

Assessing Policy Outcomes: Social and Political Biases

Mark Bovens and Paul 't Hart

In the polis, then, problem definition is never simply a matter of defining goals and measuring our distance from them. It is rather the strategic representation of situations.

Deborah Stone¹

Policy Failures as Artifacts

What prompts people to view the outcomes of certain policy episodes as failures or disasters, while playing down others, which may be perceived by some to be equally harmful or debatable? Why do many people (and policy analysts) regard President Ford's massive inoculation program against what he was told might be a new and serious swine flu epidemic as a "fiasco?" One could also view it with good reason as nothing more than an example of a somewhat overzealous "caring government." Why, for many years, was the Sydney Opera House regarded as a classic planning disaster when much larger budget escalations occurring some years later during the construction of Australia's New Parliament House failed to tarnish that project's image as an astounding success?

To pose these questions already suggests the answer: judgments about the failure or success of public policies or programs are highly malleable. Failure

Source: *Understanding Policy Fiascoes* (New Brunswick, NJ: Transaction Publishers, 1996), pp. 21–51.

is not inherent in policy events themselves. “Failure” is a judgment about events.² The thing about cases such as the swine flu inoculation program and the Sydney Opera House is that people often disagree in their judgments about them. Whether certain policy episodes are judged to be major failures strongly depends on who are doing the judging, what yardsticks they use, when they do it, and on the basis of what information. In chapter 1, we have tried to facilitate the task of analysts and commentators by making a distinction between four “layers” of meaning-making involved in the notion of policy fiascoes: The assesment of outcomes, the identification of agency, the explanation of agents’ behavior, and the evaluation of agents’ behavior. At first sight the importance of normative judgment gradually increases with each step or layer. The assessment of outcomes in terms of damage seems a rather technical affair, in which the core of the debate is about accounting procedures, the sophistication of standards, and the methodology of registration and measurement. In contrast, the evaluation of behavior, dealing explicitly with the question of who is to blame, seems a matter of full-fledged value judgments.

Unfortunately perhaps, things are not that simple. In fact, the very first layer in our conceptual scheme, the assessment of the outcomes of a certain policy episode, presupposes many value judgments. The very decision of what effects to register, when to register them, and by what means reflects the analyst’s implicit or explicit biases and strongly predisposes his or her final evaluation of these policies. This may sound like a cliché, and perhaps it is, but the implications of this cliché for the debate about highly controversial policy episodes can easily be underestimated. The problem is not merely one of research methodology; it is predominantly a problem of conceptualization and value selection. Standards of evaluation insensitive to, for example, time, institutional context, and local culture are not likely to come to be regarded as authoritative yardsticks by all stakeholders in the debate about a controversial policy episode. Transnational discussions of policy fiascoes will easily turn into *dialogues des sourds* if cultural differences in the ways people interpret social and political events are not taken into account. In this chapter, we shall first discuss three major types of biases that stem from the social context in which policy events are inevitably embedded and evaluated: biases of time, space, and culture.³ The second part of the chapter will deal with another group of assessment biases, that is, those that originate in the political context in which policy evaluation inevitably takes place.

Policy Horizons: The Time Bias

In assessing policy outcomes, where one stands often depends upon when one looks, and with what kind of temporal perspective in mind.⁴ Time is said to heal many wounds, and it can certainly alter the perceptions of the outcome of a public policy or project. A case in point is the construction of the

Sydney Opera House (1954–1973).⁵ During the conflict-ridden and traumatic implementation phase of this highly adventurous architectural project, it was considered a major fiasco. Construction took ten years longer than initially planned and the costs exploded from the 1954 tender of 7 million A\$ to well over 100 million A\$ upon completion in 1973. Significantly, the architect had walked out in midcourse following a series of confrontations with an increasingly skeptical client, the minister of public works whose party had won the New South Wales election which that year was virtually dominated by the Opera House issue. Yet this image of failure was not to last. During the late seventies and the eighties the Opera House became a major tourist attraction and symbol of Australian prowess. The original budget overruns came to be viewed in a different light. It was stressed more forcefully that most of the money for the building had not come from the public purse at all, but from a series of public lotteries. To complicate things further, in the early nineties yet another revision of judgments has been emerging as major maintenance and repair work to the building are required, the costs of which may turn out to exceed those of the original construction. Ironically, the very architectonic sophistication and innovativeness in the design and use of special forms and materials that made the building's reputation once again may compromise it. Moreover, this time it is probably not going to be public lotteries that will pay for the bills.

A prime source of analytical biases therefore involves the varieties of possible time horizons and the registration of the various effects policies have over time. The objectives of policies may vary in terms of their temporal scope (in economic policy planning, a differentiation between short-term, medium-term, and long-term policies is quite common) and temporal quality (unique/non-recurrent versus permanent/iterative policies). This affects the timing and nature of assessments about their effects when implemented. Policymakers are in fact continuously vacillating between different time horizons in setting priorities, allocating budgets, and making decisions. At the same time, many elected officials and others subject to the vagaries of the electoral cycle will be predisposed to judge policy proposals or feedback about past policies first and foremost in terms of their short-term political implications.⁶

Short-term effects are also more easy to register than long-term effects, which are likely to become intertwined with other phenomena in complex and often unintended ways. Moreover, short-term and long-term effects may in some cases be at odds with one another, the latter reversing or neutralizing the former. In general, the longer the time frame used for the assessment of policy outcomes the bigger the scope for controversy about their meaning and evaluation is likely to be. Similarly, policies aiming at nonrecurrent outcomes (such as the construction of a building, the liberation of hostages, or the policing of a soccer match) tend to be more easy to monitor than policies with iterative objectives that are constantly being renegotiated and adapted by different participants and in the face of changing circumstances (for example

maintaining the territorial integrity of Israel, France, or the Netherlands, the production and distribution of affordable energy, and the maintenance of public health).

For many policies an exclusive reliance on short-term and quantifiable indicators to measure their impact is likely to result in negative evaluations: “[In] evaluating efforts to significantly change the behavior of large numbers of people, a limited time frame is inappropriate because it neglects both the severity of the initial administrative problems and the possibility of learning by doing.”⁷ For example, Roosevelt’s resettlement program for black agricultural laborers failed to meet its short-term political objectives, yet had the latent effect of generating a black middle class which later would become the backbone of the civil rights movement.⁸ Also, when judged from a short-run perspective, government reorganization programs are almost always judged to be failures. Yet a considerably better achievement record appears when analysts adopt a long-run perspective more capable of detecting and appreciating the latent functions and long-term effects of these reorganizations.⁹ Similarly, an emphasis on short-term goals of policies leads people to regard as failures programs whose implementation has met with delays, even though these delays can be perfectly reasonable or hardly damaging to the long-term impact of these programs. Talking about a job creation program in Oakland targeted at the black youths and in part aimed to prevent a much-feared recurrence of ghetto riots, Pressman and Wildavsky observe that “[In] the light of this high standard – the creation of jobs for black people now – delay constitutes failure. But others may reasonably argue the maxim ‘Better late than never.’”¹⁰

In addition, the intended and actual lifetime of a policy or program may come to diverge.¹¹ It often happens that policies aimed at short-term effects are continued beyond their originally term of operation. This may have various reasons. Implementation may be slow, clients may have become used to them, or the policy in question may have become embraced by powerful institutional interests.¹² In any case, it requires analysts to go beyond the original mission statements of policymakers and continue to monitor the policy’s effects. Ingram and Mann give the example of the American energy policy, which shows yet another way in which time horizons can considerably change the evaluation of outcomes. In many respects the American energy policy was very successful in the 1960s. Through price controls, allocation schemes and the noninclusion of external costs, consumers were provided with inexpensive petroleum products. But seen from a longer-term perspective, the picture becomes less sanguine:

These benefits created incentives to rely on the automobile for transportation, and oil and natural gas for heating, while ignoring mass transit and coal. The success of one policy has now led to the realisation of its harmful consequences: a nation shackled to oil and natural gas and unprepared to pay the real costs that such dependence demands, i.e. subservience to foreign producers and the costs they impose.¹³

This is also an instance of what Wildavsky calls the “Paradox of Time”: past successes lead to future failures. To illustrate it, he provides the example of the ironies of achieving success in public health care which come to haunt policymakers a decade or so later. The essay’s title reflects the sense of despair policymakers may feel when they understand the paradox of time. It is called “doing better and feeling worse”:

As life expectancy increases and as formerly disabling diseases are conquered, medicine is faced with an older population whose disabilities are more difficult to defeat. The cost of cure is higher, both because the easier ills have already been dealt with, and because the patients to be treated are older. Each increment of knowledge is harder won; each improvement in health is more expensive. Thus time converts one decade’s achievements into the next decade’s dilemmas.¹⁴

In many of the cases we studied, time-related problems of evaluation manifested themselves also in another major way: a discrepancy in the time perspective of key policymakers and the dominant time frame of the people and institutions evaluating their actions. One example can be found in the case of the so-called RSV affair in the Netherlands, which reached its political peak in 1983–84. It evolved around the failure of an extensive, and in many people’s opinion excessive, flow of government subsidies to the RSV shipbuilding conglomerate. The aim was to keep this vital part of Dutch manufacturing industry alive in the face of ever increasing international competition. The “blank check” thus granted by the government to what turned out to be an internally weak, badly mismanaged set of firms which it had previously had pressured into merging, became the subject of the first formal parliamentary inquiry held in the Netherlands since the end of the Second World War. This produced a dramatic and televised spectacle in which successive ministers for economic affairs and senior civil servants from that ministry were called to testify before a special parliamentary committee.

