-empiricist inductive policy, namely, to learn nothing from past experience.
>
> In fact, however, we did not "stop working," but changed our approach. Olaf provided the basic idea, which made use of R. A. Fisher's notion of maximum likelihood; this led to the definition and theoretical elaboration

> of a concept of degree of confirmation which did not have the objectionable non-empiricist feature, but which, in contrast to Carnap's concept of confirmation, did not have all the formal properties of a concept of probability.

Following the interpretation of Rescher (1997, 351f), the new approach proposed by Helmer was not to assess the degree to which a hypothesis (or theory) is confirmed by a set of evidence, but instead to identify the hypothesis (or theory) for which the probability resulting from the set of evidence is highest, thereby applying the maximum likelihood procedure. The objective of the degree of confirmation in these earlier works was not to measure the credibility rationally attributed to a specific hypothesis, but instead to establish a procedure to select, on the basis of a given set of evidence, among a variety of alternative hypotheses or theories.

While concluding with the definition $dc(H, E) = m/n$ that later was taken up in *OEIS*, the bulk of Helmer and Oppenheim's 1945 paper represented an advance in the philosophy of science that can hardly be guessed from the final definition of the degree of confirmation. As presented in *OEIS*, the static character of this definition hides the actual computational process; and it was with regard to this process that Helmer's proposal was innovative.

According to Rescher, this idea of Helmer's is the link between the group's work on the degree of confirmation in the early 1940s to the ideas put down in the "Studies in the Logic of Explanation." By preforming their approach, Helmer's ideas had an impact on the deductive-nomological form of explanation later developed by Hempel and Oppenheim. Helmer's turn to Fisher's maximum likelihood procedure channeled the group's thinking toward the position that "the best standard of theory assessment is one that proceeds not in terms of evidential support, but rather in terms of the extent to which the theory correctly directs and canalizes our observational expectations" (Rescher 1997, 352). What do we learn from this excursion into philosophical debates of the 1940s? The concept of degree of confirmation functioned as a bridge in the central argument of *OEIS*; it linked the credibility attributed to a predictive hypothesis to the personal probability estimate of experts, thereby justifying the latter's systematic use in scientific prediction procedures. In order to function as bridge concept, however, the notion of degree of confirmation was stripped of all the intricacies and paradoxes that arose during the past decades of philosophical discussion. This is especially noteworthy as one of the authors of *OEIS*, Olaf Helmer, had formulated an essential contribution to this discussion 15 years earlier. Since this had been done deliberately, it is safe to conclude that *OEIS* was never meant to contribute to the philosophy of science literature on the degree of confirmation. Rather, *OEIS* departed from it in an innovative manner. It introduced expert opinions as additional sources of information in scientific prediction, and it did so by giving them the task to evaluate a set of evidence that is available and scientifically established. Whereas $K_{(A)}$ might be implicit or tacit, E is definitely explicit. Thus, the estimated degree of confirmation is not merely an estimation *ex nihilo*, but instead results from an assessment of available scientific evidence; it is an evaluation of intersubjectively available information, and that evaluation is based on the estimator's background knowledge.

Conclusion

Obviously, the perspective developed in "On the Epistemology of the Inexact Sciences" was based not only on the conviction that experts were far more able to deliver stable prognoses than lay people, but that for large parts of the sciences, expert opinions formed a crucial material of epistemic progress (cf. Dayé 2018). Yet, far from simply declaring expert opinions to be evidence, *OEIS* developed a more sophisticated approach to the use of experts in the inexact sciences. It claimed that the task of experts was to evaluate available evidence. Quite in line with the first Delphi study described in Chap. 3, *OEIS* allowed for the set of evidence to be flexibly adjusted and expanded over time, enabling the experts to give feedback on what they think was crucial evidence. "The predictive use of an expert," Helmer and Rescher (1959, 46; emphasis added) wrote, "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," before finally using the expert's personal probability as an estimate of the degree of confirmation.

OEIS thus paved the way to understanding the core difference between the precursor study by Kaplan et al. and the first Delphi carried out by Helmer and Dalkey. While the precursor study resulted in a *prediction*, a statement about the future made without any systematic empirical or evidential corroboration, Delphi had produced a *forecast*, that is, a statement about the future that was based in a systematic assessment by experts of available information and data. *OEIS* developed a methodological foundation for using experts as forecasters. While using a trivialized version of the degree of confirmation concept, the "intrinsic use of expertise" suggested in *OEIS* certainly represented methodological progress in comparison to the study by Kaplan, Skogstad, and Girshick, and terminological progress in comparison to the first Delphi study.

The intended audience of *OEIS*, however, was larger. It addressed the whole range of allegedly inexact sciences and showed them that they were indeed sciences. Despite this large intended audience, the impact of *OEIS*—if measured on citations—concentrated on the literature on Delphi.[15] Here, its reception consisted in ennobling Delphi and other prospective procedures involving expert opinion as scientific methods as contrasted to mere techniques of the policy analyst. A precondition for this was the presence of a comprehensive and accessible set of evidence which the experts had to evaluate based on their prior and partly implicit knowledge. This procedure, *OEIS* argued, was the optimal way to use expertise in prognosis. But as the next chapter shows, this message got lost.

At the same time, *OEIS* attempted to change the prevalent philosophy of science. It showed that the idea that explanation and prediction were logically identical held only for a minority of scientific reasoning. It urged philosophers and scientists alike to see that while both explanation and prediction are required characteristics of any field to be called a science, many branches and disciplines had to rely on experts to deliver predictions. And it wanted to achieve a general acknowledgment of expert opinions and estimations as crucial elements of philosophy of science. Regardless of whether they answered questionnaires or participated in group discussions; and regardless of whether they were capable of producing predictions, forecasts, or prospections: experts had to become integrated into epistemological thinking. Thus, while they certainly adhered to positivist tenets (cf. Brown 1977, 5; Dayé 2014, 2016), what they proposed was nothing less than a social epistemology *avant la lettre*.

Notes

1. The quotation is from a 1989 meeting of six former RAND researchers, among them Robert Specht, Hans Speier, and Albert Wohlstetter. The meeting was organized by the team of the RAND Oral History Project carried out by the Smithsonian Institution's National Air and Space Museum (cf. Collins 2002). This was also the context of the interview with Speier quoted above.
2. Helmer and Rescher also occasionally met with Rudolf Carnap to discuss issues of induction and probability (cf. Rescher 2007, Chap. 8). Carnap had joined UCLA's Department of Philosophy in 1954 after the unexpected death of Hans Reichenbach the year before.
3. This published version was then re-issued in February 1960 as RAND report R-353 (Helmer and Rescher 1960).
4. It is thus a crude misinterpretation to claim that "By the inexact science they meant the social sciences" (Andersson 2018, 87).
5. The resemblance of this argument to the one famously made by C. P. Snow in his talk on *The Two Cultures* ([1959] 2012) is striking. All three authors lament a dichotomy—exact versus inexact, scientific versus literary culture. However, they differ in their assessment of which side of this lamentable dichotomy had suffered more from it. To Helmer and Rescher, these were primarily the social sciences—to Snow, the natural sciences. While Helmer and Rescher were interested in the nature of science and thus argue on the level of epistemology, Snow talked about different social collectives, putting forth a scathing analysis of contemporary culture. In his view, the incomprehension toward science on the part of literary intellectuals, who tended to praise traditional culture, had a major impact on modern society. "That total incomprehension gives, much more pervasively than we realise, living in it, an unscientific flavour to the whole 'traditional' culture, and that unscientific flavor is often, much more than we admit, on the point of turning anti-scientific. The feelings of one pole become the anti-feelings of the other. If the scientists have the future in their bones, then the traditional culture

responds by wishing the future did not exist. It is the traditional culture, to an extent remarkably little diminished by the emergence of the scientific one, which manages the western world. This polarisation is sheer loss to us all. To us as people, and to our society" (Snow 2012, 11). In a 1957 RAND paper, "The Prospects of a Unified Theory of Organizations," Helmer appears to address exactly this form of pessimism: "There still seems to be an unfortunate attitude prevalent, according to which the social sciences constitute a sort of second-class realm of sciences which will never attain the purity, at times culminating in axiomatization, of the exact sciences. This attitude has done a great deal of harm. On the one hand it has produced an inferiority complex among the social scientists and put them on the defensive to such an extent that many among them have rationalized their intuitive procedures as a necessary concomitant of the essentially vague nature of their subject matter. In many non-social scientists, on the other hand, this attitude seems to have led to a prolonged state of resignation, a feeling that nothing can be done about a social science until that happy day in the distant future when it will suddenly turn out to have matured and someone will construct a neat unified theory with basic terms and axioms and definitions and theorems" (Helmer 1957, 6–7).

6. This line of thinking, in turn, shows peculiar similarities to the argument developed some 15 years earlier by their colleague and friend, Carl G. Hempel. In his classic text, "The Function of General Laws in History," Hempel (1942, 1965) argued that in historical arguments, laws have a function similar to the one they have in the natural sciences, albeit this function is not always acknowledged. However, there is no reference to Hempel's text in *OEIS*.
7. "Writers on the methodology of the physical sciences often bear in mind a somewhat antiquated and much idealized image of physics as a very complete and thoroughly exact discipline in which it is never necessary to rely upon limited generalizations or expert opinion. But physical science today is very far from meeting this ideal" (Helmer and Rescher 1959, 30).
8. "[T]he same formal analysis […] applies to scientific prediction as well as explanation. The difference between the two is of a pragmatic character. […] [A]n explanation is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for prediction the phenomenon under consideration.—Consequently, whatever will be

said in this article concerning the logical characteristics of explanation or prediction will be applicable to either, even if only one of them should be mentioned" (Hempel and Oppenheim 1948, 138). The same position is reflected in Hans Reichenbach's ([1944] 1965) coinage of "predictability" and "postdictability."

9. These are Scheffler (1957) and Rescher (1958). The latter explicitly refers to Hempel and Oppenheim's 1948 paper.
10. This is presumably the first published mentioning of the Delphi technique. A search on JSTOR corroborates this impression.
11. The probability is the result of dividing the average with the days per year. For the data, see http://www.nws.noaa.gov/climate.php/xmacis.php, a webpage run by the *National Weather Service* of the *National Oceanic and Atmospheric Administration*.
12. The formulation by Helmer and Rescher (1958, 23) is that given

    $E = Pa_1 \cdot Pa_2 \ldots Pa_m \cdot\ -Pa_{m+1} \cdot\ -Pa_{m+2} \ldots\ -Pa_n$ and

    $H =$ '$P(a_{n+1})$', the degree of confirmation is

    $$dc(H,E) = \frac{m}{n}$$

    This formulation, however, is missing from the published paper. Note that here the hypothesis is defined as being concerned with a single event not contained in *E*, and not as a general or even law-like statement. The same concept of hypothesis is used in Hempel and Oppenheim's (1945) paper on the "Definition of 'Degree of Confirmation," to which we will return later.
13. The white-shoes red-herring examples here are taken from later discussions, for example, between Hempel (1967) and I. J. Good (1967). The example originally used by Hosiasson-Lindenbaum (1940, 136) is "Every man is mortal," and "This chair is not mortal, and is not a man" and so on. She attributes this example to Hempel, though without any citation.
14. In formulaic expression, this condition is:

    $dc(H, E) + dc(\sim H, E) = 1$
15. However, some of the thoughts contained on its pages found their way into other branches of the social sciences. As one important instance, one should probably mention that Harold Garfinkel, in *Studies in Ethnomethodology* (Garfinkel 1967), cited at length from the parts of *OEIS* describing quasi-laws. Garfinkel drew a parallel from quasi-laws to his characterization of accounting practices: "When members' accounts

of everyday activities are used as prescriptions with which to locate, to identify, to analyze, to classify, to make recognizable, or to find one's way around in comparable occasions, the prescriptions […] are law-like, spatiotemporally restricted, and 'loose'" (Garfinkel 1967, 2).

References

Andersson, Jenny. 2018. *The Future of the World: Futurology, Futurists, and the Struggle for the Post Cold War Imagination*. Oxford and New York: Oxford University Press.

Bessner, Daniel. 2015. Organizing Complexity: The Hopeful Dreams and Harsh Realities of Interdisciplinary Collaboration at the Rand Corporation in the Early Cold War. *Journal of the History of the Behavioral Sciences* 51 (1): 31–53. https://doi.org/10.1002/jhbs.21699.

Brown, Richard H. 1977. *A Poetic for Sociology: Toward a Logic of Discovery for the Human Sciences*. Cambridge: Cambridge University Press.

Campbell, Virginia. 2004. How RAND Invented the Postwar World. *Invention & Technology* 2004: 50–59.

Collins, Martin J. 2002. *Cold War Laboratory: RAND, the Air Force, and the American State, 1945–1950*. Washington and London: Smithsonian University Press.

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

———. 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. https://doi.org/10.1177/0952695116664838.

———. 2018. The Expert as Messenger: Media Philosophy and the Epistemology of the Inexact Sciences. In *Dynamiken der Wissensproduktion. Räume, Zeiten und Akteure im 19. und 20. Jahrhundert*, ed. Wolfgang Göderle and Manfred Pfaffenthaler, 239–258. Bielefeld: Transcript.

Garfinkel, Harold. 1967. *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.

Good, I.J. 1967. The White Shoe Is a Red Herring. *British Journal for the Philosophy of Science* 17 (4): 322.

Goodman, Nelson. 1946. A Query on Confirmation. *The Journal of Philosophy* 43 (14): 383–385.

Helmer, Olaf. 1957. The Prospects of a Unified Theory of Organizations. P-1053. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P1053.html.

Helmer, Olaf, and Paul Oppenheim. 1945. A Syntactical Definition of Probability and of Degree of Confirmation. *The Journal of Symbolic Logic* 10 (2): 25–60.

Helmer, Olaf, and Nicholas Rescher. 1958. On the Epistemology of the Inexact Sciences. P-1513. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P1513.html.

———. 1959. On the Epistemology of the Inexact Sciences. *Management Science* 6 (1): 25–52.

———. 1960. On the Epistemology of the Inexact Sciences. R-353. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/reports/R353.html.

Hempel, Carl G. 1942. The Function of General Laws in History. *The Journal of Philosophy* 39 (2): 35–48. https://doi.org/10.2307/2017635.

———. 1943. A Purely Syntactical Definition of Confirmation. *The Journal of Symbolic Logic* 8 (4): 122–143.

———. 1945a. Studies in the Logic of Confirmation (I.). *Mind* 54 (213): 1–26.

———. 1945b. Studies in the Logic of Confirmation (II.). *Mind* 54 (214): 97–121.

———. 1946. A Note on the Paradoxes of Confirmation. *Mind* 55 (217): 79–82.

———. 1965. The Function of General Laws in History. In *Aspects of Scientific Explanation, and Other Essays in the Philosophy of Science*, 231–243. New York and London: The Free Press.

———. 1967. The White Shoe: No Red Herring. *British Journal for the Philosophy of Science* 18 (3): 239–240.

Hempel, Carl G., and Paul Oppenheim. 1945. A Definition of 'Degree of Confirmation. *Philosophy of Science* 12 (2): 98–115.

———. 1948. Studies in the Logic of Explanation. *Philosophy of Science* 15 (2): 135–175.

Hosiasson-Lindenbaum, Janina. 1940. On Confirmation. *The Journal of Symbolic Logic* 5 (4): 133–148.

Reichenbach, Hans. 1965. *Philosophic Foundations of Quantum Mechanics*. University of California Press.

Rescher, Nicholas. 1958. On Prediction and Explanation. *British Journal for the Philosophy of Science* 8 (32): 281–290.

———. 1997. H2O: Hempel-Helmer-Oppenheim, An Episode in the History of Scientific Philosophy in the 20th Century. *Philosophy of Science* 64 (2): 334–360.
———. 2002. *Enlightening Journey: The Autobiography of an American Scholar*. Lexington Books.
———. 2006. The Berlin School of Logical Empiricism and Its Legacy. *Erkenntnis* 64 (3): 281–304.
———. 2007. *Autobiography*. Walter de Gruyter.
Rohde, Joy. 2013. *Armed with Expertise: The Militarization of American Social Research During the Cold War*. Ithaca and London: Cornell University Press.
Scheffler, Israel. 1957. Explanation, Prediction and Abstraction. *British Journal for the Philosophy of Science* 7 (28): 293–309.
Snow, C.P. 2012. *The Two Cultures*. With Introduction by Stefan Collini. 15th printing. Cambridge: Cambridge University Press.
Speier, Hans. 1989. Introduction: Autobiographical Notes. In *The Truth in Hell, and Other Essays on Politics and Culture 1935–1987*, 3–32. New York and Oxford: Oxford University Press.
Thomas, William. 2015. *Rational Action: The Sciences of Policy in Britain and America, 1940–1960*. Cambridge, MA: MIT Press.
Whiteley, C.H. 1945. Hempel's Paradoxes of Confirmation. *Mind* 54 (214): 156–158.

6

The Boredom of the Crowd: The Long-Range Forecasting Delphi, 1963–1964

A Delayed Innovation

In contrast to political gaming, whose proponents had been instrumental in transferring the technique to other places, RAND had initially hesitated to disseminate the Delphi approach. Though originally written eight years earlier, in 1951, the report on the first Delphi study (Dalkey and Helmer 1963) was made publicly available only four years after "On the Epistemology of the Inexact Sciences" (Helmer and Rescher 1959; hereafter *OEIS*) had contained the first published mention of Delphi. This might have created a demand from outside RAND for more information about this method at a time when, as discussed in Chap. 2, RAND was broadening its research agenda. Also, RAND had already experienced the successful dissemination of other approaches to policy analysis and planning. Apart from game theory and political gaming, this applied first and foremost to systems analysis, a methodology developed by RAND engineer Ed Paxson in the late 1940s that combined elements of game theory, decision science, and expert judgment procedures. Ten years later, as former employee Daniel Ellsberg remembered, RAND had become "the sort of Vatican of the

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*, Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_6

church of systems analysis" (Interview with Daniel Ellsberg by the author, 11 May 2010, p. 1). Being the fundament of the Planning Programming Budgeting System (PPBS), systems analysis was introduced into federal budgeting when Charles J. Hitch (1910–1995), former head of the Economics Division, moved from RAND to Washington to become Assistant Secretary of Defense and Comptroller of the Department of Defense. From there, systems analysis was disseminated in various spheres of social and societal planning (cf. Hounshell 2000; Jardini 2000, 2013; Light 2003; Sapolsky 2004).

There was also another and probably more important reason that led RAND management to declassify the abridged report: its results were outdated. The invention and perfection of Inter-Continental Ballistic Missiles (ICBMs) and the successful launch of Sputnik in October 1957 had rendered the study's results obsolete, because the original study had started from the then valid assumption that atomic bombs would have to be delivered by airplanes. Thus relieved of its risk to reveal the nation's vulnerabilities, the report could now be reshaped to present the study's methodological design. With these arguments, in all likelihood, Helmer managed to convince RAND's management that disseminating Delphi might contribute to increasing the attractiveness of RAND work to organizations, agencies, and businesses in need of prognoses.

The decision to release the report of the first Delphi study was a step in a coordinated effort to foster the use of Delphi within and beyond RAND. When the report appeared in *Management Science* in October 1963, Theodore J. Gordon and Olaf Helmer had already started a Delphi study concerned with predictions into the more distant future. After 12 months of study, the "Report on a Long-Range Forecasting Study. P-2982" (Gordon and Helmer 1964) was issued in September 1964 and made publicly available via the RAND bookstore. This study appears to have been the decisive factor in the breakthrough of Delphi, especially after Helmer included it, with only minor adaptations, in his widely read book *Social Technology* (Helmer 1966): "Not until after the 1966 publication of Helmer's Social Technology," Nicholas Rescher (1997, 353) notes, "did Delphi effectively penetrate beyond the RAND Corporation orbit."[1] Upon receiving a master's degree in aerodynamics from the Georgia Institute of Technology, Theodore Jay Gordon (1930–)

became an engineer for the Missiles and Space Department of McDonnell Douglas Aircraft Company, RAND's earlier host organization. Quickly climbing the hierarchies, he was a test conductor for early Thor rocket launches (1958–1960), some of which carried vehicles of the Pioneer-type and managed to return data on space and the moon. Based on this experience, he became Douglas' chief engineer for the third stage of the Saturn rocket in Cape Canaveral, Florida. Further, together with the journalist Julian Scheer, he wrote a book about these launches entitled *First Into Outer Space* (Gordon and Scheer 1959). In his view, this led to his being listed as one of the 100 outstanding young leaders proclaimed by *Life Magazine* in 1962.

> I was selected by *Life Magazine* as an Outstanding Young Man, one of the hundred Outstanding Young Men of the Country, right, so big pictures, and I decided, because of that honor, to write a number of books. […] I had this idea that we could learn something reasonable and important about the future based on what it was that scientists were working on today. Simply ask the scientists what the hell they're working on! And what they expect to achieve by working on it. And that will give us some picture at least of what lies ahead. (Interview with Theodore J. Gordon by the author, 16 August 2013, p. 2)

One of the scientists Gordon decided to interview for his book was RAND scientist Richard Bellman, who at the time was developing linear programming (cf. Gordon 2011, 1099).

