7 Designing a Science for Decisions

The framework for risk assessment and risk management that William Ruckelshaus articulated in 1983 emphasized the separation of the two activities because these notions helped the administrator to construct a public image of the internal operations of the EPA and its legitimate mode of using science against uncertain events. It helped to show that the agency was both legitimately using the best possible science and assuming its responsibility to make protective decisions, even where and when the science, given remaining uncertainties, could not conclusively indicate what the exact risks were, and how to reduce them. In this chapter, I show that, in those years at least, proclaiming the separation of risk assessment and risk management to construct a general, acceptable image of the agency for the public was counterbalanced by a closer political management of risk assessment. By defining the conditions of the legitimate articulation of science with policy, the risk framework enabled greater administrative steering of the science. This policy influence was not felt in individual risk decisions, nor in the involvement of political appointees in the actual work of risk assessors. But it materialized via the involvement of higher-level managers of the agency in the manufacturing of guidelines for risk assessment.

As the risk framework integrated assessment with management, science became a tractable object for the managers; scientific research and scientific assessment of risk became an object of action for the managers of the agency and the architects of its decision-making process, particularly Alvin Alm, Milton Russell, and other policy analysts. It was this way precisely because science was becoming an element of the design that the agency wanted to project to its audiences. The reforms of the ORD and SAB, the institution of a risk assessment forum, and the revision of the 1976 guidelines for cancer

risk assessment—three major reforms aimed at limiting internal disputes in the agency, with enduring effects—clearly illustrate this.

Reorganizing the EPA's Science and Research

In their first month at the agency, Ruckelshaus and Alm decided to mandate the SAB to review the work of the laboratories and define a possible reorganization for the ORD. The group of five scientists included Bernie Goldstein, not yet confirmed by the Senate, but already at work as a consultant to the administrator. Ernest Cloyna, the chair of SAB, sent the conclusions of the expedited review to the administrator at the end of July 1983.[1] After having visited a number of laboratories across the country, the review group concluded that the assistant administrator should be a scientist with great stature in the scientific community, and that professional qualifications should be the primary criteria for recruitment of both the assistant administrator and the deputy assistant administrator. The group also noted excessive centralization in the ORD and a need for more involvement of research laboratories in the research planning process. The two recommendations seemed contradictory, but they were not, in the context in which the ORD was reorganized into five megalaboratories. Given the scope of their research and their internal planning processes (each had a research planning officer), these laboratories would ensure that greater decentralization did not mean fragmentation and irrelevance. The final report of the group (SAB 1983) pushed for more peer review of ORD's research and temporary recruitment of academics through the Intergovernmental Personnel Act. The number of staff at the EPA's ORD was high, in spite of massive cuts in the ORD budget and declining morale, but in some laboratories, the Reduction in Force policy pursued under Ann Gorsuch was threatening expertise, which remained thin. The report called for a fresh injection of new personnel into the laboratories.

In the report, the OHEA occupied a central place. Indeed, the OHEA and its components, the CAG and the Exposure Assessment Office, among others, had in recent years assumed "a key role in coupling science and regulatory decision making in the Agency," developing evaluations that required critical consideration of large amounts of scientific and technical data on health and environmental effects, while being attuned to "the particular needs of the regulatory office" (SAB 1983, 33). The report suggested

supporting this newly acquired role of the office, turning it into a center with the same status as the four other megalaboratories. Most of the OHEA teams would be relocated to the Research Triangle Park in North Carolina in order to ensure close interaction with the Air Office, specifically its Office of Air Quality Planning and Standards (OAQPS), its main client, as well as several leading universities in environment and health research and the National Institute for Environmental and Health Sciences. The report suggested that the CAG and the director of the OHEA maintain their location at the EPA headquarters in Washington, D.C. It thus capitalized on the experience of the CAG and of direct coordination between large laboratories and program offices rather than through the ORD chiefs. That situation was a legacy of the past: Program offices enjoyed direct relationships with labs, which in turn were directly attached to their former departments. The CAG was a new EPA-created lab that had developed a close relation with a program office, the OAQPS. It seemed to have evolved a better coupling with regulatory work than the previous research committees of the 1970s, operated by the ORD hierarchy.

The changes decided by Ruckelshaus and Alm went in the direction advocated by the SAB report. They identified research as a critical area. Ruckelshaus secured an increase of the EPA budget from Ronald Reagan's level of funding (up to $214 million in fiscal year 1986, the last appropriation that Ruckelshaus negotiated). They recruited Goldstein to revive the prestige and political importance of the office in the agency. They did confirm the OHEA's central role in the articulation of research and regulatory decision-making, as well as its presence at the EPA head office in Washington, D.C.

The risk assessment–risk management framework, as well as the internal assemblage between people, expertise, and offices that it helped institute inside the agency, reestablished an image of independence of EPA's science from arbitrary policies. In the aftermath of the then-well-publicized actions of former political appointees on the assessments of dioxin, benzene, or formaldehyde, Goldstein picked up on the language of risk assessment and risk management promoted by Ruckelshaus, to draw a line between the ORD and regulators: The ORD would cover risk assessment, while the program offices would take care of the risk management. This was definitely not intended to redesign entirely the work of the program offices. Many of these offices had scientists to make the kinds of health and environmental assessments needed for their decisions. But the dualism meant that

the ORD would certainly not get involved in risk management and that, conversely, program offices were not supposed to ask the ORD for data and studies that it already had in hand and that might help to support their preferred decision. Goldstein very publicly asserted this new generic characterization of EPA research and the ORD during his confirmation hearing both in Congress and in the media (Anonymous 1983).

The risk assessment–risk management design was also there to inform how far the ORD could concretely work to coordinate research for regulatory needs. Goldstein was particularly active on that front, suggesting to Alm that all research programs initiated by regulatory offices go through the same research committee as all ORD research programs.[2] This did not materialize. However, what did happen was that, from the mid-1980s onward, representatives of the ORD were included in the interoffice workgroups. For Goldstein, this was an essential mechanism to ensure the quality and relevance of ORD research. When properly structured in risk assessment/science policy/risk management terms, the meetings helped to identify, and agree on, the uncertainties during regulation development. Goldstein wrote back to all the ORD directors to explain how excited he was about this new ORD role, and that he would give it full priority.[3] Later, as Goldstein left the agency in 1985, he supported the creation of a new group, the Office of Technology Transfer and Regulatory Support, managed by Peter Preuss, formerly with the OHEA. The staff of this new office was "the focal point for the program offices' interaction with ORD" (EPA 1990a, 8). Its main function was to analyze and integrate scientific and technological information in the development of regulations. The group was later disbanded and replaced by other similar services. It was an important first attempt to establish a risk assessment–risk management routine to bridge the gap that existed between the ORD and regulatory offices.

