The case for a new ethic of evaluation

Jean-H. Guilmette

Background paper for remarks at the Performance Workshop
jointly organized by CIDA, IDRC and OAG

May 23 & 24 1996

In time of war, the stakes are so high that populations generally are willing to give great discretionary powers to governments and armies, allowing thus for risk taking, leading to both successes and losses. But in times of peace, populations tend to become increasingly more conservative and risk adverse\(^1\) when it concerns the usage of tax revenues. The level of acceptable risk premium may vary from society to society of course, but there is always a measure of "non rational" decision taking.

A critical question is: "how risk adverse is the general population when it judges the performance of its government, of its administration and its civil servant?" And conversely, "how deep a risk aversion runs in the bureaucracy? In the Government?" Consequently "How much of a risk premium is Parliament and the Administration ready to charge the taxpayers to avoid any possibility of errors leading to political embarrassment?" Or "how irrational has its behaviour become?"

This is not a moot point in conflict prevention or in any long term preventive activity, as it constrains the capacity of any government to deal with anticipated problems and crisis.

---

\(^1\) Definition of risk adverse and risk premium:
A group of people is offered two choices: choice one: each will receive $100. or, choice two, at the draw of a coin each will get $200 for head or nothing if its tail. Individual who choose the first alternative are defined as "risk adverse people". Subsequently, we measure the depth of risk aversion by introducing a risk premium i.e. the cost each risk adverse individual is prepared to pay to avoid risk. Each will be offered $75. vs a 50/50 chance of earning either $200 or $0.00. In this case the risk premium is valued at $25. Mathematically, the first test will yield equal results on the long term. In that sense it is equivalent; however, paying a risk premium is not an optimal choice. It is not then define as a "rational decision"; unrelated, emotional factors come to play in what should remain essentially a measurable dilemma.
In today's context, decision taking in the Federal Government is increasingly biased in favour of risk adverse decisions, and the risk premium which is considered acceptable can be very high at times. In that sense, we cannot characterize the decision making process as "rational". Fears of public criticism, when factored in, may lead to the search of "ultimate error and risk free" decisions as rare an entity as the Unicorn.

The purpose of the following reflections is to attempt to decipher the specific "system of rationality" that has come to govern Foreign Affairs and Aid decisions in the recent past. Especially as they influence the decision taking for complex problems such as, for example, conflict prevention in Africa.

I will first attempt to explain why prevention is difficult to undertake. I will show why managers refrain from taking high risk decisions after they have been subject to critical comments by evaluators and how they learn to cope in very specific manners: their selection of choices become increasingly more irrational. This is, of course, the contrary reaction expected by the evaluators whose primary intentions are to increase rationality in decision making.

Among solutions that could be envisaged, one should spend some time assessing the impact and role that evaluation plays in the build-up of such increasing risk premiums. Game analysts would suggest we move out of what is now a very bad "prisoner dilemma" toward the somewhat more efficient "rational pigs"2 A new dimension must be added to the ethical code of evaluators which would take into consideration the direct effect of the evaluator's own behaviour.

WHY?

It is generally believed that it is better to prevent than to cure. By extension, it is assumed that governments will apply such basic rules of conduct in managing State affairs. However many demands, concurrently imposed by the taxpayer, are perceived as more pressing and definite. The sum of those conflicting demands, makes it difficult to act in a preventive and efficient manner, to take the subtle decisions that are needed to prevent and to survive as a Government or as an Official.

---

2 Game and strategy aficionados will find those models of decision making well described in: John McMillan, Games, strategies & managers, Oxford university Press. London, 1992. chapter 2. For the convenience of the reader, we have photocopied the relevant pages in annex 2.
A) A typology of African conflicts

Current Foreign Policy emphasizes the need for "Diplomatic Prevention of Conflicts". This is a good case to analyse and demonstrate the very real difficulties of acting wisely in a complex situation.

Conflict prevention is generally confused with "diplomatic prevention of conflicts". An implicit assumption follows. It assumes therefore that managing conflicts is confined to the art of diplomacy. Conflict management will make use only of the tools designed for such functions. Such an approach leaves aid agencies (read development assistance institutions) boxed out of any attempt to prevent conflicts in the making. The financial resources of ODA thus becomes a mere convenience to provide increased resources and assets in the hands of diplomats and diplomatic institutions. It reduces the ability and the role of development assistance in the long term prevention of human disasters.

To help clarify the roles played by the various institutions, we have charted\(^3\) a theoretical sequence of events leading to a major catastrophe. Five phases or periods are identified and correspond to various evolutions from a "malaise" to a "war". Put simply, a typical conflict will start with a "malaise" **moment 1**, progressively evolving toward a "crisis", **moment 2**. The etymology of crisis comes from the Greek word for decision or, more specifically, "time for a decision". If no decision is taken, we move toward **moment 3**, a "conflict". As is often the case internal conflicts are denied. For obvious reasons related to the sovereignty of nations, diplomats can hardly intervene during such periods. In **moment 4**, however, the conflict degenerates into the open and usually moves across boundaries. It is then the moment when diplomatic activities may take place, in an attempt to prevent further aggravation toward a full civil war or a Genocide. Those diplomatic activities are generally complemented by Humanitarian Assistance and Peace Keeping Operations. In a symmetrical manner, when a conflict mutates into open military activities, development activities become increasingly irrelevant. Moment 5 will happen after truce and reconciliation have taken place, and time has come to reconstruct the country. ODA goes back then to its fundamental origins rooted in the Marshall and Colombo Plans.