> So I called and I said I like doing interviews. He was very courteous and said "Sure, come on over." And I went over there and thought we'll be in a conference room just the two of us, and I'd say "What are you working on, Dick?" And he'd tell me and we go on from there. But instead, it was he and I and then the room was lined with three or four other people who were there to observe, Olaf [Helmer] among them. And I said "Why are you here?" And he said "We're working on something similar to that, and we want to observe your interview technique." Technique? [laughs]—I had no technique, I was going to talk to this guy. So, at the end of that interview, which went for a while, Olaf says "Listen, I would like you to be a consultant to RAND and we'd like to pay you for that consultancy well." (Interview with Theodore J. Gordon by the author, 16 August 2013, pp. 2–3)

After arranging that he could use some of the materials for his book, Gordon accepted. This, however, was no trivial arrangement. The RAND administration held the opinion that knowledge produced by people on its payroll was RAND property. Arguing that they were produced with public money, it even kept the royalties on books published by RAND researchers (cf. interview with Hans Speier by Martin J. Collins, 5 April 1988, p. 59), unless they were completely unrelated to RAND works, like Speier's translation of a book by German seventeenth-century writer Hans Jakob Christoffel von Grimmelshausen. However, Gordon got the arrangement he sought for, and used some of the materials from the RAND study in his book *The Future* (Gordon 1965).

At RAND, the study by Gordon and Helmer was accompanied by another study designed to assess the potential of Delphi. Supported by Helmer, Bernice L. Brown carried out a study that focused on how to increase the likelihood of identifying "good" predictive experts. Originally published as "Improving the Reliability of Estimates Obtained from a Consensus of Experts. P-2986" (Brown and Helmer 1964) in September 1964, it later also found its way into Helmer's *Social Technology*. However, it was the report by Gordon and Helmer that had the most sustained impact on the then emergent future studies community. They decided to modify several features of the earlier Delphi design. Most importantly, the number of participating experts increased markedly, and they omitted the collaborative compiling of a set of empirical evidence and relevant information. None of the papers on Delphi, however, not the coeval ones nor those written later, reflected on these modifications. Rather, Gordon and Helmer claimed their approach to be in line with the previous work on the method. Because their report was widely distributed and read, their specific approach to the Delphi method became paradigmatic.

Solutions—technical as well as theoretical—to scientific problems become paradigmatic when they are understood as examples of best practice; such paradigmatic cases can also replace explicit rules.[2] It is in this sense that it appears justified to credit the 1964 study as paradigmatic: Delphi now had a shape, and the 1964 study had defined it. The technique was received as means for long-range prognosis—a reception that, in turn, crucially informed its further trajectory. The phase of invention and initial development of Delphi had ended. The methodology saw further consolidation. After the publication of the RAND studies, there was not only a bigger pool of researchers using the technique, but there was also a paradigm, an example that guided its further development. In this sense, secondly, to say that the methodology was consolidated is to say that there were less degrees of freedom to think about it.

"Fifty Years into the Future": Delphi's Baptism of Fire

The "Report on a Long-Range Forecasting Study. P-2982" (Gordon and Helmer 1964) became paradigmatic not only in its almost exclusive reliance on quantitative questionnaire items and its restriction of feedback to aggregate results from previous rounds, but also in the presentation of results and in its general narrative in which the technique was embedded. The introduction emphasized the crucial relevance of foreknowledge: "Prediction-making is a fundamental part of technological, military, commercial, social, and political planning in the modern world" (Gordon and Helmer 1964, v). However, whereas short-term prediction was common and reasonably accurate and trustworthy, many decisions required the estimation of longer periods. The extension of the time period, of course, entailed specific epistemic 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)

While nobody could reasonably expect all predictions about the world in 50 years to come true, Gordon and Helmer argued that such long-range predictions still offered orientation. "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).[3]

This "pragmatic" position, they continued, had to be maintained in face of the conceivable criticisms put forth against the procedure (cf. Gordon and Helmer 1964, vi): that it showed insufficient inherent reliability; that it produced self-fulfilling or self-destroying prophecies; or that it could never predict the unexpected. All of these points of criticism were, in some sense, justified. However, "[o]ne must judge the merits or promise of an approach such as this in terms of the alternatives available" (Gordon and Helmer 1964, vi). And as long as there were no better alternatives available, a sufficiently well established and transparent methodical structure was obviously better than mere intuition. The specific contribution of a long-range Delphi consisted in initiating a "process of sifting the likely from the unlikely among the contingencies of the future" (Gordon and Helmer 1964, 2).

The longer temporal outlook of this study—set to 50 years into the future—distinguished the 1964 study from earlier RAND efforts. Gordon and Helmer attempted to elicit forecasts of individual experts in six broad areas. These areas were (1) Scientific breakthroughs; (2) Population control; (3) Automation; (4) Space progress; (5) War prevention; and (6) Weapon systems (cf. Gordon and Helmer 1964, 2). For each of these areas, Gordon and Helmer decided to set up a panel of experts, but opted for a bottom-up approach, allowing participants to decide which panels they felt expert enough to contribute to. They contacted about 150 persons, and 82 responded to one or more consecutive questionnaires (response rate = 54.7%). Little more than half of the respondents were RAND employees (35 persons) or consultants (7 persons). The remaining 40 persons had no official connection to RAND. Six of those persons with no connection to RAND were from Europe.[4]

Obviously, Gordon and Helmer continued their attempts to recruit participants even after the first phase of their Delphi had been completed.

The papers of Stanford geneticist Joshua Lederberg (1925–2008), co-winner of the 1958 Nobel Prize in Physiology and Medicine, contain a letter signed by Olaf Helmer. Machine-typed, dated 8 October 1963, and marked—in a manner typical for RAND—as letter L-20339, the letter opened by acclaiming the accolades of Lederberg:

> Dear Dr. Lederberg:
>
> Your outstanding work in the fields of physiology and genetics has suggested to me that you might possibly be interested in a rather unusual project, which has been underway for some time, and I am writing to ask whether you might wish to participate in it at its present stage.The project is concerned with the development of better techniques for assessing the direction of long-range scientific and technological trends and their probable effect on our society and our world. The RAND Corporation, under whose auspices we are undertaking this work, is engaged in a variety of studies, all related in a general way to the security of the United States and in many instances to plans and policies regarding the relatively distant future. The particular approach in this project is a new one; it is not a technique that we have used, except experimentally, in our work thus far. […].
>
> We have been experimenting for almost 15 years with various approaches to such problems and have evolved several promising techniques. (Some of these, incidentally, have been subjected to empirical verification, with gratifying results.) The essence of the method to be employed in this present study is a controlled opinion feedback in which a panel of experts exchanges reasoned opinions anonymously and through an intermediary. This feedback tends to produce a converging group consensus. We are at present using this method to examine questions in various fields related to the future in 10 to 50 years hence; you are particularly invited to join in correspondence a panel which is investigating questions concerned with scientific breakthroughs. […]
>
> We sincerely hope that you will be intrigued by this project and will consent to join us. If you elect to participate, please send us a note to this effect. We will then mail you a questionnaire currently in preparation, which will first summarize whatever consensus seems to have been reached thus far by the panel you are being asked to join, and then invite you to give us your own appraisal of these findings. (Letter from Olaf Helmer to Joshua Lederberg, October 8, 1963, The Joshua Lederberg Papers, U.S. National Library of Medicine, Box 19, Folder 14)

While it is unclear whether Lederberg decided to join, we still get a feeling of both the effort involved in conducting the study and the way its design was introduced to potential participants. Each panel of experts was provided 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).[5] However, in order to keep all participants informed on the current state of the other panels, each envelop contained all six questionnaires. The questionnaire that the participating expert was expected to fill in was printed on paper of a different color. Nonetheless, some experts felt interested in more than one questionnaire and filled in several of them. In total, 348 questionnaires were sent back to RAND.

Obviously, the methodological procedure differed considerably from the first Delphi study (Dalkey and Helmer 1962), and this difference was caused in large parts by the decision to include a larger amount of participants. However, the 1964 Delphi also differed from the methodological considerations put forth in *OEIS*. Most prominently, it avoided the collaborative compilation of a set of evidential materials. Also, while some of the questionnaires developed by Gordon and Helmer inquired into the reasons behind the experts' estimates, these reasons were never fed back to the other panel members.

While the procedure was identical for all panels, the panel of scientific breakthroughs (Panel 1) served as an example for introducing the procedure. The first questionnaire distributed to this panel was open. After describing the general objective of the study and defining the forecasting time of 50 years, it asked the experts to list "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 then compiled in a list, with multiple nominations being listed only once. In total, the list comprised 49 possible inventions or breakthroughs. In a second step, this list was distributed to the participants with the task to estimate for each list entry a date of implementation. The questionnaire item, however, was structured more complexly. The study authors had predefined time intervals (Gordon and Helmer 1964, 7):

1963–1965	1972–1978	1997–2013
1965–1968	1978–1986	Later than 2013
1968–1972	1986–1997	Never

These time intervals were added to every item on the list, asking the participants 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, medians and quartiles were calculated for these 50% values.

It must be emphasized that all the values reported in the following by Gordon and Helmer rely on this 50% value. The consequences of this procedural decision cannot be underestimated. Above all, this meant that the values reported did not represent the year by which the experts were certain that an invention would be implemented, but rather a year in which the experts saw the chances of its implementation evenly distributed. This, in turn, weakened the study's central messages. Gordon and Helmer did not provide any justification for this 50% threshold. One might speculate that by so doing, one accounts for both discontinuous advances and the non-linear acceleration of scientific and technological innovation. However, this threshold rendered the displays of results that were included in later sections of the report potentially misleading. As shown below, the displays leaned heavily toward taking the figures as if the experts agreed that the innovation or event would certainly take place at the given point in time. Moreover, this methodological decision offered a huge problem with regard to verification. Whereas it is possible to determine the point in time in which a specific event took place or an innovation was generally implemented and accepted in society, it is impossible to determine when the chances for this event or innovation of becoming real were even. The decision to use the 50% personal probability threshold rendered the prognosis per se unverifiable, or at least did much to avoid any sensible evaluation *post festum*.

Gordon and Helmer used three exemplary items to demonstrate the further processing of data. These items were "Chemical control over heredity—molecular biology," "Popular use of personality control drugs," and "Reliable weather forecasts," items no. 19, 10, and 6, respectively. As

Table 6.1 Exemplary prediction items, results of round two. (Adopted from Gordon and Helmer 1964, 8; own calculations)[6]

Events	Median	Quartiles	Range of quartiles
19. Chemical control over heredity	1993	1982–2033	51
10. Use of personality control drugs	2050	1984–2050	66
6. Reliable weather forecasts	1975	1972–1988	16

described above, the participating experts had allocated estimated probabilities to the periods predefined by the study authors. Gordon and Helmer used the resulting distribution to determine the 50% implementation probability of each item for each respondent, and used these data in aggregation to calculate the median and the two quartiles (see Table 6.1).

These values can be read as saying that one quarter of the interviewed experts saw a 50% likelihood that chemical control over heredity would be achieved prior to 1982 ("the lower quartile"). Further two quarters estimated that this 50% likelihood would be achieved between 1982 and 2033, with the median of 1993 separating these two quarters. The final quarter expected this even distribution to become real only after 2033 ("the upper quartile"). The authors concluded that 10 of the 49 entries in the list of potential breakthroughs were more or less consensual among the participants, item 6 on reliable weather forecasts among them. How exactly they defined a "reasonable consensus" can only be guessed. Clearly, however, the range of the area between the 25% quartile and the 75% quartile is far narrower than those of the other two exemplary items (16 years compared to 51 respectively 66).

In a next, the third, round, the questionnaire singled out the ten consensual events, providing the median and quartile values and the information that this had been understood as indicating consensus among the experts. The participants were invited to indicate whether they agreed with this understanding, or whether they objected. If so, they were asked to provide reasons. For the 39 non-consensual items, the procedure differed. The "experimenters at this point used their discretion in singling out a subset of 17 items which they thought to be deserving further exploration" (Gordon and Helmer 1964, 8). This reduced list of items formed a second part of the third questionnaire. Further, brief verbal information on the distribution of the estimates was provided (e.g., "Consensus that it will occur; disagreement as to when"). In some cases, the formulations were modified "because it was felt that the ambiguity of the original phrasing, rather than any factual disagreement among the participants, might have been partly responsible for the observed divergence of responses" (Gordon and Helmer 1964, 8f). In retrospect, Gordon remembered the struggle for clear formulations as unforeseeably complicated. "Asking the question, and making it a single question rather than a multiple question, so that everybody is answering the same question and has the same understanding is essential to the whole process. […] This was a lesson that we had to learn a hard way" (Interview with Theodore J. Gordon by the author, 16 August 2013, p. 15). For instance, event no. 10 had originally read "use of personality control drugs." In the third round, it was reformulated to read "Widespread socially accepted use of non-narcotic personality drugs producing specific psychological reactions" (cf. Gordon and Helmer 1964, 9). With these modifications made, round three resulted in a narrowing of the range between the two quartiles as well as in numerical changes in the median—especially for the non-consensual items. While having been 1993 and 2050 for items no. 19 and no. 10, respectively, both settled at the year 2000 after this third round.

The fourth round followed the same procedure. Gordon and Helmer were able to move some items from the list of non-consensuals to the list of consensuals. Further, a few items were eliminated; and some were rephrased in order to avoid misunderstandings. Item no. 10, for instance, now read "Widespread *and* socially *widely* accepted use of non-narcotic drugs *(other than alcohol) for the purpose of* producing specific *changes in personality characteristics*" (Gordon and Helmer 1964, 10; emphases added). Finally, for the non-consensual items, they added a brief synopsis of the reasons of those participants who in the previous round had expressed their disagreement with the majority opinion. Thus, while for item no. 10, the "majority consensus to date" was that the 50% year was reached in 2000, the "minority opinion" was that it "[w]ill take 50 years or more, because research on psycho-pharmaceuticals has barely begun, and negative social reaction will cause delays" (Gordon and Helmer 1964, 10).

Table 6.2 Collected results for items nos. 19 and 10. (Adopted from Gordon and Helmer 1964, 8, 9, 10; own calculations)[7]

	Items	
	19. Chemical control over heredity	10. Use of personality control drugs
Round two		
Median	1993	2050
Quartiles	1982–2033	1984–2050
Range between quartiles	51	66
Round three		
Median	2000	2000
Quartiles	1989–2015	1980–2033
Range between quartiles	26	53
Round four		
Median	2000	1983
Quartiles	1990–2010	1980–2000
Range between quartiles	20	20

Again, this procedure resulted in changes in the median and the quartile values. Whereas the median of item no. 19 remained on the year 2000, it clearly dropped to 1983 for item no. 10. More important, however, for the authors was that the range of the two central quartiles also decreased significantly enough to interpret them as “reasonably narrow consensus” (Gordon and Helmer 1964, 10). A summary of the results is given in Table 6.2. As with the earlier 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 interpreted as consensus of experts. This interpretation was reasonable, in the view of the authors, and we can only guess that their justification for this conclusion was that the range between the two quartiles did not exceed 20 years.

The World of the Future

In all five expert panels, Gordon and Helmer’s efforts resulted in a chronological ordering of possible events. A more complete picture of the world 50 years ahead could then be derived by synthesizing all these separate events and their expected dates of realization. However, the authors warned against overstating the character of the ensuing statements. The future described on the following pages was a potentiality, not a pre-determined fate. While meeting standards of transparency and intersubjectivity, the study data were of anticipatory, reckoning character. The results thus consisted "of summaries of considered opinions about the future by a small group of people, each an expert on some, but not necessarily all, of the subjects under inquiry" (Gordon and Helmer 1964, 11). Furthermore, the interaction within the various panels occurred at different levels of quantification versus qualification. In some panels, like Panel 1 on scientific breakthroughs, the interaction showed a comparatively high degree of quantification. Here, the results were presented graphically in a manner that allowed for a quick impression of both the median and the distribution of the experts' estimates. For each item, a bar was drawn which reached from the lower to the upper percentile, thus depicting 50% of the answers received. The location of the median was represented by a peak in the bar's thickness, which makes the symbols look like little houses. In the graph representing the final results, the items are put in ascending order along their median, the first item thus being the one for which the point in time at which the chances of feasibility are even has been assessed to be the nearest to the date of the report's publication. The list of items is (Gordon and Helmer 1964, 13):

1. Economically useful desalination of sea water
2. Effective fertility control by oral contraceptive or other simple and inexpensive means
3. Development of new synthetic materials for ultra-light construction
4. Automated language translators
5. New organs through transplanting or prosthesis
6. Reliable weather forecasts
7. Operation of a central data storage facility with wide access for general or specialized information retrieval
8. Reformation of physical theory, eliminating confusion in quantum relativity and simplifying particle theory
9. Implanted artificial organs made of plastic and electronic components

10. Widespread and socially accepted use of non-narcotic drugs (other than alcohol) for the purpose of producing specific changes in personality characteristics
11. Stimulated emission ("lasers") in X and Gamma ray region of the spectrum
12. Controlled thermonuclear power
13. Creation of a primitive form of artificial life (at least in the form of self-replicating molecules)
14. Economically useful exploitation of the ocean bottom through mining (other than off-shore oil drilling)
15. Feasibility of limited weather control, in the sense of substantially affecting regional weather at acceptable cost
16. Economic feasibility of commercial generation of synthetic protein for food
17. Increase by an order of magnitude in the relative number of psychotic cases amenable to physical or chemical therapy
18. Biochemical general immunization against bacterial and viral diseases
19. Feasibility (not necessarily acceptance) of chemical control over some hereditary defects by modification of genes through molecular engineering
20. Economically useful exploitation of the ocean through farming, with the effect of producing at least 20% of the world's food
21. Biochemicals to stimulate growth of new organs and limbs
22. Feasibility of using drugs to raise the level of intelligence (other than as dietary supplements and not in the sense of just temporarily raising the level of apperception)
23. Man-machine symbiosis, enabling man to extend his intelligence by direct electromechanical interaction between his brain and a computing machine
24. Chemical control of the aging process, permitting extension of life span by 50 years
25. Breeding of intelligent animals (apes, cetaceans, etc.) for low-grade labor
26. Two-way communication with extra-terrestrials
27. Economic feasibility of commercial manufacture of many chemical elements from subatomic building blocks

28. Control of gravity through some form of modification of the gravitational field
29. Feasibility of education by direct information recording on the brain
30. Long-duration coma to permit a form of time travel
31. Use of telepathy and ESP [extrasensory perception] in communications

Figure 6.1 gives the corresponding graphical display of this Delphi study's results with Panel 1. As mentioned above, there is no indication that what is depicted is not certainty, but 50% confidence. Thus, to take as an example item no. 14, the median year in which the likelihood that the exploitation of the ocean bottom through mining (other than oil drilling) would have a 50% chance to become economically useful was 1989. The lower quartile was at 1980, the upper quartile at 2000. The range between the lower and the upper quartile was 20, which, as we saw above, qualified as consensus in the view of the study authors.

However, to read the results correctly, it is important to know that the graphical displays foreshortened the time scale after the year 2020: the final ten vertical lines did not denote years, but instead referred to the following intervals (cf. Gordon and Helmer 1964, 11):

2020 – 2025 – 2035 – 2050 – 2100 – 2200 – 2300 – 2400 – 2600 – 2800 – 3000 or never

Thus, for instance, the median year of item no. 29, "Feasibility of education by direct information recording on the brain," was not 2028, but 2600.

Apart from these quantitative results, the study authors also inquired into potential developments on the organizational level of scientific research and academic training. They noted "a strong consensus" concerning four developments, which the participants both consensually expected and considered desirable. These trends in the organization of science were: (1) a "reform of present modes of scientific communication through the use of automated information retrieval systems" (Gordon and Helmer 1964, 15)—a prediction we can verify today; (2) a process of "reorientation of scientific methodology toward greater interdisciplinary cooperation" (Gordon and Helmer 1964, 15)—a prediction which would

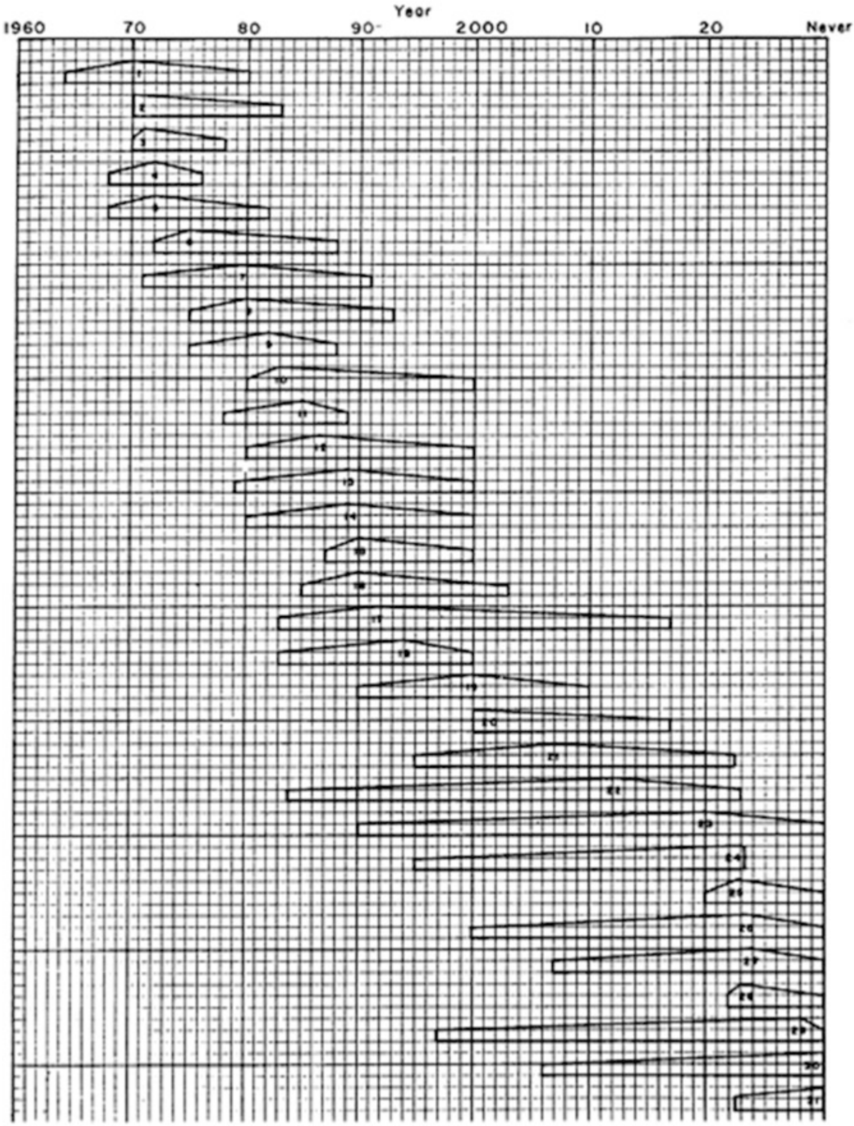

Fig. 6.1 Results of panel 1 on scientific breakthroughs. (Source: Gordon and Helmer 1964, 12; reproduced with permission)

require a sophisticated evaluation, but which has been verified at least on the discursive level; (3) "[i]ncreased emphasis on basic research in government-supported" research and development (Gordon and Helmer 1964, 15)—a prediction which is probably nothing more than wishful thinking; and finally, (4), the "[r]eformation of educational processes toward an increased interdisciplinary understanding of science" (Gordon and Helmer 1964, 15)—a prediction which would have to be assessed in a cross-national comparison, but which, at least from the European outlook, cannot be verified.