In this context, it is interesting to note that the overriding consideration on the part of the Dutch minister for economic affairs and his key departmental advisors in continuing to allocate subsidies to RSV seems to have been their desire to counter the short-term threat of a possible RSV bankruptcy. This would have meant instant unemployment for several thousand workers. The center-right coalition government at the time (1977–81) was politically committed to bringing unemployment down. The considerable increase in the number of unemployed resulting from RSV’s collapse (which in itself would be of considerable symbolic significance) would be a major blow to its prestige. At the same time, the social-democratic opposition in parliament was also pressing for the subsidies to RSV because it was committed to saving the jobs of its working-class electorate. At that point in time, taking the long-term view and making a cold-blooded analysis of the prospects for long-term survival

of the Dutch shipbuilding industry was politically inopportune and perhaps cognitively infeasible to key players on both ends of the political spectrum. Yet when the chickens came home to roost and RSV went bankrupt nevertheless, parliament responded in 1983 by the dramatic gesture of holding a special inquiry. The special parliamentary committee ended up severely criticizing the policy of the ministry for ignoring long-term issues and failing to take more stringent measures to improve RSV's capability to continue without further public subsidies. Furthermore the minister for economic affairs came under attack for withholding parliament crucial information about the true extent of RSV's problems.

There was undoubtedly a degree of hypocrisy or at least inconsistency involved in parliament's rapid turnabout with respect to RSV and the government's policy of subsidizing troubled industries. Its ultimate assessment was that the policy of granting what had amounted over the years to no less than 2.5 billion guilders (1 billion U.S. \$, at the time) in subsidies to a badly managed conglomerate in an industry experiencing structural decline was misguided. But when the case is put in its historic context one is able to see the prominent role parliament itself played subscribing to the same kind of time frame that drove government policy. For many years, parliament itself had taken a short-term perspective and supported subsidies to save jobs at RSV. Only when that policy was collapsing did parliament take a broader time perspective. From that moment on, the frames of reference of government and parliament started to diverge.

In other words, what happened was that parliament's standards for judging the outcomes of policy gradually altered as the process unfolded.¹⁵ At the time the RSV subsidies were investigated the political agenda had changed considerably. The new cabinet had introduced a very strict financial policy which led to substantial cuts in subsidies, welfare benefits, and salaries. Virtues like frugality, efficiency, and restraint had risen to prominence in the political debates and the idea of full employment slowly had been given up. The credibility of the new budgetary policy would have been damaged severely if the allegations about financial waste had gone unscrutinized. Many of the major actors in the RSV case thus found to their unpleasant surprise that they were judged in the early 1980s by different standards than those they had acted upon in the late 1970s.

Collective mood changes thus affect not only the relative success of particular problem definitions and policy alternatives, but also the kinds of standards people use to evaluate the conduct of public officials and agencies. It has been observed, for example that "all scandals are to some extent time-bound: what shocks in one era produces barely a ripple in another."¹⁶ Apparently, policymakers who fail to grasp the current and future dynamics of "the public mood" about what is and what should be the substance and form of public affairs, are likely to run into serious problems.

Policy Scope: The Spatial Bias

Temporal complexities in policy evaluation are matched by spatial ones. Here the key problem is one of delineating the geographical or social system boundaries in determining and assessing the outcomes of public policies. The spatial distribution of the costs and benefits of a policy or project generates controversy and highly “localized” judgments about its success or failure. This spatial dimension is particularly acute in the design and evaluation of regional economic policies and large-scale public projects. The very decision by a central government body to select a particular community as a special economic zone or as a target for sustained central government or supranational economic support may in itself be a source of considerable controversy. The same goes for the designation of sites for major public projects, such as the construction of ports, railways, roads, plants, museums, and government buildings. In some cases, the source of conflict is competition for the expected benefits, but in many the spatial distribution of risk and harm is at stake.

A good example concerns the siting of high-tech facilities with negative side-effects, such as airports (noise and pollution) and nuclear power stations (waste disposal and disaster potential). Conflicts over the siting of the new airports in the London and Paris regions discussed in chapter 1 look the form of community administrators and citizens protesting against regional or central government airport construction plans on or near their territories. In the French case of Charles de Gaulle Airport, such local-level protests were quashed by a powerful central government, and the airport was built. In the London region, local-level veto power in combination with other factors succeeded in getting the construction plans shelved. From a spatial point of view, multiple evaluations of these different cases are possible, depending most literally on “where one sits.”¹⁷

A more complex case presents itself when a national policy with international implications has to be evaluated. A case in point is the Japanese procrastination in revaluating the yen when the Bretton Woods international monetary system was on the verge of collapse and Japanese dollar holdings and economic strength were skyrocketing (1969–71).¹⁸ Japanese policymakers delayed revaluation for two main reasons. Firstly, accustomed to years of postwar economic hardship and tight-belled monetary policies, Japanese politicians and Ministry of Finance bureaucrats were slow to publicly recognize that the Japanese economic resurgence of the late sixties was structural and not ephemeral. Secondly, they were afraid that any rumor of an impending Japanese revaluation to save the ailing U.S. dollar and the international monetary system based on it, would encourage runaway speculation and severely hurt the Japanese monetary position and the national economy. Makeshift policies to steer some kind of middle course were failing and the crisis atmosphere, both within Japan and internationally, was mounting. When U.S. president Nixon – frustrated with the Japanese attitude and seeing no

other way out to save the dollar – announced his decision to let the dollar float on August 15, 1971, the Japanese were taken by surprise. In a crisis atmosphere prompted by the rapid devaluation of the value of the massive Japanese dollar holdings, it was eventually decided to float the yen as well. Japan's central bank sustained a considerable loss as a result of this policy, but the policymakers' fears of severe economic consequences of the yen's de facto revaluation never materialized. After the initial turbulences caused by the rapid developments, the international monetary system entered into a new stage without too many difficulties. However, Japan's international monetary reputation had suffered dearly as a result of the crisis and of the general perception that Japanese policy had been both self-seeking and ill-guided.

How to judge this episode? Again, location, or in this case perhaps more appropriately level of aggregation, is quite important. First of all, it is extremely difficult to pinpoint exactly who suffered from what in this case. Partly as a result of Japanese intransigence, the international monetary system went through a highly turbulent and traumatic period from 1969 to late 1971. Internationally, many individuals, organizations, and countries lost out when Nixon devaluated and a flurry of speculation followed his announcement. But who these people and groups were and how great their losses were is almost impossible to tell. Taking a national perspective, the Japanese economy itself became highly volatile because of the continued speculations and the apparent inability of the government to put and end to them. Speculation and signs of panic took a certain toll for which the government's policy can be held directly responsible. At the same time, however, the policymakers' strategy was directed to prevent what they saw as a worst-case outcome: a restabilization of the Bretton-Woods system at the price of a Japanese recession caused by the revaluation of the yen. The assessment of the policy outcomes in this case strongly depends on whether one takes a Japanese or a global perspective.

Policy Cultures: The Ethnocentrism Bias

The third category of social biases in assessing negative events as policy outcomes originates in cultural frames and the social expectations about government conduct that come with these frames. What classifies as a negative policy outcome and as a policy fiasco is inextricably embedded in the social "milieus" in which particular events occur, and the standards of accountability they espouse. As observed by students of political scandal, "what is a political scandal in a liberal democracy is often business as usual in other political systems."¹⁹ For that reason, a certain element of ethnocentrism is always looming in debates about policy fiascoes, particularly when comparisons between different types of episodes are made in forming or legitimating assessments. In fact, cultural biases can easily turn a comparative conference on policy fiascoes into a scholarly Tower of Babel. These cultural biases operate both between but also within nations.