The risk assessment–risk management combination similarly supported the role that the SAB acquired in the agency in the 1980s. The board aligned on this design in order to demonstrate its usefulness and become more mechanically involved with the agency's regulatory work. The staff director for the SAB,[4] Terry Yosie, actively designed a role for it, in close interaction with Ruckelshaus. Yosie had joined the EPA in July 1978, after obtaining a PhD at Carnegie Mellon University in Pittsburgh, where he studied the development of technologies for water pollution control. He joined the EPA staff for the SAB directly, moving up quickly to become staff director for

the SAB. In that position, he reported directly to Ruckelshaus and Alm, with direct and continuous access to them. These new strategic advisory functions were inaugurated by working on the recommendations of the SAB group that reviewed the organization of ORD labs. The chair of SAB and other senior members of the SAB, along with Yosie, actively pushed for the recommendations of the review group (promoting the role of the OHEA, setting both applied and fundamental research as two concurring objectives of the ORD, planning research programs better, etc.) and folding the ORD into the risk framework much more than the review group had,[5] along with the SAB as well.

Yosie argued for a more routine involvement of the board in the review of the scientific work of program offices. He outlined the plan in a memo to Ruckelshaus at the end of 1983, offering the administrator ways of "making effective use of the SAB," [6] demonstrating the benefits, for the administrator, in relying on this independent yet allied source of advice. What Yosie outlined was a system to define an agenda for the SAB that would be more independent from the variable, ad hoc needs of program offices, and more reflective of the agency's strategic knowledge needs. With greater planning of SAB review work at the highest level of the agency, the board could contribute to increasing coordination among offices, police the border between risk assessment and risk management, identify ways to upgrade research quality, and act as a special sounding board for new strategic ideas tested by the administrator.[7] In effect, he was positioning the SAB as a kind of review panel for risk assessment, as recommended by the authors of RAFG (SAB 1987), to counter the AIHC's proposal to create an external, independent, and perhaps policy obstructive science panel.[8]

In numbers, the increase in the reviewing activity of the SAB was unmistakable, with formal reviews going up from ten to seventy-seven between 1981 and 1987 (SAB 1987). In substance, it played an increasing role in the review of proposed risk assessment guidelines, as well as in the analysis, document after document, of the methodologies and interpretations made by various program offices. The reviews of the board were on several occasions instrumental to changing the orientations of the risk assessments, forcing offices to abandon their risk estimates, considering omitted scientific factors, testing nonlinear hypotheses where these appeared credible, and so on. The complicated dossiers of asbestos, ozone, arsenic, and dioxin all bear the traces of these interventions by SAB panels (Jasanoff 1995).

The Problem of Accuracy and Consistency: Toward the Risk Assessment Forum

Alm's managerial vision, predicated on RAFG, took a much firmer hold of the agency's internal scientific work. In a sense, risk management meant a management of risk assessment, and hence stronger intervention on the scientists in the agency. This did not necessarily involve day-to-day supervision of data choices and interpretations, but rather pressure to explicate analytical choices, models, and data sources, with their attendant uncertainties, so that managers could discover the nuts and bolts of decision-making and intervene if necessary. In the very first months of his new mandate, Alm kept talking about the necessary accuracy of the procedures of risk assessment. He wrote to all assistant administrators and regional administrators of the agency—whether or not they were concerned with risk assessment, like parts of the Air and Water offices at EPA headquarters in Washington, D.C.—that "The Administrator, as part of his overall goal of improving the scientific bases for Agency decisions, has taken a particular interest in improving risk assessment procedures." Therefore, everyone in the agency was required to use "the most accurate possible procedures for risk assessments."[9]

Embracing accuracy was daring and could certainly provoke and frustrate the proponents of risk assessment in the agency. First, risk assessors took it for granted that these calculations could not be accurate and precise. Ruckelshaus compared risk assessment to a spy: If you tortured him long enough, he told you what you wanted to know (Ruckelshaus 1984, 157–158). In the early 1980s, at an interagency meeting in the IRLG, a US Department of Health official had denigrated risk assessment as being as accurate as a five-year weather forecast (Barnes 1993). Risk assessors themselves knew this: Minute changes in the assumptions made, such as in the level of estimated exposure of a fictitious individual to a substance in an exposure model, could lead to enormous differences in the final estimate. Goldstein, who managed large platoons of risk assessors in the ORD, acknowledged this.[10] Risk assessment is accurate only insofar as one trusted the analytical choices and expert assumptions articulated and applied by risk assessors. Going for accuracy meant challenging this capacity and authority to make agency decisions based on their judgments.

But accuracy was also a challenging theme for Alm and policy analysts more generally. The usual approach in policy and cost-benefit analysis

(i.e., Alm's presumed perspective on decision-making, as well as that of the economists in the Policy Office) matched a culture of regularity and consistency, not accuracy. In the "mechanical objectivity" that is typical of these analytical exercises, something gains the value of being true if and when it is found to be commensurable with past and future calculations, not by correspondence to a supposed natural state (Porter 1992, 1995; Gill 2009). In approaching accuracy, Alm and the OPPE people were going beyond their usual epistemological territory to bring science into an organization for decision-making purposes.

The head of OPPE, Milton Russell, was in charge of advancing this agenda, together with Bernie Goldstein, the ORD's chief. Arriving at the agency in May 1983, Russell wrote to all assistant administrators in the agency that science was problematic in several respects. First, estimates developed in the science part were of "dubious accuracy" and sometimes "grossly exaggerated"; decision-makers and the public alike were generally unclear about the nature of the EPA estimates—how uncertain, how conservative were they? The public and Congress had trouble interpreting risk estimates and failed to understand that absolute safety was impossible. There was a great lack of accuracy in risk assessment, and the nature of the estimates (uncertainty, conservatism) used was unclear.[11] This intervention inaugurated the close, but respectful, involvement of the managers and economists of the agency in everything relating to the science underpinning regulatory decisions, and in choices in the face of uncertainty.[12]

The TITF was the channel through which projects of reforming risk assessment practices and rules were developed. Based on the discussions held in the risk assessment subgroup of the TITF, Alm announced three direct centralized actions on risk assessment: audit of risk assessment practices, creation of a risk assessment forum, revision or creation of guidelines for risk assessment. In his words, "These three actions demonstrate the Agency's commitment to scientific quality in regulatory decision making. I believe they will confirm that our scientific work has been done well in the past, and help prevent inconsistencies and deficiencies from occurring in the future. I thank you for your input to date and I look forward to working with you as we continue to improve our risk assessment process and products."[13]

In February 1984, the EPA contracted with Joseph Rodricks, formerly of the FDA, the chairman of the IRLG risk assessment group, and a member of the RAC, to perform an audit of risk assessment practices in the agency.