One panel that produced qualitative rather than quantitative data was Panel 5, on war prevention. This, in turn, required another way of presenting the results. In both the opening questionnaire in June 1963 and the final fourth questionnaire in January 1964, the experts of Panel 5 were asked three main questions. The first was to provide an estimate of the probability that another major war takes place within a 10-year and a 25-year period. Interestingly, the second estimates of January 1964 were on average significantly more positive than the first ones (see Table 6.3). In the January 1964 results, the distribution of the estimates showed a significantly higher density. Almost all values were reduced, which justifies the conclusion drawn by the authors that the participating experts now see the future more optimistic than in the first round.

Of course, the authors speculated about the reasons for this change. "While the identity of the panel membership was not stable enough to draw the conclusion […] that events of the intervening seven months had caused most of the respondents to take a rosier view of the future,"

Table 6.3 The probability of another major war. (Adopted from Gordon and Helmer 1964, 27)

	June 1963 (%)	January 1964 (%)	Difference (%)
10-year period			
Median	10	10	0
Lower quartile	3	1	−2
Upper quartile	33	20	−13
25-year period			
Median	25	20	−5
Lower quartile	5	4	−1
Upper quartile	50	30	−20

the authors argue that an "examination of the responses of those individuals who participated in both the first and fourth questionnaires did tend to confirm this hypothesis" (Gordon and Helmer 1964, 27). What had actually happened? The authors remain tacit on this point, and thus we have to rely on present-day Cold War historiography. Two major events happened between July 1963 and January 1964. First, the "Treaty banning Nuclear Weapons Tests in the Atmosphere, in Outer Space and Under Water," better known as Limited Test Ban Treaty (LTBT) was signed. Deliberations on test bans had been initiated by the Soviet Union in the early 1950s, but the details had been subject to long negotiations that endured for virtually more than a decade. Finally, the political leaders of the United States, the United Kingdom, and the Soviet Union agreed in mid-July 1963 on a test ban that allowed for underground testing, interestingly with the Soviet Union, who had preferred a more comprehensive ban, accepting the position favored by the other two atomic powers. It was signed by the representatives of the national governments on 5 August 1963 and entered into force by 10 October that year.[8] The second major event was, of course, the assassination of President John F. Kennedy in Dallas, Texas, on 22 November 1963. While it is comprehensible that the first event might have caused slight optimism among foreign policy experts, it is not clear whether and how the assassination of Kennedy could induce optimism. By January, the Warren Commission had only begun their examination of the event, and the possibility that the Soviet leadership was responsible, in one way or another, for the assassination was widespread and not yet refuted. The assumption that the experts' optimism was increased by the energetic appearance of Lyndon B. Johnson and the continuation of Kennedy's political projects which he pursued in the first months after his inauguration appears rather unlikely.

Clearly, another stream of events factually tangential to, albeit culturally interwoven with the Cold War must be mentioned. On 28 August 1963, roughly a quarter million people participated in the March on Washington for Jobs and Freedom, chanting the songs of, among others, Mahalia Jackson and Bob Dylan, and listening to Martin Luther King's famous words "I have a dream…" echoing over the Reflecting Pool below Lincoln Memorial. Though assessed critically by both conservative forces and the black nationalists around Malcolm X, these events certainly led to an increased general optimism among the moderate progressive and liberal thinking US Americans. Whether this also applied to the foreign policy experts participating in the study, must be left open.

The possibility that this "rosier view of the future" was an effect of the participation in a study that implied an intensive analysis of war prevention was not considered. One of the major tasks for the participants, described in more detail below, was to propose measures of war prevention. To see a long list of potential measures, some of which consensually assessed as being highly effective and probably implemented soon, might have had a considerable influence on the respondents.

Be it as it may, the second question concerned the conditions that could induce a major war. With regard to this question, however, it was found that the experts' opinions did not change much between the first and the fourth questionnaires. Leveling out the existing minor differences between the two rounds by calculating the average, one can summarize the probabilities estimated by the panel on four "modes of outbreak" (Gordon and Helmer 1964, 28):

Escalation of a political crisis	45%
Escalation in the level of violence in an ongoing minor war	37%
Inadvertence	11%
Surprise attack at a time when there is no ostensible acute crisis	7%

In general, these results give a picture of a community of experts who agreed that the main threats to global security were neither the proverbial accidental activation of the red button nor an unforeseeable, unilateral opening of a new conflict. Rather, the main danger arose from the possibility of escalation of already existing political or military conflicts, most urgently of course if one of the world's atomic powers was involved in it, directly or by alliance to an involved nation. As such, this result underlines the importance of global policy-making. Of course, it should be taken into account that this result is in line with the creed of the community of the "new civilian militarists" from which the study authors most probably recruited its participants. Underlining the importance of politics in the arena of global conflicts means to underline the importance of the "new civilian militarists."

As already mentioned, the third main question of Panel 5 dealt with future measures of war prevention. This was, in the view of Gordon and Helmer (1964, 28), the panel's "main assignment." In total, panel members submitted and, in later rounds, evaluated 42 proposals. Having established a list of proposals after the first round, the next rounds asked the participants to rate the "desirability" of each measure, its effectiveness if implemented, and its probability of implementation. Effectiveness, in this context, was defined specifically as the impact on lowering the probability of war.

Though these ratings came in numerical form, much substantial exchange on this issue was verbal rather than numerical, and Gordon and Helmer decided upon another form of presentation of results (see Table 6.4). The proposed measures were arranged according to their desirability rating from high to low desirability. If there was "considerable consensus" among the respondents, the respective entry was set in a frame. A question mark was used to denote measures where the participants had not been invited to assess the general desirability.

The way in which Gordon and Helmer chose to display these qualitative results looks similar to the tables resulting from earlier political gaming efforts. It gives an impression of the breadth of measures considered by the panel, while at the same time providing a ranking that follows the panel's assessment of the desirability, effectiveness, and probability of the proposed war prevention measures. To a present-day observer, some proposals are remarkable for their content. For instance, proposal no. 3 proposed that both sides, the United States and the Soviet Union, developed weapons that could not be damaged by the opponent's first strike and could then, and only then, be used for counter-measures, thus excluding the possibility that they be used for a first strike attack. Desirability, effectiveness, and probability of this measure were assessed as high, the latter even with a "reasonable consensus" among the experts. Another interesting single measure, item no. 14, considered the probability of a political coalition between the United States and the Soviet Union against China or other potentially powerful states. Interestingly, against the widely held view of the irreconcilable two powers, the probability that this option was realized was assessed as high by the panel—albeit without "reasonable consensus." However, the option of the initiation of

Table 6.4 List of war prevention measures. (Adopted from Gordon and Helmer 1964, 28–31)

	Proposed measure	Overall desirability	Effectiveness if implemented	Probability of implementation
1	Build-up of Western-bloc conventional forces	High	High	High
2	Increased security of command-and-control and retaliatory capacities	High	High	High
3	Development on both sides of invulnerable delayed-response weapons that are incapable of surprise attack	High	High	High
4	Greater political and economic unity among free advanced democracies	High	High	Medium
5	US-SU political agreement to seek peace and restrain other nations from developing nuclear weapons	High	High	Medium
6	Establishment of a standing worldwide UN police force, not subject to veto	High	High	Low
7	Improved defensive warfare techniques to reduce probability of escalation in limited wars	High	Medium	Medium
8	UN economic and military aid to areas threatened by political upheaval	High	Medium	Low
9	Development of a code of international law and establishment of effective world courts of justice and a world supreme court	High	Medium	Low
10	US-promoted rapid technological and economic advancement of under-developed nations	High	Low	High
11	Strengthening of the UN with the objective of forming a world government	Medium	High	Low
12	Bilateral US-SU arms control agreements	Medium	Medium	High
13	Studies by sociology, group psychology, and so on, seeking clues to war prevention	Medium	Medium	High
14	US-SU political association against China or other third party	Medium	Medium	High
15	Holding the status quo against even minor aggressions	Medium	Medium	Medium
16	Central-European disengagement to reduce military activity, induced by an improving SU-US atmosphere	Medium	Medium	Medium

(*continued*)

Table 6.4 (continued)

	Proposed measure	Overall desirability	Effectiveness if implemented	Probability of implementation
17	Instituting population control in all nations according to UN decisions	Medium	Medium	Low
18	Establishment of national assessment centers which would evaluate crisis situations and transmit policy statements to the potential enemy to clarify US intent	Medium	Medium	Low
19	US or SU demonstration of the intent to use force of increasing levels (in identifiable steps) in response to specific provocations	Medium	Medium	Low
20	Removal of trade barriers with Communist countries	Medium	Medium	Low
21	Development of realistic understanding among western Allies of dynamics of nuclear warfare, by techniques including joint US/Allied crisis and war gaming and systems analyses	Medium	Medium	Low
22	Settlement of the division of Germany on terms acceptable to West Germany and compatible with German membership in NATO	Medium	Medium	Low
23	Development of a cadre of international UN civil servants dedicated to world values	Medium	Low	High
24	Military alliance between US and SU (plus possibly India)	Medium	Low	Low
25	Support and promotion of a United States of Africa, Latin America, Europe, Asia	Medium	Low	Low
26	Invitation to other nations to become member states of the United States	Medium	Low	Low
27	Simulated US-SU war games, played by professional military planners of both sides (possibly with sides interchanged)	Medium	Low	Low
28	Increased cooperative economic, political, and military ventures by the US with the SU and China to promote interdependency	?	High	Low
29	Bilateral reduction of armaments enforced by UN police force	?	High	Low

(*continued*)

Table 6.4 (continued)

	Proposed measure	Overall desirability	Effectiveness if implemented	Probability of implementation
30	Strengthening of NATO alliance to insure [!] a guaranteed response to pre-stated provocation	?	High	Low
31	SU-initiated gradual improvement of political atmosphere	?	High	Medium
32	Strategic arms control (halting production but not R-and-D)	?	Low	Medium
33	Clear US statement as to which national interests are to be protected by nuclear deterrents, and orientation of our policies to that end	?	Low	Low
34	Development of a new system of international political cue "signals" which would indicate real intent to go to war unless political situation changes, such as general mobilization in the past	?	Low	Medium
35	Fostering educational and propaganda measures designed to amend or establish values of mutual toleration of various ideologies and the right to self-determination	?	Low	High
36	Sharing of technological innovations between the United States and SU	?	Low	Medium
37	Support of NATO, SEATO, and OAS to increase the number of world forums where political differences can be resolved with minimum "loss of face"	?	Low	High
38	Offer of nuclear weapons to countries which agree to support our stated national policies	?	Low	Medium
39	Organized encouragement of conscientious objection on the part of scientists to cooperation in the improvement of weapon systems	Low	Low	Low
40	Creation of buffer zones to avoid direct confrontation of major powers	?	Low	High
41	Recognition of Communist China and East Germany—creation of a realistic policy	?	Low	Medium
42	US-initiated unilateral steps toward disarmament	?	Low	Low

a unilateral disarmament process by the United States (item no. 42) was perceived as very unlikely, and in line with what had resulted in one of the political games described in Chap. 4, there was a consensus that such a measure would not be able to efficiently diminish the risk of a new major war. On a side note, we might further emphasize that this panel considered the use of war gaming approaches (items no. 21 and 27), but was skeptical toward the effectiveness and probability of these as measures of war prevention.

Some other things can be said by interpreting several of the items in combination. For instance, the potential impact of international organizations on the prevention of war on a global level was generally assessed to be both desirable and effective, but the probability of realization was generally assessed as low, most likely because the configuration of these organizations with the required authority was perceived as utopian (items no. 6, 8, 9, 11). If not on the level of formal international regulation, such organizations would have a moderate impact via their civil servants (item no. 23). Also, they could serve as bodies responsible for establishing and maintaining educational programs "designed to amend or establish values of mutual toleration of various ideologies and the right to self-determination," as item no. 35 reads.

In face of this perceived lack of effective international regulatory and juridical organizations in the near future, one measure could be the support of "under-developed nations" by the United States (item no. 10). This support should aim at increasing technological and economic standards in these countries. Interestingly, although the probability of this measure was assessed as high, the effect of this measure on the reduction of risk of war was assessed as low. In the eyes of the participating experts, the modernization of third countries did not significantly reduce the probability of a major war. Clearly, in this item, coeval modernization theory was put to test, but not questioned as a whole. While the experts considered the effectiveness of US-induced modernization in other countries, the item did not challenge the underlying ideology. The assumption of a continuum of development on which all countries or societies in the world can be located was itself not part of the question (on US modernization theory, see Latham 2000; Gilman 2003; Shah 2011; on how this theory played out in various other parts of the worlds, see Adalet 2018; Feichtinger 2011; Miralles 2015).

Finally, a group of items took up the issue of transparency in foreign policy and military strategy (items no. 18, 19, 33, 34). In view of the impression that the core tension in Cold War decision-making, and in fact some wrong strategic decisions we know of today, resulted from the lack of reliable information on the opponent and the inability, at least on the US side, to identify and establish a stable understanding of the opponent's foreign policy, these measures apparently address the root of the problem. However, the experts were not very optimistic about these measures. With one exception (item no. 34), the proposals to foster transparency were assessed as rather improbable. Game theory's core tenet that information asymmetry is one of the biggest assets in conflicts of interests had been well learned.

Methodological Lessons

While Gordon and Helmer's study thus covered a vast array of substantial issues, the authors also critically assessed the methodological lessons to be drawn from their study. Several of the issues emerged because of the decision to use a study design almost exclusively oriented on principles of quantitative social research. The first issue they addressed was the *instability of panel membership*. For a variety of reasons, it happened repeatedly throughout the study that participants dropped out. And as we have seen with Joshua Lederberg, additional experts were invited to join the panels after the first phases of research. Although some changes would not be problematic—"in fact, scientific progress in general relies on the constantly changing collaboration of many contributors" (Gordon and Helmer 1964, 57)—the authors suspected that too many changes inhibit the study's progress by counteracting the convergence effect. The authors suggested using some form of contractual agreement with the study participants as a means of diminishing the dropout rate.

The second issue addressed, *time lapse*, might have been a factor causing the already mentioned instability. "Too much time elapsed between successive rounds, the average lapse having been about two months" (Gordon and Helmer 1964, 57). A better preparation, Gordon and Helmer concluded, might have contributed to shortening the periods between the rounds, as would have—in times of postal service—the omission of overseas experts; then, the authors estimated, the time between two successive rounds could have been reduced to one month.

Ambiguous questions, the third issue on their list, has already been discussed above, and it suffices to mention it briefly and to note that the ambiguous wording of questions becomes a problem only if we considered the questionnaire to be a quantitative measurement device. If instead, as it was done in the first Delphi study, the questionnaire is conceived of as a form of communication (among a decisively smaller group of people), then the continuous specification and disambiguation of a question would have been no problem, but rather part of the solution.

Fourth, a critical remark concerned the *respondents' competence*. Even within the panels, some questions concerned several fields of expertise, and it was, "[w]ith all due regard for our eminent respondents, […] not reasonable to expect that each could be equally competent with regard to all of the areas touched upon by our questions" (Gordon and Helmer 1964, 58). Of course, this was a problem, because it entailed an inefficient use of the available expertise. The procedure blended the answers of the most competent panel members with those of the less competent members. To remedy this, Gordon and Helmer identified two possible measures. One was to define better-delimited fields of expertise, and to invite only persons who have demonstrably a high amount of expert knowledge in these fields. Another, which they stated to prefer, was to instruct participants only to answer questions which they felt competent in and to leave blank those for which their answer would be no more than a guess. An alternative option of this second measure would be to continue asking the participants to answer all the questions, but to include for each question "a self-appraisal of their degree of confidence in answering it. Precisely how this should be done is an open question which might be made the subject of a separate study" (Gordon and Helmer 1964, 58). As described below, at the time of writing, Bernice Brown, on the instigation and with the guidance of Helmer, was already involved in such a study. However, we see here a clear deviation from the program formulated in *OEIS*. Its authors, Helmer and Rescher (1958) had argued that in many relevant policy questions, several fields of knowledge are interwoven. Correspondingly, experts would have to interact to solve these questions. They would do so by collaboratively creating and evaluating a pool of empirical data and material (a set of evidence). Gordon and Helmer, in contrast, took the other direction. They proposed to have more differentiated fields and to assess a degree of self-confidence. This was required only because there was no attempt of collaboratively creating a set of evidence.

The fifth critical remark concerned the production of *self-fulfilling and self-defeating prophecies* by the Delphi methodology. One panelist had remarked that the results of Delphi might make themselves true (or wrong) precisely by being published. Apart from giving a rather harsh reply to this remark, Gordon and Helmer also took a second perspective on the problem.

> Leaving aside the implication—to which we emphatically do not subscribe—that the publication of the answers to some of our questions might in fact affect the future course of history with regard to the subject of the question (e.g. by hastening or retarding a predicted event), there still remains the possibility that a respondent's answer might be biased by his expectation (whether conscious or not) that the announcement may affect the truth of the prediction's outcome. (Gordon and Helmer 1964, 59)

This, however, was a problem of the expert's self-concept or, in our words, his or her epistemic role: "If this were so, then the respondent would cease to be acting as a pure predictor, but would in part become a would-be manipulator of the future" (Gordon and Helmer 1964, 59). There appeared to be no means to deal with the deeply personal decision whether an expert answers along moral or strictly scientific lines of thinking. Yet again, this also was a consequence of the decision to increase the pool of participants and the resulting omission of a sophisticated discussion of reasons and lines of thought behind the estimates.

An allegation also put forth by some of the study participants was that Delphi produced *consensus by undue averaging*. "The objection has been raised that the emphasis we place on the median as a descriptor of the group opinion and on the quartile range as a measure of the degree of consensus biases the outcome unduly against the far-out predictor, whose judgment may after all prove to be right while the majority opinion may be wrong" (Gordon and Helmer 1964, 60). In the view of the authors, this objection was justified, but only as regards the actual study, and not the principles of the Delphi design *per se*. They emphasized that it was an essential feature of the methodological design that those who strongly disagreed with the majority were asked to state their reasons for so doing. Subsequently, all panelists could inspect those reasons and evaluate their salience. "Thus a far-out opinion is in principle rejected only if its proponent fails to justify it before the rest of the panel" (Gordon and Helmer 1964, 61). However, the authors added that the actual study did not communicate the claim to provide reasons with much emphasis. "In retrospect," they admitted, "it seems that we should indeed have been more insistent on eliciting explicit reasons for minority opinions, and should have provided an opportunity for explicit critique of such reasons, even at the expense of an additional round if necessary" (Gordon and Helmer 1964, 61). The suggestion thus would be to include more space for qualitative communication at the expense of additional attempts to produce quantitative convergence of opinions. In assessing this suggestion, it appears valuable to relate this argument to the already mentioned study by Bernice Brown. Entitled "Improving the Reliability of Estimates Obtained from a Consensus of Experts. P-2986" (Brown and Helmer 1964), the study used Almanac-style questions on past or current facts in order to compare the estimates with true values. Here, as discussed later in more detail, it turned out that in a decisive number of cases, and despite the opinions converging, the final medians were far away from the true value. In fact, the final estimates in those cases were much closer to the initial estimates than to the true values—an effect nowadays known to social psychologists as anchoring-and-adjustment effect. If not methodological reasoning alone, then at least this finding could have triggered a move of the burden of justification from the group of deviants toward the majority, perhaps by making the disclosure of one's reasoning a general requirement, as it was done in the first Delphi study. However, neither Gordon and Helmer (1964) nor Brown and Helmer (1964) considered inviting all participants to provide reasons for their estimates. Consensus, it appears, was far too valuable to have it disrupted by the deliberate introduction of self-reflection.

Their last point of self-critique concerned the *substantive breadth* of the study, which still was not comprehensive enough. To describe the major aspects of the world of the future in a way that can inform political decisions would have required a separate panel on international relations.

