

## Doing Believable Knowledge Assessment for Policymaking: How Six Prominent Organizations Go About It

Lawrence McCray<sup>1</sup>

### A. INTRODUCTION

We would all like to ensure that public policies are fully consistent with the best available pertinent knowledge. Many observers, however, feel that our regulatory policy processes are weak in this regard. This paper derives lessons on the fusion of policy and scientific and technical knowledge from the real-world experience of six organizations that have specialized in providing *knowledge assessments* to decision-makers.

I employ the expression “knowledge assessment” for lack of a commonly used, existing term. A knowledge assessment<sup>2</sup> is a review of scientific and other accepted knowledge that is germane to the pros and/or the cons associated with public policy alternatives that are under consideration. Our hope is that by examining these six organizations we can draw some practical lessons about how broadly credible knowledge assessments are done. How do assessors avoid being influenced by powerful interests? What pitfalls are they concerned with? Does their credibility come at a reasonable cost to the policymaker, measured either in dollars or in time?

The six subject organizations, listed in rough order of the depth of their historical experience with knowledge assessment, are: the National Research Council [**NRC**—often misidentified as the National Academy of Sciences], the Office of Technology Assessment [**OTA**—1972-1995], the Consensus Development Program of the National Institutes of Health [**CDP**], the Clean Air Scientific Advisory Committee of EPA [**CASAC**], the Health Effects Institute [**HEI**], and the newest of the group, the H. John Heinz Center for Science, Economics, and the Environment [**Heinz**].

<sup>1</sup> Please send questions and comments about this paper to [lmccray@mit.edu](mailto:lmccray@mit.edu).

<sup>2</sup> A full knowledge assessment is taken to have three essential characteristics. First, it undertakes to consider, in part or in full, the available *scientific and other factual* [e.g., economic, enforcement-related] *basis* for policy choice. Second, it reflects the *collective judgment* of a set of experts in the pertinent domains of knowledge, and not just the views of a single individual. Because most policy decisions impinge on several fields of expert knowledge [physics, human biology, economics, etc.] there is a need to synthesize them, and knowledge assessment that is collective can do this. Third, the assessment admits both what the experts take to be the objective evidence and their subjective, but informed, views—courts have called this the “*reasoned feel of the expert*.” Because of the presence of the second and third factors, there

would be hazardous in typifying knowledge assessment itself as a strictly "objective" or "scientific" process, even when it is controlled entirely by scientists of broad renown.

The present review of the six organizations is restricted largely to their role as knowledge assessors, which in some cases is only one of their several functions. (For example, some of these organizations [NRC, CASAC] also sometimes lay out priorities for publicly-funded research, and some [HEI] undertake or oversee particular new research projects.)

A schematic overview of the six organizations is shown in Chart A on the following page.

Chart A: Chart A: Knowledge-Assessing Institutions by Credibility and Manageability

ORG	HISTORY	SETTING	LEADERSHIP	MODUS OPERANDI	KEY DECISIONS	INDEPENDENCE
CASAC	1977 CAA	EPA FEDERAL ADVISORY COMMITTEE ACT	7 MEMBERS 5 ACADEMIC 1 STATE AGENCY	MEETS 6 TIMES A YEAR, DEVOLVES WORK ONTO PANELS	EPA ADMINISTRATOR CONTROLS MEMBERSHIP AND PROJECTS	HOUSED IN EPA'S SCIENCE ADVISORY BOARD  OPEN WITH OCCAS. VOTING
HEI	1980 NONPROFIT	EPA AND MOTOR VEHICLE INDUSTRY	BOARD COX, EPA AND KENNEDY	EXTRAMURAL TEAM TO CONDUCT RESEARCH AND REAPPRAISALS	HEALTH RESEARCH COMMITTEE SETS PROJECTS AND PERSONNEL	PRESTIGE AND FIREWALLS, DE NOVO REVIEW
NIH CDP	1977 120 CONSENSUS STATEMENTS TO DATE	W/ OFFICE OF MEDICAL APPLICATIONS RESEARCH OMAR IN NIH	DIRECTOR NIH AND OMAR STAFF	ROUTINIZED - NONFED NON ADVOC 1.5 DAYS TESTIMONY COMPOSE AND RELEASE STATEMENTS	NIH DIRECTOR AND OMAR STAFF APPROVE TOPICS PROJECTS PERSONNEL * PUBLIC HEALTH IMPORT * SCIENTIFIC INFORMATION * NEED	ABSENCE OF ADVOCATES REDUCES PRESSURE (EXCEPT FROM NIH PROGRAMS THEMSELVES ?)
NRC	1916 WWI	WORKING ARM OF NAS AND NAE AND INST MEDIC STAFF 1100 1000 COMM 10000 EXPERTS 105TH CONG INITIATED 45 STUDIES 300-350	ELECTED MEMBERSHIP NRC BOARD OF FIVE FROM EACH ORG  STANDING COMMISSION	12-15 EXPERTS 4-6 MEETINGS 18-27 MONTHS COSTS 250k+ STAFF INFLUENCE VARIABLE	EXCOM OF GOVERNING NRC BOARD APPROVES PROJECTS COMMITTEE MEMBERS APPROVED BY SUBORGS RELEASE	FUNDERS AND SPONSORS DO NOT CONTROL COMMITTEE MAKEUP, DO NOT CLEAR RELEASE OF DRAFTS

		REPORTS/YR 2/3 KNOWLEDGE ASSESSMENT 1/3 RESEARCH PRIORITIES FOR FEDS	NS	COMMITTEE BALANCE CRITICAL	REQUIRES SIGNOFF OF COMMITTEE MEMBERS PLUS REPORT REVIEW COMMITTEE	
OTA	1972-1995	TECH ASSESSMENT BOARD OF 12 LEGISLATORS	LEGISLATIVE OVERSIGHT AND FUNDING EXTERNAL ADVISORY BOARD	18-30 MONTHS ADVISORY COMMITTEE PLUS STRONG STAFF 3-5 2 TO 3 MEETINGS LONG REPORTS	TAB APPROVAL FOR NEW PROJECTS APPOINTMENTS OF PERSONNEL BY DIRECTOR RELEASE TAB APPROVAL	NOT MUCH...BUT STILL PULLED OFF REPORTS
HEINZ	EST'D 1995	ENDOWED BY HEINZ FAMILY FOUNDATIONS	BOARD OF TRUSTEES REPRESENTING 4 SECTORS - GOV'T, BUSINESS, RESEARCH, AND ENV. ADVOCATES	STILL EVOLVING— TWO-YEAR STUDIES ARE TYPICAL	PRESIDENT HAS BROAD DISCRETION OVER PORTFOLIO, M.O.	IN THEORY, THE "FOUR SECTORS" JOINTLY CONTROL THE WORK

The Policy Context: Four Channels for Supplying Knowledge for Policy. In focusing on formal advisory groups, we are putting the spotlight on only one of four alternative channels by which knowledge can flow into the policy process. First, an *agency's staff members* can apply their educational and experiential expertise in assessing options. Two practical limitations of this channel, of course, are any external skepticism of the impartiality of such assessment and the difficulty an agency may have in continually keeping staff current with the latest in research knowledge as the relevant knowledge base continues to evolve. Second, there are commonly *notice and comment* provisions in an agency's proceedings, during which the projected policy options are laid out for comment by interested and affected parties; the elicited comments can, of course, include claims about factual inaccuracies in the proffered analysis. The real and perceived effectiveness of this channel depends on the agency's alacrity in re-examining such claims fairly. Third, and similarly, an agency may be

required to prepare and publish an *impact statement* that lays out the anticipated effects of policy choices. Although on the wane as a decision-aiding tool, agencies still find themselves subject to as many as ten distinct impact-statement requirements [examples include environmental impact statements, economic impact statements, small business impact statements, family-farm impact statements, historical-preserve impact statements, etc]. Fourth, an agency can directly retain specialized *for-profit and non-profit contractors* to perform analyses.

The six institutions covered in this paper represent the last of these four knowledge channels, but they should not be taken as representative of that class. They may be thought of as the "high end" of the class of outside knowledge providers. This class also comprises a large number of other assessors, including for-profit contractors, university policy research programs, non-profit think tanks, FFRDCs, and industry research groups. Among this larger set, our six specimens are marked by an unusual intensity of concern for their actual and perceived *independence* from particular vested interests, and thus for their reputation for *broad-spectrum credibility* as perceived by a wide array of interested and affected parties. In fact, groups like these six can be thought of as "the credibility specialists"<sup>3</sup> within the American policy process. Each prizes its reputation for staying above politics and for "getting the science right," and can point to procedural safeguards for maintaining that reputation.

The paper represents my personal reflections on a meeting held at MIT in 2000, the ***Workshop on Institutional Means for Policy-Related Knowledge Assessment***.<sup>4</sup> This two-day session<sup>4</sup> was attended by 27 persons; including 13 with first-hand experience with one or more of the six organizations, five with client/user experience as senior governmental decision-makers, and nine with faculty appointments relating to public policy processes. While the views expressed here have benefited from workshop discussions, they should not be taken as the consensus views of attendees.