Rodricks had just launched a consulting business specializing in risk assessment,[14] and he performed the three-month, $15.000 audit in that capacity. His conclusion, unsurprisingly, was that the offices were not performing risk assessment evenly, at least judged against the background of the emerging orthodoxy of RAFG. Some were performing only the first steps of the full assessment, and some offices were strong on cancer while others ignored that aspect. The audit also made it clear that offices were treating genotoxicity in different ways, which might have had a major impact on how a carcinogen was assessed.

Cases of contradicting assessments among offices remained rare, fortunately. Most of the offices' work concerned routine assessment of regulatory issues, with no need to involve any other office and no outside scrutiny. The problem of inconsistency would appear for subjects of greater public attention, where the administrator would have to choose one assessment for the whole agency among competing ones, and where contradictions or hesitations would immediately be picked up by opponents in legal challenges. So the bigger problem that the audit identified was one of risk: The agency was not protected from subsequent conflicts among offices because different offices used different guidelines, or no guidelines at all, and because there was no common review mechanism for all offices.

A more long-term effect of the codification of the agency as risk assessment/risk management was the creation of a new cross-agency institution, authorized by Alm, called the Risk Assessment Forum (RAF). The forum was expected to fulfill four functions, some of which referred to the suggestion by the RAC to create a board to review risk assessment policies and guidelines published by agencies: "(1) review risk assessments upon the request of the Administrator, Deputy Administrator, Assistant Administrators, or Regional Administrators; (2) provide a mechanism for interchange on science issues in risk assessment; (3) advise the Administrator and Deputy Administrator on precedent setting cases and important risk assessment issues; and (4) recommend revisions or updates to the risk assessment guidelines, as appropriate."[15] It would be chaired by the main risk assessment group of the ORD, the OHEA, comprising representatives of all risk assessment groups from program offices, as well as representatives from the Office of General Counsel and the OPPE. It was topped by a Risk Management Council in 1986 by Ruckelshaus's successor, Lee Thomas, and populated by the institution's top "risk managers" to check on science-policy positions decided in the RAF.[16]

The RAF soon demonstrated its usefulness, in coordinating the process of constructing agency-wide methods of assessing risks, and of constructing shared decisions in the agency. The forum was the site of invention of a bureaucratic technique that was long used to demonstrate the agency's way of establishing risks to the outside, the uncertainty factor, but only because it was a place in which controversies surrounding uncertainty and its perception could be safely translated.

As mentioned in chapter 3, there was no formal guideline for determining safety levels for noncarcinogenic chemicals until that time. The area was simply too broad to be amenable to standardization: too many chemicals, causing too many different adverse effects. The result was that the methodologies adopted by the various offices of the EPA were even more diverse than in the closely framed exercise of carcinogen risk assessment. Some offices determined their rule for a chemical based on the evaluation of its most critical health effect (e.g., an effect on the kidneys), defined as the observable effect occurring at the lowest dose. Other offices picked one effect that they deemed more important or worthy of attention than another, regardless of the levels measured in experiments. Some offices measured these effects over the lifetime of a person to determine the daily dose at which one could safely be exposed, and others at less than the lifetime, considering that lifetime measurement produced overly conservative results. There were many more variations as well.[17] The result was that on a single chemical, there could be as many as thirty or forty risk estimations and so-called acceptable daily intake (ADI),[18] which was the source of an organizational conflict, as well as a major risk for the credibility of the agency.

That conflict started in November 1984, when the Office of Toxic Substances complained that scientists in the ORD duplicated evaluations that it had already performed. The problem surfaced on several occasions that year, leading to a rather hostile exchange of memos between the two offices, with the administrator and deputy administrator copied in. The OPPE stood in between, appalled by the new problem of consistency across the agency's science. Donn Viviani, head of the Regulations Analysis branch, wrote to all other chiefs of the OPPE: "Intermedia transfer, conflicting and/or duplicative assessments, different regulatory control levels, inconsistent health numbers, lack of coordination, sound familiar? They should, every one of these "integration" issues has been previously identified and studied, recommendations presented and, in at least several cases, some 'fix' implemented."

For him, the problems were emerging again because program offices had no incentive and interest to share information and work, and were getting no credit for integration. The memo spoke of program offices as "fiefdoms." [19]

To stop the conflict, a small group of four people from all concerned offices was set up, becoming the so-called ADI policy group in the RAF. Agency toxicologists worked through all existing ADIs to analyze the source of the many differences and to think about ways of resolving them. The authors of RAFG had already seen the problem: There were a lot of choices made in the course of calculating an ADI, and any of these choices could be manipulated for regulatory offices to attain the kind of result they wanted—not necessarily in terms of setting a level of risk, but in terms of justifying the amendment of adopted rules, or justifying not doing so (NRC 1983). If a decision-maker thought that it was better not to change an ADI, even though a new study indicated a new lowest adverse effect level (LOAEL; adverse effects of the substance appeared at a lower dose, so presumably more frequently), she or he would compute the difference between the LOAEL and the dose found in humans, called the *margin of safety,* or *safety factor*. The argument would be that, somehow, it was sufficient, even though the computed margin differed from the one that had to be applied conventionally to the LOAEL to determine the science-based ADI.

RAFG, in fact, provided the key to analyze this problem. It was not that regulators always directed the science, or that science was unable to respond to regulators' needs. It was that, in critical cases, any member of the relevant regulatory office (scientist, lawyer, or decision-maker) could play with any of the parameters of a risk assessment—the intermediary world of "science policies." That generic notion helped to depict a constituent problem for the whole agency and legitimized a standardizing effort, led by this newly formed, cross-agency ADI policy group. The group moved to delete the language of safety factors and of ADI. Safety and acceptability were judgments, so the terms empowered regulators to formulate these judgments without consideration for the conventional methods defined by scientists. It standardized a set of "uncertainty factors" instead, to apply in order to create a scientific "reference dose." A variety of uncertainty factors could be chosen, depending on the level of available data.