The Problem of Improvement, Revisited

Although the Gordon and Helmer study contributed almost all of the substance of the Delphi paradigm, the parallel study carried out by Bernice Brown under the auspices of Helmer (Brown and Helmer 1964) should not be ignored. For one, it added some relevant methodological aspects to the picture, especially by introducing the issue of evaluating the experts' predictions. It did so by reference to the study by Abraham Kaplan et al. (1949); in contrast to the other studies, however, this reference now engendered a substantial interest in one of the problems treated there. Brown and Helmer's study used some elements of the earlier study's design. The report by Brown and Helmer, entitled "Improving the Reliability of Estimates Obtained from a Consensus of Experts. P-2986," reported the outcome of an experiment which combined two techniques of expert interrogation, Delphi and, following the example of Kaplan et al., "the computation of a consensus based on self-appraised competence ratings" (Brown and Helmer 1964, 1). The experiment aimed to contribute to consensus research by addressing three methodological research questions, or "desiderata":

> (i) to cause convergence of opinions in the sense of shrinking the opinion spread as expressed by the interquartile range; (ii) to cause convergence in the sense of more closely approximating the true value by the median; and (iii) to find a formula for determining a consensus that would be a more reliable estimator of the true value than the group median. (Brown and Helmer 1964, 2)

The experiment included 23 respondents, all of them RAND researchers (cf. Brown and Helmer 1964, 3), and probably all based in Santa Monica. The respondents received a questionnaire comprising twenty questions. Of these, 18 "were of a kind for which the answers can be found in the World Almanac" (Brown and Helmer 1964, 3); the remaining two questions could in principle be answered by means of mathematical computation, but were of a relatively high complexity so that solving them required some effort. Interestingly, the questionnaire did not contain a single predictive item. Furthermore, as the authors acknowledged, a consequence of this design choice was that the study participants "were relatively inexpert with regard to the questions posed" (Brown and Helmer 1964, 12). This becomes instantly clear when one reads any one of the questions, for example, question no. 4: "What was the average price received by the United States farmer for a bushel of apples in 1940?"

The participants were asked not to use any reference materials and not to take longer than a few minutes for each question. Like the earlier Delphi studies—and unlike the study by Kaplan et al.—there were no predefined answer categories. The participants had to give their estimates in numerical form. However, comparable to the study by Kaplan et al.—and unlike the earlier Delphis—the respondents were invited to additionally submit some measure of their confidence in their own estimate. Whereas Kaplan and his colleagues had decided on a design that asked for distributing subjective probabilities to four alternative answers, the Brown and Helmer study attempted to accomplish this via a self-rating. The expert was invited to evaluate, on the natural scale from 1 to 4, "his [or her] own degree of expertise on each question" (Brown and Helmer 1964, 4).

The questionnaire with the same 20 questions was distributed four times. In the second round, the participants were informed of the median and the interquartile range (the absolute range between the lower and the upper quartile of the answers if ordered) of the first round's estimates. The participants were asked to reconsider their initial answers and, in those cases where they were located outside the interquartile range, to provide a reason. These were then summarized for the third round, and participants were invited to provide brief counter-arguments in case they had some. Again, in the third round, median and interquartile range was provided for each answer, and the participants were asked to alter or restate their answers. The fourth-round questionnaire contained, apart from the 20 questions and the usual information on the distribution of the answers from the previous round, summaries of all the arguments collected in rounds two and three.

Table 6.5 gives the results of the first and the final fourth round for six exemplary questions used in the study. The questions were selected in order to provide an overview both over the variety of subjects of the questions and of the variety of results. All in all, a continuous convergence of

Table 6.5 Exemplary questions and results from Brown and Helmer's 1964 study. (Adopted from Brown and Helmer 1964, 3, 4, 6; own calculations)

Exemplary questions	Median and interquartile range: Round 1	Median and interquartile range: Round 4	True answer	d_M (%)	d_T (%)
1. How many randomly selected persons must there be in a group so that the probability is ½ that at least three of them have their birthday on the same day of the year?	122 50–549	122 90–400	"87"	0	−28.7
4. What was the average price received by the US farmer for a bushel of apples in 1940?	100 50–180	100 60–140	56	0	−44
5. What is the distance in geographical miles between Cape Town and the geographical point antipodal to Los Angeles?	2500 2000–5000	2500 2130–3500	1950	0	−22
8. What was the total tonnage, in millions, shipped through the Port of New York in 1962?	40 16–200	35 20–42	145	−12.5	+314
14. How many dozens of ordinary lead or graphite pencils did RAND (Santa Monica) buy in 1963?	2,000 1000–3000	1,800 1000–2000	1841	−10	+2
17. What is the basic fly-away price of a DC-9, in thousand dollars (including seats, radios, and galleys)?	5500 3500–8000	3500 3500–4500	3067	−36.4	−12.4
20. In the alphabetical listing of the current Santa Monica telephone directory the mid-name is "Lancaster"; the Manhattan telephone directory's alphabetical listing begins on p. 17 and ends on p. 1803; what is the number of the page on which "Lancaster" appears?	900 890–910	900 900–910	916	0	+1.8

the interquartile range could be observed. Two additional measures have been calculated to assess the results of the study. The relative difference between the first-round median M_1 and the fourth-round median M_4, d_M, is used to describe how much the opinions moved over the multiple interrogation phases. The relative difference between the fourth-round median M_4 and the true value T, d_T, is used to describe how much the fourth-round opinions (as described by the median) would have had to move in order to meet the true answer. In formal notation, these measures read:

$$d_M = \frac{M_4 - M_1}{M_1}$$

and

$$d_T = \frac{T - M_4}{M_4}$$

Reconsidering the three desiderata (or research questions) stated in the introduction, Brown and Helmer concluded for the first research question that a convergence of opinion took place in the sense that the estimates increasingly approximated each other and, as a consequence, the interquartile range either continuously diminished or at least never increased. However, the subsequent second research question was whether this convergence tended to approach the true answer. Here, Brown and Helmer drew differentiated conclusions. The decrease of the interquartile range in some cases led to the fact that while it included the true value in the first round, it did not do so in later rounds. "While the first-round interquartile ranges were large enough so that 13 out of the 20 contained the true value, in Rounds 2 and 3 only 10 still contained it, and in Round 4 the number of interquartile ranges containing the true answer had decreased to 7 out of the total of 20" (Brown and Helmer 1964, 8). In the majority of cases, however, the median had moved toward the correct cases, so that, if comparing the medians of the first and the last rounds, the medians of the fourth round were closer to the true answers in 13 of the 20 cases. One median remained stable throughout all four rounds of interrogation, and for six questions, the first-round median had been closer.[9]

Of course, the improvement produced by the procedure of repeated interrogation was far from spectacular. This links to the third research question and to the problem of improvement as it was originally described by Kaplan et al. (1950): What can be done to achieve better, in the sense of more correct and reliable, results? Brown and Helmer wondered whether there was an alternative to taking the median of the group's estimates. Seven of the 23 participating experts had actually performed better than the median. With regard to situations where the true value is unknown, the question was how to identify these better-performing experts. The study authors hoped to achieve this via the self-ratings. If there was a strong positive correlation between the performance and the self-reported degree of expertise, the latter could be used as a criterion for excluding those with lesser expertise.

In the questionnaire, the instruction aimed at soliciting the self-rating read as follows:

> Write one of the numbers 1, 2, 3, or 4 indicating how relatively confident you feel about your answer, using 1 for the most confident. More specifically, imagine that the answers from all respondents (chosen from various departments of the RAND research staff) are ranked according to their distance from the true answer; then your number should indicate whether you think your answer falls in the first, second, third, or fourth quartile in this ranking. (Brown and Helmer 1964, 10)

For each question, then, Brown and Helmer defined a subgroup of those who declared themselves as having a high confidence. Across the questions, this subgroup comprised approximately a third of the whole group, thus around six or seven persons. In a manner echoing Paretoian sociology, this subgroup was called E, because it "represents in a sense the (self-appointed) elite among the experts" (Brown and Helmer 1964, 10). Using only the fourth-round estimates of the elite resulted in significantly better results. Compared to the fourth-round median of the entire group, the fourth-round median of the elite was closer to the true answer in 15 of the 20 questions. In one case, the medians were identical. And in the remaining four cases, the entire group median came closer to the true value than the elite's estimate. Furthermore, the use of the elite's medians "produced an estimator which performs 50% better […] than the very best individual participant" (Brown and Helmer 1964, 11f).

These, of course, were positive results for the proponents of the methodology. A similar picture would probably emerge when calculating the correlation between the self-ratings and the actual performance. This topic was briefly discussed by the authors. They reported that to describe the actual performance of a participating expert, they had constructed a variable which was based on the distance of each answer to the true value. Thus, a distribution resulted, and it was then assessed for each question and each participant in which quarter of this distribution his or her answer was located. Thus, the actual performance variable also had four properties (1, 2, 3, and 4), with 1 denoting those answers which showed the least distance to the true value. The rank correlation between the two variables turned out to be positive, showing an average of 0.3 across all questions.

The authors' summary was generally positive. A convergence of opinions could be observed, and it could further be determined that the medians tended toward the true value. Furthermore, "the use of self-appraised competence ratings in forming a consensus appeared to be a powerful tool for increasing the reliability of the group estimates" (Brown and Helmer 1964, 12). This was, by all means, an exaggerated conclusion. What they had factually done is to ask the same bulk of dummy questions again and again, adding only minor additional information from round to round. The effect of fatigue must have been considerable. As a consequence, the results were highly stable. Consider for instance the fate of question no. 1. Although the answer to this question—"How many randomly selected persons must there be in a group so that the probability is ½ that at least three of them have their birthday on the same day of the year?"—can in principle be calculated, the group's consensual answer (if we take the median to represent it), was overrating the true answer by 140%. Moreover, the median did not change at all over the four rounds. Whereas it is true that in the complete list, the majority of the medians did move toward the true value, they did so only slightly and, for the most part, remained at a significant distance of the true value.

The most impressive example from our list that can be used to corroborate this critique is question no. 8—"What was the total tonnage, in millions, shipped through the Port of New York in 1962?" The initial median was at 40, and it slightly declined throughout the procedure to finally amount to 35, which leads to a d_M of −12.5%. The correct answer, however, was 145. The final estimate still had underestimated the correct value by a factor of 4! Or, in other words, the final median would have had to move in the other direction and to increase significantly, namely d_T = +314%! The conclusions which, for their disturbing implications, were not drawn by the study authors but could have been were, firstly, that once the first set of estimates was collected, there was only slight movement; and secondly, that this phenomenon was more likely to emerge due to fatigue caused by the monotonous reiteration of the same tasks.

The Diffusion of the Delphi Paradigm, 1963–1969

As explained at the outset of this chapter, the two studies discussed were part of a larger effort to boost the use of the Delphi technique, an effort that started around 1962. More precisely, they formed the core of those efforts that went into developing the design and methodology of the technique. Apart from that, Olaf Helmer and his team also took steps to disseminate the technique. In July 1962, Dalkey and Helmer had managed to receive the security clearance to publish an abridge version of the report of the first Delphi study from 1951. The subsequent dissemination, however, forgot about the initial Delphi study. In July 1963, Olaf Helmer confronted the participants of the Third International Conference on Operational Research in Oslo, Norway, with the question of whether or not operations research (OR) can be regarded as a science. Drawing on some thoughts published in OEIS, and thus providing a "qualified 'yes'" to his question (Helmer 1963, 1), Helmer continued by introducing the two studies underway at RAND, the long-range forecasting study (Gordon and Helmer 1964) and the experiment on improving the reliability of estimates (Brown and Helmer 1964). Another paper on Delphi, this time entitled "Convergence of Expert Consensus through Feedback" and published by RAND as P-2973 (Helmer 1964), followed the same structure: no mention of the first Delphi study, but an attempt to relate some general issues of *OEIS*—without ever reaching the philosophical depth of *OEIS*—to the two current studies. Helmer presented this paper in September 1964 at the 10th Annual Meeting of the Western Section of the Operations Research Society of America (ORSA) at Honolulu, Hawaii. Helmer's widely read book, *Social Technology* (Helmer 1966), included the two study reports from 1964 in slightly modified form. A chapter on Delphi that was published in 1967 in the Italian journal *Rivista Italiana di Amministrazione Industriale*, listed by RAND as "Analysis of the Future: The Delphi Method. P-3558" (Helmer 1967a), referred only to the long-range forecasting study. A contribution submitted to *Science Journal* and listed by RAND as "The Future of Science. P-3607" (Helmer 1967c) focused on reporting the estimations on the development of science gathered in the long-range forecasting study. In November of that year, Helmer addressed the board of the Air Force Advisory Group (AFAG) with a similar paper on Delphi, now entitled "Systematic Use of Expert Opinions. P-3721" (Helmer 1967b).

While Helmer was preparing to leave for the Institute for the Future that he co-founded with Paul Baran, Theodore Gordon, and Arnold Kramish in 1968 (see Chap. 7), Norman C. Dalkey continued the dissemination activities on behalf of RAND. In October 1967, he presented the method to the participants of the Second Symposium on Long-Range Forecasting and Planning, Alamogordo, New Mexico (Dalkey 1967). He introduced Delphi at the National Meeting of the American Chemical Society in San Francisco, California, in April 1968 (Dalkey 1968b), 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 1968b). Finally, an article by Bernice Brown, 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 (Brown 1968).

In their descriptions of the Delphi procedure, all these papers are quite similar to each other. The following description, taken from Dalkey (1967, 4), can be considered as 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.

This quotation describes neatly the specific form of Delphi that was disseminated in the 1960s. As regards methodological reasoning, the Delphi paradigm that emerged around this definition relied exclusively on the two studies from 1964 (Gordon and Helmer 1964; Brown and Helmer 1964), thereby ignoring 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). The objective of Delphi still was to produce a consensus, but this was attempted not so much by using the iteratively distributed questionnaires as means of communication among the study participants, but rather as a tool to retrieve measurement data. 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 Dalkey and issued by RAND for public relations: "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 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).

With the years to come, however, various points of critique have been directed toward the Delphi paradigm, and alternatives have been explored—outside, but also within RAND, where Harold Sackman was commissioned in 1973 to write a critical analysis and evaluation of Delphi. Sackman's conclusions were devastating. He found

> considerable evidence that results based on the opinions of laymen and "experts" are indistinguishable in many cases; aggregated raw opinion presented as systematic prediction; technical shortcomings, such as untested and uncontrolled halo effects in the application of Delphi questionnaires; unsystematic and non-replicable definition and use of "experts;" manipulated group suggestion rather than real consensus; ambiguity in results stemming from vague questions; acceptance of snap judgments on complex issues; and the virtual absence of a vigorous critical methodological literature even though hundreds of Delphi studies have been published. (Sackman 1974, v)

As a consequence, Sackman suggested that "[e]xcept for its possible value as an informal exercise for heuristic purposes, Delphi should be replaced by demonstrably superior, *scientifically rigorous* questionnaire techniques and associated experimental procedures using human subjects" (Sackman 1974, vi; emphasis added).

Were Sackman's allegations correct? Had Delphi's inventors, despite their training in philosophy of science, failed to construct a method that met scientific standards? Of course, Sackman's evaluation was deliberately systematic, not historical, in the sense that he applied those criteria he himself perceived and selected as valid. The standard that Sackman applied is the one of psychology and sociology in the mid-1970s, and he shows no understanding for the pragmatic stance that characterized the early Delphi proponents—the conviction that doing something imperfectly was better than doing nothing. In their publications, Delphi inventors had repeatedly stressed that Delphi might be of help in situations where there is no knowledge available to base policy decisions on. In such a context, Delphi might be pragmatically the best choice; in the absence of truth, a good estimate is better than nothing. Also, Sackman had not concerned himself with the position developed in *On the Epistemology of the Inexact Sciences* (*OEIS*), which offered a very distinct definition of science unlike the one assumed by Sackman. Thus, his critique might appear exaggerated. And at the same time, it did not identify the most fundamental flaw of the Delphi paradigm.

As shown above, its inventors were highly ambivalent with regard to the nature of Delphi as a means of knowledge production: was it a scientific method or a technique of analysis? Was truth its final objective, or was it to improve decision-making? Sackman decided to evaluate Delphi as a scientific method. And *OEIS* had been clear about this: if it followed its characteristics, Delphi had the right to be treated as a scientific method. The 1964 Delphi attempted to avoid this question. Instead of shaping its argument along the lines of the epistemology of the inexact sciences, it again, where feasible, relied on pragmatic considerations to defend the choice of procedure. In some passages, the 1964 study was treated as a piece of science—in others, as a piece of interesting and speculative, albeit systematic policy advice.

Set against the standard formulated in *OEIS*, however, the 1964 study was no science. Since it omitted the (collaborative) setting up of a transparent and accessible set of evidence, it lacked a crucial characteristic required to call it part of a scientific endeavor. But beyond losing the claim to be scientific, the omission of a collaborative composition of a set of evidence had an utterly destructive consequence for the whole methodology. Without such a set of evidence, the interpretation of the convergence as “consensus” lost its justification. Since the experts could not agree on the epistemic value of a set of evidential material, what were they expected to reach consensus upon? Or, to put it in other words, what is the likely psychological effect of being repeatedly asked to revise one’s answers in the face of the opinion of the majority? Does such a procedure result in consent? Or, rather, in a mixture of annoyance and fatigue? Without an evolving set of accessible evidence, the convergence that should be a rational result of the technique became a mere artifact of its procedure. While in their long-range forecasting study, Gordon and Helmer preferred to speak of forecasts instead of predictions to indicate that the study applied a more open process of sketching several possible futures (cf. Gordon and Helmer 1964, 1), a closer analysis of their procedure corroborates the claim that the epistemic role of the experts in their study was not to forecast, but to *predict*—just as it had been the case with RAND's first expert prediction study by Kaplan et al.

Conclusion

Delphi—and *OEIS* in particular—can indeed be seen as continuing the program of the interrelatedness of science and politics that characterized the positivist (and early neo-positivist) movement. Already in Vienna, logical positivism was not a mere scientific movement, but was understood as a project oriented toward improving society. Since politics were to become a scientific project, science was intrinsically political. After the emigration of the movement's European key members to the United States, relations to the prevailing pragmatists were soon established. The separation of the political from the scientific that so crucially determines our current picture of positivism was a result of the fear of McCarthyism and anticommunist persecution, as it reinforced the tendency of the refugee members of this scientific community to detach themselves from the political discourses (cf. Reisch 2005).

However, even Delphi, despite having been created by proponents of logical empiricism, did not conform with all its central tenets. Prominent among these was of course the tenet of empiricism, meaning the systematically collected experience is the only source of new knowledge. By conceptualizing (tacit) expert opinions as a crucial element in the production of knowledge, Delphi—and *OEIS*, for that matter—deviated from this central tenet (Dayé 2016).

OEIS plays a curious role in the history of Delphi. It has been widely cited, yet its explication of a specific epistemic role of the expert did not affect the Delphi paradigm in any substantial way. This is surprisingly true even for those Delphi studies carried out at RAND in the 1960s, prime among them of course the long-range forecasting study from 1964. The same holds for a series of further studies carried out under the auspices of Helmer and Dalkey (Gordon and Helmer 1964; Brown and Helmer 1964) and for the way Delphi was disseminated in these early years (Helmer 1963, 1964, 1967a, 1966; Dalkey 1967, 1968b, 1968a). None of these texts seriously considered the methodological principles developed some years earlier.

Several factors might have contributed to this turn. On a micro level, the personality of Olaf Helmer could have played a role. As Rescher remembered, Helmer was "impatient of detail—and the active writing up of his researches was for the most part left to his collaborators" (Rescher 1997, 349). With Rescher leaving RAND in 1957 for a university career, the second author of *OEIS* was not at hand any more. Though he returned to issues related to prediction and social prognosis later in his career (Rescher 1969, 1998), his publications from this period do not include anything on Delphi. Olaf Helmer was attracted by the emerging field of futurology and, especially in *Social Technology* (Helmer 1966), the book that introduced Delphi to the public outside RAND, he emphasized the practical use of the method without delving deeply into epistemological issues. The same applies for the remaining Delphi co-inventor, Norman Dalkey, especially for the series of textbook-like introductions to Delphi mentioned above that Dalkey wrote collaboratively with a range of other RAND researchers (Dalkey 1969; Brown et al. 1969; Dalkey et al. 1969).[10]

On both meso and macro levels, several interconnected developments have confronted the new methodology. On the meso level of RAND, Helmer and Gordon had expected that the inclusion of non-RAND experts would increase the persuasiveness of the technique. Clearly, this required either to invite these experts to stay at RAND for the duration of the study, rendering it a very expensive method, or to rely on current communication technology and accept its restrictions as regards the interactive structuring of a shared set of evidence. Furthermore, the macro-level trend to raise the degree of quantification within the social sciences (Platt 1996; Steinmetz 2005) also had important proponents at RAND, most notably in the Mathematics Division to which Helmer, Dalkey, and Rescher belonged. Quantification was increasingly understood as the main avenue to objectivity within the social sciences. This understanding was not only a phenomenon internal to science, but instead emanated to the public sphere (Porter 1995) which led to an increased demand for quantitative studies. And while in *OEIS*, they had argued against precision as the defining characteristic of science, their critique certainly did not concern quantification.

In addition to such "external" developments, the turn to bigger samples also offered "internal" solutions to open methodological problems with the Delphi technique. This was so mainly in relation to the aforementioned problem of the two ways to define the expert. Though one can, as Kaplan and his colleagues suggested, start setting up databases on the predictive success of experts and thereby construct some measure of their predictive ability, one must perforce start with the expert as a sociological category, that is, as the person who is seen by a community as possessing expert knowledge. Though not exploring this issue in detail, Delphi inventors probably were aware of it. In combination with the acknowledged problem that someone who has a detailed knowledge of a given field does not necessarily have the competence to predict, it suspended a question mark over the process by which experts are selected. A person being falsely defined as expert and lacking the required competences would introduce a bias. And the smaller the sample, the higher the distortion of the overall results. Enlarging the pool of participating experts might have appeared as a way to level out this risk.