First of all, the assessment of the extent or nature of the failure of a project or policy can be biased by differences in national policy cultures. The very concept of policy or organizational failure is culturally loaded. Even within the contemporary Western liberal democracies, significant cross-cultural variations can be detected in people's judgments about the performance of their rulers. In shorthand: British "quiet desperation," Belgian mockery and cynicism, Dutch Calvinism, or American puritanism for example, make for quite different bases of judging policy events.

An illuminating example is the Rainbow Warrior affair. In 1985, the French secret service, in an attempt to keep Greenpeace from protesting against the nuclear tests at Mururoa, bombed and sank its flagship, the Rainbow Warrior, in the harbor of Auckland, New Zealand, thereby killing a Dutch photographer.²⁰ If one compares the evaluations of this action in the three countries, important differences become visible. In Holland and New Zealand, the action was considered a major failure in moral terms. In both countries, but with particular ferocity in New Zealand, the illegal and illegitimate nature of the operation were stressed. In France, on the contrary, the action was also considered a failure, but for quite different reasons. The legitimacy of the operation was hardly questioned, but many observers and politicians felt embarrassed about the awkward, incompetent way in which the operation had been carried out, leading to its exposure and the resultant international outrage that hurt France's national interests. Nevertheless, the responsible minister Hernu whose position was compromised under international pressure and who there upon resigned was wildly cheered at the conference of his ruling Socialist Party for having stood up for the *force de frappe*.

Secondly, cultural frames can strongly bias the assessment of outcomes even within one political system. Anthropologists tell us that cultural pluralism is a feature of all social situations. Different cultural biases may therefore also manifest themselves within the context of a superficially homogenous organizational or political culture. In fact, this cultural prism works both ways; what people identify as policy failure or success will also say something about the way they see the world and the institutions they live with.²¹

This more subtle shape of the ethnocentrism trap can be illustrated with the help of cultural theory as developed by Thompson, Ellis, and Wildavsky, using the work of Mary Douglas.²² They emphasize that in every community of some size elements of at least four different cultural ideal types can be found. At the core of this cultural theory lies the grid-group typology developed by Douglas. The group dimension refers to members' experience of the group as tightly versus loosely bounded. High-group cultures thus have strong boundaries, endure through time, and have internal, group-based forms of structuring. The grid dimension refers to the number and salience of the kinds of rules and social categories people use to control behavior. Combining the two dimensions yields four ideal types: hierarchy (high group, high grid), individualist (low, high); egalitarianism (high, low); fatalism (low, low).

Various authors have applied this cultural theory framework to the domain of policy debates about technology and the environment, and institutional reactions to technological failures and accidents. They have argued persuasively that people intuitively select the risks they face, the dangers they are concerned about, and the criteria by which they judge government performance in dealing with these risks.²³ Schwarz and Thompson formulate it as follows:

Our policy-analytic approach is premised on the idea of a small number of competing and culturally-dependent selection biases. Each distinct cultural orientation is seen to involve an appropriate way of selecting and vindicating how a policy issue is defined, what options and consequences are taken into account, and which evaluative criteria are to be seen as credible.²⁴

Cultural theory can be applied usefully to illustrate how people with different sets of values are likely to call different kinds of events “failures,” and how they tend to view and evaluate a single set of events in different ways.

In evaluating public policies, individualists will tend to focus on economy and efficiency. They will approve of whatever policy they feel contributes to the efficient allocation of societal resources. In many cases, this would be policies that minimize the extent of government intervention in social affairs, but individualists would also applaud policy initiatives that entail efficient and competitive public-service delivery. As a consequence, when asked to identify policy failures, they are most likely to point to cases of inappropriate government intervention, waste of resources, and bureaucratism (red tape, inflexibility).

In contrast, people inclined towards hierarchy, emphasizing the importance of detailed rules for achieving social order, will be particularly keen to ensure government operates in accordance with legal norms and organizational rules. The hierarchs tend to focus less on fundamental questions about the proper ends (and limits) of governance, as on the proper discharge of politically validated duties. They are troubled most by cases where officials deviate from accepted standards of political and administrative conduct, and compromise the public trust vested in them and their offices. When called to identify failures, hierarchs will be likely to focus primarily on instances of unlawfulness, insubordination, and other forms of disruption of bureaucratic routine and predictability.

Egalitarians will eschew both the archetypes of big government and market-oriented government and will instead seek a just government, one that protects the collectivity against the multiple and universal risks that threaten it, and offers equal outcomes to all. They will find failure most of all in cases where government fails to live up to its protective role, as in technological accidents, economic breakdowns, and natural disasters, as well as in cases where government policies condone or contribute to extreme inequalities.

Fatalists, finally, are the odd people out. Their preoccupation with fate and luck will lead them to consider the question of judging success and failure of governance largely as irrelevant. For them policy fiascoes are just another unalterable fact of life.

It should be noted that, to some extent, one cultural cluster's failures are the other's successes. This is particularly so for egalitarians versus individualists. For example, the swine flu program of the Ford administration discussed earlier would presumably trouble individualists most. They would regard it as an example of overzealous government protectiveness, implemented in a heavy-handed and overly bureaucratized way. They would have wanted to wait with any measures until the existence of a serious epidemic could be confirmed with greater certainty and then take swift action in response. Hierarchs would probably support the spirit of the program but would be troubled by the bureaucratic politics that surrounded it, the politicization of scientific expertise it entailed, and the fact that the president's decisions – misguided or not – were not implemented more effectively. Egalitarians would probably be supportive of the program's objectives, which reflected a sense of public responsibility to protect the population nationwide against danger from illness. If anything, they would be concerned about the dominant role of epidemiological experts in the development of the program.

It follows from this differentiation in cultural types evaluating the success and failure of public policies, that the specific normative "label" which a particular policy episode eventually acquires in the political system, may depend on the particular configuration of cultural biases in that system. Whether, to what extent, and for what reasons a policy episode becomes viewed as a failure is a matter of a political process of attribution and blaming in which cultural biases compete and sometimes amalgamate.

This brings us to the second category of biases that may affect the ways in which we look at and assess policy outcomes. Assessing the outcomes of a policy episode in terms of failure or success does not only reflect temporal, spatial, or cultural perspectives, it is also, and perhaps predominantly, a political matter. Labelling policy outcomes as negative events is never done in a vacuum. To understand why key institutions and forces in the political system come to view a policy episode as a failure, one needs to take into account the political dynamics that give rise to these judgments and the political context in which they are made.²⁵ Public policymaking is a form of political action, conducted by office-holders, be they elected legislators or appointed administrators, operating in the context of large-scale, complex bureaucracies, in interaction with other public and private interest groups. In these political and institutional arenas, stakeholders with different interests and perspectives meet and interact. Conflict, whether overt or implied, is part and parcel of the politics of policymaking. Especially when a policy episode has generated a certain amount of controversy along the way, the evaluation process tends to take the form of a struggle about what is to become the

dominant interpretation and assessment of the policy's results. This produces a context for policy evaluation that harbors not only social but also distinctly political sources of bias.

Public Perception Biases

Peter Hall, in his book on what he calls "great planning disasters," used public perception as a measuring rod in the selection of his case studies.²⁶ He considered a planning disaster "any planning process that is perceived by many people to have gone wrong" (later he adds that a great disaster is one that "cost a lot of money by almost anyone's standards").²⁷ This is simple and straightforward, but in the end equivocal criterion. Public perception, which is a direct function of the extent and nature of media coverage of certain policy episodes, can be a highly partisan guide. First of all, some negative effects of programs or policies may never make it to the public agenda at all because widely accepted ideological premises prevent their recognition, because appropriate conceptual categories to identify them are lacking, or because dominant political coalitions have an interest in framing the issue in a different way. For example, in many countries the negative effects of certain agricultural, policies and major infrastructural projects on the environment and wildlife have only recently come to be noticed as a legitimate cause for concern.