The concept of *uncertainty factor* had an enormous advantage over that of *safety factor*: it was an explicitly judgmental and qualitative notion, not a

realistic one like that of *safety factor* (which tried to account for true biological differences between species). It was also more contextual because levels of uncertainty could vary from one risk assessment to another. It was a way to get "a better approximation of how a particular chemical should be considered,"[20] unencumbered by the shadow of the truth of what the chemical really, precisely does in the body of an animal or human. The group also decided on the creation of an electronic registry of reference doses for hundreds of chemicals that various program offices should take as a basis for their decisions, the Integrated Risk Information System (IRIS).[21] The group did not reduce the range of possible choices in scientific assessment of risks, but it explicated and standardized this range, helping to establish a more regulative design to routinely agree on chemical values, for IRIS.

So, alongside the guidelines—which were an essential instrument for the agency to be auditable by the courts, but did not strictly standardize the practices across all program and regional offices[22]—the RAF, as an institution to organize a cross-agency or external review of the assessments performed in the agency, worked as an institution that generated choices for decision-makers. It did so in a way that was accountable, or at least readable from the outside, thanks to the commonly accepted risk assessment/science policy/risk management grid. It was soon hailed by many as a very astute and important new institution that helped to avert many controversies. Dorothy Patton, who administered the RAF for several years starting in 1985, found that offices knew they had an interest in referring to the RAF issues on which they lacked expertise and on which they felt other offices would have competence or could offer opportunities for publicly challenging the agency: "Basically, the offices that had the responsibility to issue a standard, and were under pressure and lobbying to do so, on arsenic for instance, were looking for stronger arguments, reassurances, better science, by coming to the forum."[23] The overall opinion on this committee was that it did become a sort of neutral space or no man's land between offices, and that many issues were sorted out there, thus averting and avoiding conflicts among offices, disciplinary cultures, and science policies. At the same time, the ability to converse among offices was at the cost of a kind of cultural homogenization, with the main platform for discussions there being toxicology and the hazard identification–hazard characterization part of the risk assessment orthodoxy, at least in the first few years of operations.

Managing the Revision of the 1976 Guidelines

The utility for the RAF to develop risk knowledge for the whole of the agency, and not for one office, was also tested through the process of revising the 1976 guidelines—and probably even more radically than for the establishment of noncancer guidelines, given the political sensitiveness of the cancer issue. These guidelines were interim, and had mostly been applied by the CAG to develop health assessments or criteria documents on behalf of regulatory program offices for water and air. In the meantime, the CAG had expanded inside the OHEA and was replicated in several other groups (exposure assessment group, mutagenicity assessment group, etc.), that had also drafted guidance. But many of these documents had simply not been finalized and officially applied in the agency. The guideline-making process was reignited, with a view to definitely put them to use across the agency to ensure consistency.

The difference with the earlier efforts to establish guidelines was that the OPPE was directly involved in developing them, which meant a closer involvement by the heads of the agency, and Alm specifically. Russell pointed the way forward:[24] More guidelines should be developed to implement a different approach of quantitative risk assessment, particularly of exposure assessment. Guidelines would be of enormous benefit, to improve coordination across the agency (without having to resort to a vain centralization of risk assessment activities in one office, such as the ORD), and to limit discretion at lower levels and clarify the agency's "science policy." This new bureaucratic theme of accuracy and explicitness in risk assessment, combined with economists' favorite idea of what constituted "efficiency" and "consistency," corresponded to the CAG's ambition to expand the use of quantitative risk assessment across the agency. Betty Anderson had always been clear that the 1976 interim guidelines, and other pieces of guidance developed thereafter, would need to be regularly updated as science progressed. In the meantime, RAFG had convincingly made the case for developing more guidelines, in particular to explicate experts' uncertainties and judgments.

Starting in July 1983, Elizabeth Anderson took the lead for that project, and leaders for the development of every new guideline were nominated. Each of them would manage a group with representatives of the various offices. The project was designed to be a short, but intense six-month

effort, helped by the accumulation of experience in the OHEA. The process took longer than expected. The twelve full-time equivalents and $600,000 of extra funds requested by the OHEA to perform the work were never obtained. Because the OHEA was unable to divert resources from the routine and time-consuming tasks of developing health assessment documents or exposure and risk assessment documents for the Air and Water offices, the work was perceptibly delayed. The participation of other offices in the process was just sufficient, though patchy. Key offices were absent from some of the meetings, or they were slow to respond to the request for comments on drafts. The OPTS, an important player in the agency concerning risk assessment, and one of the main targets of the planned extension of the application of guidelines, did not attend the meetings or submit written documents. Only after the draft was completed did the office raise concerns in a lengthy document. On several occasions, the leaders of the workgroup asked the assistant administrators and Al Alm to weigh in and stress the importance of cooperation and hard work. While the redevelopment of guidelines was in no way conflictual, the inclusion of a larger set of people from across the agency amplified the "differences of opinion on several important issues."[25] The standard methods and criteria for carcinogen risk assessment stabilized by the CAG were discussed on four levels: terminology, presentation of exposure calculations, weight of evidence, and, once more, extrapolation to low doses.

Terminology: Deploying a Standard Knowledge Representation

One of the first issues to come up was terminology. An official from the OPTS maintained that in the first meetings, "it was apparent people were using terms in different ways"—something that the confusing definitions of exposure, health, or risk assessment in existing documents could only confirm. He soon developed a "Regulatory Decision-Making Nosology" memo to "help demystify matters."[26] It straightforwardly transcribed the definitions laid out in RAFG, showing how the existing scheme used in and around the CAG (two qualitative questions, followed by a quantitative risk assessment) would profitably be replaced by a reference to RAFG's generic scheme for risk assessment, comprising four components: "hazard identification, dose-response assessment, exposure assessment, and risk characterization. The first of these, hazard identification, is a qualitative

risk assessment. The three remaining components comprise quantitative risk assessment."[27] His assistant administrator endorsed the adoption of this vocabulary in the drafting of the guidelines.