Those definitions, as simple as they are, help clarify the different fields for research and for policy making related to "conflict prevention". Clearly Development Assistance's role is in the long term actions, those most difficult to link to a distant outcome.

What happened before a malaise grew into a war? What could have been done at that moment by each society? What, then, can Aid Agencies do to assist those groups in dealing with difficult situations? When is the time for them to act preventively? Should it be when a conflict is still localized and the separation of actors into enemy camp is not completed? When existing conflict resolving mechanisms are still at play? In the absence of models,
those are fundamentally most difficult questions. Those are however, the only real and worthwhile questions for an Aid Agency, and for its staff.

B) Ignorance is the first impediment for action

During the first decades of African independence, Aid Agencies (Multilateral, Bilateral and NGO's) undertook an extraordinary array of projects. Infrastructures were built, schools and universities were inaugurated, vaccination programs, as well as many other health and social activities, were launched etc. Some economic growth ensued, some benefits was shared, but it is generally accepted that the results are mixed when measured from an economic point of view.

But what of the impact of all those actions from the perspective of the security of African Nations? What can be truthfully concluded today? What conclusions should be drawn from the absence of civil war in countries like Mali or Ghana? Should it be argued that aid reduced the probability of conflicts where Aid was significant? Could it be that, like Monsieur Jourdain who wrote prose unknowingly, ODA help prevent conflicts when attending to its ordinary business? Or, is it true, as many analysts and journalists believe, that Aid had a neutral effect at best, and some of the time it played into the nascent conflicts and augmented them? How?

The first factor which impedes conflict-preventive action, is the general ignorance of its root causes and the subtle mechanism that have been at play for centuries. Many people hold opinions about such questions, but there is little solid evidence. The community interested in those matters has collected little data and undertaken too few researches to document the causalities at play. The first and almost only way to overcome this "knowledge gap" is to be found in the regular and dedicated attention to impact evaluations. Building up the knowledge base about what works and what doesn't, agglomerating data and analysis until some causal relationships become clearer.

The fascination for "input" evaluations, and for measuring compliance toward policies and procedures has created a single-mindedness4. This single-mindedness, reinforced by rewards and sanctions, has obscured for the managers the need to build a real knowledge base. For the time being, development doctrine remains overly dependent on ideologies and anecdotal wisdom as opposed to the rigorous, incremental pursuit of conclusions drawn from accumulated empirical evidence. The mass media prefers anecdotes to analysis as they are easier to use and can be charged with essential "entertainment value". In due time, even the most hardy,

---

4 The recent evaluation on the Rwanda Genocide was a first of its kind. The DAC members should be complimented on such an initiative and encouraged to increase efforts to gather intelligence from those collective endeavour. It should be noted however, that the Development Assistance in the years prior to 1990 is still to be assessed in term of its impact ($3,5 billions remains unnoticed...?!). It is commonly claimed that Agencies invested millions in helping along South Africa ease into democracy: did it really have such an impact as we all want to believe, or was it neutral? Rigorously speaking, the satisfaction derived from the electoral process remains unsubstantiated.
scientifically-trained managers can't resist the trend: they will start believing anecdotal evidence and disregard the search for facts.

C) Reading into the past

When dealing with new problems, an institution cannot apply existing procedures. Each crisis in Africa is "new". It has been observed by the evaluating team which assessed the donors response to the Rwanda Genocide⁵, that actors saw the crisis through the lens of the earlier Somalia crisis. What obscured the UN response in the months preceding the eruption was the "shadow" of the Somalia crisis. It is feared by the author that the International Community is again repeating the same error, as it seems to be viewing the Burundi drama as a repeat of the Rwanda situation.

This tendency to fight the last war is predominant; managers, conscious that they are being observed in their compliance to systems and procedures, will tend to ensure that whatever comes out of the previous evaluation will be replicated in the design for the following project. This is especially pronounced when moving between phases.

The evaluation process should be extra-sensitive and underscore the inherent dangers of applying "lessons learned" mechanically.

D) The Mikado strategy

Generally, Government policy analysts search for standard responses, for systems and procedures to deal in advance with predictable events. It may be true that conflicts are going to happen in Africa over the next few years, and in that sense they are "predictable events". However, their nature is very specific and their characteristics are highly fragmented. Each conflict has its system of roots, its own context and very specific rules of engagement. Various African societies respond differently to stimuli and conflict resolution mechanisms are very diversified in the way they function.