An example of this is the reinforcement of the river dikes in the Netherlands in the eighties and nineties. This is a mammoth project of fortifying and heightening hundreds of miles of river dikes to guarantee maximum security from flooding. It was initiated by the powerful national agency for waterworks on the basis of its new flood protection standards. At first, the project was perceived as a politically unproblematic and primarily technical issue – the long tradition of fighting floods and claiming land in this country that partly lies below sea level has virtually assured popular support for any kind of waterworks. It nevertheless became clear after some years and many miles of new dikes, that the safety-first approach of the agency resulted in the complete destruction of the landscape, fauna, and culture that had developed around the existing small, dwindling river dikes. Many ancient houses, sometimes entire townships, and many habitats for rare plants and animals were replaced, or are about to be replaced, by huge concrete dikes coming straight off the safety designers' drawing boards, not allowing for any vegetation or dwellings. These cultural and environmental effects had never been taken into account in the decision making and evaluation processes. The key government agency, Rijkswaterstaat, a stronghold of engineers within the Dutch ministry of transport with a multimillion dollar budget, as well as regional actors in the political arena had simply framed the issue, often probably even unconsciously, as one of flood protection full stop. The design

process was therefore governed only by the trade-off between the desire for appropriate safety and the financial costs of achieving that. Only recently, and not without slow but increasingly fierce political infighting, have local and environmental groups managed to swing parts of public opinion towards the idea that other legitimate values and concerns are at stake too. If one takes these into consideration, these groups have shown, the current mode of implementing the anti-flooding policy produces catastrophic results. As a result of this maneuvering, a compulsory environmental audit has now been included into the decisionmaking and planning procedures for those stretches that have not been reinforced yet.

Political maneuvering and competing evaluation frames do not remain confined to cases of unintended effects and hidden tragedies. Even those “fiascoes” that from the first instance indeed are “perceived by many people to have gone wrong” can be the product of political expediencies or media distortions. In fact, many public and political scandals do not originate in failures of policymaking as defined in this book. Many of them have little to do with public policymaking but evolve around the faux pas and indiscretions of individual politicians and bureaucrats. A recent major reference compendium on political scandals since 1945 is filled to a large extent with cases of extramarital affairs, espionage, obscure deaths, and many instances of corruption.²⁸ Individual tragedies usually make for better newspaper reading than the technical and intricate details of policy processes that might have gone astray.²⁹ Media attention in general is often biased towards the spectacular cases, preferably those that can be personalized.

Moreover, the accounts of those policymaking episodes that do reach the front pages and the nine o'clock news are often strongly biased. Much public attention depends on how these events become labelled in the press and by political opinion leaders, who, in turn, are influenced by vested interests and political expediencies. In other words, to paraphrase Edelman, a policy failure, like all news developments, is a creation of the language used to depict it; its identification is a political act, not a recognition of a fact.³⁰

Consequently, whether an episode of problematic policymaking appears on the political agenda and turns into an “affair” does not only depend on the nature and extent of the damage done, but more predominantly on the political situation in which several actors might find it opportune to label a situation as a failure and expose it to the press.³¹ Scandalogy, that is, research into the genesis of “affairs” and scandals, is therefore first of all a matter for rhetoricians and political sociologists.³² The danger here is that policy analysts and other observers unwittingly allow their initial and subsequent assessment of policies to be biased by media-amplified criticisms and condemnations of policymakers voiced by politically interested parties and actors. By the same token, political mythologies about seemingly successful policies – for example about the “Swedish model” of running an economically strong welfare state or

the “Japanese model” of industrial harmony and efficiency – may desensitize observers and analysts to their drawbacks and darker sides.

Political Failures versus Program Failures

Important in securing political support for policy programs is the degree to which policymakers succeed in making the program look good. They must seek to provide different political stakeholders with such information, and present the information in such a way as to secure their support. This means that the political fate of a policy or program depends not only on its substance but also upon its appearance. Political judgments about the success or failure of a policy are mostly made on the basis of impressions about the effects and costs of the program, that is, predominantly on images and symbols.³³ These images might, but need not reflect the program’s substance and operational reality as experienced by those directly involved with it. To put it bluntly: when a policy is judged to be a failure, this may reflect a failure of public relations as much as a failure to deliver the goods.

An example of this is U.S. postwar unemployment policy. Its image was distorted by its interference with preexisting and ongoing political games, institutional arrangements, and policy initiatives – in particular those that arose from the War on Poverty. Both policies were developed and administered by a myriad of public agencies, and were subject to complex political maneuverings within and between the White House and Congress. The main result was that both policies suffered from a lack of coherence, which resulted in a lack of coordination between various initiatives. In some cases, different programs were administered simultaneously, working at cross-purposes. In other cases, the multitude of agencies and programs competing for political support and budgetary appropriations led to a shortage of “critical mass” necessary to be able to achieve the original ends. Due to institutional fragmentation and the vagaries of the political process, these social programs intended to be sweeping in scope and speed began to “incrementalize” and ultimately generated disenchantment.³⁴ The sense of urgency which prompted the unemployment policy got lost in the legislative and administrative morass, as did its coherence. But does this mean it should be regarded as a failure?

Mucciaroni, who analyzed this policy episode, argues it is necessary to distinguish between political failures and program failures. A program is a political failure when it lacks political support and momentum necessary for its survival as a prioritized area of government activity. A program failure occurs when the program fails to have an impact on target populations, or produces unintended and unwanted effects. Hence programs that do have a certain positive social impact may still be labelled failures by many observers using a predominantly political success/failure criterion. Many of them will look at the political and administrative process rather than at the substantive

outcomes of a policy in forming their impressions of it, even though the two might not necessarily be closely related:

[T]he reality and perception of failure have had more to do with program design, administration, and service delivery than they have with the substantive impacts on the targeted populations. Thus, performance is important in understanding the political failure of employment policy, but performance mainly in the sense of how employment policy was conducted rather than what it accomplished.³⁵

Mucciaroni then explains that the perception of failure was mainly focused on the many adjustments in the program's design and implementation that were required to keep it afloat in a very complex and conflict-ridden social and institutional environment:

Policy has shifted from social to economic goals, from training to job creation, from efforts to assist the long-term unemployed living in the nation's cities to the cyclically unemployed and those displaced by structural economic change. Administrative arrangements have alternated from centralization, to decentralization, to a reassertion of centralization, before giving way most recently to another attempt at decentralization. Funding has risen and fallen less in response to any policy rationale than in reaction to economic crises and shifting political winds.³⁶

It then becomes ironic to observe that the same political actors and institutions which have contributed to these displacement effects and continuous adjustments of means and ends, have come to judge the policy a failure precisely for that reason! It seems the roots of the problem lie as much within the political process itself as within the program and its achievements. To be sure, employment policy's achievement record was not splendid. But it was certainly not as bad as the widespread perception of it as a failure seemed to suggest:

More than any other domestic program of the 1970's, CETA [the Comprehensive Employment and Training Act, as key program designed to integrate and administer the various components of employment policy, MB/PtH] came to symbolize the exhaustion of the liberal agenda and the corruption of its activities. . . [But] the CETA story is one not so much about the failure of a program as it is about a set of political institutions. It should not be confused with the notion that CETA failed to achieve, to one degree or another, several of its substantive objectives. While few of its programs were undisputed and overwhelming successes, few were abject failures. The gap between actual results and the dismal perceptions of performance suggests that the place to look for the sources of CETA's difficulties is not in its statistical evaluations but in its political-institutional framework.³⁷

The underlying cause of CETA's problems appears to lie in the fact that in the crisis-ridden atmosphere of the 1973–74 economic downturn, it became a kind of garbage can for all kinds of short-term initiatives and programs addressing various of the baffling problems that Washington policymakers suddenly saw themselves confronted with. Lacking an adequate understanding of emerging phenomena such as stagflation, policymakers went into a burst of activity and kept piling new programs on top of CETA's already overpopulated landscape.

Mucciaroni's distinction between political and program failures enables one to see further shades and biases in the assessment of outcomes of policies and programs. A failure in political terms does not necessarily stand for a failure in terms of the program's objectives (even though political failures may erode the political support for a program and may thus in the long run cause a program to fail). Vice versa, failures to reach substantive objectives need not necessarily result in political failures.