One such problem emerged in the distinction between the concept of dose-response assessment and exposure assessment. The ambiguity was that there is no agreed-upon definition of the point on or in the body where exposure takes place. In this context, *exposure* means contact with the chemical. But does it mean contact of the visible external envelope of the body as a whole, or exposure of a particular organ, or of the cell? Toxicologists extended their jurisdiction as far as they could, to the whole body, claiming that they could analyze, quantitatively, the effects of a chemical on the physiology and biology of the organism taken as a whole. This scale defined their territory of expertise. The exercise of exposure assessment concerned what happens outside this boundary. It works with the applied, or external doses of the chemical, while the internal doses was the purview of toxicologists. The separation between applied and internal doses is the line that defines the respective domains of exposure people and dose-response people in the agency. It could be seen as a bureaucratic schematization of the body, which allowed the articulating the various expertise to be applied in a standard decision-making process. Exposure calculations came after dose-response assessment.

It was thus, at that very time, in 1983, that the agency was reorganized by a set of categories that henceforth defined the discrete competences and roles of its scientists, fitting them into a broader organizational scheme to turn out decisions. Just like RAFG had subsumed the various scientific analysis of the hazard under a generic notion of "risk assessment," so too the memo described previously and the guidelines put the emphasis on dose-response assessment—and, implicitly, quantitative, probabilistic risk assessment—to embody the agency's expertise. While it might seem natural to move in that direction, given the precision and order in the integrated scheme, it was by no means an innocent move. As the audit of risk assessment practices would later reveal, along with a survey of risk assessment in regional offices, the quantitative aspect of risk assessment was not the easiest part of the job, and certainly not the most practiced one. Most offices were doing what was now called *hazard identification,* demoted to an initial step in a broader scheme, whereby decisions were to derive from a contained quantification of the risk. This change was facilitated by the authority that

this knowledge representation had in addressing controversies. Promoting this formalism was not just a prescription for embracing quantitative risk assessment among reluctant staff; it was also a shift toward another method of treating uncertainty—one that was more bureaucratic, in as much as it was applicable without deep scientific knowledge, and thus applicable across the entire agency. RAFG's anatomy of risk assessment, replicated in the agency in the "nosology" memo mentioned above,[28] carved out a number of recognizable, commodified knowledge items that officials across the agency, however closely involved in the particulars of a risk assessment, or however knowledgeable in the science, could extract from a risk assessment document.

One of the striking aspects of the risk assessment guidelines eventually published in 1986, was that they included detailed, suggested formats for the presentation of scientific information, accumulating each of these elements of information. Uncertainties were an explicit component of these knowledge schemes. Hence, the big change, from 1976 to 1986, was that risk assessment documents would include explication of choices made in the scientific analysis (e.g., on exposure assessment, as discussed later in this chapter), rather than justifying preferences for one or the other option in terms of best professional judgment. Itemizing knowledge in this way was an effective method for a bureaucracy because, if well delineated, these items would work to define categories of people and groups, organizing them in a broader organizational mechanism for making decisions. And indeed, the guideline revision efforts implied going one step further from RAFG, by specifying the boundaries between these elements of knowledge and the associations among them in order to ensure that no more confusing organizational knowledge politics were created in the agency.

Manufacturing Risk Presentations

The ways in which exposure assessment was generally handled in the agency's health and environmental assessment documents constituted the second problem that the guideline working group had to address. First, the behavior of a substance in the environment was the source of multiple variations and uncertainties: the nature of a chemical substance changes as it is transported, degraded, and so on. To account for these dynamics, exposure assessors are forced to rely on models. For instance, as arsenic was emitted

into the air, the EPA used dispersion models to be able to calculate, with an unavoidably high degree of uncertainty or variations, the dose at which populations in the environment of the point-source would be exposed. Second, exposure assessment would seek information about the lifestyles, location, and consumption practices of various subpopulations at a level of detail that no existing database could match. To compensate for the patchy monitoring data, assessors modeled qualitatively the situation of three fictitious individuals: the exposure of a "typical person" (undefined), that of a person living close to the source of the hazard, and the "maximally exposed individual"—a person exposed to the hazard 24/7, for an entire lifespan.

In practice, risk assessors tended to be selective in their presentations of the estimate, reporting only the one that was based on the notion of the maximally exposed individual. This reduced the ability of anyone to subsequently participate in the choice of the most useful calculation for addressing the regulatory question. The selective presentation of exposures invisibilized an operational choice, and indeed even a moral dilemma that the EPA as a whole was to face in its decisions: either taking measures to protect the small number of individuals who suffered most (e.g., by removing a plant emitting extremely high doses of a dangerous chemical at a given site) or reducing average emissions over the country to reduce average levels.[29] Ruckelshaus was personally concerned about the equity behind these choices, and that translated into Alm and Russel's request, in the process of revising guidelines, to eliminate the practice of exclusively advancing data on the maximally exposed individual to present real-life monitoring data or presentations of ranges of exposures and uncertainties instead.[30] The suggestion to risk assessors not to use only worst-case exposure, but also "realistic exposure scenarios," had in fact been the object of a dedicated memo of Alm to all assistant administrators.[31]

The revised text did not address the issue at length and made no clear statement about the relative value of data and models. It emphasized the uncertainties and sophistication needed to treat the problem of exposure assessment, and it referred readers to a separate guideline on exposure assessment. The resulting guidelines (EPA 1986b) handled the problem by shifting from a black-boxed expert judgment to a different, algorithmic model. The guidelines stipulate to work with monitoring data, and then construct real-life monitoring profiles, shedding light on more vulnerable people; where insufficient, to model exposure in a probabilistic way; to

perform a sensitivity analysis of the model; and to present all uncertainties involved in the modeling as part of the results of the exposure assessment.

Having articulated that sequential reasoning, the authors of the guideline expounded it through a flowchart—a formal visual technology of uncertainty reduction that minimized the reliance on trust in the judgment, competence, and experience of remote risk assessors in another office (see figure 7.1).

Governing Judgment

The methodological concept of WOE was the third aspect in line, and it was discussed far more extensively. There too, Alm and other policy analysis staff in the agency showed an appetite for these arcane concepts that were supposed to structure the judgment of risk assessors. The interest and availability of high-level administrators for the science used in their agency varied widely, but Ruckelshaus and Alm were deeply interested and involved. Alm, in particular, was a voracious reader, including of scientific documents.[32] According to another senior scientist of NCEA, he was respectful, asking many questions and trusting the competence of his staff instead of imposing his directives. In those days, he requested and received dozens of memos clarifying the science of risk assessment and the dilemmas involved from his most senior and experienced scientific staff. Betty Anderson had produced one such memo outlining what a WOE was. The topic had become a bone of contention between the CAG and the OPPE that was reflected in an important amendment of the 1976 guidelines that derived directly from the controversial case of perchloroethylene (PCE), also known as *tetrachloroethylene* or *perc*.