<sup>3</sup> Our MIT colleague Sanford Weiner introduced this useful term in our early project discussions.

<sup>4</sup> For more information on the workshop agenda, attendees, and handout materials, contact me at [lmccray@mit.edu](mailto:lmccray@mit.edu). The workshop was supported by funds from the MIT Consortium on Environmental challenges and by discretionary funds from the MIT Center for International Studies. It was principally designed by me and was chaired by former Undersecretary of Energy Thomas Grumbly.

## **B. THE SIX "CREDIBILITY SPECIALISTS"—SOME HIGHLIGHTS**

Note: A profile for each of the six organizations is found in Appendix 1 on page 23; it covers each organization's history, setting, leadership, modus operandi, key process decisions, independence measures, and points of uniqueness.

### **1-- National Research Council<sup>5</sup>**

The NRC [frequently misidentified in the press and elsewhere as the National Academy of Sciences] is the operating arm of three honorific private societies: the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. It is by far the largest, oldest, and best-known of the six organizations reviewed in this paper.

The NRC issues 300 to 350 reports a year and has revenues approaching \$200 million per year. It is estimated that between 40% and 60% of its studies can be termed knowledge assessments—most of the rest seek to advise agencies on their scientific research programs. It is typical for an NRC study to involve a committee of 12-15 experts who meet 4 to 6 times over 18-27 months and spend from \$250K to \$750, with the variation in cost highly dependent on the level of staffing.

The NRC's procedures put a premium on safeguarding the organization's independence from intervention by vested interests, including government sponsors. For example, sponsors cannot specify who shall serve on study committees, and do not have a role in evaluating draft reports. Within the NRC process, a vigorously autonomous Report Review Committee must complete its review before a report can be released. The entire NRC program is subject to oversight by a Governing Board that is drawn from the three elected memberships. The organization is best seen as a creature of the broad US research community. Many scientists consider the NRC to be very effective; typical is the judgment, expressed at the Workshop by Gilbert Omenn, that the NRC represents "the gold standard" for integrity and quality. Scientists tend to think of the NRC as "their organization."

<sup>5</sup>

Disclaimer: I was employed at the NRC from 1975-1977 and from 1981-1998, eventually serving as the founding director of its Policy Division. In the latter period I enforced all NRC policies and procedures, and originated some of them. The reader can reasonably assume that it remains somewhat unnatural for me to see NRC practices as deeply flawed, despite my earnest wish to see the organization objectively.

Among the six subject organizations, the NRC has the broadest span of expertise [with the arguable exception of economics, it covers most fields of science and technology]. It enjoys a reputation, particularly among journalists, for independence and credibility. Its long history, and the number and the prominence of its past studies, contribute to this reputation.

Some questions that arise about the NRC include whether its highly articulated procedures result in studies that are too slow and too expensive, whether it is too tightly linked to the university sector, and whether its control by the research community weakens its appreciation of practical policy realities. Conservatives have questioned whether a liberal academic bias is found in the NRC's work, particularly in assessments of social programs and policies.

## **2 --Office of Technology Assessment [1972-1995]**

The OTA, established by Congress in 1972, had grown to command an annual budget of about \$20 million and a staff of 140 persons when Congress eliminated it following the Republican electoral sweep of 1994.

OTA was, at that point, producing about 50 reports a year, with a typical assessment costing from \$300,000 to \$600,000 and taking 18 to 30 months to complete. For each study report, an advisory committee was appointed to supply review comments, and a strong professional staff of three to five persons was required to pay close attention to such comments. It was not unusual for OTA to engage outside contractors to conduct studies under OTA staff supervision, a practice not found among the other five knowledge assessors.

The OTA program was overseen not by substantive experts but by a Technology Assessment Board [TAB] comprising 12 legislators selected so as to achieve a balance of input from the two houses of Congress and the two political parties. TAB approved new starts and all draft reports. Thus, unlike some of the other subject organizations, there was no attempt to sequester OTA from politics. OTA was unique among the six groups in serving a single client, the US Congress.

The perceived quality of OTA reports varied widely. Some observers [including myself] who were familiar with OTA reports and NRC reports on similar subjects saw them as similar in quality, and sometimes better focused. Others, however, perceive OTA reports as subject to the subjective bias of OTA staff members.

Why did Congress terminate OTA? A range of views is heard. OTA's last Director, John Gibbons, has explained that the new Republican Congress of 1995 needed to prove its ferocity by eliminating at least one established institution, and that OTA, tightly allied to no power base, was left uniquely undefended as its leaders scanned the Washington horizon for likely targets. Others believe that [despite the presence of conservatives on TAB itself] OTA reports showed a systematic liberal tint that the new Hill leadership simply wanted to eliminate.

### **3 -- Clean Air Scientific Advisory Committee [CASAC]**

CASAC's primary role is to review the scientific quality of EPA analyses relating to the six "criteria" air pollutants, including particulate matter, tropospheric ozone, and nitrogen oxides. CASAC operates under the Federal Advisory Committee Act [FACA], and is in fact one of 30 FACA committees now serving EPA. Its annual direct cost budget is estimated at \$2 million to \$3 million.

CASAC comprises seven members, and works through subcommittees. CASAC membership is proposed by the EPA Administrator with input invited from the White House. Membership of CASAC and its panels is largely drawn from

universities and from independent research laboratories, but is subject to a set of partisan balance requirements specified by FACA.

CASAC meets six times a year. FACA rules require that its meetings be open to the public, announced in advance, and always attended by a "Designated Federal Official." Its work schedule is driven by EPA's statutory requirement to review periodically the underlying scientific basis for its air standards.

Unlike the other five subject organizations, CASAC does not perform original knowledge assessments, and instead restricts itself to the evaluation of the quality of assessments that are drafted by agency personnel.

Is CASAC truly independent from partisan and political pressures? The answer to this question is mixed, but it can be said that it is a lot more independent than most FACA Committees are. While it is true that CASAC depends entirely on EPA officials for its budget and staffing, and is under at least nominal control by its designated government official, it is also true that it enjoys obvious autonomy in deciding when and if to release its reports. In fact, in recent years, the Committee has not infrequently been seen to be at odds with EPA leadership over the scientific content of EPA's draft assessments. It is not unusual for CASAC to send draft EPA assessments back for more work. CASAC's stature as an independent voice is often said to stem from the close attention paid by powerful Congressional leaders, and from CASAC's formal placement in EPA's prestigious Science Advisory Board, which is directly responsible to the EPA Administrator, rather than within the air program office, whose staff drafts the assessments.

There are many hundreds of FACA committees advising federal agencies. Many are simply groups representing parties at interest in agency programs, and few appear to have the scientific stature or the tradition for independence that CASAC demonstrates.

Among our six subject organizations, CASAC [along with the Heinz Center] is at the extreme in the openness of its procedures.

#### **4 -- The Heinz Center**

The H. John Heinz Center for Science, Economics, and the Environment was established in 1995. It has targeted three areas for initial emphasis: environmental reporting, global change, and sustainable oceans and waterways. The Center has had an operating budget in the neighborhood of \$4 million per year and a staff of 15.

Two characteristics distinguish the Heinz Center from other knowledge assessors. First, there has been a special attempt to bring in economic expertise

issues as well as other scientific expertise. Second, the Center is carefully built on the premise that four sectors need to be directly and systematically involved in its knowledge assessments: industry, environmental groups, government, and scientists. The Heinz Center's Board of Directors, itself, pointedly comprises representatives of the four sectors—in about equal strength.

The Center's deliberate "four sector" approach had made its knowledge assessments different from the other five subject organizations in two important ways. First, with all sectors participating in the program, the Heinz Center conducts its business in the open, feeling no particular need to shield itself from untoward access or influence by agencies and powerful private interests—such parties are, effectively, already at the table. Second, having a distinct but minority position in the program, academic and other subject-matter experts do not exert substantive control over the Center's everyday assessments, as is the tendency in the other five programs [assuming one credits the role of the Advisory Committees in OTA projects.] At Heinz, the scientists are on tap, but not on top.

Several of the Center's evolving study processes generally resemble those of the NRC, where the Center's first President and several of the Heinz Center's early senior staff members had had prior experience.

At the time of the MIT Conference, the Heinz Center's program had yet to issue any assessments, so it had yet to build a national reputation. Since then, its initial reports (on coastal management and environmental monitoring) have been published, and observers remain optimistic that it will have something unique to contribute to the field of knowledge assessment. The Center has, in the past year, undergone the kind of internal reforms that often accompany innovative ideas, and a good proportion of its early leaders have now moved on. However, it is reported that the Center remains dedicated to its original "four sectors" approach, which is its chief hallmark, and which recommends it for continued scrutiny.

## **5 -- NIH's Consensus Development Program**

Since 1977, the National Institutes of Health [NIH] has completed well over 100 knowledge assessments in its formal Consensus Development Program [CDP]. These consensus statements typically address a particular medical technology or practice: in a recent year, for example, the Program issued reports on breast cancer screening, acupuncture, and treatments for heroin addiction.