The Political Paradox of Achievement

Judgments of success and failure are directly related to the expectations people use to judge programs, policymakers, and institutions of governance. There is a case to be made that in the first three decades since the Second World War, societal expectations about governance and its benefits have increased sharply in many Western countries. In many cases, governments have actively promoted these rising ambitions, perhaps without realizing that they also entailed increasingly high and stringent standards for judging their actual performance. From the mid-sixties onwards, many governments were confronted with the effects of the growing discrepancies between societal demands and expectations on the one hand, and the limits of progressive government on the other. Political observers of many ideological persuasions have coined this development in different ways, for example as the crisis of the welfare state, government overload, or as a legitimation crisis. Seen from this perspective, modern government is destined to "fail" in many of its reformist and redistributive activities:

One result of the social activism of the 1960s was that Americans grew disillusioned with government. As criticism of social conditions rose, so did the level of government activity; but regardless of how much government had achieved in the way of social improvement, it had led the public to expect more. The more government tried to do, the plainer became its inability to do very much, or at least all that its leaders said it would. Both the right and the left attacked it for failing to fulfill its promises.³⁸

Inevitably, this general disenchantment with government shapes people's judgments about specific policies or programs. Cynicism about politics and bureaucracy – tellingly defined by one of its most prominent students as a

type of organization unable to learn from its mistakes – among citizens and policy evaluators alike does little to offset this.³⁹ This combination of the rise of government activity and a growing disenchantment with it, constitutes what one might call a “paradox of achievement.” This paradox is, again, best captured in Wildavsky’s observation about the medical sector: “doing better, feeling worse.”⁴⁰

To make things worse, government agencies in many areas of public policy are faced with conflicting demands. Patients, medical professionals, and other lobbies demand sophisticated and costly health programs which cover many illnesses and defects that hitherto had been considered a stroke of bad luck. At the same time and sometimes even by the same persons, the government in general is expected to balance and preferably to reduce its budget. Similarly, governments currently are pressed to reduce acid rain and other forms of environmental pollution. Yet simultaneously, in a different arena, other powerful interests argue against a levy on energy or any other burdens that might reduce the competitiveness of national industries or increase the costs of living. As policy spaces become denser, “as more and more policies aim to solve more and more difficult problems, the consequences become harder to predict and the chances for failure increase.”⁴¹

Thus progressive ideals and/or rationalist standards concerning government planning and policymaking can only produce predictably gloomy evaluations of overwhelmingly incrementalist planning and policy realities. The day-to-day reality of public policymaking is one in which the objectives of programs are continuously modified to accommodate the limitations of the available means and the impact of multiple social constraints that become apparent during program implementation.⁴² In the political world of public policymaking problems are not so much solved as they are elaborated, redefined, and superseded.⁴³ Why then do so many observers and analysts continue to adhere to standards that are clearly beyond the grasp of even the most sophisticated and powerful policymakers? At one level, this may be a function of a professional role conception as critics of the powers that be. But as we have tried to show, many of these biases may go unrecognized by all concerned. Part of the explanation then can also be found in the following paradox: The logic of democratic electoral politics appears to dictate that policymakers themselves are in many ways actively involved in raising expectations and espousing myths of rational, just, and omnipotent governance that help to create the conditions for their own political failure.

Goal-Based Biases

One way to circumvent at least part of the paradox of achievement is by simply evaluating government by its self-confessed standards, namely the official goals of policies and programs as contained in policy documents and

other authoritative written or verbal statements of purpose. Using official goals as standards not only allows for a convenient anchor for assessing policy effectiveness and efficiency, it can also be viewed as a form of democratic control. It is a way of taking government at its word.⁴⁴ Accordingly, goal-based assessments of program effectiveness have constituted an important if not the dominant current of public policy evaluation studies. Of these, the great majority were critical, and revealed many discrepancies between policy-as-planned and policy-as-implemented.⁴⁵

In response to these findings, questions have been raised about their significance: do they tell us that most of the time government does not work as it should even by its own standards, or do they reflect the inherent limitations of this type of methodology of outcome assessment? This has led to a debate about the strengths and weaknesses of goal-based versus goal-free evaluation.⁴⁶ In this respect, an important issue concerns the linkages between goals and values. Goals and goal-related evaluation criteria are always based on underlying social and political values. In many societies a plurality of value systems exists (one might call them ideologies or cultural biases), which could make for quite different expectations and specifications of the goals of public policies. Which values, and perhaps more acutely, whose values, in fact come to guide the official goals of a policy, depends partly on the differential appeal of the various value sets to individuals and groups in society, and partly on the structure and process of political power in that society. Should this process of selective value articulation be taken for granted by fiasco researchers, or should it in itself be a topic for evaluation and reflection? This touches upon one of the most difficult normative issues in contemporary policy analysis.⁴⁷ In addition, the observer should be aware that the particular set of values that drives the official policy goals need not be the same as those dominating the post-hoc societal and political assessment of the outcomes of the policy, as we have seen earlier in this chapter.

From a political perspective on public polity, several other well known and important limitations of goal-based evaluation stand out. The common denominator among them is that, for various reasons inherent in the complex organizational and political nature of contemporary governance, official goals have little of the normative or precriptive strength imputed to them by most goal-based evaluation researchers: "the goals of policy are often not what they seem to be, and it is a mistake to take stated purposes too literally."⁴⁸ It follows that official goals cannot easily be translated into a reliable and relevant standard to base evaluations on. Analysts who overlook the complex and ambiguous character of goal-based policy evaluation fall victim to the "official objectives bias," and produce analytically coherent but politically naive and bureaucratically irrelevant assessments of past policies. The following complications should be taken into account.

First, policy goals are often vague. They allow for multiple interpretations and, consequently, for different operational standards of evaluation. As

observed by Stone: “The inescapable ambiguity of political goals means that they are more like moving targets than fixed standards.”⁴⁹ From a political perspective, such ambiguity makes sense. It dampens controversy and allows various internal and external stakeholders the flexibility to present favorable interpretations of what the official policy is going to be to their different constituencies, while at the same time providing a semblance of coherence in government policy. From the perspective of assessing policy outcomes, however, ambiguous goals are not a facility but a problem.

Second, seen from the perspective of technical feasibility, policy goals often appear to be set unrealistically high or low. Agency myths and highly ambitious policy statements often do not take into account the costs and limitations of implementation. But this in itself is no reason to fault them, because their main rationale tends to be not substantive but symbolic. They often fulfill the strategic function of mobilizing broad support and the symbolic function of providing a vision for the future, projecting a leader’s commitment, or indicating the general thrust of the government’s direction.⁵⁰ Sometimes laws are enacted and policies endorsed which clearly will never be able to reach their stated objectives. But this does not turn them into instant failures. They can be meant to convey a message to particular groups and interests, to take the pressure off a particular issue, to delineate boundaries, or to instigate a change of social attitudes. Anti-discrimination policies, equal protection clauses, and environmental laws often fall into this category. In contrast, excessively modest goals are usually formulated in a more defensive spirit: in anticipation of political criticism or to prevent making commitments that may come to haunt the policymakers. In both cases the stated goals in all probability do not reflect the actual policies of the government.

Third, policymakers pursue many types of goals, either simultaneously, or, as revealed by Cyert and March’s studies of strategic decision making, sequentially.⁵¹ But the various goals and subgoals they pursue may well be at odds with one another. Classic examples include the police or the corrections system. The police are continuously confronted with competing demands on their limited resources and conflicting policy imperatives:

- which types of crimes to “go after” in force and which ones to treat in a low-key manner;
- enforcing the law or keeping the peace;
- trying to prevent crime and trying to control it;
- working to bring down actual levels of victimization or seeking to reduce the public’s fear of becoming victimized (the two turn out to be at best only indirectly related);
- focusing on crime alone or taking a more broad perspective of community service.

Selecting a profile implies trading off some concerns against others.⁵² Perhaps even more acutely problematic, managers of a corrections system need both to incarcerate and rehabilitate convicted and often seasoned criminals. This they have to do amidst growing public concerns about crime and safety, disenchantment with the performance of the criminal justice system, and competing pressure groups for and against liberal corrections philosophies.⁵³ Task structures like these harbor inherent goal conflicts. Moreover, different stakeholders within these systems may pursue different (sets of) goals. As a consequence, the formulation of agency objectives and program principles involves a major political and bureaucratic balancing act, which is bound to be vulnerable to political if not substantive failure.⁵⁴

A fourth and final major problem with goal-based evaluation is that it assumes that policy goals remain stable throughout the policy process. This assumption is inaccurate at best. A number of detailed studies of decision making and policy implementation show that goals are highly dynamic properties. In fact, there is much to be said for a view of public policymaking as a continuous process of adjustment in which goals are discovered, specified, modified, and subsequently abandoned in favor of new ones. The more sophisticated empirical theories of the policy process all share a view of goals as dynamic rather than static.⁵⁵ For this reason, these theorists argue against any rigid prior specification of criteria for policy effectiveness and related specification of evaluation procedures.⁵⁶ Brewer and DeLeon argue that programs can rarely be developed with complete foreknowledge of their eventual outcomes and effects. To maintain this fiction and to impose “neatly prespecified criteria” may, in fact, be counterproductive and act as a deterrent to learning and good policy. They quote March:

The prior specification of criteria and the prior specification of evaluational procedures that depend on such criteria are common presumptions in contemporary social policymaking. They are presumptions that inhibit the serendipitous discovery of new criteria. Experience should be used explicitly as an occasion for evaluating our values as well as our actions.⁵⁷

All of this complicates the task of goal-based evaluation in various ways. The official goals may be a tenuous political and bureaucratic compromise among policymakers faced with conflicting yet politically equally inevitable tasks and claims. In many cases, the official goals constitute a tactical façade for hidden goals that cannot be articulated openly, for the very fact of their publication may diminish their chances of being attained. Also, substantive goals may be less important than symbolic ones; the same goes for long-term and unspecified versus short-term and explicit goals.