PCE was a popular solvent used in the dry-cleaning industry, the regulation of which was under consideration by the Water Office, the Air Office, and the Office of Toxic Substances simultaneously. OHEA was working on a risk assessment of the substance on behalf of these offices. In April 1978, its CAG had published a preliminary health risk assessment report for the substance. In October 1980, another group of scientists, also from OHEA, worked with the Water Office to set an ambient water quality criterion. In 1983, the process for regulating the substance accelerated, and the CAG revised its health assessments. Back in 1976, the CAG had used no other way of classifying the level of hazardousness than a simple characterization

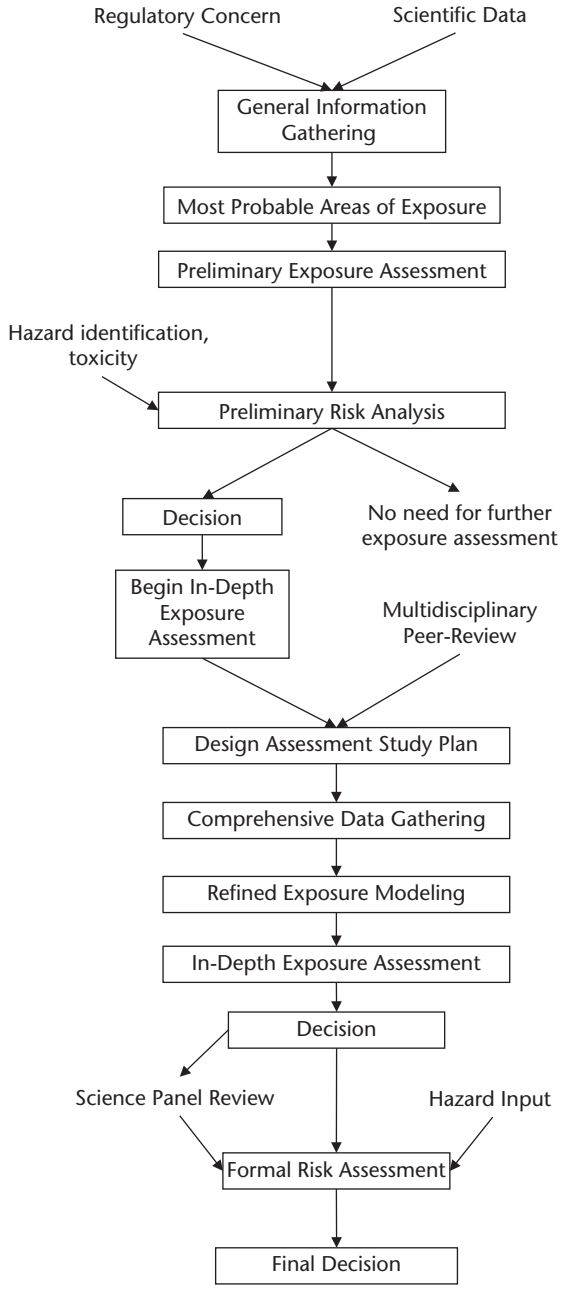


Figure 7.1
Decision path for exposure assessment (adapted from EPA 1986b).

as either substantial or suggestive. By 1982, it had evolved to consider that the scheme used by the International Agency for Research on Cancer (IARC) was the only one with substantial institutional use and international acceptance (EPA 1982). The scheme devised by CAG on the basis of the IARC scheme, a class-based aid to judgment, was as follows:

- A substance for which there is "sufficient evidence" to support a causal association between the exposure and cancer will fall into group 1: carcinogenic.
- Where evidence is "almost sufficient" or "suggestive,"[33] the chemical will be put in group 2: probably carcinogenic to humans. Group 2 is divided into higher or lower degrees of evidence: 2A and 2B, respectively.
- Where evidence is "inadequate" or "limited," the substance should be considered "possibly carcinogenic to humans": group 3. [34]

CAG's conservatism was reflected, specifically, in the way that it defined and used group 3. To IARC, a group 3 chemical is a substance that "cannot be classified as to its carcinogenicity to humans" because of inadequate or limited evidence. The CAG, in contrast, argued that there was a continuum between "limited" and "almost sufficient" evidence, so it argued that a labeling as "possible carcinogen" was more appropriate than as "cannot be classified." CAG's revised group 3, in essence, seemed to recognize uncertainties and the possibility of false positives a little less straightforwardly than IARC's group 3. And it certainly contributed to make the entire scheme more regulation-forcing: If a chemical is a possible carcinogen, it could possibly be regulated.[35]

In spite of the limited evidence, CAG placed PCE in group 3. This was not acceptable for the Water Office, which was in charge of actually deciding on regulatory measures for this chemical and dealing with the concerned industries. Paul Milvy wrote to Al Jennings, the director of the Office of Standards and Regulations in the OPPE, to complain that the CAG had produced three different estimates of the risk of the substance between November 1983 and March 1984, finally concluding, on the basis of WOE, that the most conservative scientific view was to consider the substance a probable human carcinogen. There was too much variation there to be able to establish a credible risk standard. More important perhaps, the Air Office disagreed with CAG's traditional conservatism, complaining that CAG distorted the classification scheme of IARC to fit with its conservative policy.

CAG placed PCE in group 3 of IARC "because the available evidence corresponds to the conservative scientific view that PCE is probably carcinogenic in humans." For Milvy, this original determination by CAG was "totally wrong." For the Water Office, CAG was distorting a tool for objectifying judgment, and thus forced the regulation of PCE. The fact that CAG modified its initial assessment only days later made things worse. This lack of continuity, in the context of conflictual relations with offices applying different policies, could only cause controversy. As Milvy noted, "because a simple, honest and straightforward approach was not adopted [by CAG], the resulting blatant contradiction fueled the controversy."[36]

This was precisely where PCE was a close call, as Russell noted in his memorandum to all offices contemplating a regulation of this substance. Russell asked to have a meeting on this as soon as possible.[37] The chemical was already being reported on in the press, and the EPA was unable to determine what action to take with regard to it. The Air Office interpreted the "limited evidence" classification as indicating that the substance could not be considered a carcinogen, so it should not be regulated. On the other hand, the Water Office, the hazardous waste office, and the Superfund office considered that these chemicals for which there was "limited evidence," and so were "potential" carcinogens, should still be regulated.