The CDP operates out of the office of the Director of NIH, who approves each new consensus review. Each report emanates from a national conference of outside experts on the chosen subject. Planning for a conference is typically

shared between the CDP staff and the staff of the pertinent Institute of NIH. This means that there is no extensive attempt to isolate the process from NIH agencies that may have established positions on the issues under examination. The direct cost of a review is on the order of \$200,000 over a total period of 9 to 18 months. The typical annual output has been 3 to 5 CDP consensus statements, implying a direct cost budget of \$1 million or less.

Uniquely among the six subject organizations, the NIH reviews are completed in a single meeting of the review committee, a meeting held at NIH over a period of a few days. Considerable advanced effort is made to define precise and specific questions for the panel to answer. These review panels are described as "nonfederal, nonadvocate individuals knowledgeable in the field." A "public representative" is included, sometimes to represent the viewpoints of patients. The panel typically listens to presentations during the first phase of its solitary meeting, and it writes and delivers a short consensus statement to the press and public at the end. NIH excludes members from panels who have published work directly addressing the subject of the review, often relying heavily on experts from adjacent areas of expertise. This is a unique practice among the six knowledge assessors.

Owing partly to pressures on the Program following a CDP report saying there is insufficient evidence to recommend annual breast cancer screening for women under the age of 50, the CDP was subjected to a fresh review in 2000, and the Program was also reviewed by the NRC/Institute of Medicine. The IOM report was generally favorable, but encouraged consideration of holding multiple meetings of review panels, rather than depending on a single gathering.

## **6 -- Health Effects Institute [HEI]**

HEI was established in 1980 as a special-purpose organization to help "remove science from the adversarial process"<sup>6</sup> in the area of motor vehicle pollution. The new organization was established jointly by the highest officials at the EPA and the automobile industry, and equal funding from these two sources has been a hallmark of HEI.

Over the years, the bulk of HEI's \$6 million annual program has been devoted to the creation of a body of specific new research findings that are deemed credible by both EPA and industry. This is accomplished through extramural projects, most of them conducted by university researchers. However, it has also conducted a few "special studies" that amount to knowledge assessments. In recent years, it has conducted about one such assessment per year, including a review of mortality owing to particulates, of asbestos contamination, and of the health effects of magnetic fields.

HEI has organized itself to provide a high degree of independence from its

powerful sponsors. Individuals with national standing, including Harvard's Archibald Cox, Stanford's Donald Kennedy, and NAE's Robert White have been visible in HEI's leadership, and an autonomous review committee of experts carefully crosschecks each new study. The program influence of financial sponsors, after they provide funding, is highly circumscribed, and individuals with any financial connections to interested parties are not involved. The involvement of representatives of public interest groups in HEI work has been rarer than in other some subject organizations, including the NRC.

For many years, HEI's budget was flatlined at \$6 million, with EPA and industry providing precisely matched funding. Some observers linked this budgetary stagnation to the perception that HEI was more tuned to academic topics than to policy-relevant studies that EPA is concerned with.

### **C. HOW TO ORGANIZE FOR KNOWLEDGE ASSESSMENT—THE SEARCH FOR GENERAL LESSONS**

The current project originated as one part of a broader effort to understand credibility as a factor in public policy – especially in regulatory policy. One way to understand credibility, we reasoned, was to examine a set of actual institutions that had achieved reputations for being credible across a range of partisan viewpoints. In time, we decided that our six chosen groups' knowledge assessment functions were the best places to start.

Our purpose is inductive; having examined each of the six high-end knowledge assessors, we want to draw some inferences about what they have in common, how they are prospering as a class, and what they foretell for future improvement in fitting knowledge and power together in cases where issues are factually complex and controversial. If our goal is to improve the quality of the knowledge brought to bear when policies are set, can these six organizations together show us the best pathways?

<sup>6</sup> Thomas Grumbly,

Because this inductive exercise involves, in part, asking “what is the secret of these organizations' success?” we need to reflect briefly on what is meant by “success.” We do not claim that the groups are models of effectiveness, of efficiency, of objectivity, or of influence; that is because it is too difficult [of course] to measure these things. Their “success,” it is enough to say, is in their *persistence*; they have undertaken to provide assessments that are both sound and beyond the control of partisans, and [for the most part] they *lasted*. Their clients, who paid for assessments <sup>7</sup>only because they needed results that were broadly credible, kept coming back.

Ideally, of course, we might hope to correlate an organization's relative credibility with its operating procedures in order to discern an optimal organizational design.

In practice, however, we have no reliable measure of organizational credibility; it may be true that some of the organizations have achieved higher general-purpose reputations for independence than others have, but we have no sound way to rank them. So we are left with that one fact: they have persisted over time.

(Reminder: By restricting our attention to organizations that are dedicated to achieving broad-spectrum credibility, we are looking at the small range at the “high end” of the industry, and at organizations that are likely most useful to sponsors when a knowledge assessment was sought for a particularly difficult or controversial subject.)

**Point 1. There is, alas, no simple recipe for achieving organizational credibility. The six groups share surprisingly few central structural or procedural special features.**

When we started our examination of organizational credibility, we imagined that if we worked hard enough, we would surely uncover simple rules for attaining broad-spectrum credibility. Some possible conjectures that occurred to us:

[a] The key to credibility is in the careful handling of funding; if, in reality or perception, the sponsor of a study actively manages a knowledge assessment, credibility is damaged. Steps must be taken to keep funding sources at some distance from the assessment process. It is poisonous to be tied to funding by governmental and private interests.

The careful reader may argue that OTA no longer persists, but must admit that it did ply dangerous waters over the span of five Presidents. The careful reader may well ask whether the Heinz Center is perhaps still “too young to fully qualify as ” persistent:”

[b] The key to credibility is giving full autonomy to subject-matter experts in the conduct of a study. Steps must be taken to isolate expert consensus-building from the reality or perception of partisan private and government pressures. It is poisonous to be seen as subject to partisan influence in designing and conducting a knowledge assessment.

[c] Inasmuch as the National Research Council, as the largest, most diverse, and most famous knowledge assessor, had long established a reputation for performing credibly, other organizations will be found to have adopted NRC procedures, at least in part, in order to attain credibility.

In fact, however, we find more diversity than congruence among the six programs. The following is the spread of findings for a range of organizational features variables: We find no convincing, central commonality.

Isolation from funding sources? Only the NRC and HEI take deliberate and visible steps to isolate an assessment from control by funding sources whose

interests may be affected. At the Heinz Center, funders can actively help manage a project. At NIH, CASAC, and OTA, assessments have been undertaken with public funds, and sometimes those funds were obtained from those federal programs with clearly-perceived stakes in the outcome.

Arm's length arrangements for vested interests? Only HEI has striven to build a firewall between vested interests and the study process; it excludes those with research funding from EPA or industry from participating. NRC not infrequently appoints to its study committees a small number of persons who are employed by parties at interest. OTA and the Heinz Center have welcomed participation by parties at interest.

Independence from appointed policy officials? OTA, HEI, and the NRC have undertaken to protect their studies from influence by federal agencies with a stake in the findings. However, neither NIH nor CASAC features such safeguards; in the former case, the relevant NIH Institute participates fully in planning a consensus project, and in the latter case, EPA [although not its relevant program office] directly oversees the study process, assigning staff and deciding on the project budget and committee memberships.

Autonomy for the experts? Led by NRC and HEI, four of the subject organizations attempt to isolate the experts from outside influence as they deliberate. OTA and the Heinz Center, however, have not. At OTA, the elected politicians on the governing Technology Assessment Board reserved the right to determine whether reports are good enough to release, and the review process for drafts was managed by staff supervisors without highly specialized expertise. At the Heinz Center, key study design questions are worked out, directly, *among* representatives of the contesting interests.

Fierce *de novo* review of drafts? The NRC and HEI have instituted independent reviews of draft reports. However, review requirements are weaker at OTA and CASAC, and they are absent in the NIH Consensus Development Program.

In general, then, the six organizations have found quite different paths to the credibility they have achieved. The broad strategy followed by NRC and HEI emphasizes giving autonomy to the experts and trying to isolate them from pressure. CASAC follows the same strategy, but without the luxury of achieving demonstrable independence from EPA's political appointees. NIH successfully capitalizes on the high public esteem with which NIH is regarded, and does not fret about isolating its process from federal programs that have something at stake. The Heinz Center casts its lot with a strategy of *inclusion, not isolation*—its credibility depends not on removal from influence, but on prior commitment by the full set of influence groups to abide by the assessment.

It is OTA that presents the most intriguing apparent anomaly. While OTA did not

survive the conservative political tide of the 1994 national elections, it did accrue enough credibility to weather two decades of political winds, a feat that should not be taken lightly. And it did so with very few of the assets that we might have taken as essential to survival; it had no protection from partisan pressures, no persuasive dependence on the prestige of nationally-recognized experts seen to be above politics [such experts served on advisory committees, but knew that OTA staff had no obligation to incorporate their comments on draft assessments], and did not rely on highly demanding review standards and procedures. How, then, can we explain OTA's skein of success? Two factors suggest themselves. First, the powerful nearby legislative forces could cancel one another out, providing that an assessment was astutely designed. And second, OTA's long-time Director, Jack Gibbons, may have possessed unusual skills of management and persuasion. In all likelihood, one might think, an exposed program like OTA's should have foundered on party politics and ideology well before 1995.