The Anglo-French Concorde project constitutes one example of how the acknowledgement that policymakers may have been striving after long-term, latent goals even at immense short-term costs, alters the over-all assessment

of a policy episode. In 1962, the British and French governments signed a treaty for the joint design, development, and manufacture of a commercially viable supersonic airliner, the Concorde. This official objective was never fully reached. The Concorde was taken into production, but at great cost to the British and French taxpayer. It is currently being operated only halfheartedly by two government-backed airlines, British Airways and Air France. No other airlines have bought Concorde and the planes have not been allowed to fly outside of very narrow, transatlantic routes. Many airlines and airport authorities have been reluctant to embrace Concorde because of the noise produced by its flying at great speed. Many studies of the Concorde project have emphasized these points and have consequently condemned it as a monumental failure.

And indeed, in a narrow sense, the project has failed. A number of key objectives internal to the project have not been met. However, the assessment of the Concorde project begins to change for the better if one takes into account that both the French and the British governments adopted the project pursuing not only the commercial goals of the project itself, but also, and perhaps much more strongly so, a number of indirect, strategic political goals.⁵⁸ The French wished to reestablish and modernize their ailing aircraft industry and needed access to the state-of-the-art technology that was being successfully developed and exploited by the then superior British aircraft industry. For their part, the British were perfectly happy to accommodate the French in this respect even though it meant sharing a risky joint project and, in time, helping one's own future competitors. They were after something far greater: proving their willingness and ability to be a reliable partner in transnational ventures in order to get the French to lift their veto against a British entry to the European Community. Seen from this broader angle, it can be argued that both governments managed to accomplish these latent goals by initiating and continuing against the odds with the Concorde project.

The point here is not to argue that goal-based evaluation is useless and should be abandoned. Rather, we have sought to illustrate that goal-based outcome assessment cannot be taken for granted as an unproblematic, neutral, and static methodology for policy evaluation. Using it to evaluate complex and controversial policy episodes is insufficient in itself to properly negotiate the cognitive and value complexities inherent in the task. We have pointed out that underneath the seemingly simple procedure of comparing outcomes to preestablished goals, there remain multiple ways of identifying and weighing official policy goals, each of which may generate quite different assessments. Just as with any other method of outcome assessment and policy evaluation, goal-based analysis needs to be contextualized in order to be useful. Failure to put the results of goal-based assessments into a proper institutional and contingent perspective implies a serious underestimation of the degree to which the selective articulation of policy goals is an integral part of the politics of policymaking.

Action-Oriented Biases

In labelling a particular policy episode as a failure, investigators or evaluation researchers may fail to reflect on what outcomes might have resulted had other courses of action been followed. This constitutes what might be called the “neglected alternatives bias.” It is a bias because by failing to assess – even if this can usually only be done counterfactually – the consequences of alternative options, the analyst may not fully appreciate and discount the ambiguities and dilemmas the policymakers were facing at the time they had to make a decisions. The analyst may be quite pervasive in arguing that the policy that was pursued had highly negative effects. But in this he or she may have failed to consider fairly the possibility that any other course of action than the one actually selected by policymakers would have resulted in consequences that were even worse. Or, to recapitulate terminology used in the previous chapter, the analyst evaluates the present without reference to a number of other possible presents. As we will argue in chapter 4, in some cases an argument can be made that policymakers found themselves in a situation of “tragic choice” in which either of the choices they had would result in a significant amount of social harm. If this is the case, and even though the policy outcomes were highly negative, it would be misleading to label these episodes in terms of failure.

Secondly, a related phenomenon is what one might call the “action onus bias.” This bias comprises two tendencies: one entailing selective attention and the other selective judgment. The dimension of selective attention refers to the tendency for analysts to focus on policies that imply proactive intervention and expansion of governmental activities, and to pay less attention to policies of nonintervention and contraction.⁵⁹ Activities seem to be much more visible than nonactivities or contraction-based policies. Who ever seriously evaluates cutbacks and terminations but opposition parties? Which policy analysts challenge the predominantly financial and economic rationales of deficit reduction from which they derive both their impetus and legitimation? These policies are, in essence, chiefly inward-oriented (putting the government household back in order), and it is suggested that they should be evaluated in these same inward-looking terms. By doing so, the indirect external, social consequences of these cutbacks and terminations are often overlooked. The policies are couched in technical jargon and presented with an air of necessity that few tend to take issue with. Contrast this with the amount of critical analytical attention paid to interventionist infrastructural, social, and environmental programs. These programs are outward looking. They are designed to effect social changes and their potential successes and failures in doing so generally stand out much more clearly. The social, political, and ecological rationales underlying them tend to be a far better and less “technical” subject for analytical scrutiny as well as public and political debate.

The dimension of selective judgment in the action onus bias stems from cultural predispositions of many people in Western societies to value action as opposed to abstinence. Policymakers who try but fail tend to be judged less harshly than policymakers who are seen to have failed to try. In other words, errors of omission – though they are less frequently noted by the operation of selective attention – tend to count for more than errors of commission. Perhaps anticipating this bias, U.S. health officials confronted with scanty information about a possible new swine flu epidemia were keen to pursue a nation-wide inoculation program as soon as possible. They feared that if they would decide to adopt a conservative policy of “sit and wait” now, this could be construed as negligence if the epidemic would assume major proportions later.⁶⁰

A nice illustration of this general phenomenon of differential action-oriented valuation comes from Feldman and Milch’s study of airport construction projects. It is worth quoting them in full:

There are two distinct levels at which different outcomes, however temporary they may be, must be assessed. Decision makers everywhere wanted to build. According to their criteria, authorities in Dallas-Fort Worth, Paris and Montreal were successful, whereas their counterparts elsewhere failed to achieve their objectives.

The adoption of these criteria reflected a commitment to the fulfillment of strategies. The strategy chosen everywhere for solving the problem of anticipated air travel demand was the construction of civil aviation infrastructure. The intermediate goal of construction effectively supplanted the long-term goal of meeting demand; alternative solutions, therefore, could not be considered until and unless there would be no construction. Success was measured by building infrastructure, not by success in meeting travel demand.

The second level appropriate to a judgment of policy success involves the consequences of government accomplishment and the fulfillment, in the most efficient and just manner, of the original goal. According to these criteria, success and failure must be reversed. The long-term goal, to meet the demands of the travelling public for the rest of the twentieth century, has not been compromised in the cities where construction has not taken place.⁶¹

Their analysis shows not only the biases at work in evaluating nonevents and negative decisions, but also highlights some of the limitations of framing outcome assessments in terms of the official goals of policymakers discussed earlier. It is interesting to see how the political dynamics of evaluation produced a differential assessment of the costs and benefits of positive” versus “negative” outcomes. In those cases where the original construction plans were called off, the opportunity costs and alleged destruction of capital investments were highlighted and considered a main reason for calling these episodes failures.

At the same time, far less attention tended to be paid to the construction costs of those airports that did get built, and the considerable efficiency losses resulting from operating multiple airport facilities in one region. Upon closer inspection, it might be argued that in each of the three cases studied by Feldman and Milch where construction went ahead as planned, the negative outcomes and costs overshadowed the benefits.⁶²

Challenges for Understanding Policy Fiascoes

Important as these various types of social and political biases are in shaping judgments about policy events, the challenge they pose to understanding policy fiascoes is ambiguous. One may, for example, view these biases as distortions of what ought to be a proper, balanced assessment of social events as policy outcomes. It follows from this view that what we need is a form of policy evaluation which skirts around or transcends these biases. Implicit in this position is a role conception of policy evaluation as an arbiter: a neutral, rule-driven, expert authority able to adjudicate various competing claims about how to evaluate the outcomes of certain policy episodes. The main challenge to policy analysis then becomes to find ways to perform this role adequately, which means a need to develop systematic approaches to outcome assessment enabling the analyst to stand aloof from, but at the same time understand, the social and political context of the events that are being assessed.