The March 1983 memo from the Air Office firmly challenged the IARC scheme. Was it not forcing these kinds of mistakes by artificially placing chemical substances in one group, "when in fact nature has provided a continuum of potency"? At the same time, would a new consistency-promoting classification inside the agency not reduce the ability of individual offices to engage with audiences, whether scientific, industrial, or other, thereby finally increasing the potential for controversies? The memo hit the bull's-eye: This was a perfect example of the high potential of a scientific assessment to cause major confusion outside the agency and to stimulate criticism of its decisions, if not legal challenges, at the very time when the OPPE was pushing an all-out policy of increasing agencywide consistency. Representatives of the offices concerned (Air, Water, ORD, and OPPE) met a few days later for what turned out to be a difficult, heated meeting. They agreed that consistency was a necessity, but they continued to disagree on whether the IARC scheme was the best classification and whether picking a classification was a scientific or a managerial task. Reporting the conclusions of this meeting to Alm and Ruckelshaus, Russell

questioned the IARC classification: the agency never formally adopted the scheme, but the OHEA was using it as a matter of habit. The "policy implications" and "regulation-forcing dimension" of the scheme needed to be thought through, otherwise "programmatic conflicts will result and Agency policy will be shaped by default. Perhaps worse, the Agency will look like it doesn't know what it is doing and in the long run, lose credibility."[38]

During this episode, the OPPE and its particular synoptic view of the EPA, as well as pressure to streamline the science to make it conducive to shared decisions, seemed to dominate the ORD. Russell asked Elizabeth Anderson, the chief of CAG, for a clarification of the reasons of its use of the IARC scheme. In her response, Anderson aligned on the risk assessment–risk management framework, showing how the WOE method underpinning the classification of carcinogens fit into the risk assessment–risk management framework. She clarified that the WOE approach did not produce a decision by itself, nor was it regulation-forcing; in her words, it could not provide cutoff criteria to decide whether to initiate decision-making. This was a matter of risk management, and risk management needed to rely on other considerations on top of whether the chemical had the property of being carcinogenic (such as magnitude of exposure and size of the health risk). Along with these other pieces of information, "the weight of evidence that the chemical is a carcinogen supports or tempers the decision to initiate risk management or not." In her view, one could not black-box a risk assessment based on WOE in order to decide whether and how to manage the risk. No standard, general risk management strategy could be devised.[39]

Anderson essentially gave in to OPPE's attempt to redefine the WOE bureaucratic technology. Eventually, Russell suggested using a set of criteria other than the IARC because, in practice, they resulted in only two possible regulatory options: to regulate or not to regulate. More flexibility was needed, as were more precise criteria and more sophisticated weighing of the evidence. Alm concurred, as did the lawyers of the Office of the General Counsel of the agency, which put pressure on the OHEA to incorporate into the guidelines some elements of a WOE approach that would replace the IARC criteria. Eventually, the method was refined to be more sensitive to a variety of regulatory options. Five groups were defined:[40]

- A—Known human carcinogen (sufficient evidence from epidemiological studies or other human studies)

- B—Probable human carcinogen (sufficient evidence in animals and limited or inadequate evidence in humans)
- C—Possible human carcinogen (limited evidence of carcinogenicity in animals in the absence of human data)
- D—Not classifiable (inadequate or no animal evidence of carcinogenicity)
- E—Evidence of noncarcinogenicity for humans (no evidence of carcinogenicity in at least two adequate animal tests in different species or in adequate epidemiological and animal studies)

Taking Control of Extrapolation

The OPPE weighed in, finally, to discuss dose-response assessment and the CAG precautionary linear model of extrapolation—the Linearized Multistage (LMS) model—and the upper confidence limit. Alm requested a memo from Goldstein to explain, didactically, the scientific issues underpinning carcinogen risk assessment. Goldstein, who had Roy Albert, the former chair of CAG (see chapter 2), as a professor when he was an undergraduate, summarized the key debates in the scientific community (as concerned threshold, the variety of cancers, genotoxicity, and other topics) to conclude and give support to the use of more conservative ways of assessing these risks: "The traditional public health approach has been that in the absence of a reasonable degree of certainty, it is proper to accept the more conservative assumption of a linear no threshold relationship."[41] In most of its quantitative cancer risk assessments, the CAG was supplementing the use of the multistage model, borrowed from biostatistician Kenny Crump, with another convention of extrapolation: the so-called upper bound limit of 95%, a statistical confidence limit used to "force the multistage model to provide a linear term." The alternative method consisted in calculating a maximum likelihood estimate. The two methods produced variable results, the upper-bound confidence limit being the one that produced the most precautionary results. Both were consistent with the CAG's conservatism because the upper-bound limit was by definition higher than the most likely estimate, and that maximum likelihood estimate would be higher than an average likelihood estimate.

The OPPE questioned both, or at least aimed to qualify their use, notably because these implicit policies were sometimes considered inadequate

for particular chemicals. On various occasions in the preceding years, the reviews of CAG health assessments by the agency's SAB led to acrimonious debates about the quality of the assessment and the opportunity of applying the usual linear no-threshold model to substances that could very well be interpreted as displaying a threshold. For Terry Yosie, SAB's staff director in these years, "the board was very strong on its recommendations that the EPA continue to diversify its expressions of risk probability" and stop relying on one model only, increasingly allowing nonlinear responses.[42] In 1980, PCE posed such a problem, as did formaldehyde. In December 1984, in a steering committee meeting on the substance, the OPPE representative underlined that there was a 500-fold difference in the estimated risk of cancer from exposure to formaldehyde, depending on whether one used the upper 95% confidence limit or the maximum likelihood point estimate of risk at the relevant doses. He required to present both risk estimates and to state explicitly which one would be followed, and for what policy reason. Arsenic was perhaps the most contentious of all, as the Water Office had decided to prove that this carcinogenic substance was a threshold substance, and thus to set a precedent of departing from the CAG's default linear model (Powell 1999).