In bridge, or in chess, analysts have studied various responses that can be planned in advance depending on the various moves of the adversary. But African conflict prevention has no similarity to such games. It is maybe analogous to the Mikado game, where one drops various colored sticks on the ground and picks up as many as one can without moving any other sticks. Points vary according to the color of each stick; the winner is the one who adds most points. In such case there is no precise action that can be preplanned. A priori, no activity is more commendable than any other. The player must assess the way the sticks interlock, start with the obvious moves, and proceed from one successful action to another.

The only significant factor to consider in Mikado strategy is "relevance". To remain relevant, one needs to look at the specific situation, use experience, know-how, knowledge, rigorous and disciplined analysis, and a real measure of intuition to design and tailor an adequate and

⁵ Howard Adelman, "Early warning and conflict management", in Joint evaluation of emergency assistance to Rwanda, study 2, p.70.
appropriate course of action. However, over the years, domestic based policies have progressively reduced the area reserved for relevance. In some countries, the regulatory environment has been increased by the introduction of laws such as the environmental impact assessment, and by policies and priorities. However, nothing a priori suggests that the combination of such laws, policies and priorities constitute the mix that is required to prevent future conflicts in Africa.

E) The ambiguous role of Policy accountability

Policies deal with generalities and with generic situations. As such, policies can become the worst enemy of relevance: specific actions deal with specific problems, and compliance to policies and program priorities may not provide the answer.

This raises a major problem within risk-adverse societies. Relevance, specificity, and intuition are all concepts fraught with risks. Errors may happen. Individuals may chose actions whose appropriateness is difficult to demonstrate unless one shares the same background of experience and the same body of knowledge. The need for trial and error is more often than not perceived as a luxury. A crisis of confidence may result from abandoning standard procedures and approved.

Most analysts of government behavior, from the Auditor General to journalists and academics, wish to measure compliance to approved policies. They demand facts, figures and dollars figures, that will clearly demonstrate that the government acts precisely in line with its avowed policies. This propensity to tie governments to generic behavior and pre-set policy responses, compounded with the propensity to avoid risks, reduces the ability of aid agencies to deal early with potential conflicts.

Clearly, political correctness, which is one particular form of policy compliance, further limits the capacity to analyze and to act appropriately.

However, one can hardly miss seeing how much damage has been done to the Government machinery through the mediatization of the post facto search for "value for money". The general public has been led to believe there exists, as in bridge, an "objective optimal game strategy leading to correct decisions"; this injects into the public service a risk-minimizer mentality and an acceptance of ever increasing risk premium. This adds to costs and reduces, rather than increases, the probability of success.

Preventing conflicts or taking long-term action rests on a sophisticated and subtle analysis intertwined with risk appreciation. The Mikado analogy demonstrates the limit to the predictability of any single action or of any series of actions. As a corollary, it suggests the futility of trying to reconstruct single events and link them to any final outcome. Coaches review critically significant moments of past games to improve the technique of players. However such an evaluative learning process stops short of creating a winning scenario, as changing one single event changes the rest of the game, maybe for the better, maybe for the worse. Most significantly, the general public is never invited to those "locker room" evaluation cum training sessions.

Proclaiming to an entertainment conscious media, the existence of "value for money" may not be the best way to induce thriftiness and strategic behavior.
F) The democratization of Foreign Affairs

Until the late sixties, many analysts of International Affairs considered that Foreign Policy partially escaped the democratic process. Decisions were taken by an elite of Officials, in concert with a select group of interested parties (journalists, academics, interests groups and business). The issues were generally considered to be of little interest to the taxpayer or the electorate as the media did not dwell on those complex and generally uninteresting issues. Domestic affairs were the center of the democratic debate. The Seventies have opened up all our democracies to the World, and in the Eighties subjects that were ignored previously, attracted the attention of electronic media.

What to do about conflicts in remote countries is currently discussed openly, forcing Governments to take public stands on all matters. In the context discussed previously, this obligation to take decisions in uncertain circumstances, and to be seen as pleasing the public, limits seriously the decision making.

In the years that preceded the last election in South Africa, the Canadian Government (CIDA, External Affairs, in conjunction with Canadian interested parties, IDRC, NGO’s and many experts), created a committee to define the aid that could be extended to South Africans to foster the process of democratization and to enable the black majority to quickly acquire the knowledge and the skills associated with running a democratic state. The most common desire of the members of this group was to attract as little attention as possible from the press: they preferred to work efficiently, yet secretly, rather than reap the glory of the media attention.

There is, therefore, an apparent conflict between democracy and quality decision.

G) The Nasa dilemma

In 1986, the Challenger space shuttle, carrying six astronauts and commissions worth hundreds of millions of dollars, exploded. It was shown that the accident was predictable, therefore, avoidable. The cause of the tragedy was the O-ring, and the defect had been identified by NASA employees, but no preventive measures had been taken. The authorities decided that the launch should not be delayed. The dilemma seemed insoluble: If NASA delayed the launch to replace the defective joints, everyone would know exactly how much the delay cost, but it would be impossible to prove how much money would have been saved because the precaution would have destroyed the proof that there had been cause for worry.