The existence of so many distinct—yet workable—paths to achieving broad-spectrum credibility suggests an unexpected conclusion: *The attainment of sufficient organizational credibility is easier to accomplish than it might seem.* The public is, perhaps, less skeptical about the corruptibility of expert assessments than we thought. It does not insist, for example, on distancing the six assessors from partisan pressures.

For almost any policy subject, a large number of parties may potentially serve as attackers if they disagree with the outcome of a knowledge assessment. In fact, however, what we see is considerable tolerance for imperfect processes. And unless the knowledge assessor steps directly on the toes of a vested interest, the interest may be inclined to give the benefit of the doubt to the assessor. The roots of this tolerance are not clear to us at this point, but tolerance seems rather more common than bickering and outrage about flawed credibility.

## **Point 2. The groups find some subjects just too hot to handle.**

In discussing the programs of the six subject organizations, Workshop attendees repeatedly referred to the need to avoid "third rail" issues. As with the third rail in an electrically-powered transport system, touching such an issue could be fatal. One cannot simply "speak truth to power."

This was an unexpected Workshop theme. There is no discussion of third-rail realities in the six organizations' self-descriptions; academic writings do not emphasize them; and the topic was not raised in the material prepared for the Workshop.

But the actual cases of third-rail phenomena are graphic and persuasive. For instance, some observers say that a key cause of OTA's demise was its report on missile defense systems, which was very unpopular with conservative

lawmakers. The NRC for several years avoided studies on the feasibility of the Reagan Administration's Strategic Defense Initiative ("Star Wars"). It understood the risk to the organization if an assessment came out in opposition to such a high-priority and high-profile program [especially when opposition to the concept was suspected among the membership of the NAS itself, a membership that might well be portrayed as having the progressive leanings generally associated with universities].

And the NIH's Consensus Development Program was rocked by its experience with a consensus statement on breast cancer screening, one that concluded that there is not sufficient evidence to recommend annual mammography screening for women under the age of 50. That report was openly opposed by the Director of the National Cancer Institute the very hour it was reported. It was also denounced by patient groups, and it led to an explicit threat to cut the NIH budget on the part of the Senate Appropriations Committee Chairman. It is probably not coincidental that the CDP program was subjected to new assessments in the wake of this controversy.

Beyond these cases, it is clear that there are many issues, often cited colloquially as "hot-button" issues, for which political realities make a knowledge assessment impossible, or at least very tricky. Abortion, Agent Orange, EMF emissions from overhead power lines, seat-belt safety, and AIDS are some other examples.

The point is not that knowledge assessors lack the courage to challenge the conventional wisdom or unfounded public policy. In fact, the NRC, with exceeding care, completed studies [sometimes using its own funds] that treated formerly untouchable topics including the AIDS epidemic, acid rain, and global warming. HEI took on asbestos risk and airborne particulates. Other knowledge assessors have done the same, and their reputation for independence has been duly credited when they have. However, it remains true that some issues remain off-limits for these organizations, and that the organizations know it. So although the numbers of well-recognized "third-rail" issues are not huge, it seems fair to observe that all knowledge assessors are acutely aware of them, and that they cast a shadow over the way that less dangerous studies are performed.

While knowledge assessors appear to know what issues are too hot to touch, it is not clear that one can define them in the abstract. This subject warrants more attention, because it can define the outer limits of the art of knowledge assessment.

**Point 3. Independent knowledge assessment is not a booming sector. The demand for highly credible knowledge assessments is not increasing, and it's not clear why. One deterrent to the use of extramural knowledge assessments by federal sponsors may be that they bring a risk to an agency—and to its publics—that the associated policy debate will escalate**

**and become intractable.**

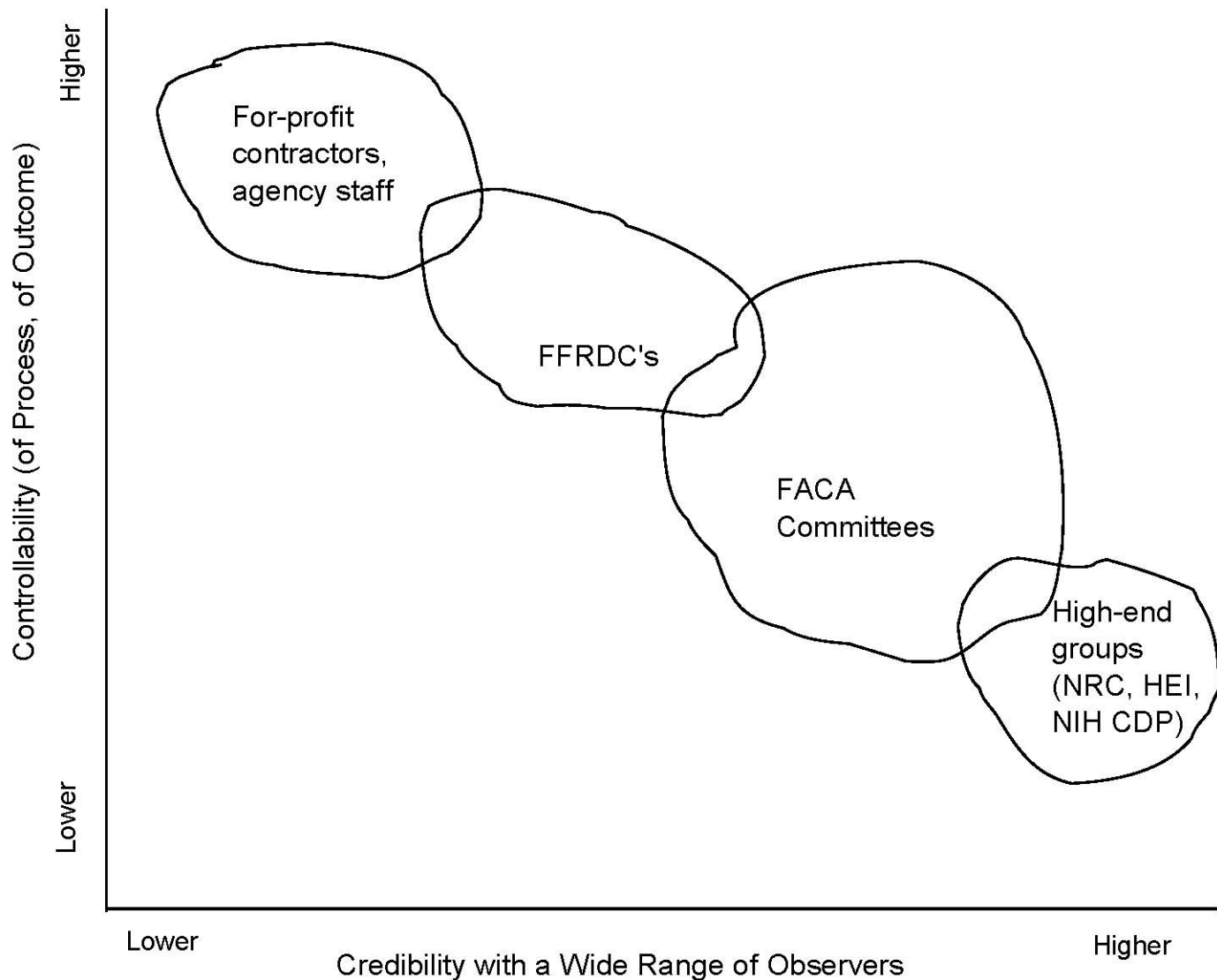
One might have surmised that the demand for credible knowledge assessments is expanding to reflect the complexity of modern life. An ever-larger fraction of policy issues rest in part on specialized scientific and other factual bases that are beyond the grasp of laymen. Regulatory policymaking seems ever more contentious, and ever more demanding of detailed analyses to address the major concerns of a growing set of feisty interest groups. While it is true that the size of federal programs is not increasing, their complexity is still growing. Shouldn't these trends lead to a rising insistence on pristinely believable knowledge assessments?

If one considers the six "credibility specialists" as a class, one does not see a particularly robust sector. OTA, the second largest of the six entities, is gone, and Congress does not appear to feel a need to replace it. The NRC program is running roughly at steady state, as measured by its overall staffing strength. The HEI experiment seems to be working, but it did not expand for many years, and it has not as yet served as a model for other new special-purpose assessors. CASAC's central role in assessing EPA's science has not spread to other FACA Committees. The Heinz Center is a fresh development, but it seems fair to say that it is still finding its niche, and its long-term role is not yet established. One cannot conclude, in fact, that the federal demand for credible assessments is on the rise.

One plausible way to explain this apparent contradiction is to say that knowledge assessment is increasing over time, but not among the high-end organizations. That is, the growth in assessment has come to the benefit of for-profit contractors that work closely with the agencies, and whose schedules and outcomes are easier for agencies to predict, or even to influence.