At the other end of the spectrum lies the extreme relativist position. This would argue that these biases are inherent to any kind of claim about the nature of social and political reality, and that they are particularly marked in normative claims purporting to evaluate that reality against a certain set of criteria. In this view, policy evaluators, like other participants in debates about social and political events, are inevitably affected by these biases. These pervade the kinds of questions they ask, the selection of what are taken to be the relevant facts, and the choice and prioritization of norms and standards they use in evaluation. Since no one can completely divorce himself from the social environment in which knowledge is gained and judgments are formed, policy scientists and evaluators – whether they acknowledge it or not – play a role as advocates. They cannot aspire to be the kind of “legitimate value judges” that traditional science models suggests they should. The methodological apparatus of modern science can at best be used only to explicitize not transcend the social biases they bring to bear in assessing policy outcomes. In practice it is often used to obscure them, in order to preserve special status.

Underlying these different perspectives on how to interpret the challenge posed by these social and political biases in outcomes assessment are epistemological differences. The former position fits in with a positivist, the latter with an interpretivist epistemology. However, each of these epistemologies can be interpreted with different degrees of orthodoxy. This is

fortunate because neither extreme outlined above seems to offer a promising way forward. The former provides us with an ideal that fewer and fewer students of evaluation see as realistic, while the latter may lead us to wonder why we should bother at all.

As suggested in chapter 1, we find ourselves most comfortable with a modified-interpretive position which basically argues that all knowledge about social affairs is based on limited information and social constructions. It accepts that public policymaking is a particularly complex object of inquiry which at best allows only for highly contextual and contingent generalizations, let alone predictions. Policy science is conceived of here as a set of conventions designed to generate and continually reassess more or less systematic and verifiable knowledge about public policy. It is taken to be one among other forms of making sense of complex and controversial policy episodes, one that like all the others is pervaded by both epistemological assumptions and social and political values.⁶³ For that reason, policy analysis that rests on these conventions of science is not superior to other forms of sense-making.⁶⁴ Yet its single most important asset, and in it also the key to its practical relevance, lies, we believe, in its institutionalized commitment to facilitate enlightenment, reflection, and dialogue.

From this perspective on policy evaluation, the social and political biases identified in this chapter are indeed challenges. They challenge us to continually reexamine the basis of our reality claims and normative judgments. They imply that meaning-making about policy events may slip into myth-making if we forget that the initial judgments we form about events and policymakers are likely to take a narrow, contemporaneous and politically opportune time perspective. These initial judgments tend to guide the socially organized process of remembering and forgetting, that we call history-making. The biases also imply that the same set of events can be assessed very differently, and legitimately so, from different cultural, spatial, and institutional positions.

Slowly we are beginning to come to terms with these challenges. Contemporary historians, for example, are building up a tradition of revisiting momentous, controversial, and conventionally regarded as politically catastrophic policy events such as the Munich agreements of 1939 and the Suez crisis of 1956. Benefitting from hitherto classified new archival materials, they ask poignant questions about these events, including what has become the dominant political and/or historical appreciation of these episodes. Benefitting also from the passage of time, they are naturally better able to assess the events in question from a more balanced temporal perspective, differentiating more sharply between short-term and longer-term effects.⁶⁵ Benefitting also from the fact that time heals many wounds, both historians and political scientists have set about bringing together key players or scholars from opposing sides in historical confrontations such as the Cuban missile crisis, in order to get a sharper feel for the differential effects of space and

culture in evaluating these events.⁶⁶ Also, political scientists are beginning to explore the importance of temporal dynamics in constraining the performance of policymakers as well as in shaping the kinds of social expectations and values that their contemporaries are likely to employ in judging their success or failure.⁶⁷ Policy scientists are beginning to take cultural pluralism as a starting point in developing conceptual frameworks for policy evaluation.⁶⁸ And finally, policy analysis themselves become increasingly aware of the need for them to be politically astute, both in avoiding being used politically when shaping their analysis and in making sure the results of their work become politically relevant.⁶⁹ In the concluding chapter we will return to these emerging ways of coping with social and political biases in assessing policy outcomes, and try to reflect on what they may offer in the way of helping us to better understand the contextuality and the institutional dynamics of policy fiascos.

Notes

1. D.A. Stone, *Policy paradox and political reason*, Glenview: Scott, Foresman and Company 1988, p. 106.
2. C. Hyatt, L. Gottlieb, *When smart people fail*, Harmondsworth: Penguin 1988, p. 20 and pp. 31–33.
3. Compare for example D. Elazar, *The American mosaic: The impact of time, space, and culture on American politics*, Boulder: Westview 1993.
4. An important preliminary distinction in these matters is between clock time and social time: these are different entities and may diverge considerably. See R.H. Lauer, *Temporal man: the meaning and uses of social time*, New York: Praeger 1981, who observes that planning and policymaking can be viewed as attempts to bring the drifts of social time under a form of rational control – and who may use time factors and timing as important instruments of such control (pp.113–117). For an elaboration of this point, see: Y. Dror, *Policymaking under adversity*, New Brunswick: Transaction 1986, pp. 5–7, 35–37, and 174–175; also: P. A. Sabatier, What can we learn from implementation research?, in: F. X. Kaufmann, G. Majone and V. Ostrom (eds.), *Guidance, control and evaluation in the public sector*, Berlin: De Gruyter 1986, pp. 318–319.
5. J. Baume, *The Sydney Opera House affair*, Sydney: Halstead Press 1967; P. Hall, *Great planning disasters*, London: Weidenfeld and Nicolson 1980; A. Kouzmin, Building the New Parliament House: An Opera House revisited?, in: G. Flawker (ed.), *Working papers on parliament*, Canberra: Canberra Case Series in Administrative Studies, Canberra College of Advanced Education 1979; J. Yeomans, *The other Taj Mahal: What happened to the Sydney Opera House*, London: Longman 1968.
6. R.A. Hays, Perceptions of success or failure in program implementation: The “feedback loop” in public policy decisions, *Policy Studies Review*, 5, 1985, 51–67, in particular p. 57.
7. G. Majone, *Evidence, argument, and persuasion in the policy process*, New Haven: Yale 1989, p. 182.
8. L.M. Salomon, The time dimension in policy evaluation, in: *Public Policy*, 27, 1979, 121–153.
9. Compare JG. March, J.P. Olsen, *Rediscovering institutions*, New York: Free Press 1989, pp. 53–116.

10. J.L. Pressman, A. Wildavsky, *Implementation: How great expectations in Washington are dashed in Oakland*, Berkeley: University of California Press (3d ed.) 1984, p. 123.
11. A. Hoogerwerf, Policy and time: Consequences of time perspectives for the contents, processes and effects of public policies, *International Review of Administrative Sciences*, 56, 4, 1990, 671–692.
12. See, for example, H. Kaufman, *Are government organizations immortal?* Washington: Brookings Institution 1967; B. Hogwood, B. Guy Peters, *The pathology of public policy*, Oxford: Clarendon Press 1985.
13. H.M. Ingram, D. E. Mann (eds.), *Why policies succeed or fail*, Beverly Hills: Sage 1980, p. 14.
14. A Wildavsky, *Speaking truth to power: The art and craft of policy analysis*, New Brunswick: Transaction Books 1987, p. 283.
15. See also K.S. Cameron, D.A. Whetten, Organizational effectiveness: One model or several?, in: K.S. Cameron, D.A. Whetten (eds.), *Organizational effectiveness: A comparison of multiple models*, Orlando: Academic Press 1983, p. 13.
16. A. Kuntz, From Spiegel to Flick: The maturation of the West German Parteienstaat, in: A.S. Markovits, M. Silverstein (eds.), *The politics of scandal: power and process in liberal democracies*, New York: Holmes and Maier 1988, p. 151.
17. Referring to Miles's Law: "where you stand depends on where you sit", made famous by Allison as a key proposition of his paradigm of governmental politics. G.T. Allison, *Essence of decision: Explaining the Cuban missile crisis*, Boston: Little, Brown 1971.
18. R.C. Angel, *Explaining economic policy failure: Japan and the international monetary crisis of 1969–1971*, New York: Columbia UP 1991.
19. A.S. Markovits, M. Silverstein, Introduction: Power and process in liberal democracies, in: Markovits and Silverstein (eds. 1988), *op. cit.*, p. 9.
20. See R. Shears, I. Gidley, *The Rainbow Warrior*, Sydney: Sphere Books 1985.
21. Compare also March and Olsen (1989), *op. cit.*
22. M. Douglas & A. Wildavsky, *Risk and culture: An essay on the selection of environmental dangers*, Berkeley: University of California Press 1982; M. Thompson, R. Ellis, and A. Wildavsky, *Cultural theory*, Boulder: Westview Press 1990.
23. For a detailed empirical case study, see R.P. Gephart Jr., L. Steier, and T. Lawrence, Cultural rationalities in crisis sensemaking: A study of a public inquiry into a major industrial accident, *Industrial Crisis Quarterly*, 4, 1990, pp. 27–48. See also M. Schwarz, M. Thompson, *Divided we stand: Redefining technology, politics and social choice*, London: Wheatsheaf 1990.
24. Schwarz and Thompson (1990), *op. cit.*, p. 62.
25. See P. Bachrach, M.S. Baratz, *Power and poverty: Theory and practice*, New York: Oxford UP 1970.
26. Hall, (1982), *op. cit.*,
27. Hall (1982), *op. cit.*, pp. 1–2.
28. L. Allen et.al. *Political scandals and causes célèbres since 1945: A reference compendium*, Harlow: Longman 1990.
29. See M. Edelman, *Constructing the political spectacle*, Chicago: Chicago UP 1988, chapter 5.
30. Edelman (1988), *op.cit.*, p. 31.
31. On the political functions of labelling policy events as successes or crises, see M. Edelman, *Political language: Words that succeed and policies that fail*, New York: Academic Press 1977.
32. See W. Klose, *Skandal und Politik: Tin Kapitel negativer Demokratie*, Tübingen: Katzmann 1971; M. Schmitz, *Theorie und Praxis des politischen Skandals*, Frankfurt: Campus 1981; and Markovits, Silverstein (eds. 1988), *op. cit.*
33. Compare Stone (1988), *op. cit.*