Considering the above, to anticipate and abort a conflict is a process which, to the extent that it is successful, will usually obscure that proof of its own success. It becomes extremely difficult to overcome this quandary in the recently introduced notion of "result based management". Such an approach is most useful to measure "measurable activities", pardon the tautology. One needs to prove that one action leads to one precise result. In long term crisis prevention, such a clear linkage is generally impossible to demonstrate. Furthermore, other incidents may happen as a
direct and visible consequence of the preventive action, obscuring further the raison d'etre of the decision.

One can view the export of management, decision-making and control systems as transplants: the body will either accept or reject them, often after a considerable amount of struggle and suffering. Every value system carries with it an intrinsic bias, a natural drifting force. When exported and integrated into another cultural milieu, it will interact with local values. Depending on the tendencies inherent in the two systems, the exported system may amplify and exacerbate the defects of the recipient system.

One might be tempted to believe that what protected the long lasting excellence of the most famous Mayo Clinic is that for 140 years it has remained concerned with results. Follow-up practices mean that teams ask such questions as “Did our work help cure the patient?” At another level, the same team will ask: “Are current operating procedures the best available? How can they be improved?” And, thirdly, the team must analyze the tools used: “Should they be re-evaluated to produce better results, or simply to be more efficient and less costly?” This is the proper way to deal with results.

The new principle called “results-based management,” however, has made its way to Canada from the major American schools. It consists, in part, of another fashionable technique aimed at providing technocrats with jobs and prestige. Jourdain has always written prose without knowing it. Those described by Pamela Pitcher⁶ as craftsmen (and a fortiori, the visionaries) have always been extremely conscientious about attaining results. Posing the question again today may initiate a very intelligent and productive deliberation, but what we see in the end is merely the superficial and mechanical application of a new slogan. The time will come when this “transplant” is replaced by another. Measurement of results certainly constitutes an important criterion, but is not in itself sufficient to maintain excellence. An analysis of natural drift tendencies will show why.

The American system has historically done everything possible to avoid any type of discretionary power. The “contract” that binds the government and its principals, or the State and a supplier, must be free of all subjective judgment. Accordingly, the target is to establish “honest contracts.”⁷ However, this system will easily fall prey to procedural and legal drift, because the goal of objectivity will be pursued past the stage of realism; the result is a very specific manner of procrastinating. It is the American sense of risk that regularly helps correct the excesses of

---

⁶ See *Artists, Craftsmen and Technocrats* by Pamela Pitcher, HEC Press, Montreal, 1994. This remarkable study of the evolution of a great Canadian financial institution divides senior management into three groups: artists, who have a vision and build up the enterprise; craftsmen, who know how to make things, and technocrats, who deal with processes. Unfortunately, experience shows that the latter have a huge destructive capacity. More gregarious than the other two groups, they are efficient predators. They will hunt down the artists because they do not understand how they think, and enslave the craftsmen. In the end, having destroyed these two groups, they have no other choice but to destroy the company as well.

legalism and the exaggerated concern with short-term, measurable results. As market forces are allowed to operate fairly freely in the United States, bankruptcies, mergers and takeovers make it possible on a regular basis to eliminate the victims of "rampant proceduritis."

In the absence of this natural counterbalance—i.e., risk tolerance—we may see the agonizing spectacle of an institution goings round in circles, the prisoner of a logic it cannot understand, incapable of finding answers within its natural environment to the problems plaguing it. This is often what occurs in the Canadian Public Service, when consultants are used as intermediaries to directly import unadapted organizational models designed for private enterprise in the United States.

The results-based doctrine thus provides an ideal alibi for avoiding any kind of risk-taking: results must be tangible, measurable and predictable. The theory makes it possible to divide any human activity into distinct little steps, and to preserve the privilege of recanting in the face of the slightest changes in policy. New approaches and long-term commitments are thus totally marginalized, leaving the way open for only that which is familiar and capable of providing immediate gratification.

NASA’s dilemma is related to the difficulties involved in demonstrating a cause-and-effect link between preventive action and the absence of a disaster. The analysis tools provided by "results-based management", in particular its emphasis on short term and clearly identifiabe causal relationships are not well suited to this type of endeavour.

Clearly, we are deeply into the dysfunctionalities of the prisoner's dilemma. What is good for the community, or for the long run, is contradicted by what is good for the individual, or for the short term.

Thus, over the years, layer after layer of well intentioned improvements have been introduced in the management of the Aid business in Canada and in most other OECD member countries. The net result is that it is now virtually impossible to attempt to prevent conflict. In sum, ideas intended to foster a rational and accountable system may lead to irrational behavior on the part of aid managers and Government officials. This may also contravene declared policy.

The response to the NASA dilemma

The lesson learned by the auditors from the NASA crisis is that three key strategic elements had been lost over the years, thus depriving this institution, of the needed elements to take good decisions in uncertain circumstances: hands on knowledge, excellence, and leadership.