It is helpful to consider the formal demand for knowledge assessments. If one arrays the set of knowledge-assessing institutions on a notional chart whose axes are credibility and manageability [See Chart B on the following page], one appreciates a central tradeoff that a policy official faces when selecting a knowledge assessor. The organizations under consideration in this paper here are high in independence and high in credibility, and many other assessors are more subject to direct oversight but are less credible. So if it needs to farm out an assessment, an agency faces a choice: If it wants credibility, it must cede some control over the management and the timing of the assessment, but if it wants to exert control, it must cede some credibility. It cannot readily find stones to slay both birds at once.

Chart B: A Sponsor's Tradeoff, Credibility vs. Controllability in the Choice Among Knowledge Assessors



In finding that agencies have not supported more work by the high-end knowledge assessors in the recent past, we are led to postulate that they have most frequently opted for manageability in their knowledge assessments, and not for increased credibility. (An alternative explanation, that the overall volume of knowledge assessments has diminished, is conceivable, but seems unlikely.) And that choice, presumably, reflects the fact that the immediate publics that the agencies face—the regulated firms, interested NGOs, etc.—have not in fact demanded the extra credibility that the more independent assessors can provide. They find they can live with the assessment work done by the agencies' own contractors. So, presumably, can Congressional overseers and appropriators who could question the impartiality of an assessment.

A second tradeoff attends the placement of knowledge assessments, and thus affects the demand for high-end assessments. As a practical matter, turning to an organization like the NRC or HEI for a knowledge assessment removes a

policy issue from ordinary business routines of an agency. Going to an independent assessor can lead to an escalation of the decision to a venue that brings risks to those involved.

Commissioning an independent knowledge assessment is likely to alter the bureaucratic politics of an agency's policy decision. It is worth considering how. The mechanism is, arguably, escalation.

A somewhat simplified view might describe three dominant venues in which an agency's policy issues may be resolved. Each venue has its own politics, its own media coverage patterns, and its own procedures.

The first-level, which we might call the "normal" venue, is subject to the routine procedures and time requirements of the relevant agency, the Administrative Procedure Act, and the agency's governing statute.<sup>8</sup> The players are the immediate interested and affected parties that follow the agency's program most closely. The relevant media are the trade press for the regulated business sector.

At the second level, the "intermediate" venue, matters pass from the agency's routine procedures and are referred elsewhere. This referral may take the form of a special review by the NRC or another assessor, intervention by the White House, or another source. When such a referral is made, the agency's decision clock is suspended [presumably with the approval of the Congress]. Later, a report is made. The report may be covered by national media, a broader national debate may ensue, and additional interested and affected groups may decide to become involved.

At a higher third level, the "abnormal" level, there is direct and continual involvement by the legislature and national media. Congressional hearings may take place, and the contested issue may be debated on *Nightline*, the *Today Show*, or even in late-night comedians' monologues. In the "abnormal" venue it may be difficult to determine what the decision schedule is. Sometimes it will become unclear who will make the decision, and by what criteria.

<sup>8</sup> Because the process [by definition, I think] involves the same set of interveners, I include judicial challenge as an aspect of the "normal" decision venue.

A key question, obviously, is how and when issues escalate from lower venues to higher ones. The dangers of escalation to an agency—and to its immediate publics—may come with a decision to pursue an independent knowledge assessment.

One would assume that agencies will prefer to keep issues in the "normal" venue, where events remain most predictable and commitments to their many publics can be kept. In that case, they might be expected to oppose the referral of a knowledge assessment, for example, to a high-end assessor, for doing so would amount to moving the issue to the "intermediate" venue.

Similarly, the routinely active interested and affected parties may, for their own reasons, wish to keep the issue in the "normal" decision-making venue. If any group becomes disaffected, it may endorse escalation to the "intermediate" venue [by launching a publicity campaign, or by directly lobbying the legislature to enact a referral]. However, such groups may be tempted to willingly trade their freedom to defect—to argue for escalation, in exchange for a key point it wants the agency to incorporate into its proposed policy.

If it is true that an agency and its immediate publics may prefer to inhabit the "normal" venue, that fact would explain the tolerance for knowledge assessments performed under the agency's own control. To refer the matter to the more independent "credibility specialists" like those in our sample will subject the issue to a new timeline, the entrance of new advocates, and a possible loss of any ground gained in "normal"-venue deliberations. As regulatory programs become more complex, much of the increase in demand for new knowledge assessments is absorbed by contractors who are under some administrative control.

This change-of-venue perspective may help explain the recent stagnation of the class of knowledge-assessing organizations.

It also may illuminate two other findings listed above. First: why is there so little commonality [point 1, above] among the way the six subject organizations go about their business? One reason may be that, for both the agency and whatever force is driving the escalation to an outside assessor, the main point is to place the matter outside agency control; whether or not the assessor receiving the referral is slightly or emphatically independent. That is, the *differences* among the high-end assessors are not particularly significant to their audiences, and thus not a point of competition among them.

Second: If a knowledge assessment metaphorically touches the third rail [point 2, above], the issue thereby escalates from the "intermediate" venue to the "abnormal" venue. This can undermine the credibility and the utility of the assessment, and perhaps as important, this will result in further loss of control by the sponsoring agency and at least some of its publics.

It is not infrequently argued that agencies will arrange for independent ["high end"] knowledge assessments in order to stall a decision, or to deflect responsibility from itself. While these motivations may have come into play, they are subject to some limitation. Regarding delay, an agency faces also delay with an internal, or more controllable, knowledge assessment—although the time for the assessment may be less predictable with an independent review. Regarding the deflection of responsibility, this is not a trait that agencies commonly exhibit [witness the prevalence of conspicuous "turf fights" within and among contiguous agency programs], and commissioning an independent knowledge assessment can only *temporarily* relieve an agency of its responsibility to act.

**Point 4. Policymakers have a common critique: assessments are not relevant enough to policy realities, and they take too long.**

The Workshop heard from several persons who had, as senior policy officials, received advice from one or more of the six subject organizations. Attendees also had received a summary of views from Federal sponsors of knowledge assessments, which is found below as Appendix 2 [page 36].

One common complaint was that some of the assessors, particularly those controlled by researchers, tend to stray from the key policy questions, dwelling instead on [1] that portion of the relevant science that is certain and [2] the current need for research funds. [Example: In its middle years, complaints were lodged by EPA officials that HEI was neglecting policy issues in favor of funding studies of mainly academic interest.]

One way that the NRC and the Heinz Center have tried to maintain a practical policy focus in some of its studies is by naming former federal officials to study panels. [An interesting question, but one that is beyond my present scope, is whether the credibility specialists are in fact weaker in this area than are the other assessors that an agency may turn to instead.]

It is common for clients to decry the length of time [and, often, the associated costs] that independent knowledge assessors take in completing their work. How much quicker and cheaper the agencies' less-independent alternatives are is unclear . . . although the *uncertainty* of the time of delivery may be less for assessors that the agencies control.

Meanwhile, NIH's Consensus Development Program has completed over 100 consensus statements, each one based on a *single meeting* of a panel at which a report is released to the public on the meeting's final day. One key to the CDP process appears to be the delineation of fairly narrow—but policy relevant—questions for the assembled experts to address.

Clearly, the CDP process cannot be applied to all assessments; some truly do require an open-ended deliberative process, and thus the serial exploration of new ideas as they emerge among study participants. Other assessments, however, might easily follow the CDP model; [a] the careful posing of relevant questions that can be answered in one sitting, [b] a few months to appoint and coordinate a review panel, and [c] a short resulting statement. Is the ubiquitous "NIH" ["not invented here"] syndrome preventing the replication of the NIH success with quicker assessments?

**Point 5. There is an absence of highly-credible "on-tap" forms of advice.**

When one examines the range of programs performed by the six subject organizations, a gap reveals itself. These six organizations all provide what might be termed postal-style answers; that is, a decision-related question is posed, the assessor takes it under deliberative advisement, and somewhat later a formal answer is provided. The client has to wait for the postman to ring.

There appear to be few functional equivalents to the Federally-funded Research and Development Centers [known inelegantly as FFRDC's] that are employed to provide "on-tap" advice to national security organizations. The US Air Force can turn to the prestigious Rand Corporation for short-term, ad hoc help from relatively independent experts; EPA, for example has no parallel arrangement. Is this gap one that should be filled, or can agencies like EPA obtain all the outside help they really need from their routine contractors? There are a few domestic policy areas featuring FFRDC-like arrangements, including the relationship between the FAA and the Mitre Corporation. Should EPA consider establishing a similar setup?

The creation of high-end on-tap sources of knowledge would also, of course, address the need for quicker responses to federal risk-reduction program needs.

**Point 6. There is a recent trend toward public openness in conducting assessments**

Several of the assessing organizations report a trend toward conducting more open assessments. The Heinz Center holds meetings that are open to public and press. Given the inclusive nature of its program, Heinz officials feel that openness is the best way to gain the confidence of the four sectors it aims to serve. OTA's representatives also reported progressive comfort with conducting its assessments in the open. As a result of a successful legal challenge in the late 1990s, the NRC was induced to shed more light on its deliberative processes, and now must announce meetings in advance and reveal the names of proposed study panel members for public comment.