Given that the OHEA applied the technique most of the time, it turned out to be an implicit policy that needed to be clarified in a new framework of thinking that distinguished more strictly between risk assessment, science policies (assumptions applied in the risk assessment), and risk management. As in the case of the WOE approach and IARC classification scheme, the CAG resorted to arguments that had a wide scientific consensus: the plausible upper bound "is generally recognized in the scientific community as a reasonable approach."[43] This kind of argument, now that a framework was in place that explicated the knowledge base used in the agency and established rules for coordinating judgment, was no longer valid. Anderson could not simply mention scientists' convention now that there was a public expectation, under the framework, to see how one consideration potentially leads to policy, decision-making, and which decisions are made.

These documents sketched out important changes in the practices of risk assessment in the agency, as well as an emerging new policy: Against the reliance on the LMS model only, the OPPE defended the idea that a variety of models may be used to estimate the risk of cancer at low doses and that, where possible, the most likely estimates, as well as upper- and

lower-confidence limits, should be provided. Only a lack of data can justify the exclusive use of the upper-limit estimate of risk. Here, in essence, Alm and the OPPE were getting closely involved in the fashioning of what RAFG had called a "risk assessment policy," reformulated in the agency as *science policy*. For the first time in the young history of the EPA, a functional policy office, staffed by economists and policy analysts implementing the administrators' policy orientations, was getting involved in the details of science, redesigning it into an instrument of decision-making in the name of the controversy that it risked creating otherwise.

After extensive discussions, the CAG agreed to use the different methodologies side by side in a given assessment, and to compare their results. The 1986 guideline was also slightly less assertive than the 1976 one so far as the linear no-threshold model is concerned: "No single mathematical procedure is recognized as the most appropriate for low dose extrapolation in carcinogenesis" (EPA 1986a, 12), and could be applied to assess the suitability of models case by case, including a rationale justifying the use of one in particular: "in the absence of adequate information to the contrary, the linearized multistage procedure will be employed" (ibid., 13). No standard exposure assessment method was preferred, but the guideline indicated that a "cumulative dose received over a lifetime, expressed as average daily exposure prorated over a lifetime, is recommended as an appropriate measure of exposure to a carcinogen" (ibid., 13).

The notable change from the 1976 guideline, overall, was that these guidelines were structured according to the "more explicit terminology" (ibid., 2) of RAFG, rather than going by the two-question structure of 1976 or the qualitative versus quantitative assessment distinction made since. The 1986 guidelines structured informational bricks, components of a mechanically produced decision. They did so for unknowns, or nonknowledge, giving them a place in the scheme as *science policies*, a term that was absent from previous documents. In November 1984, Ruckelshaus wrote to program managers to indicate that the guidelines were soon to be published in the Federal Register, to "increase consistency in use of risk assessment and underline importance of scientific analysis itself as a factor in regulatory decision-making"[44]—a formula that was drawn from the 1985 guideline for risk assessment of the OSTP, showing how CAG lost autonomy not only to the OPPE, but also to external design actors trying to bear on the agency's way of making rules.[45] The CAG's early risk assessment guidelines

had finally been translated into an agencywide position stressing the need for quality, accuracy, and consistency[46] and acknowledging that the judgments of experts constituted science policies. They had come to embody the new coherence between the agency's external, official mission and the value of reducing risks to health; its internal functioning and overall knowledge of risks, based on experimental toxicology and medicine; and a growing interest in epidemiology and exposure data, as well as recognition of the existence of uncertainties and of the possibility of reducing them through expert judgments and estimates.

Conclusion

This chapter has listed the initiatives concerning the reorganization and management of risk assessment in the mid-1980s, involving the leadership of the agency. During that period of time, the leaders of the agency have gotten involved in the formalization of what one may term *bureaucratic technologies* (the elaboration of guidelines, processes of cross-agency consultation, and distribution of responsibilities) so that the agency to be able to more surely turn out unified—and hence defensible and ultimately credible—computations of risk.

To be sure, these technologies were already in the making, and their importance was already sensed in some part of the EPA. Guidelines for the assessment of hazards and risks had existed for nearly ten years. The need for coordination between ORD scientists and regulatory offices was also recognized long ago. But this time around, the risk assessment–risk management assemblage provided a common point of reference to reorganize the work of scientists from the top down. More than a mere reorganization, the policies implemented here aimed to characterize the work of scientists in a completely different way, creating a clear classification of tasks and elements of knowledge needed to articulate a risk assessment that could be accepted across the agency and by top decision-makers. The risk design deposited a new set of identities in the organization in such a way that the various kinds of scientists became the components of an organizationwide system to guide the production of scientific information and interpretation of risks and uncertainties in order to make shared decisions. The science of risk became an object of this reordering of the organization. It became part of an agencywide architecture designed to limit internal dissensus and to

increase the EPA's capacity to present uncontestable statements of risk to external audiences.

The particularity of that period is that the top managers of the agency, supported by the economists and decision scientists of the agency, orchestrated the work. They led the designation of the generic knowledge and practices defining the agency as a whole, giving it its identity as a knowledge-based policymaking organization. They could do so legitimately only because the political circumstances—the loss of credibility that the agency had suffered due to the improper use of science during the years of Gorsuch and inconsistencies among the multiple assessments of the same chemical—gave these people the necessary legitimacy to intervene deep into and across the organization to redefine the work of each entity involved.

The Science of Bureaucracy

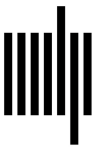
Risk Decision-Making and the US Environmental Protection Agency

By: David Demortain

Citation:

The Science of Bureaucracy: Risk Decision-Making and the US Environmental Protection Agency
By: David Demortain
DOI: 10.7551/mitpress/12248.001.0001
ISBN (electronic): 9780262356671
Publisher: The MIT Press
Published: 2020

The open access edition of this book was made possible by generous funding and support from MIT Libraries



The MIT Press

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Demortain, David, author.
Title: The science of bureaucracy : risk decision-making and the US Environmental Protection Agency / David Demortain.
Description: Cambridge, MA : The MIT Press, [2019] | Series: Inside technology series | Includes bibliographical references and index.
Identifiers: LCCN 2019010651 | ISBN 9780262537940 (pbk. : alk. paper)
Subjects: LCSH: United States. Environmental Protection Agency--Management--History. | Risk management--Government policy--United States--History. | Environmental policy--United States--Decision making. | Environmental policy--United States--History.
Classification: LCC TD171 .D46 2019 | DDC 363.700973--dc23
LC record available at https://lccn.loc.gov/2019010651