Over the years NASA, (and so it is for some aid agencies), abandoned progressively all its hands-on capacity to become instead "process managers". The knowledge acquired was passed on and desegregated to various suppliers and consultants. But nothing can replace intuition, based on
systematic usage of sound scientific traditions, and practical experience to take decisions in uncertain and untested circumstances. Furthermore, an institution needs a critical mass of people who share related and common experiences to arrive at a solid "intuitive" consensus of what is to be done.

In the 70's, following experts advice on sound management practices, NASA abandonned the recruitment of "excellent" people as it seemed to reduce its financial viability. All those "Nobel prizes" were relegated to universities and research centers. Their absence was cruelly felt, as their credibility and their acute sense of analysis was not available at the time of most critical decisions. Their high sense of self confidence and self esteem would have been essential to read through the many paradoxes and false savings typical in such complex problems. Deprived of their overpowering presence, how can NASA explain to a demanding, yet ignorant American public, the subtleties of risk assessments and of costly decisions in rocket launching?

Lastly, for each industrial sector of industry there must exist a lighthouse industry. It sets the pace and the standards, it exercises leadership, and placates the general public. Without such leadership, doubts and fears arise. In the confidence gap thus created, anybody who claims expertise may attract the attention of the media and pretend they know the true way of doing things. Bombarded by these know-it-alls, inundated by their contradictions, the general public does not know whom to trust and is suspicious of any decisions. Having lost its leadership over the Industry, NASA could not take the right decision and hope to survive.

**What can evaluators do for their part?**

To renew the current code of ethics

The following considerations are premised on the conviction that ethical behavior is measured not solely on its pure intentions, but also on the adaptation of one's action to its consequences. Once known, remedial action must replace ineffective behavior.

A new ethic of evaluation must be developed. It must define standards as well as clarify the limits and limitations of post facto judgments. It must then move outward and contribute, if not inspire, to the drafting of professional codes of ethics for all professions which deal with evaluations: I refer to auditors and evaluators, of course, but to academics and media as well when they play the role of evaluators.

Some points for consideration for the confrontation between "real-evaluation" and evaluation theory:

a) The difference between Audits and Evaluations is subtle but very real. It happens all too frequently that this border is blurred, intentionally or inadvertently, leading to disastrous effects.

The more advanced views on "performance assessment" consider that all those activities form part
of an integrated panoply of instruments useful to the managers. Endowed with a variety of good
information, no doubt that the responsible manager will take rational decisions. This doctrine finds
abundant use in small, privately owned and managed corporations. The eyes of the owner are
omniscient, and its personnel interests generally correspond to those of his firm. Profits at year end
are there to ensure integrity and lucidity in appraising business practices. Unfortunately, this does
not deal adequately with the bureaucratic environment where a profound cleavage exists between
individual interests and those of the Administration. Furthermore, a very different rapport exists
between the obligation to comply to rules, and the bottom line. In the private sector, the bottom
line generally comes first, while in the Public service, obedience is praised more highly than the
free interpretation of rules and regulations.

Audits verify compliance to rules and regulations. As such, individual accountability is at the
center of auditors preoccupation. Managers know it and expect it: when signing a document with
a clear audit trail, most managers will ensure respect for delegations of powers, compliance with
Government regulations and laws. Those are the accepted rules of the game. On the other hand,
evaluation deals with corporate accountability. As long as good faith is not challenged, evaluators
deal with questions about the state of the art, the corporate behavior, the corporation choice of
personnel, the training it provides its employees, the knowledge it renders accessible, the routine it
encourages, etc. Ultimately, evaluation measures the practices and wisdom of a ministry, of the
Government and of the Parliament.

Any ambiguity between those two contexts of performance assessment breeds instant fear, and
provokes recoil into managers. Evaluation conclusions will be doctored if individual blame may
result from its publication (internal or external). In such circumstances, evaluations will be
rigorously neutered. This will tend to compromise the evaluation process and substitute cynicism
for probity. Faced with outside criticism, scapegoats are often sacrificed: eventually the "clever
goats" build the necessary safeguard. It should come as no surprise to the sagacious observer that
objectives are fuzzy, milestones as flexible as rubber bands, and budgets increased by the additions
of redundant studies and padded with innumerable contingency funds.

**Years of adaptation on the part of managers as well as ample evidence that evaluation is still
not truly part of the learning process suggests those two activities must be kept absolutely air
tight from each other, contrary to theory and logic.**

**B) Care must be taken to avoid giving the impression, to the media, to the general public and
to managers, even inadvertently, that complex problems can be resolved simply and error
free.**

The risk premium costs that ensues can tax resources and reduce quality decisions to the extreme.
In some OECD countries, taxpayers have been led to expect unrealistic achievements. As a result,
a wave of charlatans and "ambulance chasers" prospect for jobs as "Counsellors of the Impossible"
and "Makers-of-rain". Management gurus, who claim that the application of recipes such as "result
base management" will resolve all uncertainties and eliminate risks in the decision process, are
misleading the public and, in my view, acting unethically. Moreover, the gradual effect of those
assaults on the credibility of Governments are reducing its' leadership role. In many instances,
today's questioning does not only relate to the size of the State, but to the rationality of its existence. This breeds irrationality and poor decisions.

c) Goals and Objectives form the basis of evaluations and project management...in theory.