In the past, knowledge assessors were chary of open proceedings. It was deemed important that assessors could informally examine daring or exploratory ideas in isolation from the press and strong interest groups. Now, however, there is a noticeable absence of concern among the credibility specialists that this trend will undermine expert deliberations.

One has the impression that we are seeing a continuing ripple-effect from the "sunshine in government" initiatives of the 1970's. Both public organizations and assessors are finding ways to operate effectively in the public eye.

**Point 7. The knowledge-assessing sector lacks self-awareness.**

Very few of the attendees at the Workshop seem to have thought about the common functions of the six subject organizations. Nor have most students of science policy considered these groups as a *class*.

The Workshop was organized on the premise that the nation, by using these six knowledge assessors and others, had in fact conducted a natural experiment that might now be evaluated. Which organizations, and which types of organizations, have proven most successful? Should OTA be replaced? Should Sigma Xi enter this field? Should other agencies replicate the carefully-balanced HEI model? Should FACA be adjusted in light of experience with CASAC and other similar cases? At this point, neither the practitioners nor the academic observers appear to see these as riveting questions.

## **APPENDIX 1**

### **ORGANIZATIONAL PROFILES**

March 2000

- 1. National Research Council page 23**
- 2. Office of Technology Assessment ['72-'95] page 26**
- 3. Clean Air Scientific Advisory Committee page 28**
- 4. Heinz Center page 30**
- 5. NIH Consensus Development Program page 32**
- 6. Health Effects Institute page 34**

#### **1 The National Research Council [NRC]**

*URL:*

*History:* The NRC was established in 1916 as the United States neared entry into World War I and there was a felt need to coordinate and expand research in this country. The National Academy of Sciences, established in 1863, had conducted a few consensus advisory studies for the government over the years,

and the NRC took over that role. In 1964 the NAE was established and joined NAS in managing the NRC, and in 1970 the Institute of Medicine [IOM] followed suit.

*Setting:* The NRC is the “working arm” of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine, each of which is an independent honorific organization organized to acknowledge the best individuals in their field of specialization. The NRC, with a staff of 1,100, operates over 1,000 committees and involves over 10,000 volunteer experts [fewer than 20% of whom are elected NAS/NAE/IOM members] in its studies. The NRC’s annual revenues now approach \$200 million, of which about 85% comes from contracts with federal agencies [13 agencies provide more than \$1 million per year]. The 105<sup>th</sup> Congress asked the NRC to do 45 separate studies. In all, between 300 and 350 reports are produced annually, with the National Academy Press publishing over 200 of them. The fraction of studies that may be called knowledge assessments is perhaps two-fifths to three-fifths, most of the remainder being studies intended primarily to set out research priorities for federal agencies.

*Leadership:* Governance of the NRC may be best understood in light of the elected members of the Academies’ strong interest in maintaining control over work done in their name. The NRC is chaired by the NAS President and vice-chaired by the NAE President, both elected by their membership and holding full-time positions. The NRC’s Governing Board comprises five members sent by NAS, five by NAE, and two by IOM. Linkage between the Governing Board and particular studies is provided by a two-level hierarchy [mostly organized along disciplinary lines] of standing commissions and their boards. For example, a study on an environmental topic is most likely subject to management by the Board on Environmental Studies and Toxicology [BEST] and its parent Commission on Life Sciences.

*Modus Operandi:* An organization as large and diverse as the NRC resists tight generalization. In general, studies involving a knowledge assessment will involve a study committee of 12-15 experts that meets 4-6 times over 18-27 months and costs from \$250,000 to \$750,000 [depending largely on staffing strength]. The study committee can, and usually does, determine its own study approach, can re-negotiate its charge, and can decide whether its members draft a report or rely on staff for drafting. Members’ direct costs of participation are reimbursed, but that is normally their only compensation for service. Committees undergo “potential sources of bias” procedures, and direct financial interests are avoided, but few other hard rules govern the selection of study members. Concern about committee balance—in reality and in perception—often dominates the selection process. Most study committees draw a majority of members from academic institutions. It is not unusual for an NRC committee to include members whose current or past employer has a stake in the results of the studies, but all members are asked to serve as individuals. Historically, the NRC was not

subject to FACA, but recently the Act was modified to open parts of the study process to public inspection.

*Key Decisions:* The Executive Committee of the NRC Governing Board formally approves all **new projects**. The **appointment of committee members** is approved by the NRC Chairman, often with the advice of the NAE/IOM Presidents. An NRC report is **released** only upon joint signoff by all committee members and by the strongly autonomous Report Review Committee, which is customarily chaired by an elected official of the NAS and which oversees an often complex review by many knowledgeable individuals, none of whom were involved in the study.

*Independence:* Sponsors of NRC studies are told that they can expect no role in the placement of studies within the institution, no role in the choice of committee members, will be excluded from the committee's executive sessions and its drafting process, and cannot expect to know the study's results until shortly before its public release. Other outside interested and affected groups are also held to these restrictions. The fact that ultimate control rests with the three autonomous honorific societies also contributes to the NRC's perceived independence. The NRC will conduct industry funding for studies, providing that at least matching sums are added. The NRC's dependence on future contracts from sponsors is seen as potentially compromising its independence.

*Uniqueness:* [1] The NRC is by far the largest, oldest, and complex of the organizations we are considering. [2] The NRC is also perhaps the most tightly linked to the academic research community. [3] The NRC has the broadest substantive domain of our six groups, and, with the possible exception of economics, effectively covers the full range of expertise in science and technology.

*Some Questions:* [1] Does the NRC's regard for its reputation for credibility prevent it from adopting novel approaches and ideas in its studies? [2] Is the highly-articulated NRC study process compatible with sponsors' need for timely and inexpensive work product?

## **2 The Office of Technology Assessment [OTA; 1972-1995]**

*URL:* See legacy site at <http://www.wws.princeton.edu/~ota/>

*History:* OTA was established in October 1972 for the special purpose of informing Congress on the impacts of emerging technologies. It was originally expected to do "technology assessment," narrowly defined. One motivation was that legislators felt increasingly "outmanned and outgunned" by executive agencies with large technical staffs. About 5 years later, some key resignations led to some rethinking of OTA procedures, and a new phase began with the

appointment of Russell Peterson [1978] and then Jack Gibbons [1979] as OTA directors. By 1995, OTA had a professional staff of 140, an annual budget of over \$20 million, and was producing about 50 reports per year. Congress eliminated OTA as of September 1995.

*Setting/Leadership:* OTA was governed by its Technology Assessment Board, comprising a carefully balanced set of 12 legislators drawn from both houses and both parties. The Board appointed an outside Technology Assessment Advisory Council, which acted essentially as a visiting committee, and which was a blend of academic, corporate, and government members.

*Modus Operandi:* Typical OTA assessments took 18-30 months and on the order of \$300,000 - \$600,000 in direct costs. Each project involved an advisory committee that was chosen so as to cover both the substantive range and the major parties at interest in the study subject. A strong professional staff of three to five persons was required to pay close attention to guidance from the advisory committee, which met two or three times and played its major role in reviewing draft material [reports were usually 200 to 300 pages in length]. However, staff were not required to comply with reviewer suggestions. It was common for OTA to retain outside contractors to provide major assistance to a study. OTA studies were barred from including self-labeled recommendations, instead often providing one or more preferred "options". Some at OTA saw OTA report-based *testimony* as perhaps more influential than its written *reports*.

*Key Decisions:* **New projects** required explicit TAB approval, although disapproval was very rare. Staff tried to routinely respond to bipartisan interest by a sponsoring committee of Congress in the process of defining new projects. **Appointments**—the choice of staff, selection of contractor, and assembly of a roster of advisors—were reportedly most frequently handled at the level of the cognizant Assistant Director. **Release** of an OTA report required formal TAB approval.

*Independence:* OTA managed to produce reports without formally sequestering itself from political or other pressures. How this was accomplished merits some discussion.

*Uniqueness:* [1] OTA reports were not, in precise terms, consensus studies, in that the permanent staff had sole final authority for the work. However, the risk of an advisory committee member either resigning or complaining publicly did present inducement to accommodate reasoned views [if not others] in re-drafting. [2] In contrast with NRC and NIH, OTA frequently relied on corporate and university contractors in the study process, with staff then playing a synthesizing role. [3] OTA's sole clientele, ostensibly, was the Congress.

*Some Questions:* [1] How have congressional committees compensated for the loss of OTA: have CRS and GAO adapted in order to fill the gap? [2] Could

anyone but Jack Gibbons have found a way for OTA to succeed in producing good studies in such an over-charged environment, and with so few protections against opposition? [3] Should other assessors make fuller use of corporate contractors, as OTA sometimes did?