34. See P. Schulman, *Large-scale policymaking*, New York: Elsevier 1980.
35. G. Mucciaroni, *The political failure of employment policy, 1945–1982*, Pittsburgh: Pittsburgh UP 1990, p. 12.
36. Mucciaroni (1990), *op. cit.*, p.12.
37. Mucciaroni (1990), *op. cit.*, p. 185.
38. M. Derthick, *New towns in town: Why a federal program failed*, Washington: Brookings Institution 1972, p. XIII.
39. The definition is from Michel Crozier, *The bureaucratic phenomenon*, Chicago: Chicago UP 1964.
40. Wildavsky (1987), *op. cit.*
41. Ingram and Mann (eds. 1980), *op. cit.*, p. 18.
42. H. van Gunsteren, *The quest for control: A critique of the rational-central rule approach*, New York: Wiley 1976.
43. Wildavsky (1987), *op. cit.*, p. 57ff.
44. A. Hoogerwerf (ed.), *Succes en falen van overheidsbeleid* [Succes and failure of public policy], Alphen: Samsom 1983.
45. Ingram, Mann (eds., 1980), *op. cit.*; Hoogerwerf (ed., 1983), *op. cit.*
46. See M. Scriven, Goal-free evaluation. In: E.R. House (ed.), *School evaluation: The politics and process*, Berkeley: McCutchan 1973, pp. 319–328. See also the highly informative discussion by W.R. Shadish, Jr., T.D. Cook, and L.C. Leviton, *Foundations of program evaluation: Theories of practice*, Newbury Park: Sage 1991, pp. 73–118.
47. See, for example, F. Fischer, *Politics, values and public policy*, Boulder: Westview 1980 and E.B. Portis, M.B. Levy (eds.), *Handbook of political theory and policy science*, Westport: Greenwood 1988. Shadish, Cook, and Leviton (1991), *op. cit.*, pp. 46–52; 455–463 discuss this issue in terms of a challenge for evaluators to develop coherent theories of valuing.
48. Ingram, Mann (eds., 1980), *op. cit.*, p. 20; See G.D. Brewer, P DeLeon, *The foundations of policy analysis*, Homewood: Dorsey Press 1983, pp. 328 ff; H. Bressers, Analyse en evaluatie van beleidseffecten [Analysis and evaluation of policy outcomes], in: H. Blommesteijn, H. Bressers, and A. Hoogerwerf (red.), *Handboek beleidsevaluatie* [Policy evaluation handbook], Alphen: Samsom 1984, specifically pp. 127–130; B.Hogwood, L. Gunn, *Policy analysis for the real world*, Oxford: Oxford UP 1984, pp. 150–169 for critiques.
49. Stone (1988), *op. cit.*, p. 195.
50. Y. Dror, Visionary political leadership: On improving a risky requisite, *International Political Science Review*, 9, 1988, pp. 7–22.
51. R.T. Cyert, J.G. March, *A behavioral theory of the firm*, Englewood Cliffs: Prentice Hall 1963.
52. M.H. Moore, Police leadership: the impossible dream? In: E.C. Hargrove and J.C. Glidewell (eds.), *Impossible jobs in public management*, Lawrence: Kansas UP 1990, pp. 72–102.
53. E.C. Hargrove, J.C. Glidewell (eds., 1990), *op. cit.*, p. 3.
54. Partly for that reason, running a police department or a corrections system have figured prominently in a comparative analysis of “impossible jobs” in public administration. E. Hargrove, J.C. Glidewell (eds. 1990), *op. cit.* Others include the management of mental health and social security systems, as well as the management of the AIDS epidemic.
55. See, for example, M.D. Cohen, J.G. March and J.P. Olsen, A garbage-can model of organizational choice, *Administrative Science Quarterly*, 17, 1972, 1–25; Pressman and Wildavsky (1984), *op. cit.* Brewer and DeLeon (1983), *op. cit.*; D. Hickson, R.J. Butler, D. Cray, GR. Mallory, D.C. Wilson, *Top decisions*, Oxford: Blackwell 1986.
56. For example, Brewer, DeLeon (1983), *op. cit.*, p. 332.

57. J.G. March, Model bias in social action, *Review of Educational Research*, February 1972, pp. 413–29, as quoted in Brewer and DeLeon (1983), *op. cit.*, p. 332.
58. See, for example, E.J. Feldman, *Concorde and dissent*, Cambridge: Cambridge UP 1985; P.W.G. Morris and G.H. Hough, *The anatomy of major projects: A study of the reality of project management*, New York: Wiley 1987, pp. 195–197.
59. See however, for example, C.C. Hood, A. Dunsire, *Cutback management in public bureaucracies*, Cambridge: Cambridge UP 1989.
60. A.M. Silverstein, *Pure politics and impure science: The Swine Flu affair*, Baltimore: Johns Hopkins UP 1981, chapter 3.
61. E.J. Feldman, J. Milch, *Technocracy versus democracy: The comparative politics of international airports*, Boston: Auburn House 1982, p. 117.
62. Feldman and Milch (1982), *op. cit.*, pp. 118–122. As the authors point out themselves, and as argued at length in chapter 2, this assessment might shift again if one extends the time frame used for evaluation. Air travel did reach new highs during the eighties, and pressures on the London, Toronto and other regions where building plans were stalled did increase, while in the other three cases, passenger increases were easily absorbed and even viewed with great relief.
63. R.A. Heineman, W.T. Bluhm, S.A. Peterson, E.N. Kearny, *The world of the policy analyst: Rationality, values, and politics*, Chatham, NJ: Chatham House 1990, pp. 1–6.
64. See, for example, D.K. Cohen, C.E. Lindblom, *Usable knowledge: Social science and social problem solving*, New Haven: Yale UP 1977; Majone (1989), *op. cit.*; R. Tong, *Ethics in public policy*, Englewood Cliffs: Prentice Hall 1986.
65. W. R. Louis, R. Owen (eds.), *Suez 1956: The crisis and its consequences*, Oxford: Clarendon Press 1989.
66. J.G. Blight, D.A. Welch, *On the brink: Americans and Soviets reexamine the Cuban missile crisis*, New York: Hill and Wang 1989; for a pointed piece of revisionism about the same crisis, which was long held to be a policy fiasco for the Soviets, see R.N. Lebow, J. Gross Stein, *We all lost the cold war*, Princeton: Princeton UP 1994.
67. S. Skowronek, *The politics presidents make: Leadership from John Adams to George Bush*, Cambridge, Mass: Belknap Press of Harvard UP 1993.
68. C.C. Hood, A public management for all seasons? *Public Administration*, 69, 1991, pp. 3–19. See also R. Hoppe, A. Peterse, *Handling frozen fire: Political culture and risk management*, Boulder: Westview Press 1993.
69. Compare, for example, Heineman, Bluhm, Peterson, and Kearny, (1990), *op. cit.*, pp. 63–64; 113–118; 173–174.