In practice, however, these goals and objectives are written under a cloud of ambiguity. Many things happen to them on their way to the market!

First, the planning stages have increasingly become the responsibility of head offices. Away from the field reality, where a multitude of actors freely interact and conflicting interests are at play, projects are aseptic as they are almost test-tube generated. The process is increasingly lengthy, divorced from field reality but immersed in headquarters' preoccupations and dictates. To secure project approval, managers must kowtow and blend into current policy agendas, fads, whims, imbedded with power struggles and restricted definitions of "political correctness". At the end of this obstacle course, project objectives have acquired many meanings, some of them "unspoken". When, eventually, the financial resources of the project become available to the implementers, the slan seldom fits with reality. It must be changed, most often without proper documentation.

Does that constitute, as it is often argued, an obstacle to evaluation? How do we recreate, in the midst of layer after layer of confusing documentation, the true and authentic goals and objectives of a project or of a program?

If for no other reason than that, it should be recognized that evaluation is first and foremost an art, of the most sophisticated form indeed. The first activity is to "reconstruct" reality, as Sherlock Holmes did. It shares very few attributes with scientific endeavor. As such, it should be practiced only by the very best individuals, endowed with rich professional experience, recognized for the quality of judgment, intuition, perspicacity and, most of all, for their love of the subject matter and for its practitioners. The pedagogical role of evaluators is such that it should be reserved for the very best of us. As a consequence, fewer evaluations may possibly be undertaken; however, the result should be increased quality and higher standards of the resulting evaluations.

d) Increasingly the Aid recipient has been unempowered, first from the design, then from the implementation, and eventually from the evaluation processes.

This is certainly more true in Africa and less so in Asia and in Latin America. It varies considerably as an outcome of administrative procedures, especially the obligation to report to the recipient as is the case in International Organisations. It may be explained by the desire to secure "objective assessors". It may be the result from tied purchasing policies. Whichever; but it demonstrates a lack of confidence in the recipient's capacity to deal with its own development needs. Moreover, it breeds apathy on its part in those crucial matters.

The result is reduced pertinence and relevance, both of the projects and of the evaluations. The emphasis centers on what pleases headquarters, as opposed to the results as defined and needed by the end user. Methodologies are invented in the West, reflecting values and priorities which are characteristics of its systems, its environment, and sense of priorities. Scientifically, this reduces the range of methodologies available to assess and define true results. A gender and a cultural bias
has invariably resulted from such systematic abuse of dominant methodologies

e) to replace the current fascination surrounding base line studies with supporting countries data bases.

Base line surveys have become the darling of evaluators. Nevertheless, aid donors agencies have given minimal support to the creation of indigenous data bases, and management information systems. Poor countries need more than any other a sound base upon which to make strategic decisions and choose investments strategies. The amount of money earmarked for base line evaluation may be put to better use. Collectivelly we should tried to bolster programs that support indigenous data bases and information systems. It would serve also as a mean to further empower aid recipients.

Finally, by way of conclusion, some pertinent observations by former Nobel prize winner, Francois Jacob (medecine, 1965).

"A première vue la science parait moins ambitieuse que le mythe par les questions qu'elle pose et les réponses qu'elle cherche. De fait les débuts de la science moderne datent du moment où aux questions générales se sont substituées des questions limitées; où, au lieu de se demander: "comment l'Univers a-t-il été créé? De quoi est faite la matière? Qu'elle est l'essence de la vie? On a commencé à se demander: "comment tombe une pierre? Comment l'eau coule-t-elle dans un tube? Quel est le cours du sang dans un corps? Ce changement a eu un résultat surprenant. Alors que les questions générales ne recevaient que des réponses limitées, les questions limitées se trouvèrent à conduire à des réponses de plus en plus générales."

To foster development, experts and decision makers need answers to many simple and "limited" questions rather than large ideologically based statements and doctrines. For example we may start our quest for relevant information with questions like:

- "How does rural electrification affects fertility rates? Does it work the same way in urban slum areas? How is social peace related to infrastructure? Does a well structured and maintained communication system enhance or reduces inequalities? How and why does agrarian reform and the establishment of clear property ownership generally allows for sustained periods of economic growth? What are the social effects of irrigation schemes?"

Ottawa, may 23, 1996

---

8 François Jacob, Le jeu des possibles, Essai sur la diversité du vivant, Fayard 1981.
TYPOLOGY AND GRAPHIC REPRESENTATION OF THE EVOLUTION OF CONFLICTS

Period of development interventions
(1) crisis
(2) denied conflict
(3) open conflict
(4) war
(5) reconciliation and reconstruction
Period of humanitarian aid

Evolution over time

intensification of the conflict

Annex 1
Playing Games as Games

GAMES IN the technical sense of situations involving interdependent decisions can also be games in the everyday sense of the word. In this chapter we will see some games that work both as brainteasers and as introductions to some of the main ideas of game theory. The most effective way to learn game theory is to do it: for this reason, you should try to reason each of these games through before reading the solution that follows.