### **3 CLEAN AIR SCIENTIFIC ADVISORY COMMITTEE [CASAC]**

*URL:* none found – try SAB’s site at <http://www.epa.gov/science1/index.html>

*History:* CASAC was established under the Clean Air Act Amendments of 1977, which charged it to review the criteria documents for “criteria” air pollutants, to advise the EPA Administrator on research needs, and perform other functions. While housed under EPA’s Science Advisory Board, CASAC has operated somewhat independently, reporting directly to the Administrator and to Congress. CASAC’s most visible role has been in the review of comprehensive Criteria Documents for specific pollutants—notably stratospheric ozone and particulate matter—that are prepared by EPA staff. CASAC works through special panels assembled for each Criteria Document.

*Setting:* CASAC operates under the Federal Advisory Committee Act, which specifies many details of its structure and procedures. CASAC is one of about 30 FACA committees at EPA, and one of eight that are required by statute. The annual costs of running CASAC have been reported as in the range of \$2 million to \$3 million.

*Leadership:* CASAC comprises seven members. Its current Chairperson, Dr. Joe Mauderly, is at the Lovelace Respiratory Research Institute in Albuquerque. Of the other six members, five are at academic institutions and one is with a state agency.

*Modus Operandi:* CASAC meets roughly six times a year. Its work is regularly assigned to its panels – for example, its Ozone Panel included six members and ten paid consultants. A majority of these are from universities or independent research labs. All CASAC committee meetings fall under FACA requirements, which call for open meetings, advanced notice of each session, and the presence of an EPA employee who serves as the Designated Federal Official.

*Key Decisions:* As with all FACA committees, ultimate formal authority for CASAC rests with EPA leaders via the designated official. CASAC **membership** is formally determined by the EPA Administrator and the EOP, and must comply with FACA balance requirements. Many of CASAC’s **new projects** are initiated by EPA when it moves to review existing standards for the six criteria pollutants – nominally every five years for each pollutant. In each case, CASAC’s elaborate review role and schedule are keyed to the agency’s deadlines for action. **Release** of CASAC reports is decided by CASAC itself.

*Independence:* While CASAC is formally dependent on political appointees at

EPA with respect to appointments, budget, and staffing, it is perceived to have established a reputation for autonomy in reaching its conclusions, and there have been times when its views and those of EPA leadership were not congruent. The fact that CASAC is housed in the SAB, and is at arm's length from EPA's program offices, is likely a factor in its culture of independence.

*Uniqueness:* [1] CASAC has clear statutory authority under the Clean Air Act for much of its work. [2] CASAC conducts much of its work in the open, as is specified under FACA. [3] At one point, CASAC panelists conducted open voting on matters before the Committee, something that is rare in science advice.

*Some Questions:* [1] Few FACA committees appear to have the relative independence and scientific standing of CASAC. How has CASAC managed this? Does it stand as a model for other science-intensive regulatory programs? [2] Some observers note that CASAC has used many members with prior experience with NRC studies: is that a factor in its traditions? [3] Given CASAC's track record on assessments of particulate matter pollution, why has the monitoring of EPA's PM research been given to the NRC, and not CASAC, and why was HEI, rather than CASAC, asked to review the results of the "Six Cities" study?

#### **4 H. John Heinz III Center for Science, Economics, and the Environment [the Heinz Center]**

*URL:* <http://www.heinzctr.org/>

*History:* In late 1995 the Vera I. Heinz Endowment provided seed money to supply an endowment and operating funds for a new center established in the name of the late H. John Heinz, Senator from Pennsylvania. The Center launched three programs, in environmental reporting, global change, and sustainable oceans, coasts, and waterways. Its first work products are now beginning to appear.

*Setting:* The Heinz Center stresses a "four-sector" approach – it seeks to systematically involve the four major sectors of the environmental policy community [private industry, environmental groups, government, and academia] in its programs. One of the Center's stated objectives is to "create and disseminate non-partisan policy options for solving environmental problems." Thus, it sees itself not just as a forum but a venue in which, for the first time, all the sectors can collaborate in reaching new outcomes.

*Leadership:* A 12-person Board of Trustees provides guidance for the Center. It is chaired by John Sawhill, and its membership manifestly includes about equal representation from the four sectors. William Merrell is the President of the Heinz Center. The Center has a staff of about 15 persons.

*Modus Operandi:* Center officials stress that the Center is not bound to any one

project template, and that it intends to try a variety of approaches to see what works best. At this point, there is a tendency to compose a steering committee for each approved project, and, increasingly, for that steering committee to carefully balance the perspectives of the four sectors. The Center's meetings are typically open to all, including the press. All approved projects have at least one outside sponsor, and some have a consortium of funders. It may be noted that key Center personnel have extensive experience at NRC and OTA, and both the culture and processes of these organizations may emerge at the Center.

*Key Decisions:* **New projects** are approved by the President, with advice elicited from Board members. Steering committee **members are selected** with input from the Board. **Release** of reports is decided by the President, with the advice of members of the pertinent steering committee.

*Independence:* While a measure of independence is provided by the Center's endowment and by the national reputation of the Chairman, the Center seeks to function mainly by inclusion, and its perceived independence from interested groups is thus beside the point.

*Uniqueness:* [1] Academic experts have a definite voice in Center operations, but unlike most of the others in our review, they do not dominate the process. [2] At least nominally, the Center emphasized the role of economics more than the other groups.

[3] The Center relies on inclusion, rather than separation, to prevent undue influence by interested and affected groups.

*Some Questions:* [1] Will the Center be able to attain new syntheses and identify "non-partisan policy options through its intensive collaboration with interests, or will there be irresistible pressure from the advocacy group to dilute novel ideas when they emerge? [2] It is too early to assess the product and effectiveness of the Center's new approach? When will a meaningful assessment be possible?

## **5 The NIH Consensus Development Program [CDP]**

*URL:* <http://odp.od.nih.gov/consensus/about/about.htm>

*History:* Beginning in 1977, NIH has held consensus development conferences to evaluate biomedical technologies and practices and disseminated the resulting consensus statements to the public and to health professionals. Some examples: In 1997 CDP completed work on breast cancer screening for women aged 40 to 50, on acupuncture, on treatment of heroin addiction, and on three other subjects. A total of 120 consensus statements have been issued over the years.

*Setting/Leadership:* The CDP is located in the Office of Medical Applications of Research [OMAR] in the Office of the Director of NIH. In general, consensus conferences are organized and conducted with close cooperation between

OMAR [which has a total staff of 7 persons] and the pertinent NIH Institute.

*Modus Operandi:* CDP's MO is highly routinized. NIH selects panels of "nonfederal, nonadvocate [U.S.] individuals knowledgeable in the field," which meet for 1.5 days of public testimony and then immediately compose [in closed session] and release their succinct consensus statements. Staffing is provided by OMAR staff and staff from the pertinent NIH institute. Panel members themselves compose the statements. From 9 to 18 months of advanced planning are typically required, and the direct costs of a statement/conference are in the \$200,000 range. The process is not subject to FACA requirements, as the results are held not to represent advice to the government.

*Key Decisions:* **Project approval** is provided by the NIH Director and OMAR staff. The 3 criteria for acceptance include public health importance, adequacy of scientific information, and need, which is often manifest in visible controversy or debate. After this approval is granted, OMAR works with the approved panel chair to **select panelists**, aiming for a broad representation among different sectors, including researchers, health care providers, methodologists, and a public representative. Researchers who have published on the exact subject topic are excluded from the panels out of concern that their presence would signal that NIH had prejudged the outcome. **Public release** of a consensus statement is approved by the review panel itself, but is promised during the conference.

*Independence:* The absence of advocates from the panels helps to isolate the panel from partisan pressures. Isolation from interested policy officials is partial: the NIH institutes are somewhat removed from control [OMAR being in the Director's office], but institute staff participates in planning and is normally present in drafting sessions. Perceptions of undue influence by the NIH Director seem not to have occurred [as it has for EPA panels].

*Uniqueness:* [1] The CDP process can be much faster than our other exemplars. Panelists accept the normally strict condition that a statement must be written in real time during the consensus conference. [2] Perhaps driven by the need for immediate closure, OMAR works very hard to define a discrete list of concrete questions for the panels to address. The charges to other consensus groups are, typically, more vague. [3] Unlike other consensus programs, CDP relies on experts from "adjacent fields," rather than on those who have worked on the problem directly. [4] Consensus statements are terse, compared to study reports from counterparts like the NRC and OTA.

*Some Questions:* [1] The CDP reports having held only one consensus conference since late 1998, and the program itself has recently been subjected to internal review. Does the medical community now sustain its confidence in the program? [2] Have the results of consensus conferences affected clinical guidelines over time? [3] What are the pro's and con's of more systematically

covering medical costs in the consensus conferences? [4] What process makes sense for ensuring the currency of past consensus statements?

## 6 Health Effects Institute [HEI]

*URL:* <http://www.healtheffects.org/>

*History:* HEI was established in 1980 “to provide high-quality, impartial, and relevant science on the health effects of pollutants from motor vehicles.” A non-profit corporation, HEI resulted from the conviction at the highest levels within EPA and the automobile industry that a new institution was needed to instill mutual trust where little existed. One HEI representative explained that “by removing science from the adversarial process, one can both improve the science and clarify the basis of conflict.”