The history of the creation and early development of the Delphi method is thus one of a path sketched but, in the end, not taken. Out of a concern with the epistemological justification of their work, Delphi developers formulated a philosophical foundation for the systematic use of expertise in the inexact sciences. This foundation defined the epistemic role of the expert as evaluating, based on his or her implicit knowledge, a set of available and explicit evidence related to a hypothesis about a future event. The epistemic role of the expert as described in *OEIS*, thus, implied that she was able to *forecast* from an available set of evidence. However, a few years later, RAND researchers implicitly and tacitly returned to the earlier conceptualization and expected experts to come up with *predictions* without establishing a parallel discussion of the reasoning behind these predictions. This was not only a regression in epistemological terms. Moreover, it pulled the rug from under Delphi's core methodological claim (cf. Dayé 2018). It was not justified anymore to interpret the convergence of estimates as consensus among experts; rather, Delphi had become a means of producing conformity. Apart from informing about the distribution of answers from the previous round, no new information was introduced. In *OEIS*, as well as in the first Delphi study reported in

Chap. 3, the methodological structure allowed for an expert consensus on how to assess the available set of evidence. In contrast, the long-range forecasting study, and all the other studies that followed its design, only allowed the expert to consent—or dissent—with figures provided by other experts. Hence, the aggregated opinion of the majority had become the only impulse available for changing one's opinion.

Notes

1. German Delphi practitioner Michael Häder (2006, 351, 2009, 15) reports that a total of 14 experiments had been conducted at RAND prior to the publication of Gordon and Helmer's "Report on a Long Range Forecasting Study" (1964). Although I inspected a vast array of relevant literature, including the source Häder refers to (Linstone and Turoff 1975, 10), and read all the available RAND writings on Delphi from the first two decades after its inception, I could not verify this figure. When asked how many Delphi studies RAND had carried out prior to his arrival, Gordon stated: "All I can do is give you an impression. A lot of informal experimentation: let's try this, does this work, let's try that. And formal reports: only a few" (Interview with Theodore J. Gordon by the author, 16 August 2013, p. 7).
2. This is the second meaning of the term "paradigmatic" described in Thomas Kuhn's Postscript to the *Structure of Scientific Revolutions* (Kuhn 1976).
3. Andersson (2018, 82) suggests that "the conflation of time and space" in the notion of long-range (instead of long-term) forecasting was "a result of experimentation not only with Operations Research (OR), but also with systems analysis, at RAND." This is not true. The term "long range forecasting" was well established in meteorology already at the turn of the century (cf. Pietruska 2018, 108–155).
4. Two of the European participants, "Professor Dennis Gabor and Monsieur Bertrand de Jouvenel" (Gordon and Helmer 1964, ix) are thanked by name in the acknowledgments 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 twentieth-century futurologists (e.g., Jouvenel 1967). 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. He also wrote extensively on the future (most importantly Gabor 1963).

5. These dates were mentioned in relation to Panel 5, the panel on war prevention. Considering the authors' practice to send all the questionnaires to all the participating experts, it appears justified to generalize these dates onto the whole study.
6. I have added to the items their numbers from the final list (see below).
7. For the sake of space, the items are not fully entered into the table.
8. Cf. http://www.state.gov/t/isn/4797.htm, accessed 28 June 2017.
9. Brown and Helmer (1964, 5–8) also introduced another way of comparing the correctness of the median values. They define an area of 25% around the true value, which they call "ballpark," and determine how many medians are located in the ballpark of each question. This was true for six of the first-round medians, and for nine of the fourth-round medians.
10. In an interview, Helmer shared some recollections on Dalkey. After Helmer had left RAND for the IFTF, "Dalkey pretty much devoted himself to pursuing research on Delphi for a number of years. So I think that became his main interest. I don't know if he contributed anything—I don't mean to talk down—but I think in a way because of his interest, he contributed just through that. Pursuing and maintaining an interest. I don't think he contributed very many original ideas to this, but on the other hand he was very conscientious from a scientific point of view. And so was quite careful in applying some ideas and improv[ing] the methods" (Interview with Olaf Helmer by Kaya Tolon, 3 June 2009, p. 7).

References

Adalet, Begüm. 2018. *Hotels and Highways: The Construction of Modernization Theory in Cold War Turkey*. Stanford, CA: Stanford University Press.

Andersson, Jenny. 2018. *The Future of the World: Futurology, Futurists, and the Struggle for the Post Cold War Imagination*. Oxford and New York: Oxford University Press.

Brown, Bernice B. 1968. Delphi Process: A Methodology Used for the Elicitation of Opinions of Experts. P-3925. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3925.html.

Brown, Bernice B., and Olaf Helmer. 1964. Improving the Reliability of Estimates Obtained from a Consensus of Experts. P-2986. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2986.html.

Brown, Bernice B., Samuel Cochran, and Norman C. Dalkey. 1969. The DELPHI Method, II: Structure of Experiments. RM-5957-PR. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM5957.html.

Dalkey, Norman C. 1967. Delphi. P-3704. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3704.html.

———. 1968a. Experiments in Group Prediction. P-3820. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3820.html.

———. 1968b. Predicting the Future. P-3948. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3948.html.

———. 1969. The DELPHI Method: An Experimental Study of Group Opinion. RM-5888-PR. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM5888.html.

Dalkey, Norman C., and Olaf Helmer. 1962. An Experimental Application of the Delphi Method to the Use of Experts. RM-727/1-ABRIDGED. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM727z1.html.

———. 1963. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science* 9 (3): 458–467.

Dalkey, Norman C., Bernice B. Brown, and Samuel Cochran. 1969. The DELPHI Method, III: Use of Self-Ratings to Improve Group Estimates. RM-6115-PR. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM6115.html.

———. 1970. The DELPHI Method, IV: Effect of Percentile Feedback and Feed-In of Relevant Facts. RM-6118-PR. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/research_memoranda/RM6118.html.

Dayé, Christian. 2016. 'A Fiction of Long Standing:' Techniques of Prospection and the Role of Positivism in US Cold War Social Science, 1950–1965. *History of the Human Sciences* 29 (4–5): 35–58. https://doi.org/10.1177/0952695116664838.

———. 2018. How to Train Your Oracle: The Delphi Method and Its Turbulent Youth in Operations Research and the Policy Sciences. *Social Studies of Science* 48 (6): 846–868. https://doi.org/10.1177/0306312718798497.

Feichtinger, Moritz. 2011. Modernisierung als Waffe—'Strategische Dörfer' in Malaya und Algerien. In *Macht und Geist im Kalten Krieg*, ed. Bernd Greiner, Tim B. Müller, and Claudia Weber, 359–375. Hamburg: Hamburger Edition.

Gabor, Dennis. 1963. *Inventing the Future*. London: Secker and Warburg. https://www.worldcat.org/title/inventing-the-future/oclc/1015162277?referer=di&ht=edition.

Gilman, Nils. 2003. *Mandarins of the Future. Modernization Theory in Cold War America*. Baltimore and London: The Johns Hopkins University Press.

Gordon, Theodore J. 1965. *The Future*. New York: St. Martin's Press.

———. 2011. Obituary—Olaf Helmer, Futures Thinker. *Technological Forecasting and Social Change* 78: 1099–1100.

Gordon, Theodore J., and Olaf Helmer. 1964. Report on a Long-Range Forecasting Study. P-2982. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2982.html.

Gordon, Theodore J., and Julian Scheer. 1959. *First into Outer Space*. New York: St. Martin's Press.

Häder, Michael. 2006. *Empirische Sozialforschung. Eine Einführung*. Wiesbaden: VS Verlag für Sozialwissenschaften.

———. 2009. *Delphi-Befragungen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Helmer, Olaf. 1963. The Systematic Use of Expert Judgment in Operations Research. P-2795. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2795.html.

———. 1964. Convergence of Expert Consensus Through Feedback. P-2973. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2973.html.

———. 1966. *Social Technology. Contributions by Bernice Brown and Theodore Gordon*. New York and London: Basic Books.

———. 1967a. Analysis of the Future: The Delphi Method. P-3558. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3558.html.

———. 1967b. Systematic Use of Expert Opinions. P-3721. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3721.html.

———. 1967c. The Future of Science. P-3607. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P3607.html.

Helmer, Olaf, and Nicholas Rescher. 1958. On the Epistemology of the Inexact Sciences. P-1513. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P1513.html.

Hounshell, David. 2000. The Medium Is the Message, or How Context Matters. In *Systems, Experts, and Computers: The Systems Approach in Management and Engineering, World War II and After*, ed. Agatha C. Hughes and Thomas P. Hughes, 255–310. Cambridge, MA and London: The MIT Press.

Jardini, David. 2000. Out of the Blue Yonder: The Transfer of Systems Thinking from the Pentagon to the Great Society, 1961–1965. In *Systems, Experts, and Computers: The Systems Approach in Management and Engineering, World War II and After*, ed. Agatha C. Hughes and Thomas P. Hughes, 311–358. Cambridge, MA and London: The MIT Press.

———. 2013. *Thinking Through the Cold War: RAND, National Security, and Domestic Policy, 1945–1975*. Smashwords Ebooks.

de Jouvenel, Bertrand. 1967. *The Art of Conjecture*. New York: Basic Books.

Kaplan, Abraham, A.L. Skogstad, and Meyer A. Girshick. 1949. The Prediction of Social and Technological Events. P-93. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P93.html.

———. 1950. The Prediction of Social and Technological Events. *Public Opinion Quarterly* 14 (1): 93–110.

Kuhn, Thomas S. 1976. *Die Struktur wissenschaftlicher Revolutionen*. Frankfurt am Main: Suhrkamp.

Latham, Michael E. 2000. *Modernization as Ideology. American Social Science and "Nation Building" in the Kennedy Era*. Chapel Hill and London: The University of North Carolina Press.

Light, Jennifer S. 2003. *From Warfare to Welfare. Defense Intellectuals and Urban Problems in Cold War America*. Baltimore and London: The Johns Hopkins University Press.

Linstone, Harold A., and Murray Turoff. 1975. *The Delphi Method. Techniques and Applications*. London and Reading, MA: Addison-Wesley.

Miralles, Carles Sirera. 2015. Neglecting the 19th Century Democracy, the Consensus Trap and Modernization Theory in Spain. *History of the Human Sciences* 28 (3): 51–67. https://doi.org/10.1177/0952695115579588.

Pietruska, Jamie L. 2018. *Looking Forward: Prediction & Uncertainty in Modern America*. Chicago: University of Chicago Press.

Platt, Jennifer. 1996. *A History of Sociological Research Methods in America, 1920–1960*. Cambridge, UK: Cambridge University Press.

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

Reisch, George A. 2005. *How the Cold War Transformed Philosophy of Science: To the Icy Slopes of Logic*. Cambridge: Cambridge University Press.

Rescher, Nicholas. 1969. Delphi and Values. P-4182. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P4182.html.

———. 1997. H2O: Hempel-Helmer-Oppenheim, An Episode in the History of Scientific Philosophy in the 20th Century. *Philosophy of Science* 64 (2): 334–360.

———. 1998. *Predicting the Future. An Introduction to the Theory of Forecasting*. Albany: State University of New York Press.

Sackman, H. 1974. Delphi Assessment: Expert Opinion, Forecasting, and Group Processes. R-1283-PR. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/reports/R1283.html.

Sapolsky, Harvey M. 2004. The Science and Politics of Defense Analysis. In *The Social Sciences Go to Washington. The Politics of Knowledge in the Postmodern Age*, ed. Hamilton Cravens, 67–77. New Brunswick, NJ and London: Rutgers University Press.

Shah, Hemant. 2011. *The Production of Modernization. Daniel Lerner, Mass Media, and The Passing of Traditional Society*. Philadelphia, PA: Temple University Press.

Steinmetz, George, ed. 2005. *The Politics of Method in the Human Sciences. Positivism and Its Epistemological Others*. Durham and London: Duke University Press.

7

Conclusion: The Strength of Epistemic Hopes

The Future Studies Movement

This is roughly where the period of interest to this book ends. The paradigms of the two methods had been shaped, and the epistemic roles of the experts consolidated. Now, a process of diffusion set in, and in one way or another, RAND affiliates contributed to this process. Both techniques turned out to be "dual use" techniques, techniques that easily crossed boundaries between social spheres (cf. Price 2016). Despite having been created to produce knowledge for decision-making in the military, their "operational areas" were much broader and extended into various realms of socially relevant issues. The future studies movement was about to take up on a global scale: Gordon's *The Future* (1965) and Helmer's (1966) *Social Technology* were immediately followed by Bertrand de Jouvenel's *The Art of Conjecture* (1967) and *The Year 2000* by Herman Kahn and Anthony J. Wiener (1967). Since 1964, Daniel Bell had been directing the Commission for the Year 2000 installed by the American Academy of Arts and Science, an effort eventually leading to the publication of Bell's *The Coming of Post-Industrial Society* (1973), a few years after Alvin Toffler's bestselling *Future Shock* (1970). These books joined forces

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*,
Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1_7

with earlier works from European authors like Robert Jungk's *Die Zukunft hat schon begonnen* (1952), Jean Fourastié's *La civilization de 1975* (1959), Fritz Baade's *Der Wettlauf zum Jahre 2000* (1960), and Dennis Gabor's *Inventing the Future* (Gabor 1963) to form the intellectual basis of the future studies movement, a movement that, as Andersson (2018) shows, was international from the very beginning (see also Cornish 1977, esp. 78–92). And Delphi and political gaming quickly emerged as the first techniques of prognosis around which the movement assembled.

Realizing this potential, Olaf Helmer, Theodore J. Gordon, and two other former RANDites—Paul Baran and Arnold Kramish—created the Institute for the Future (IFTF) and managed to affiliate to Wesleyan University. Here, they started to apply Delphi and other techniques of prospection they had developed, like cross-impact analysis, to problems outside the more military applications that RAND had required them to focus on. Political gaming happened to generate interest from another New England academic institution, the MIT. With the support of RAND researcher Paul Kecskemeti, Lincoln P. Bloomfield from MIT's Center for International Studies (CENIS) began to use political games as a method of both research and teaching. This and the ensuing section describe the further trajectories of the two techniques. Then, the chapter turns to a concluding discussion of the epistemic hopes and the social position attributed to the Cold War expert, and their means to justify, maintain, and if possible corroborate the trust put in them.

Delphi Moves from West to East (and Back Again)

Immediately after the 1964 long-range forecasting study, Gordon and Helmer invented another widely known method of future studies, cross-impact analysis, which in the years following its publication was tested and then systematically applied by the Central Intelligence Agency (CIA) and other intelligence agencies (cf. Heuer and Pherson 2011, 107) and became part of the standard toolkit of future research.[1] However, Helmer became increasingly estranged from the restriction of RAND's focus on problems of military relevance. While RAND was broadening its research areas, it seemed reluctant to put social problems on its agenda. Helmer remembered:

> 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 Olaf Helmer by Kaya Tolon, 3 June 2009, p. 9)

Several RAND researchers shared Helmer's conviction, among them Gordon, physicist Arnold Kramish (1923–2010), and computer network pioneer Paul Baran (1926–2011). They perceived both the demand and the social urgency to broaden the use of their techniques of prosgnosi. As Gordon emphasized: "This was in the '60s, you know, there was a big upheaval in the social attitudes in the '60s. … Wouldn't it be good if we could take the techniques that we're developing here at RAND and use them to solve social problems?" (Interview with Theodore J. Gordon by the author, 16 August 2013, p. 4). Plans crystallized, and ultimately Baran managed to secure small grants from the Arthur Vining Davis Foundation and later from the Ford Foundation to foster the establishment of the Institute for the Future (IFTF). They sought a university affiliation and visited various places in the country—among them the University of Texas, but "nobody liked the idea of living in Texas" (Interview with Theodore J. Gordon by the author, 16 June 2013, p. 5). Eventually, they received an invitation from Wesleyan University president Edwin Deacon Etherington, which they accepted. Baran, Gordon, Helmer, and Kramish quit their jobs, and in 1968, they moved to the East Coast.

Two years later, however, Helmer, Baran, and Kramish decided to relocate the institute back to California; it settled in Palo Alto, where it still resides, and Roy Amara became its director. Gordon stayed on the East

Coast, yet he continued to do future studies at a company that he founded, the Futures Group. It 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 Chairman of The Futures Group after 20 years, but continued to work as senior advisor.[2]

The move of the IFTF appears to be related to a rather bumpy start for the institute on the East Coast; beyond that, the Bay Area certainly was a region in the United states where "inventing the future" was the daily bread of many (F. Turner 2006; McCray 2013). Helmer nonetheless left the institute three years later to join the University of Southern California, which just had established a (short-lived) Center for Futures Research. In 1977, he spent some time at the International Institute for Applied Systems Analysis (IIASA) in Laxenburg, Austria, where he taught courses on Delphi and other techniques of prognosis and wrote a few papers on the topic.

Meanwhile, Japanese government officials had followed the RAND and IFTF efforts and worked toward carrying out their own Delphi. The task fell to the National Institute of Science and Technology Policy (NISTEP), which, with help from Theodore Gordon, had its first Delphi carried out in 1971. Focusing on technological innovations, it has been carried out regularly since then, the ninth time in 2009. NISTEP's Delphis attempt to assess when specific technological innovations will be made, when they will be implemented in Japanese society, how important their impact will be on Japanese society and which sectors of the knowledge economy will be the driving forces behind each innovation.[3] Smaller, but methodologically comparable regular Delphi studies on science and technology development and policy are carried out in Germany by the Fraunhofer Institut as well as in other countries. Further, Delphi is widely used in market research (e.g., Deutsche Post AG 2009), public opinion and media research (Ferguson 2000), or intelligence analysis (Heuer and Pherson 2011). Though there exists a considerable variety, Delphi studies today follow the 1964 paradigm by establishing estimations via an iterative assessment of expert opinions without direct interaction among them (Linstone and Turoff 1975; Häder 2009).

Gaming at Endicott House and Beyond

As already mentioned in the previous chapter, RAND did not intensively pursue political gaming after the completion of the fourth round. It was concluded that the efforts and resources required to carry out gaming exercises in a manner that satisfied the epistemic needs of the involved experts were in no reasonable relation to the expected results. In order to function as a serious tool of strategy testing, political games would have to continue over several months. The participating experts were withdrawn from almost all other activities for the duration of the game. This, in turn, rendered it highly unlikely that high-level experts would participate in such a comprehensive political game. Furthermore, it was still probable that many of the same insights would result from traditional forms of strategy analysis. It was agreed that "[i]ntermittent rather than continuous gaming activities appear to us […] the most productive way of combining the benefits of research and gaming," and decided that RAND would not, at least in the near future, "plan further large-scale use of political gaming in the Social Science Division" (Davison 1958, 12).

However, RAND's social scientists had already begun to disseminate their experiences, and their approach to political gaming met with high interest. It was developed mainly with two objectives, one concerned with using political gaming as pedagogical device in university training, the other continuing the use of political gaming as a technique in the analysis of international conflicts, real or potential. As RAND was preparing the later rounds of political gaming, the leading staff members received numerous declarations of interest in the new technique. Shortly after the completion of the fourth game, they started presenting their work to external audiences (cf. Davison 1958, 9ff; Goldhamer and Speier 1959, 80ff). Between 1956 and 1958, Goldhamer presented political gaming in several lectures at the Army War College. In summer 1956, Speier introduced the Division's work at a Social Science Research Council summer institute in Denver. Around the same time, Goldsen sketched RAND's experiences at various gatherings with students and faculty members at Yale. In the following year, Goldsen presented the technique at a conference in Princeton (June 1957), and Speier lectured about it at Stanford's Center for Advanced Studies in the Behavioral Sciences. In June 1959, Speier was again invited by the Social Science Research Council to present political gaming at a conference in West Point (NY). Three months later, Goldhamer presented a paper on the technique at the annual meeting of the American Political Science Association held in Washington, DC.

Additionally, several informal talks and gatherings took place, most notably with Harold Guetzkow and Richard Snyder from Northwestern University as well as with several academics from MIT. Guetzkow first independently followed the idea of applying a gaming approach to the simulation of international relations (Guetzkow 1959; Guetzkow and Jensen 1966). When, in 1956–1957, he spent a sabbatical at Stanford's Center for Advanced Study in the Behavioral Sciences (CASBS), he developed a combined manual and computer simulation called Inter-Nation Simulation (INS) "which became a widespread vehicle for pedagogy and research in world politics, focusing on simulated decision making in a hypothetical world" (Ward and Guetzkow 2009, 15). Hans Speier, who had been a CASBS fellow the previous year and was in close contact with representatives of the Ford Foundation that sponsored the Center, was a member of Guetzkow's study group (cf. Guilhot 2017, 200–201). Guetzkow's interest was theoretical: he followed the mathematical thrust of game theory and used role-playing to "validate theories about the structure of international politics" (Bloomfield 1960b, 59). To achieve this aim, Guetzkow chose not to have players represent those countries dominating international relations in the real world, but instead took care "to ensure that no known country is actually represented, so that the results may refer to the theoretical structure [of international relations] rather than to the players' notions of how nations are supposed to react" (Bloomfield 1960b, 59).