THE PRISONER'S DILEMMA

A pair of transients, Al Fresco and Des Jardins, have been arrested for vagrancy. They are suspected of complicity in a robbery, but the evidence is inadequate to convict them. The district attorney interrogates them in separate cells and offers each the following deal. "If you confess and your friend does not, you will be released and your friend will have the book thrown at him; and the other way around if he confesses and you do not. If both confess, both will receive moderately long sentences. If neither confesses, both will be convicted of a minor vagrancy charge." Specifically, the promised jail sentences, in months, are as in the following table, with the first number in each pair representing Al's sentence and the second Des's. (A minus sign is put in front of each of the numbers to remind us that, in the reverse of the usual interpretation, here more is worse.)

<table>
<thead>
<tr>
<th></th>
<th>Confess</th>
<th>Don't Confess</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Al</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confess</td>
<td>(-8, -8)</td>
<td>(0, -15)</td>
</tr>
<tr>
<td>Don't Confess</td>
<td>(-15, 0)</td>
<td>(-1, -1)</td>
</tr>
</tbody>
</table>

What does rationality dictate that our players do? (Or, is there honor among thieves?)

What is the logic of the situation represented in this table? Imagine Al's reasoning process; should he confess or not? If Al believes, for whatever reason, that Des is going to confess, then (reading down the first column) Al sees he has a choice between 8 months in jail (if he confesses) and 15 months in jail (if he does not confess). Confessing is clearly his best strategy in this case. If, on the other hand, Al believes Des will not confess, then (reading down the second column) Al sees his
choice is between going free (if he confesses) and 1 month in jail (if he does not confess); again, confessing is best for him. Although, to work through the logic of the situation, Al had to make conjectures about what Des would do, in the end these conjectures were irrelevant. Regardless of what Des does, Al's best action is to confess. Now notice that Des is in exactly the same situation as Al, so Des reaches the same conclusion; Des also rationally confeses. Thus the equilibrium of this game, the outcome of simultaneously rational decisions by both of the players, has both confessing.

This looks paradoxical. Compare the confess/confess equilibrium (which results in 8 months' jail for each) with the alternative of neither confessing (which results in 1 month's jail for both). Both would be better off if neither confessed. There is a contradiction between what is individually rational and what is collectively rational. The pursuit of individual gain results in both being worse off than they need be. Each would be better off if they could succeed in cooperating. But they can-not: if one decided to act in their mutual interest by not confessing, then, by the logic already given, it is in the interest of his rival to confess.

Let us introduce some economics jargon to describe this situation. Call an outcome efficient if there is no alternative outcome that would leave some players better off and none worse off. To reverse this definition, an outcome is not efficient if there is another outcome that the players unanimously prefer. In the prisoners' dilemma, the (-1, -1) outcome is efficient (as, for that matter, are the (-15, 0) and (0, -15) outcomes). We have found that the equilibrium of the prisoners' dilemma game is not efficient.

What the prisoner's dilemma shows, in the words of the mathematician Robert Aumann, is that "People who fail to cooperate for their own mutual benefit are not necessarily foolish or irrational; they may be acting perfectly rationally." The prisoners' dilemma is a paradigm for many diverse business and economic interactions. For example: Two firms are competing to sell the same product; the logic of profit maximization forces each to set a low price when both would earn more profits if each set a higher price. Two nations trading with each other are driven by rational, national-interest calculations to erect trade barriers when both would be better off if they were eliminated. Fishermen overfish their common fishing ground and destroy their industry, to everyone's loss.

A potential buyer and a potential seller bargain so hard that they fail to reach agreement on a price, despite the fact that there exist prices such that both would have been better off had the sale occurred.

**THE RATIONAL PIGS**

Two pigs, one dominant and the other subordinate, are put in a box. There is a lever at one end of the box which, when pressed, dispenses food at the other end. Thus the pig that presses the lever must run to the other end; by the time it gets there, the other pig has eaten most, but not all, of the food. The dominant pig is able to prevent the subordinate pig from getting any of the food when both are at the food. Assuming the pigs can reason like game theorists, which pig will press the lever?
For the sake of definiteness, let us attach some hypothetical numbers to this game. Suppose 6 units of grain are delivered whenever the lever is pushed. If the subordinate pig presses the lever, the dominant pig eats all 6 units; but if the dominant pig pushes the lever, the subordinate pig eats 5 of the 6 units before the dominant pig pushes it away. To fill in all the possibilities, we must make some assumption about what happens in the unlikely event that both press the lever simultaneously. Suppose the subordinate pig can run faster so, if both press, it gets 2 units of the food before the dominant pig arrives. Finally, suppose pressing the lever and running to the other end requires some effort, the equivalent of one-half a unit of food. Our game then is represented by the following matrix (with the first number in each pair being the subordinate pig's amount of food and the second the dominant pig's):