While most of HEI’s resources have been devoted to extramural research, it has also conducted a number of “special reports” [averaging roughly one report per year] that are essentially knowledge assessments. Examples of Special Studies include past reviews of the literature on mortality from PM and the effects of magnetic fields, and the current review of the Six Cities data.

*Setting/Leadership:* HEI receives most of its funding from EPA and the motor vehicle industry, traditionally in matching allotments of \$3 million a year. HEI has a prestigious Board of Directors chaired by Archibald Cox and including former Stanford President Donald Kennedy, former EPA Administrator Douglas Costle, and former NAE President Robert White. [There are no corporate officials or environmental advocates on the Board.] HEI also features two standing high-level scientific committees, one to develop the research program and one to “evaluate and interpret” HEI-funded work.

*Modus Operandi:* HEI’s Special Studies generally follow the same process as its new laboratory research projects. An extramural team is chosen to conduct the review, and HEI’s review committee prepares commentary that stands as HEI’s organizational position on the subject.

*Key Decisions:* **New projects** and **study personnel** are decided by the Health Research Committee, whose nine members [Bernard Goldstein is chairman] are scientists at universities or independent labs. **Public release** is governed by the ten-member Health Review Committee, comprising mostly university scientists. [For Special Studies, the Board may add other reviewers]. HEI keeps the Review Committee strictly separated from its Research Committee.

*Independence:* HEI tries hard to achieve independence in reality and in perception. Funders have had a carefully limited role in the development of specific program plans, and individuals with contracts at EPA or industry are

generally avoided. The prestige of the Board and the two standing committees reinforce HEI's reputation for integrity. The special arrangements for *de novo* review of draft reports also contribute.

*Uniqueness:* [1] More than the other organizations we are assessing, HEI limits the direct participation of individuals with partisan views in its deliberations, including participants from environmental groups and corporations. [2] Similarly, its insistence on maintaining arm's-length separation from its sponsors led some in EPA to question whether HEI's work has centered on academic interests to the detriment of policy relevance.

*Some Questions:* [1] Is it significant that the number of Special Studies has not increased significantly over the years in which HEI has established its reputation for scientific quality? [2] Recently, perhaps in part because of the recommendations of an outside panel, HEI has begun to engage sponsors and partisans—especially in its formal long-term planning process. Has this experience affected HEI's credibility? Its perceived policy-relevance? [3] Over the years, HEI has contemplated expanding beyond the field of mobile-source pollution; is it a good idea to capitalize on its institutional reputation and enter areas where it has less depth of scientific expertise?

L. McCray 3/23/00

## APPENDIX 2

### Knowledge Assessments: The Views of Federal Users

[**Note:** this appendix summarizes user opinions collected prior to the 2000 Workshop.]

**1. On Organizational Credibility.** Each person was asked to comment on the relative credibility of the six organizations in our Workshop sample, and on the relative credibility of this set of organizations, as contrasted with the credibility of for-profit contractors, university centers, etc.

No respondent chose to explore this question. One said "*all four are mostly credible,*" and related that to their explicitly setting balance as a goal and their separation between funding source and those doing the work. Another placed the NRC in a large group of "more credible" institutions [alongside specific for-profit contractors], placed OTA in a large group of "less credible" organizations [alongside trade groups and advocacy groups], and offered no views on the others.

**2. On the Overall Need for Better Assessments.** Almost all said, or implied, that there is capacious room for improvement in the supply and use of credible knowledge assessments.

One person observed that, despite the growing technical complexity of regulation, *“it is notable that these organizations do not represent a growth sector. OTA is gone, the NRC is not growing fast, and the proliferation of prestigious knowledge assessors is minimal.”*

One person said that the most important need is for broader reviews of policy, and not for the narrow review of particular risks. He thinks that an assessor’s freedom to reconsider the question it is asked, and possibly to re-frame it, is an important feature, and one that helps ensure that assessments escape any constraints knowingly or unknowingly applied in forming a study charter.

**3. On the Nature of Federal Demand.** Several respondents encourage us to consider, realistically, the nature of demand as well as supply in thinking about advice to agencies.

One points out that only a few political appointees really have an interest in getting the best analysis. One said, *“Policy officials generally don’t give a fig about knowledge. They care about politics. That’s how they get to be public officials.”* Another said, *“Appointees are action-oriented, not analysis-oriented. They are concerned about their legacies, and they just don’t want to be remembered as the guy who studied our problems.”* Similarly, one person said that tight deadlines and the pressures for short-term success militate against a culture of careful learning about issues.

One person said, *“Is credibility really important? Look at how agencies really spend their money for analysis. They do in-house work, they line up for-profit contractors . . . The money they spend on the groups you’re calling the Credibility Specialists accounts for, what, less than 1 percent of what they spend? Why is that, if they really put any premium on the broad credibility of their decisions?”*

One respondent noted that the demand for independent analysis may actually be declining now, in that environmental programs themselves seem relatively stagnant. One person said he is afraid that the current EPA administration values analysis less than its predecessors, and another offered the same view of the Executive Office of the President.

#### **4. On the Limitations of Knowledge Assessors.**

- **Feeding Hand.** Several people, without prompting, pointed to what might be called the “Feeding Hand” bias in assessor organizations. As a user of information, one said, *“I wanted objective work, but I was always afraid that the organization’s dependence on its sponsors for future work would affect the product.”* Another said, *“Academics, as well as for-profit contractors, have huge conflicts. They usually depend on government*

*funding, which is known to dry up if one takes a position contrary to an interested government agency.” However, another noted, “I can’t really say that groups should ignore the practical issue of continuing support; some would say that OTA’s fatal problem was that it didn’t pay enough attention.”*

- **Policy Relevance.** There was concern, mostly associated with “the scientific organizations,” that assessors do not retain a clear focus on the pertinent policy questions. Said one, “*Policy-makers are overwhelmed with information. The real need is for better focused studies that are framed in a way that is compatible with the political framing of the issue.*” He added, “*My sense is that the studies that go wrong usually do so because the Committee got lost in details that in the end were unimportant to the actual policy choices.*”
- **Multidisciplinary Staffing.** One observer said that knowledge assessments typically require comfortable familiarity with more than one discipline, and that our educational institutions don’t usually produce such individuals. This leads to unevenness in the assessments.

## 5. On the National Research Council:

- **Credibility.** Considering the relative credibility of the named organizations, one respondent said, “*You know, it really depends. Some are most credible with some advocacy groups, some with others. I’d say that the NRC is most credible to the media, but I’m not sure I can generalize beyond that.*” Another, addressing the NRC’s limitations, said, “*All of these undermine the NAS’ scientific credibility, though loss of political credibility hasn’t occurred. Yet.*”
- **Costs.** The NRC isn’t seen as a generic brand. One said, “*For many projects, it is far too slow and expensive – and the ‘study committee’ format is not well-suited to the project.*” Another said: “*The world’s most expensive consulting firm. Some of the best dinners I have ever enjoyed have been at the NRC.*”
- **Unevenness.** Contrasting OTA and NRC, one commenter said, “*OTA has an advantage over the NRC in that panel members are not expected to do the writing of the report . . . . You need a staff that is both knowledgeable and responsive. This is only sometimes true at the NRC. Often Academy staff are either not willing, too busy, or too lacking in background to contribute much.*” Another said, “*Performance is highly variable depending on the panel composition [especially the chair], the staff, and the charge. The NRC does best when the charge stays away from policy*

*disputes.”*

**6. On OTA.** Comments on OTA were largely confined to the Office’s general reputation for quality, and there is noticeable diversity on that matter. One commenter said, *“Some people say that NRC is better than OTA was, but I’m not so sure.”* A second observer said, *“I was very impressed with OTA. They asked the right questions, and they didn’t force consensus.”* A third said, *“OTA placed great authority in the hands of staff, but on the two occasions I served there, the staff ‘authors’ were terrific—both knowledgeable and responsive.”*

At the other extreme, however, the appearance of political stewardship [despite OTA’s obvious safeguards against partisanship] seemed to color views of OTA. One observer said, *“In the 1970’s I thought that OTA studies were slow, but good. Later, they seemed to be politically-oriented.”* Another said, *“With rare exception, OTA’s performance was highly partisan and not too subtle about it. OTA was used effectively for these partisan purposes. That’s why it is dead. I have found OTA to be useful only as a guide to what the next political crusades will be about.”*

**7. On the Heinz Center.** Because it is such a young organization, few seem to have formed a clear picture of it. One individual, however, said, *“Yes, now I remember. It seemed like an interesting concept. Shouldn’t we have heard something from it by now?”*

Another comment: *“What I liked about the Heinz Center design was the relative emphasis on involving economics in the work. But from its website, I wonder if they’re really doing that.”*

**8. On CASAC/SAB.** Few commenters are familiar with CASAC. One was, however: *“CASAC’s credibility exceeds its performance. The SAB process is inherently dysfunctional. Its charge is to ascertain whether an EPA document is true: this is not an appropriate task for peer review. The process is dominated EPA staff, both by EPA’s design and by the SAB’s acquiescence. EPA frequently forces CASAC to do its job under unrealistic time constraints.”*