The interest of MIT scholars in political gaming was raised by RAND scientist W. Phillips Davison, when he was appointed visiting professor at MIT for the academic year 1957–1958. In one of his courses, a graduate seminar on international communications, Davison applied a modified version of RAND's political game as a teaching device (cf. Davison 1958, 10; Goldhamer and Speier 1959, 80). As a "result" of this experience, Davison (1958, 11) noted,

> Professor Lucian Pye—also at M.I.T.—became interested in using the political gaming technique in connection with a senior course in American diplomacy. Professor Pye, assisted by Davison and Professor Warner Schilling […], adapted the RAND technique to his course requirements and introduced some ingenious innovations. […] During the game Professors Pye and Schilling edited a 'World Newspaper,' which was distributed to all participants at the start of each session. This included "news" contributed by each team (i.e., what each country wished to have known about itself and its policies), as well as editorials and a "James Reston column" written by the instructors. The latter column often included "leaks," references to domestic reactions, etc. This experiment appeared to be highly successful in a pedagogical sense.

The use of political gaming at MIT as a means of student training raised the interest of Lincoln P. Bloomfield from MIT's Center for International Studies (CENIS). As part of a project he directed on the United Nations, Bloomfield began planning for a gaming exercise which became known as the Endicott House game, named after MIT's country estate which was selected to host the game. Before turning to a more detailed description of CENIS's Endicott House game, I want to describe briefly two student games devised by Bloomfield and his colleague, political scientist Norman J. Padelford. These student games were carried out after the CENIS game, and "[t]he experience gained in the Endicott House game served as a guide in planning the two student exercises" (Bloomfield and Padelford 1959, 1109).

The two student games took place in the first half of 1959 and were both concerned with the coeval Berlin crisis. "The first involved 90 undergraduates from the International Relations course at M.I.T., and was concentrated in a period of one week culminating in a model UN General Assembly lasting the equivalent of one day" (Bloomfield and Padelford 1959, 1105). The second game involved 40 political science students from Harvard, Yale, Dartmouth, and MIT. This game "was played over a six-week period, including two weeks devoted to diplomatic intercourse in which formal notes were exchanged between the seven states represented and missions sent back and forth between the teams. It was brought to a climax with a two-day Foreign Ministers' meeting" (Bloomfield and Padelford 1959, 1105–1106). The teaching staff had prepared a collection of materials prior to the start of gaming, which included, amongst other papers, the Potsdam Agreement on the division and reconstruction of Germany by the three occupying powers. In addition, the teams were invited to emulate governmental decision structures by installing the roles of foreign ministers, advisors, opposition leaders, public opinion, and others. Moves were made in written form, although open deliberations amongst the nations represented were possible. Taking up the idea by Pye and Schilling, a game-newspaper "was prepared and issued at intervals in each game to inform the players of public moves by other teams, to insert 'news' items desired by various parties, and to circulate new 'facts' or information released by the Umpires" (Bloomfield and Padelford 1959, 1110). In the first student game, W. Phillips Davison and Ithiel de Sola Pool, a member of CENIS, acted as Umpires. In the second game, this part was taken over by John C. Campbell, Director of Political Studies at the Council on Foreign Relations in New York City, and David E. Linebaugh, former First Secretary of the American Embassy in Bonn, Germany.

Like the earlier analysis game, the first student game took place at Endicott House. However, Bloomfield and Padelford (1959, 1111) noted that the surroundings in many ways counteracted their attempt to simulate real deliberation procedures:

> If the teams are forced to operate in too close proximity to one another, as by sharing offices or having to engage in diplomatic conversations in crowded hallways, this can deter or complicate effective role-playing and the maintenance of adequate security. Moreover, it can unrealistically speed up the dimensions of time: e.g., an instruction given to an ambassador to call on a Foreign Minister can be completed, including a report "home," within ten minutes or so. […] [T]he ease of negotiating with all countries in the world simply by knocking on a few doors up and down a corridor was, in retrospect, far too great. It produced both a lack of realism and often an unmanageable compounding of demands upon the key teams.

Despite the effort spent on the use of political gaming in the classroom and the progress made in procedural questions, observers remained very skeptical. For instance, Bernard C. Cohen had conducted a game in one of his courses at the University of Wisconsin, and it turned out that some of the central claims of the proponents of political gaming as a pedagogic device—for instance the increased interest of students in the subject—could not be corroborated. There simply was a lack of evidence that there was something like a net gain of interest among the students in political gaming as compared with a regular class. "It might even turn out to be the case that the professors who direct the games are more interested than the students—and thus they tend to see chiefly the evidence that confirms their wisdom in conducting the exercise" (Cohen 1962, 371). For instance, headcounts carried out randomly in the Wisconsin experiment had resulted in the observation that "twenty to twenty-five students were enjoying an unexplained absence" (Cohen 1962, 371).

When devising the student games, Bloomfield had already made first experiences with the gaming technique in a game he organized at MIT's Endicott House. The game was part of an ongoing research project concerned with the United States' interests in the UN. In the preparation phase, Bloomfield discussed his plans with several staff members of RAND, and he was given access to the informal records of the RAND games (cf. Davison 1958, 12; Goldhamer and Speier 1959, 81). It was decided that Paul Kecskemeti would participate in Bloomfield's game, which finally took place in September 1958 at MIT's Endicott House (Bloomfield and Padelford 1959; Bloomfield 1960b). In the official CENIS reports, the Endicott House game was called POLEX, an acronym of Political Exercise, and was the first of a series of games directed by Bloomfield.

The scenario used in POLEX was that there emerged an international crisis after the demise of the head of government in Poland. It was a future scenario, having the game start one year ahead of the real time. Although it was initially conceived optimal to use five full days for gaming, this was reduced to three days after considering the decreased likelihood to recruit high-level experts for a longer period. In the end, participants formed ten teams (cf. Bloomfield 1960b, 62), and—like later the undergraduate students—were advised to introduce within their teams some kind of division of labor by installing a variety of roles.

One major difficulty encountered in the preparation of the game was the selection of the problem. An adequate problem would have to meet several characteristics, Bloomfield and Padelford (1959, 1106) explained: the simulated crisis would have to be grave enough to ensure a level of interaction among the participants that was not artificial, but realistic; it would have to be open for political and diplomatic solutions rather than invoke military action; it should match the expertise of the available political scientists and area specialists; and, finally, it must show substantive relevance to the overarching research project it was a part of, namely to Bloomfield's project on the interests of the United States in the UN. All ideas for potential problems were evaluated along these criteria (Bloomfield and Padelford 1959, 1106):

> Among the alternative problems initially considered for the Center game were: revolt in East Germany, Communist take-over in Syria, an India-Pakistani war over Kashmir, nationalization of the Panama Canal, civil war in the Union of South Africa, Chinese attack upon Quemoy and Matsu, renewed Arab-Israeli war, British-Yemeni hostilities, Soviet attack on Yugoslavia, Indonesian attack on West New Guinea, Chinese attack on Hongkong, and a Polish change of regime. Given the criteria specified above, the choices were ultimately narrowed down to the last alternative.

In the preparation, the game directors had produced a set of materials which included (1) the problem of the game; (2) a background paper on historical, geopolitical, and economic issues; (3) an overview of the functions and procedures of the United Nations (UN); (4) a paper on "The Armed Forces of Poland and Soviet Forces in Poland"; and finally, (5) a paper that described the hypothesized state of the world one year ahead. Only the last piece was disseminated among the participants prior to the start of the game.

The initial schedule of POLEX was that the game started with a one-hour briefing session, followed by a two-hour slot which would allow participants to enter their roles. After lunch, teams should begin with communication and make moves. It was expected that toward the end of the first day, there would be agreement to call a UN meeting (or general assembly) to promote a diplomatic solution of the crisis. This meeting would be held on the second day, with diplomatic deliberations and negotiations continuing for the first half of the third day. The final afternoon could then be used for "post-mortem" evaluation of the exercise. The primary aim of Bloomfield and his colleagues was not to produce substantive results on the problem, but rather to test whether Endicott House type games were worth the effort. Although there were some deviations to the plan in the actual game, the basic structure remained intact (Fig. 7.1).

In general, Bloomfield acknowledged the pioneering role of RAND scientists, but emphasized that the Endicott House type games, apart from minor procedural aspects, were different in at least two aspects. To

Fig. 7.1 Work session of CONEX I at Endicott House, photograph, ca. 1968. (Photograph with seven men sitting around table. The man on the left is smoking a pipe and the third from the left adjusts his eyeglasses, undated, Lincoln P. Bloomfield Papers, MC 326, Box 11, CONEX I, folder 1/2. Massachusetts Institute of Technology Libraries, Department of Distinctive Collections, Cambridge, MA)

clarify the first, he argued that one had to discern two types of political games, the "reality game" on the one hand, and the "normative or optimal strategy type" on the other (Bloomfield and Padelford 1959, 1112). In reality games, like POLEX, *all* participants were advised to make those moves they assessed as the most plausible for the countries they represented. In contrast, normative games invite more eagerness to experiment. RAND's political gaming, for instance, had allowed for more freedom in the selection of strategies on the part of the US team (see Chap. 4). The purpose of not restricting the selection of strategies to the most plausible was, at least in principle, to identify an "optimal" strategy out of a set of alternative strategies. Related to this first difference was a second, which concerned the core of the game—what was to be simulated? In the Endicott House type game, Bloomfield and Padelford ascertained, it was the interaction among several bodies and nations that was simulated. The RAND games, they argued, had been more concerned with the evaluation of "specific and detailed strategic moves and countermoves" (Bloomfield and Padelford 1959, 1114). This was also the reason for the difference in the number of teams employed in the two types of games: whereas, in addition to the referees and the Committee on Nature, RAND games were played by three to a maximum of four teams, Endicott House type games, with their emphasis on interactions, had decisively more teams, for example, ten teams in the first POLEX.

Like his colleagues at RAND, Bloomfield dismissed the idea that political gaming procedures, however they were designed, could be used for prediction. To him, the notion that the future can be predicted by the use of gaming was an "ultimate temerity" (Bloomfield 1960b, 61). His justification for this position is worth quoting at length (Bloomfield 1960b, 61):

> There are variables and intangibles in any human situation which cannot comprehensively be anticipated if only because of the sheer cussedness of mankind, a quality which confounds the most elaborate prognostications. And, perhaps most serious of all, there is no assurance that culture-bound Americans, however dramatic their role-playing, are going to react identically with the kinds of mentality they are attempting to simulate, particularly those shaped in drastically different cultures and subject to pressures

> which we may find it impossible to evoke. Perhaps we should introduce the Stanislavski method into gaming. At all events, this is an imperfection which cannot be lost sight of.

Bloomfield continued to devise political games for virtually his entire career. Though it took place once again at Endicott House, POLEX II was not an Endicott House type game in the sense deployed by Bloomfield and Padelford, but instead followed more closely the principles developed at RAND. It dealt with the interests and strategies of the two superpowers toward Iran, had only two teams, and allowed the US side to play not a realistic, but a "'deviant' or 'optimal' strategy"(Bloomfield 1960a, 2). From 1962 to 1964, Bloomfield ran four games in a series called POLEX-DAIS and, together with Barton Whaley (Bloomfield and Whaley 1965), a further three in a series called DETEX. In parallel, Richard E. Barringer and Barton Whaley organized an evaluation study of the MIT games and invited all 130 participants of the first eight MIT political games completed at this time (POLEX I and II, POLEX-DAIS I, II, III, IV, and DETEX I and II) to fill in a questionnaire and, incidentally, to participate in a follow-up interview. Like the results from the POLEX-DAIS series, the results of this evaluation were published in *Orbis* (Barringer and Whaley 1965). In the second half of the decade, a further series of games, now entitled CONEX I to IV, was conducted by CENIS at Endicott House (cf. Bloomfield et al. 1970). Although accompanied by scattered claims from academic political scientists to clarify the methodological foundations of gaming instead of "attempting to look into the black box of political life" by looking into "another black box—political gaming" (Schwartz 1965, 693), Bloomfield as well as the decision-makers funding his gaming exercises remained convinced of the value of the approach (see Fig. 7.2).

Generally speaking, gaming as a technique of analysis and teaching was taken up in many places. But MIT's CENIS was certainly the place where political gaming, as opposed to military gaming, was pursued most intensively. However, comparable games were also conducted at the Pentagon. Air force officer and RAND consultant William Jones had introduced gaming to Pentagon officials, and "politico-military" or "crisis games" were conducted since 1961 under the auspices of the Joint Chief

Fig. 7.2 "What do you mean you are going on strike?," photograph, 1968. (Lincoln P. Bloomfield is standing in the middle, face to the camera (Lincoln P. Bloomfield Papers, MC 326, Box 11, CONEX I, folder 1/2. Massachusetts Institute of Technology Libraries, Department of Distinctive Collections, Cambridge, MA))

of Staff's *Joint War Gaming Agency* and, later, in Pentagon's *Studies, Analysis and Gaming Agency* (SAGA).[4] Referring to recollections of David Halberstam, Brewer and Shubik (1979, 106) report that "[d]uring a game [on the situation in Vietnam] devoted to exploring the likely impact of the United States' bombing, it appeared quickly and decisively that North Vietnam would be especially resilient and resistant to bombing. The impact," they concluded, "of the game and its findings on actual policy choices is moot." Pentagon continued to conduct crisis games throughout the 1960s and apparently the 1970s. Former and present RAND researchers were repeatedly invited.[5] It was attempted to keep this concealed from the public. However, trouble arose when the information about the games was leaked. Pentagon officials were forced to resentfully defend the idea that adult men play games instead of thinking earnestly about the nation's security.

Throughout the 1960s and 1970s, gaming was also used outside the areas of military strategy and foreign policy. For instance, games were developed for approaching problems related to city and regional planning. In the more famous cases of San Francisco and Pittsburgh, the application of simulation techniques formerly developed for military usage was restricted to operational issues (Light 2003, 55ff). However, there were several attempts to construct urban development games where, in contrast to the San Francisco and Pittsburgh simulations, "more than one set of decision makers [was] considered as players whose free will is capable of influencing outcomes" (Brewer and Shubik 1979, 36). Probably the best known of these early games is CLUG, the Community Land Use Game, which was developed by Allan Feldt in 1963. "The game was originally intended for graduate students but it has been used successfully with undergraduate and high school students as well as professionals, local officials, and businessmen" (Feldt 2010). Gaming approaches have since been on the agenda of urban planners, although almost always with educational intent (Taylor 1971).

Trust, Hope, and the Uses of Ambivalence

The techniques of prognosis discussed in this book sounded out the potential of the idea to use experts for prognostication, taking the exploration in a range of directions. However, they agreed on three core theses. The first thesis was that decisions—in foreign policy, but also in many other branches of social life—required foreknowledge. While this problem in itself was not new, the culture of insecurity had furnished it with a hitherto unknown urgency (Chap. 1). Not only that stable foreknowledge led to better, more reasonable, and more efficient decisions in the present. Moreover, good decisions in the present were perceived as paramount to securing the future existence of Western culture. This is where the expert entered the scene. The second thesis upon which all our proponents agreed was that the expert was an essential and indispensable element in the production of foreknowledge. The epistemic hope related to the expert thus was that she could help society cope with the omnipresent insecurity. It also implied that better than lay people, the expert was able to push back on personal motivations and needs and provide undistorted, "truthful" assessments (cf. Dayé 2018b). And third, the construction of transparent and structured techniques appeared to be the best way to have this hope realized. Because experts were human beings and thus fallible, techniques that took their inspiration from the field of scientific social research offered the option to ask collectives of experts, thus countering the risk of bias by single outlier opinions.

What exactly such techniques should look like, however, was controversial. In the focus of these controversies was the epistemic role of the expert, that is, the expectations about what and how experts knew things, and by which means they would provide this knowledge. The main camps of this controversy can be described by looking at the alleged epistemic capacities of experts: some techniques implied that experts were able to predict, that is, to come up with reasonable *predictions* based on what they—explicitly and tacitly—knew. In this camp, we find the earlier, non-RAND studies by Cantril (1938) and McGregor (1938) as well as the "precursor" study to Delphi by Kaplan et al. (1949). However, as Chap. 6 showed, even the long-range forecasting Delphi study carried out by Gordon and Helmer (1964) and the parallel study by Brown and Helmer (1964) belonged to this camp, as well as the subsequent canonization of the Delphi paradigm. In this camp, the epistemic expectations toward the expert emphasized prediction, and an inquiry into her reasoning was not feasible. You would not ask an oracle to justify its statements.

In the second camp, the epistemic role of the expert assumed an ability to *forecast*, and adequate procedures had to provide for an accessible set of evidential materials, upon which the experts based their estimations, as well as for a channel to discuss the rationales behind the estimations. This was formulated as a principle requirement in "On the Epistemology of the Inexact Sciences" (Helmer and Rescher 1959), and had been implicitly followed by the first Delphi study a few years earlier (Dalkey and Helmer 1963).

Finally, however, a third camp settled around the conviction that the complexities of the social world overwhelmed, to a certain degree, the prognostic capacities of experts, and that instead of stable estimates or forecasts, one could only strive for a tentative, impression-like *prospection*. However, since the techniques applied to generate this prospection fostered the collaboration among experts, a social collective supported this prognosis, making it more reliable than a single individual's view of the future. Moreover, by simulating real-world processes, the interactive structure fostered the confrontation of various standpoints, thereby broadening the knowledge of the participants and, on the collective level, further increasing the stability of the prospection.

Apart from this trisection, RAND prognosticators also held different views on the question whether and how experts should interact. The proponents of RAND's Mathematics Division claimed that direct interaction had an almost certain effect of distorting the prognostic results. The members of the Social Science Division (SSD), however, saw the direct interaction of the experts as a crucial source of realism in their games. This was a consequence of the divergent disciplinary backgrounds of the proponents (Dayé 2014, 2016). The Delphi team received their academic training in a specific school of analytical philosophy of science, logical empiricism. The political gamers, on the other hand, had strong roots in the sociology of knowledge (Bessner 2014), a program developed by Hungarian-German sociologist Karl Mannheim (1893–1947) who had been Hans Speier's mentor in Heidelberg, had taught at the London School of Economics (LSE) while Goldhamer had been there, and was married to the sister of Kecskemeti's wife.

These different background philosophies also had a consequence that was more fundamental for how the proponents conceived of the nature of the factors that determined the future. In its paradigmatic form, Delphi focused on developments. The prognosis it delivered was a synthesis of all these developments. Life in 2014 was to be shaped by innovations in science and technology, population growth and counter-measures, the spread of automation, explorations in space, and future wars and weapons systems. At least, these were the factors deemed of crucial relevance to decision-makers. Political gaming, on the other hand, did not focus primarily on factors, but on decisions and actions. Thus, it offered a more open prognosis. In other words, Delphi offered a picture of the future that was shaped by technological and ecological factors; this future should then be used for actual decision-making. In contrast, in simulating the possible development of political conflicts, gaming emphasized the agency of decision-makers. It was their actions that shaped the future. The factors deemed relevant in specific decision situations were assessed as part of the ex post analysis. In some sense, thus, the worldviews implicit in the two RAND techniques were fundamentally distinct. Delphi bore signs of technological determinism, whereas political gaming emphasized human agency. One sign of the strength of the epistemic hope is that these fundamental differences were if not downplayed, then at least not acknowledged to a degree that would have led to a more critical reflection on the underlying methodologies (cf. Dayé 2019).

In both cases, however, the epistemic hope was that the expert would help the world cope with the culture of insecurity. This hope, however, was a general one. It was not restricted to the scientific members of the Cold War policy networks and not even to the networks as a whole. It was a hope that, although not always stated explicitly, was widely shared by all sorts of people. The hope that they could and would make the world a little less insecure was the fundamental source of any form of trust attributed to the social figure of the expert in the Cold War era. While it was contested, the narrative structure weaving together the notions of hope, trust, and the social figure of the expert was effective and stable. It was effective to a degree that allowed our scientific actors and their audiences to continue, against available evidence, to believe in the predictive superiority of experts as compared to lay people. Since no empirical argument justified this, their continued belief indicates the psychological forces at work. It thus provides another way to assess the strength of the epistemic hope in experts.

In a comparable way, both analysts and audiences remained rather uncritical toward the self-referential elitism involved in providing expertise. Members of the ruling elite asked scientists to provide answers to the issues at stake, and thus moved them closer to the circles of decision-makers. The answers they gave consisted, to a large degree, in justifying the use of experts as the only reasonable, rational thing to do. As Olaf Helmer put it in his widely received book, *Social Technology*,

> reliance on the use of expert judgment, though often unsystematic, is more than an expedient: it is *an absolute necessity*. Expert opinion *must* be called on whenever it becomes necessary to choose among several alternative courses of action in the absence of an accepted body of theoretical knowledge that would clearly single out one course as the preferred alternative. (Helmer 1966, 11; emphasis added)

Further, the narrative weaving together hope, trust, and the expert was stable because many influential actors profited from its presence. Decision-makers profited from the promise that their decisions could be based on sound empirical evidence. The public profited from the assurance that science was there to level out the irrationalities that sometimes dictated the decisions by the elite in government and military. And the scientific experts profited from the authority accredited to them. Moreover, they actively sought to create, justify, maintain, and corroborate the trust toward experts. The analysis presented on these pages clearly indicated that ambivalence was the most important discursive strategy employed by RAND's scientists to create and justify trust.

Sometimes, ambivalence is the cause of a feeling of insecurity. In a discursive setting, however, ambivalence also has the capacity to put an actor in a position to make statements that suggest clarity without neglecting the unclarity of the issue at hand. Such a statement can thus achieve the comforting effect of security without involving an outright lie. One example of such a statement was mentioned in Chap. 6. Experts participating in the long-range Delphi study were asked to estimate when they thought the chances of realization for a specific innovation were at 50%—yet, the graphic displays did not reflect this, thus suggesting not a likelihood, but "certainty."