<table>
<thead>
<tr>
<th></th>
<th>Press</th>
<th>Don't Press</th>
</tr>
</thead>
<tbody>
<tr>
<td>Press</td>
<td>(1.5, 3.5)</td>
<td>(-0.5, 6)</td>
</tr>
<tr>
<td>Don't Press</td>
<td>(5, 0.5)</td>
<td>(0, 0)</td>
</tr>
</tbody>
</table>

To solve this game, let us anthropomorphically endow the pigs with deductive capabilities. Consider first the subordinate pig's reasoning. "Suppose I predict that the big pig will press the lever. Then I get 1.5 if I press and 5 if I don't. If, on the other hand, I predict it will not press, I get -0.5 if I press and 0 if I don't. Thus regardless of what it does, I am better off not pressing than pressing." Now imagine the dominant pig's thought process. "If I predict the small pig will press the lever, I get 3.5 if I press and 6 if I don't. If I predict it will not press, I get 0.5 if I press and 0 if I don't." Our dominant pig now does seem to face a dilemma. Its best response is different depending on what it conjectures its rival will do: it should not press if it conjectures the subordinate pig will press; but it should press if it conjectures the subordinate pig will not. How can it decide what its best action is? To resolve this quandary, the dominant pig must put itself in the shoes of the subordinate pig. Doing so, it sees, as we saw, that the subordinate pig's best action is unambiguous: don't press. Thus if it presumes that the subordinate pig is rational, it knows it should use its best response to its rival's not pressing: thus it is in its interest to press the lever. (Notice that the dominant pig's chain of reasoning is more complex than the logic we encountered in the prisoners' dilemma, in that its choice of action depends on its pre-dication of what its rival will do.) Rational behavior, therefore, indicates a surprising conclusion: the dominant pig presses the lever, and the subordinate pig gets most of the food. Weakness, in this case, is strength.

Animal behaviorists have actually conducted experiments of this sort. In most experimental trials, the dominant pig did actually push the lever. The pigs behaved like game theorists. Of course, few would argue that the pigs actually went through this intricate chain of reasoning. But the outcome was as if they had. There is an important general lesson here. Over a sequence of plays, it is possible to converge on the rational outcome by trial and error, ending up with a situation that looks identical to what rational play would generate. In the pig experiments, what probably happened was that the small pig tried pressing the lever. Eventually it learned that it got nothing for its trouble and ceased pressing. Then the large pig, after seeing that it was getting no food, found that the only way for it to get some was to press the lever itself. The pigs thus "discovered"
the rational solution to the game. Game-players do not have to be mathematicians in order to succeed in playing rationally.

In contrast to the prisoners' dilemma, the pigs game generates no conflict between individual rationality and collective rationality. The outcome could not be changed without making one of the players (the subordinate pig) worse off. The dominant pig or an outside observer might question the fairness of the outcome (the pig that does the work gets the smaller return); but there is no alternative that the players unanimously prefer.

Similar games arise in less trivial circumstances. Consider, for example, how OPEC worked when it was an effective cartel. OPEC's success was probably due to the willingness of its largest member, Saudi Arabia, to act unilaterally to keep oil prices high. When one of the smaller members increased its output of oil, the Saudis reduced their own in compensation. This was not altruism on the Saudis' part; rather, it arose from the logic of the situation. The Saudis were in an analogous position to the big pig. Both the Saudis and the smaller producers knew that OPEC would collapse unless the Saudis limited their own production; and the smaller producers took advantage of this, getting a free ride on the Saudis' efforts. Saudi Arabia captured for itself a large enough share of the benefits of the high prices that it was rationally willing to bear a disproportionate share of the cost of maintaining the cartel.

Consider also incentives within an international alliance, the North Atlantic Treaty Organization. As a measure of burden-sharing, let us compute for each member country the ratio of military expenditures to national income. Then, for the sake of comparison across countries, let us divide this number by the alliance-wide average. The resulting number will exceed one if the country's share of the costs of collective defense exceeds to its income share; it will be less than one if the country's defense expenditures are a smaller fraction of its income than the average. For the period 1950 to 1984, the United States's relative defense burden so defined was 1.27. All other countries in the alliance had relative defense burdens of less than one (Britain had 0.96, Greece 0.96, France 0.89, Portugal 0.88, Turkey 0.87, Netherlands 0.66, Germany 0.65, Norway 0.59, Belgium 0.58, Italy 0.50, Canada 0.49, Denmark 0.45, Luxembourg 0.20, Iceland 0.00). Only the largest member of NATO, the United States, contributed more than the average share to collective defense. The logic of this might be analogous to the pigs, with the United States playing the dominant pig's role. (Of course, both NATO and OPEC involve much more complex interactions than the pigs game, so this can only be part of the story.)