Ambivalence was used by RAND's scientists—or, for that matter, by many scientists engaged in the Cold War effort—to support the public hope in science and the trust in scientists' capabilities. The strategic use of ambivalence concerned various dimensions, of which I discuss three: the differentiation of the observers and the observed, the nature of the produced statements, and the epistemological status of the proposed procedures.

RAND's techniques of prognosis put scientists into an argumentative setting that was obviously circular (see Chap. 1). Decision-makers hoped that scientists could deliver relevant knowledge, and the addressed scientists themselves turned their eyes on scientists. The methodological answer provided by the techniques of prognosis was self-referential, and this was not so much a methodological problem as it was a social one. Seeking to maintain the trust they received from elites and the resulting closeness to decision-makers, RAND researchers successfully established a discursive differentiation between scientists and experts. They managed to introduce this differentiation into the discourse without having to clarify it, and used this differentiation to claim a separation of observer and observed and thus, in their epistemological conception, scientific objectivity, and neutrality. This differentiated them from other futurists who were very outspoken about their double role as interpreters and constructors of the future—among them those who W. Patrick McCray (2013) aptly described as "visioneers," people who both envision how the future will be changed by revolutionary technologies and set to develop those very technologies. RAND researchers might have well been aware of their constructive function, and on occasions in their writings, they also referred to the phenomenon of self-fulfilling prophecies (Merton 1948). Yet still, they insisted on the juxtaposition of scientists crafting the prognoses and experts delivering input. That this differentiation had no factual basis was not decisive for the success of the discursive strategy.

The second dimension of ambivalence that Cold War social scientists used to justify and corroborate the trust they received concerned the status of their findings. As analyzed in the preceding chapters, they held varying understandings of the natures of their prospections. Some conceived of them as predictions, statements about future events that were made without any reference to evidence; others thought of them as forecasts, statements about future events that were based on an assessment of available evidence; and some treated them as prospections, by which they meant not statements about the future *per se*, but statements defining those factors that were most likely be shaping the future. That these were not clearly differentiated might have been a consequence of the lack of previous knowledge. Delphi and political gaming rank among the first techniques developed within this nascent field, and only with the publication of the 1964 Delphi report did the movement of future studies gain momentum (cf. Andersson 2012; Tolon 2012). Also, the lack of differentiation here might be a consequence of the philosophical tradition of those who concerned themselves most fundamentally with the epistemological implications of their doing, which tended to discern explanation and prediction, but not different forms of prediction. Be it as it may, the resulting ambivalence about the very nature of the statements that their techniques of prospection produced again made it easier to maintain the trust and hopes that was placed on the scientists' shoulders.

The third and final dimension of ambivalence concerned the epistemological nature of the proposed procedures. Here, our proponents shifted between understanding their procedures as scientific methods on the one hand and analytic techniques on the other (cf. Dayé 2018a). Initially, RAND researchers explicitly set out to develop methods that produce scientific results. In their contemporary perspective, the crucial criterion for science might be intersubjectivity (cf. Chap. 5), but of course, the ultimate ideal was truth. This was obvious from Kaplan et al.'s attempt to compare predictions with reality 20 weeks after the interrogation, and it was obvious from *OEIS*. Taking up a distinction fundamental to the philosophy of science of their time, to strive for truth was the crucial difference between a scientific method and an analytic technique, which was to be judged by its usefulness. Whenever the argument concerned their epistemic authority and their self-image, the register they used was that one of science, of intersubjectivity, and ultimately of truth. But, on the other hand, whenever they deemed it appropriate, they could draw back to the hill of usefulness and describe their procedures as mere techniques. This is the main motif behind the argument that expert opinions are the only source decision-makers can turn to when requiring foreknowledge of events that do not allow for statistical extrapolation or prediction by causal laws. When asked whether their research was scientific or not, their reaction was twofold. They gave a "qualified 'yes,'" as Olaf Helmer put it in a talk mentioned in Chap. 6 (Helmer 1963, 1); but virtually in the same breath, they asserted that anyone working in this craft "is, of necessity, a pragmatist, interested primarily in effective control of his surroundings and only secondarily in detailed understanding of all the underlying phenomena" (Helmer 1966, 5).

This ambivalence concerning the epistemological status of their procedures also encroached on the interpretation of the findings. Was it true that "limited weather control, in the sense of substantially affecting regional weather at acceptable cost" would be realized between 1987 and 2000? Or was it a useful thing to consider in decision-making that limited weather control might perhaps be realized by then? The proponents of Delphi and, albeit to a lesser degree, of political gaming used both registers as they saw fit, thus creating an ambivalence that left them more comfortable with regard to the high expectations laid upon them. They had reached the ears of the powerful, and some degree of ambivalence seemed to allow them to juggle their ambition to produce valuable results with their wish to stay close to the decision-makers.

To be sure, this notional and theoretical ambivalence has to be discerned from the factual ambivalence involved in any attempt at prognosis. This point becomes clear when we compare the story told on these pages with the recent account by Eglė Rindzevičiūtė (2016) of the prognostic studies carried out by researchers at the already mentioned International Institute for Applied Systems Analysis (IIASA) in the 1970s and 1980s. Located in Schloss Laxenburg, a castle a few miles outside of Vienna, Austria, the IIASA was the first international research organization sponsored jointly by Eastern and Western powers (Levien 2000; Riska-Campbell 2011; Duller 2016). It became the place where systems analysis, an approach initially developed at RAND (see Chap. 6), was applied to issues of global relevance. In their systems analyses, IIASA researchers combined various simulation and modeling techniques to estimate environmental risks related to, for example, the nuclear winter, or acid rain. However, the epistemology informing systems analysis, with its roots in cybernetics, acknowledged the factual ambivalences involved in the craft of scientific prognosis. Systems analysts, both from the East and from the West, were sincere about the problems involved in their doings, and cautious about the statements they produced. In contrast to accounts that emphasize the "closed world" character of the Cold War, her thesis is that the broad international consensus on the nature of global problems, on the difficulties involved in scientifically exploring them, and on the factual restrictions and ambivalences inevitably resulting from attempting prospection in fact opened the Cold War world. "This new epistemology undermined both the Marxist-Leninist view of stage-driven development and high modernist beliefs in control" (Rindzevičiūtė 2016, 207). It imposed limits on governmental optimism regarding the naturalness of economic and societal development, but more importantly regarding the ability to control the future.

Obviously, the ambivalences effective in this case were different than the one examined in this book. While later researchers were outspoken about the limitations and restrictions of their studies and were thus trying to introduce ambivalence into the thinking of decision-makers, the earlier researchers described in this book used notional ambivalence to conceal the limitations and restrictions of their studies to maintain the impression of trustworthiness. While the epistemology involved in the IIASA studies functioned as a "version of an organized skepticism" (Rindzevičiūtė 2016, 207) toward the plans, visions, and convictions of decision-makers, RAND studies used the notional ambivalence to convince decision-makers of their universalism and their disinterestedness.

Experts, Trust, and Liberal Democracy

For a certain period, and with varying degrees of presence in everyday debates, the public, decision-makers, and scientists have shared the epistemic hope that science could steer power into producing good policy. In the case of the scientists, this hope was strong enough to soothe the uneasiness that the inconsistencies involved in their doing, and the ambivalences on which their position rested, created. While there are new epistemologies at work today, we are still in the age of the expert and we still hope that research will foster good policy. Nonetheless, we sometimes long for the times when (social) science was closer to decision-makers, when there was trust in science. If there is something to learn from this study of the Cold War expert, then that trust comes at a cost. In this particular case, it came at the cost of theoretical and notional clarity. Ambivalence resulted from the attempts to secure the attention of the elites. A few decades later, the path taken by IIASA's cyberneticists was to be clear about the factual ambivalences involved in scientific prognosis. Yet while this path might have maintained the trust on behalf of the decision-makers, public trust disappeared. Science lost its authority as a source of societal reason, and the price to pay for having the ear of the decision-makers was the restriction of the audiences to those in power.

Of course, this loss of public audience has disadvantageous effects on the social status of the scientist, but the true problem of this development lies in its consequences for the shape of democracies. This problem has received its most comprehensive treatment by Stephen Turner (2001, see also 2003, 2014). To make his point, Turner proposes a typology that differentiates experts according to their relation to and the nature and size of their audiences. Table 7.1 summarizes this typology. The first type mentioned is the scientist expert. Scientists occasionally discuss their findings in the public sphere and sometimes receive invitations to participate in policy-making processes. Since university scientists are paid by tax money, there exists the social expectation that they participate in public debates as soon as the issues treated there concern their field of specialty—a specific aspect of the Mertonian norm of communism. Nonetheless, their primary audience is not the public, but their fellow scientists. This is also where the authority of their claims to knowledge originates. In Turner's words (2001, 131), the authority of specific scientific fields is "more or less democratically acknowledged." Their public authority is a result of their scientific authority. They are seen as legitimate representatives of their scientific discipline, and the public acknowledges the authority of the science they represent. For these type I experts, the public is audience as well as legitimator and subsidizer.

In contrast, Turner considers the case of a sectarian preacher. Just like type I experts address their peers, the audience of preachers consists of members of a specific thought collective. However, the authority of these type II experts does not extend beyond the boundaries of what Turner calls the "restricted audience," the community of believers. His or her

Table 7.1 Types of experts according to Turner (2001)

	Audience	Authority acknowledged by	(Financial) Support from
Type I	Peers	Public	Public
Type II	Restricted audience	Restricted audience	Restricted audience
Type III	Created audience	Created audience	Created audience
Type IV	Public/subsidizers	Public	(Philanthropic) organizations
Type V	Individuals with discretionary power	Peers	Public

cognitive authority is legitimated only within this community. Both type I and type II experts rely on pre-existing audiences. In contrast, type III experts rely on their abilities to create and maintain their own audiences. Examples offered by Turner are Dr. Ruth (Westheimer) and the massage therapist: both depend, in their authority as well as, in consequence, economically, on the judgment of their customers.

For these three types, Turner (2001, 132) argues that they "have a place in the scheme of liberal democracy." While type I expertise is seen to be neutral, the liberal state attempts to be neutral toward the other two. However, other types of experts challenge the idea of the liberal state more markedly. Turner describes two further types of experts who both are "subsidized to speak as experts and claim expertise in the hope that the views they advance will convince a wider public and thus impel them into some sort of political action or choice" (S. Turner 2001, 133). Type IV experts, whose historic roots Turner finds in the activities of philanthropic foundations like the Russell Sage Foundation in the early twentieth century, are concerned with affairs of public relevance. Still, they do not address the public directly, but instead try to persuade policy-makers as well as other persons and organizations who might have an interest in promoting and realizing their idea. These addressees, so the experts expect, are to provide financial support and at the same time lend authority to the experts' claim. Whereas the expertise of the type I expert might be policy-*relevant*, the expertise claimed by type IV experts is inherently policy-*oriented*. From the point of view of liberal democracy theory, this type of experts entails considerable dangers. They have been chosen as representing a view that is not necessarily their own and earn money by doing so, whereby the sources of funding are usually kept secret. Their financial dependency on political organizations calls into question the truth-value of the ideas they are paid to promote, which seems comparably unproblematic when ideas are selected according to their scientific merit.

Type V experts are a historical development of the fourth type. Its "primary audience is not the public, but individuals with discretionary power, usually in bureaucracies" (S. Turner 2001, 135). Turner discusses public administration as a well-known example of this type. The expertise of public administration professionals is accepted and highly valued within their community. Quite often, these experts are involved in legislation procedures or what Sheila Jasanoff calls "regulatory science" (Jasanoff 1990). They are entitled to make decisions on state issues, and these decisions are consequential for the public; moreover, the experts are paid by the public, but the public has neither accepted nor legitimated them. Usually, it does not even know them. The combination of being invisible to the public eye and having the power to decide about public affairs marks the point where the relation of democracy and expertise becomes problematic. "The expert who is a threat is the expert who exerts influence through the back door of training and validating the confidence of professionals, and whose advice is regarded as authoritative by other bureaucrats but not by the public at large" (S. Turner 2001, 140). In Turner's analysis, this threat arises not from any specific feature of knowledge, but from the social communication network that formulates expertise and seeks to generate an impact.

> [T]he difficulties that have concerned theorists of democracy about the rôle of expert knowledge must be understood as arising not from the character of expert knowledge itself (and its supposed inaccessibility to the masses), but from the sectarian character of the kinds of expert knowledge that bear on bureaucratic decision-making. (…) The authority of the expert whose expertise is not validated by public achievements is the authority that comes into conflict with democratic processes. (S. Turner 2001, 140)

Hence, due to their specific patterns of communication and influence, experts of the fifth type are most clearly in structural conflict with the principles of liberal democracy. They take public money, but do not communicate publicly, and like sectarians, they seal their knowledge claims off from the criticisms coming from outside their immediate reference group. Unlike sectarians, however, their decisions have an impact on those outside these inner circles.

Reconsidering the historical developments with this typology in mind, we can think of RAND in the 1950s as an organization of experts of both types I and IV. In the words of Joy Rohde (2013), these experts are part of a gray area, an interstitial field stretched between the principles of academic knowledge production and the demands of the new customers in government and military agencies. When, as a consequence of the increasing public odium described in Chap. 1, the gray area dissolved, one part stayed oriented toward academia, whereas the other, larger part went into the protective covers of contract research institutes or the agencies themselves. There, however, they were withdrawn from any form of academic quality control: their texts were not published in scientific journals, but instead "distributed in limited numbers to the agency that paid for the work and to other interested research outfits and federal agencies. They were not peer reviewed, nor were they widely available to scholars even when they were unclassified" (Rohde 2013, 133). These parts of knowledge production became the arena of type V experts.

In the 1970s and 1980s, organizations like RAND and IIASA certainly were among those organizations that continued to follow the demands of scientific knowledge production and dissemination, thus continuing to understand and present themselves as type I or type IV experts. Nonetheless, they suffered from the loss of status that resulted from both the public attacks on the social position of the expert and from the sincerity with which they openly addressed the factual ambivalences involved in their work. However, they saw it as the price that was to be paid for the continued attention of the decision-makers. Today, presumably, the price to pay might take yet another form. The world changed, and so did the structures and conditions of knowledge production, as well as, more generally, the place of knowledge (or truth) in policy-making. Which form the price will take that science has to pay in order to be trusted, is not yet clear. Most probably, however, it will not even be up to the scientists to decide on this. Almost certainly, though, it will imply a reshaping of the constellations in which knowledge, power, and the public sphere struggle to make sense of the current culture of insecurity.

Notes

1. The invention of cross-impact analysis was fostered by a contract with Kaiser Aluminum & Chemicals Corporation to design a future-oriented game to celebrate the firm's twentieth anniversary. The game, simply

called *Future*, came out in 1966 (cf. Interview with Theodore J. Gordon by the author, 16 August 2013, pp. 3–4). Helmer had already collected experience with designing games. Together with Lloyd Shapley, he developed *Summit* which was published in 1961 by Cameo Games (cf. Interview with Martin Shubik by the author, 2 September 2011, p. 3).
2. Cf. http://www.millennium-project.org/about-us/planning-committee/ted-gordon/, (accessed 17 July 2019).
3. In 2009, 2900 experts participated in the survey. Based on their input, NISTEP's ninth Delphi Survey forecasts 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 (cf. NISTEP 2010, 12).
4. After resigning from government service, Jones continued to devise political-military games at RAND (e.g., Jones 1986).
5. Albert Wohlstetter, then already at the University of Chicago, participated in a JWGA crisis game in spring 1967 (Albert & Roberta Wohlstetter Papers, Hoover Institution Archives, Stanford, California, box 134, folders 13 and 14).

References

Andersson, Jenny. 2012. The Great Future Debate and the Struggle for the World. *American Historical Review* 117 (5): 1411–1430.

———. 2018. *The Future of the World: Futurology, Futurists, and the Struggle for the Post Cold War Imagination*. Oxford and New York: Oxford University Press.

Baade, Fritz. 1960. *Der Wettlauf zum Jahre 2000. Paradies oder Selbstvernichtung*. Oldenburg and Hamburg: Gerhard Stalling.

Barringer, Richard E., and Barton Whaley. 1965. The M.I.T. Political-Military Gaming Experience. *Orbis* IX (2): 437–438.

Bell, Daniel. 1973. *The Coming of Post-Industrial Society: A Venture in Social Forecasting*. New York: Basic Books. https://www.worldcat.org/title/the-coming-of-post-industrial-society-a-venture-in-social-forecasting/oclc/7997065063&referer=brief_results.

Bessner, Daniel. 2014. Weimar Social Science in Cold War America: The Case of the Political Game. In *More Atlantic Crossings? European Voices in the Postwar Atlantic Community*, Bulletin of the German Historical Institute Washington DC, Supplement 10, ed. Jan Logemann and Mary Nolan, 91–109. Washington, DC: German Historical Institute.

Bloomfield, Lincoln P. 1960a. *Political Exercise II—The U. S. and the U. S. S. R. in Iran*. Cambridge, MA: Center for International Studies, Massachusetts Institute of Technology.

———. 1960b. Political Gaming. *United States Naval Institute Proceedings* 86 (9): 57–64.

Bloomfield, Lincoln P., and Norman J. Padelford. 1959. Teaching Note: Three Experiments in Political Gaming. *The American Political Science Review* 53 (4): 1105–1115.

Bloomfield, Lincoln P., and Barton Whaley. 1965. The Political-Military Exercise. *Orbis* VIII (4): 854–870.

Bloomfield, Lincoln P., Cornelius J. Gearin, and James L. Foster. 1970. Anticipating Conflict-Control Policies: The 'CONEX' Games as a Planning Tool. C/70-10. Arms Control and Local Conflict. Center for International Studies, Massachusetts Institute of Technology, Cambridge, MA.

Brewer, Garry D., and Martin Shubik. 1979. *The War Game: A Critique of Military Problem Solving*. Cambridge, MA and London: Harvard University Press.

Brown, Bernice B., and Olaf Helmer. 1964. Improving the Reliability of Estimates Obtained from a Consensus of Experts. P-2986. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2986.html.

Cantril, Hadley. 1938. The Prediction of Social Events. *Journal of Abnormal and Social Psychology* 33 (3): 364–389.

Cohen, Bernard C. 1962. Political Gaming in the Classroom. *The Journal of Politics* 24 (2): 367–381.

Cornish, Edward. 1977. *The Study of the Future. An Introduction to the Art and Science of Understanding and Shaping Tomorrow's World*. Washington, DC: World Future Society.

Dalkey, Norman C., and Olaf Helmer. 1963. An Experimental Application of the Delphi Method to the Use of Experts. *Management Science* 9 (3): 458–467.

Davison, W.P. 1958. A Summary of Experimental Research on 'Political Gaming'. D-5695-RC. RAND Corporation Archives, Santa Monica, CA.

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

———. 2016. 'A Fiction of Long Standing': Techniques of Prospection and the Role of Positivism in US Cold War Social Science, 1950–1965. *History of the Human Sciences* 29 (4–5): 35–58.

———. 2018a. How to Train Your Oracle: The Delphi Method and Its Turbulent Youth in Operations Research and the Policy Sciences. *Social Studies of Science* 48 (6): 846–868. https://doi.org/10.1177/0306312718798497.

———. 2018b. The Expert as Messenger: Media Philosophy and the Epistemology of the Inexact Sciences. In *Dynamiken der Wissensproduktion. Räume, Zeiten und Akteure im 19. und 20. Jahrhundert*, ed. Wolfgang Göderle and Manfred Pfaffenthaler, 239–258. Bielefeld: Transcript.

———. 2019. Die Blindheit der Auguren: Delphi, Political Gaming und das Phänomen der wechselseitigen Nichtbeachtung. In *Geschichte der Sozialwissenschaften im 19. und 20. Jahrhundert: Idiome—Praktiken—Strukturen*, Sozialwissenschaftliche Schriften 51, ed. Uwe Dörk and Fabian Link, 265–286. Berlin: Duncker & Humblot.

Deutsche Post AG. 2009. *Delivering Tomorrow: Kundenerwartungen im Jahr 2020 und darüber hinaus. Eine globale Delphistudie*. Bonn: Deutsche Post AG. http://www.dp-dhl.com/de/presse/mediathek/dokumente/delphi-studie_delivering_tomorrow_kundenerwartungen.html.

Duller, Matthias. 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. https://doi.org/10.1177/0952695116667882.

Feldt, Allen. 2010. Re. GAMES: Community Land Use Game (CLUG). *H-Urban* (blog). http://h-net.msu.edu/cgi-bin/logbrowse.pl?trx=vx&list=h-urban&month=1005&week=c&msg=0wZzcSmQBmexlFv13eDW%2BQ&user=&pw=.

Ferguson, Sherry Deveraux. 2000. *Researching the Public Opinion Environment. Theories and Methods*. Thousand Oaks, CA: Sage.

Fourastié, Jean. 1959. *La civilization de 1975*. Paris: Presses universitaires de France.

Gabor, Dennis. 1963. *Inventing the Future*. London: Secker and Warburg. https://www.worldcat.org/title/inventing-the-future/oclc/1015162277?referer=di&ht=edition.

Goldhamer, Herbert, and Hans Speier. 1959. Some Observations on Political Gaming. *World Politics* 12 (1): 71–83.

Gordon, Theodore J. 1965. *The Future*. New York: St. Martin's Press.

Gordon, Theodore J., and Olaf Helmer. 1964. Report on a Long-Range Forecasting Study. P-2982. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2982.html.

Guetzkow, Harold. 1959. A Use of Simulation in the Study of Inter-Nation Relations. *Behavioral Science* 4 (3): 183–191.

Guetzkow, Harold, and Lloyd Jensen. 1966. Research Activities on Simulated International Processes. *Background* 9 (4): 261–274.

Guilhot, Nicolas. 2017. *After the Enlightenment: Political Realism and International Relations in the Mid-Twentieth Century*. Cambridge, UK: Cambridge University Press.

Häder, Michael. 2009. *Delphi-Befragungen. Ein Arbeitsbuch*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Helmer, Olaf. 1963. The Systematic Use of Expert Judgment in Operations Research. P-2795. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P2795.html.

———. 1966. *Social Technology. Contributions by Bernice Brown and Theodore Gordon*. New York and London: Basic Books.

Helmer, Olaf, and Nicholas Rescher. 1959. On the Epistemology of the Inexact Sciences. *Management Science* 6 (1): 25–52.

Heuer, Richard J., Jr., and Randolph H. Pherson. 2011. *Structured Analytic Techniques for Intelligence Analysis*. Washington, DC: CQ Press (Sage).

Jasanoff, Sheila. 1990. *The Fifth Branch. Science Advisers as Policymakers*. Cambridge, MA and London: Harvard University Press.

Jones, William M. 1986. On the Adapting of Political-Military Games for Various Purposes. N-2413-AF/A. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/notes/N2413.html.

de Jouvenel, Bertrand. 1967. *The Art of Conjecture*. New York: Basic Books.

Jungk, Robert. 1952. *Die Zukunft hat schon begonnen. Amerikas Allmacht und Ohnmacht*. Bern and Stuttgart: Alfred Scherz.

Kahn, Herman, and Anthony J. Wiener. 1967. *The Year 2000: A Framework for Speculation on the Next Thirty-Three Years*. Macmillan.

Kaplan, Abraham, A.L. Skogstad, and Meyer A. Girshick. 1949. The Prediction of Social and Technological Events. P-93. The RAND Corporation, Santa Monica, CA. https://www.rand.org/pubs/papers/P93.html.

Levien, Roger E. 2000. RAND, IIASA, and the Conduct of Systems Analysis. In *Systems, Experts, and Computers. The Systems Approach in Management and Engineering, World War II and After*, ed. Agatha C. Hughes and Thomas P. Hughes, 433–462. Cambridge, MA and London: The MIT Press.

Light, Jennifer S. 2003. *From Warfare to Welfare. Defense Intellectuals and Urban Problems in Cold War America*. Baltimore and London: The Johns Hopkins University Press.

Linstone, Harold A., and Murray Turoff. 1975. *The Delphi Method. Techniques and Applications*. London and Reading, MA: Addison-Wesley.
McCray, Patrick. 2013. *The Visioneers: How a Group of Elite Scientists Pursued Space Colonies, Nanotechnologies, and a Limitless Future*. Princeton and Oxford: Princeton University Press.
McGregor, Douglas. 1938. The Major Determinants of the Prediction of Social Events. *Journal of Abnormal and Social Psychology* 33 (2): 179–204.
Merton, Robert K. 1948. The Self-Fulfilling Prophecy. *The Antioch Review* 8 (2): 193–210.
NISTEP. 2010. *The 9th Delphi Survey (Summary)*. Science and Technology Foresight Center, National Institute of Science and Technology Policy. http://www.nistep.go.jp/achiev/l_all-e.html.
Price, David H. 2016. *Cold War Anthropology: The CIA, the Pentagon, and the Growth of Dual Use Anthropology*. Durham and London: Duke University Press.
Rescher, Nicholas. 1997. H_2O: Hempel-Helmer-Oppenheim, an Episode in the History of Scientific Philosophy in the 20th Century. *Philosophy of Science* 64 (2): 334–360.
Rindzevičiūtė, Eglė. 2016. *The Power of Systems: How Policy Sciences Opened Up the Cold War World*. Ithaca and London: Cornell University Press.
Riska-Campbell, Leena. 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*, Commentationes Scientiarum Socialum 75. Helsinki: The Finnish Society of Science and Letters.
Rohde, Joy. 2013. *Armed with Expertise: The Militarization of American Social Research During the Cold War*. Ithaca and London: Cornell University Press.
Schwartz, David C. 1965. Problems in Political Gaming. *Orbis* IX (3): 677–693.
Taylor, John L. 1971. *Instructional Planning Systems: A Gaming-Simulation Approach to Urban Problems*. London: Cambridge University Press.
Toffler, Alvin. 1970. *Future Shock*. New York: Random House. https://www.worldcat.org/title/future-schock/oclc/906096031&referer=brief_results.
Tolon, Kaya. 2012. Future Studies: A New Social Science Rooted in Cold War Strategic Thinking. In *Cold War Social Science: Knowledge Production, Liberal Democracy, and Human Nature*, ed. Mark Solovey and Hamilton Cravens, 45–62. New York: Palgrave Macmillan.
Turner, Stephen. 2001. What Is the Problem with Experts? *Social Studies of Science* 31 (1): 123–149.
———. 2003. *Liberal Democracy 3.0: Civil Society in an Age of Experts*. London and Thousand Oaks, CA: SAGE Publications.

Turner, Fred. 2006. *From Counterculture to Cyberculture: Stewart Brand, the Whole Earth Network, and the Rise of Digital Utopianism*. Chicago and London: The University of Chicago Press.

Turner, Stephen. 2014. *The Politics of Expertise*. New York: Routledge.

Ward, Michael D., and Daniel Guetzkow. 2009. Harold S. Guetzkow, 1915–2008. *ASA Footnotes* 37 (1): 15.

Index[1]

A

Academy of Moral and Political Sciences, 17
Addams, Jane, 21
Amadae, S. M., 18, 25, 66, 119n3
Amara, Roy, 207
Andersson, Jenny, 10, 151n4, 199n3, 206, 224
Applied Mathematics Panel (AMP), 19, 71n1, 80
Arrow, Kenneth, 26, 81
Arthur Vining Davis Foundation, 207

B

Baade, Fritz, 206
Bachelard, Gaston, 8, 9
Bacon, Francis, 15
Bailey, Dana K., 95
Bales, John Freed, 82
Baran, Paul, 192, 206, 207
Barringer, Richard E., 217
Bavelas, Alex, 96
Behavioral Science Program (BSP, FF), 98, 121n19, 123n24
Bell, Daniel, 205
Bellman, Richard, 159
Berelson, Bernard, 121n19
Berkner, Lloyd, 96
Berlin School of Logical Empiricism, 56
Bessner, Daniel, 5, 93, 94, 97, 122n19, 130, 131, 221
Bloomfield, Lincoln P., 31, 116, 117, 206, 210–218

[1] Note: Page numbers followed by 'n' refer to notes.

C. Dayé, *Experts, Social Scientists, and Techniques of Prognosis in Cold War America*, Socio-Historical Studies of the Social and Human Sciences,
https://doi.org/10.1007/978-3-030-32781-1

Bohn, Lewis, 111
Brandeis University, 95
Brewer, Garry D., 113, 116, 119n7, 218, 219
Brodie, Bernard, 2
Brown, Bernice L., 160, 182, 184–190, 192, 193, 196, 197, 200n9, 220
Bruner, Jerome S., 96
Bundy, McGeorge, 121n17

C

Cameo Games, 232n1
Campbell, John C., 212
Cantril, Hadley, 43, 44, 67, 68, 71n2, 220
Carnap, Rudolf, 5, 55, 56, 147, 148, 151n2
Center for Advanced Study in the Behavioral Sciences (CASBS, Stanford), 122n19, 131, 210
Center for International Studies (CENIS), 98, 121n19, 206, 211–213, 217
Central Intelligence Agency (CIA), 98, 206
Cochran, Samuel, 193
Cohen, Bernard C., 213
Coleman, James S., 27
Collbohm, Franklin R., 26, 98, 102
Collins, Martin J., 23, 24, 71n1, 80, 81, 120n12, 121n16, 129, 151n1, 160
Columbia University, 17, 81
Committee on Nature, 100, 101, 216
Commons, John R., 17
Consensus, 4, 55, 63, 65, 104, 163, 166–168, 171, 176, 180, 183–185, 190, 193–195, 198, 199, 226
Cooperative General Culture Test, 49
Crossley, Archibald, 42

D

Dalkey, Norman Crolee, 5, 53–63, 65–70, 73n11, 73n14, 150, 164, 168, 191–194, 196, 197, 200n10, 220
Dantzig, George, 81
Dartmouth College, 211
Davison, W. Phillips, 83, 110, 209–213
de Sola Pool, Ithiel, 212
Degree of confirmation, 137–150, 153n12
Department of Defense (Pentagon), 24, 94, 158
DeWeerd, Harvey, 95, 103, 107, 111
Douglas Aircraft Company (McDonnel Douglas Aircraft Company), 4, 23, 159
Dresher, Melvin, 81, 82
Dylan, Bob, 174

E

Einstein, Albert, 147
Eisenhower, Dwight D., 30, 111, 122n23
Elias, Norbert, 22, 28
Ellsberg, Daniel, 34, 157, 158
Ely, Richard T., 17

Endicott House, 31, 209–219
Epistemic hope, 7, 205–231
Epistemic role, 7, 8, 11, 66, 68, 70, 118, 183, 196, 198, 205, 220
Etherington, Edwin Deacon, 207

F

Fat Man (Atomic bomb detonated over Nagasaki), 1, 58
Federal Communications Commission (FCC), 94
Federal Contract Research Centers (FCRCs), 131
Feldt, Allen, 219
Fisher, R. A., 146–148
Flood, Merrill, 71n1, 81, 82, 120n8
Ford Foundation, 98, 110, 121n19, 207, 210
Forecast, 9–11, 42, 44, 54, 70, 113, 118, 150, 162, 165, 166, 195, 196, 198, 220, 221, 224, 232n3
Foreign Broadcast Intelligence Service (FBIS), 94
Foreign Service Institute (FSI), 110
Foresight, 9, 118
Fourastié, Jean, 206
Franklin, Albert B., 110
Fraunhofer Institut, Germany, 208
Future (game), 232n1

G

Gabor, Dennis, 199–200n4, 206
Gaither Report (Report of the Security Resources Panel of the President's Science Advisory Committee, 1957), 65
Gaither, Rowan H., 65, 121n19
Gallup, George, 42
Garfinkel, Harold, 153n15, 154n15
Garthoff, Raymond, 111
Girshick, Meyer Abraham, 42, 43, 45–47, 49–51, 56, 67, 70–71n1, 150
Gliksman, J., 107
Goldhamer, Herbert, 5, 92–94, 98–108, 110, 111, 114–117, 122n20, 122n21, 122n22, 123n25, 130, 209, 210, 213, 221
Goldsen, Joseph M., 98, 102, 103, 107, 110–116, 209
Goodman, Nelson, 145
Gordon, Theodore Jay, 158–162, 164–169, 171, 173–179, 181–185, 192, 193, 195–197, 199n1, 205–208, 220
Gouré, Leon, 107
Guetzkow, Harold, 210

H

Häder, Michael, 199n1, 208
Halberstam, David, 218
Halpern, Abraham M., 107, 108, 111
Harvard University, 81, 82, 95, 97, 211
Heald, Henry, 110
Helmer, Olaf, 5, 24, 43, 44, 53–70, 72n7, 73n11, 80, 87–90, 93, 132–145, 147–150, 151n2, 151n3, 152n5, 152n7, 153n12, 157–169, 171–177, 181–193, 195–197, 199n1, 199n4, 200n9, 200n10, 205–208, 220, 222, 223, 225, 235n1

Hempel, Carl Gustav, 5, 55, 56, 132, 135, 136, 144–148, 152n6, 153n8, 153n9, 153n12, 153n13
Hirschberg, Fritz, 55, 72n7
Historical epistemology, 8
Hitch, Charles, 24, 158
Hitler, Adolf, 92
Hoag, Malcolm W., 111
Hoeffding, Oleg, 108, 111
Horowitz, Irving Louis, 18, 120n11
Hosiasson-Lindenbaum, Janina, 144, 153n13
Hungarian Revolution, 31, 95, 123n26
Hunt, Victor Myron, 93, 95, 98, 102, 103, 107, 110

I

Ibsen, Henrik, 55
Institute for the Future (IFTF), 192, 200n10, 206–208
International Institute for Applied Systems Analysis (IIASA), 208, 226, 227, 231
Inter-Nation Simulation (INS), 210

J

Jackson, Mahalia, 174
Janis, Irving, 27
Janowitz, Morris, 27
Jasanoff, Sheila, 230
Joe-1 (officially RDS-1, first atomic bomb detonated by the Soviet Union), 65
JOHNNIAC, 82, 119n6
Johnson, Alvin, 93
Johnson, Lyndon B., 30, 174
Jones, William, 217, 232n4
Jouvenel, Bertrand de, 199n4, 205
Jungk, Robert, 206

K

Kahn, Herman, 32, 33, 108, 111, 131, 205
Kaiser Aluminum & Chemicals Corporation, 231n1
Kaplan, Abraham, 42–54, 56, 58, 67–70, 70–71n1, 71n3, 150, 185, 186, 189, 196, 198, 220, 225
Kaplan, Fred, 2, 3, 18, 25
Kecskemeti, Paul, 5, 29–33, 93, 95, 101–103, 107, 111, 206, 213, 221
Kelly, Burnham, 96
Kennan, George, 121n17
Kennedy, John F., 174
King, Martin Luther, 174
Kirchheimer, Otto, 27
Kitchen, Jeffrey C., 111
Kluckhohn, Clyde K. M., 96
Köhler, Wolfgang, 55
Kramish, Arnold, 111, 192, 206, 207
Kriegsspiel, 78, 83
Kris, Ernst, 93
Kubrick, Stanley, 28
Kuhn, Thomas, 199n2
Kuklick, Bruce, 18

L

Lakoff, Andrew, 6, 10, 118
Landon, Alfred, 43

Lang Kecskemeti, Elisabeth, 31
Lasswell, Harold D., 121n13
Lazarsfeld, Paul F., 27
Lederberg, Joshua, 163, 164, 181
Lederer, Emil, 93
Leites, Nathan, 103, 107, 111
LeMay, Curtis E., 5
Lexington Field Station, 96
Limited Test Ban Treaty (LTBT), 174
Linebaugh, David E., 212
Little Boy (Atomic bomb detonated over Hiroshima), 58
Livermore, William Roscoe, Major, 78
London School of Economics (LSE), 94, 221
Luce, R. Duncan, 82, 119n4
Lwów–Warsaw school of logic, 144

M

Malcolm X, 175
Mallard, Grégoire, 6, 10, 118
Manhattan Project, 96
Mannheim, Karl, 4, 5, 93, 95, 221
Manstein, Erich von, 92, 93
March on Washington for Jobs and Freedom (1963), 174
Marquis, Donald, 96, 121n19
Marshall, Andrew, 103, 107, 110
Massachusetts Institut of Technology (MIT), 43, 95, 97–98, 116, 117, 206, 210, 211, 213, 215, 217, 218
McCombe, Leonard, 32
McCray, W. Patrick, 10, 208, 224
McGregor, Douglas, 43, 44, 48–50, 53, 67, 68, 71n2, 220
McKinsey, J. C. C., 82
McNamara, Robert, 34
Merton, Robert K., 8, 21, 22, 121n17, 224
Military worth, 71n1
Millikan, Max, 96, 98
Mills, C. Wright, 20, 21
Missile Gap, 65
Modernization, 180
Mood, Alexander McFarlane, 85, 86, 90–92, 120n10
Morgenstern, Oskar, 26, 78–82
Morison, Elting, 96
Morris, Charles W., 95
Morrison, John A., 96
Mosteller, Frederick C., 82

N

Nash, John F., 26, 82
National Defense Research Council, 56, 80
National Institute of Science and Technology Policy, Japan (NISTEP), 208, 232n3
National Research Foundation (NRF), 24
National Science Foundation (NSF), 20
Naval War College, Tokyo, 85
New School for Social Research, 93
Northwestern University, 210

O

Office of Naval Research (ONR), 24
Office of Strategic Services (OSS), 20
Office of War Information (OWI), 20, 94, 95
On Thermonuclear War, 32, 33

Oppenheim, Paul, 56, 132, 135, 136, 144–148, 153n8, 153n9, 153n12
Oppenheimer, J. Robert, 21, 22

P

Padelford, Norman J., 211–214, 216, 217
Page, Edward Jr., 111
Paxson, Ed, 80, 120n9, 157
Pearl Harbor, 85
Pentagon Papers, 34
Phillips, Thomas R., 29, 30
Pierce, John R., 96
Planning Programming Budgeting System (PPBS), 158
Platon, 16
Pool, Ithiel de Sola, 27, 212
Prediction, 9, 11, 11n1, 41–70, 133–137, 139–143, 145, 149, 150, 152–153n8, 161, 166, 171, 173, 183, 185, 194–198, 216, 220, 224, 225
Price, Don K., 26, 205
Princeton University, 43, 71, 81, 119, 132, 147, 209
Prognosis, 5–11, 23, 41, 42, 132, 197, 206, 208, 221, 223, 224
Project Troy, 95–98, 121n19
Prophecy, 11n1
Prospection, 9, 34, 118, 150, 206, 221, 224–226
Purcell, Edward M., 96
Pye, Lucian, 211, 212

Q

Queens College, 132

R

RAND Economics Division, 24–26, 158
RAND Mathematics Division, 4, 11, 26, 56, 71n1, 80, 84, 197, 221
RAND Social Science Division, 4, 11, 24–27, 31, 77, 94, 98, 111, 116
Realism, 84, 88, 89, 91–102, 115, 212, 221
Reichenbach, Hans, 5, 22, 55, 56, 151n2, 153n8
Rescher, Nicholas, 5, 55, 56, 71n3, 130, 132–139, 141–143, 147–149, 151n2, 151n3, 151n5, 152n7, 153n9, 158, 182, 193, 197, 220
Research and Development Board (RDB), Department of Defense, 24, 26
Reynolds, Malvina, 28, 33
Rindzevičiūtė, Eglė, 10, 226, 227
Rockefeller Foundation, 93
Rohde, Joy, 30, 131, 230, 231
Roosevelt, Franklin D., 43, 84
Roper, Elmo, 42
Roth, Alvin E., 130
Russel, Bertrand, 33
Russel, Richard B., 30

S

Sackman, Harold, 194, 195
San Jose State College, 95
Scheer, Julian, 159
Schelling, Thomas, 26, 82
Schilling, Warner, 211, 212
Schnitzer, Ewald, 107, 111
Seeger, Pete, 28, 33

Selznick, Philip, 27
Shapley, Lloyd S., 82, 87, 130, 232n1
Shils, Edward, 27
Shubik, Martin, 82, 113, 116, 118n1, 119n4, 119n7, 120n9, 123n25, 130, 131, 218, 219, 232n1
Simon, Herbert, 26
Skogstad, A. L., 42, 43, 45–47, 49–51, 56, 67, 70n1, 150
Snow, C. P., 151n5, 152n5
Snyder, Richard, 210
Social Science Research Council (SSRC), 20, 209, 210
Sonderfahndungsliste G. B., 55
Specht, Robert D., 78, 83–90, 151n1
Speier, Hans, 5, 24, 77, 92–95, 97, 98, 102, 103, 107, 108, 110, 111, 113–117, 120n12, 121n13, 121n16, 121–122n19, 123n24, 123n25, 129, 130, 151n1, 160, 209, 210, 213, 221
Stanford University, 29, 71n1, 94
State Department, 18, 34, 94–97, 107
Stimson, Henry, 1
Stouffer, Samuel A., 82
Strategic Surrender, 27–32, 95
Summit (game), 96, 232n1
Symington, Stuart, 30
Systems analysis, 80, 157, 158, 199n3, 226

T

Theory of Games and Economic Behavior, 78, 79
Thomas, William, 71n1, 80, 133
Toffler, Alvin, 205
Tolon, Kaya, 44, 54, 55, 200n10, 207, 224
Total War Research Institute (Japan), 83
Truman, Harry S., 65
Tucker, Robert C., 111
Turner, Stephen, 208, 228–230

U

Umpires, 91, 100, 212
The Unexpected Revolution, 95
University in Exile, 93
University of Berlin, 55
University of California, Berkeley, 95
University of Chicago, 55, 56, 94, 232n5
University of Göttingen, 79
University of Heidelberg, 93, 221
University of London, 55
University of Massachusetts, Amherst, 94
University of Southern California, 208
University of Texas, 207
University of Toronto, 94
University of Vienna, 79
University of Wisconsin, 213

V

Voice of America (VoA), 96
von Grimmelshausen, Hans Jakob Christoffel, 160
von Hilgers, Philipp, 78
von Neumann, John, 26, 78–82, 119n5, 119n6, 129

W

Wainstein, Eleanor Sullivan, 103
Wallerstein, Immanuel, 16, 27
War Department, 95
Ward, Frank L., 17, 210
Wauer, William, 72n7
Weaver, Warren, 71n1, 80
Wesleyan University, 206, 207
Whaley, Barton, 217
Whiteley, C. H., 145
Wiener, Anthony J., 205
Wiesner, Jerome, 96
Williams, John D., 24, 56, 71n1, 80, 81, 84, 94, 119n5, 207
Wilson, Woodrow, 15
Wohlstetter, Albert, 120n11, 130, 151n1, 232n5
Wolf, Charles Jr., 21, 111

Y

Yale University, 2, 209, 211
Youden, William J. "Jack", 72n4