

Centre de Sociologie de l'innovation
Ecole des Mines, Paris

Rigas Arvanitis

L'évaluation de la recherche aux Etats-Unis
et l'utilisation de mesures quantitatives de
la science

Rapport de recherche
pour le
Centre de Prospective et
d'Evaluation
Ministère de l'industrie
et de la recherche

Juin 1983

Avertissement au lecteur pressé

Le présent rapport est le résultat :

- d'une enquête menée aux Etats-Unis auprès de scientifiques et responsables scientifiques de la NSF et du NIH, ainsi que d'administrateurs de la science au GAO, au Congrès et de personnalités de la National Academy of Sciences;
- d'un examen de la littérature spécialisée sur les thèmes de l'évaluation et de la productivité de la recherche.

Etant assez volumineux, nous avons jugé utile de présenter les principaux résultats de cette enquête et de la revue de la littérature au chapitre 1. De plus, chaque chapitre est introduit par un résumé afin de faciliter la lecture de ce rapport.

SOMMAIRE

Chapitre 1	L'évaluation de la recherche : pratiques et études évaluatives aux Etats-Unis	1
Chapitre 2	L'évaluation et la sélection des projets dans les entreprises	26
Chapitre 3	La productivité des chercheurs : les résultats des travaux sociologiques	72
Chapitre 4	L'évaluation de la science et la politique scientifique	119
Chapitre 5	L'évaluation de la recherche au NIH et à la NSF	152
Annexes		250
Table des matières détaillée		254

REMERCIEMENTS

Je tiens à remercier toutes les personnes qui ont eu la bonté de m'accorder les entretiens nécessaires pour mener à bien cette enquête aux Etats-Unis. Je remercie tout particulièrement Robert Brainard qui m'a ouvert les portes de la NSF et dont les conseils m'ont permis de mettre au point l'enquête. Je remercie Robert Wright et Jennifer Bond des "Science Indicators" pour leur empressement, leur aide précieuse et leur amitié. Egalement, Harry Piccariello de l'évaluation staff de la NSF, et Helen Gee de l'Evaluation Office du NIH m'ont fourni des fils directeurs pour étudier l'évaluation. Le professeur Aizenman et Irvine Pikus m'ont permis de "sentir" ce qu'est la NSF, la recherche universitaire américaine et bien évidemment l'évaluation. Les discussions que j'ai eu avec eux furent inestimables.

Je dois également remercier toutes les personnes qui à la NSF, au NIH, au GAO et au Congrès m'ont ouvert les portes et m'ont accordés leur attention et leur patience. Je remercie tout particulièrement Francis Narin et toute l'équipe de Computer Horizons.

Philippe Roqueplo a tenu à ce que j'effectue ce travail dans les meilleures conditions; qu'il en soit remercié. Enfin, Michel Callon, Bruno Latour, Jean-Pierre Courtial, serge Bauin et Bill Turner par leurs commentaires et leur amitié m'ont fourni le support moral et intellectuel sans lequel ce travail n'aurait pas été le même.

CHAPITRE 1

L'EVALUATION DE LA RECHERCHE : PRATIQUES ET ETUDES EVALUATIVES AUX ETATS UNIS

Examiner l'évaluation de la recherche aux Etats-Unis peut paraître une tâche démesurée. Nous examinons dans ces pages les pratiques d'évaluation des projets de recherche et des programmes financés par la National Science Foundation (NSF) et le National Institute of Health (NIH). Ceci nous permet de regarder ce domaine très vaste dans deux agences qui financent surtout des recherches fondamentales exécutées dans des universités pour la plupart. Mais l'intérêt principal de notre étude fut le débat sur les indicateurs quantitatifs de la recherche scientifique.

1. OBJECTIFS ET PLAN DE CETTE ETUDE

En effet, les indicateurs statistiques dans la politique scientifique (la "policy analysis" plutôt que la politique scientifique elle-même) ont particulièrement été développés aux Etats-Unis. De plus aucun autre pays n'a vu proliférer autant d'études d'évaluations que les Etats-Unis. En conséquence, il nous semble important d'essayer de faire le point sur les méthodes d'évaluations quantitatives. Nous avons cherché à rassembler le maximum d'informations sur ces travaux d'évaluation et en même temps nous avons voulu rappeler le mécanisme d'évaluation de la recherche dans les agences fédérales pour repérer l'influence des études d'évaluation sur les pratiques de l'évaluation.

La NSF paraissait être l'endroit idéal pour mener une telle enquête car elle finance des recherches fondamentales sur l'ensemble du territoire américain, elle a une histoire qui met à jour les liens complexes qui lient la communauté scientifique au pouvoir politique et elle a puissamment incité les travaux quantitatifs de "recherche sur la recherche".

Le NIH fut choisi parce qu'on y a vu une prolifération d'études d'évaluations en tous genres. La direction du NIH a beaucoup incité le développement de ces travaux, parallèlement au développement d'opérations de diffusion de l'information scientifique et de planification des recherches.

Nos enquêtes ont été menées pour examiner comment concrètement se déroulaient ces études d'évaluation, et comment s'effectuait l'évaluation de la recherche dans ces agences. Les pratiques de l'évaluation des recherches dans ces agences sont décrites au chapitre 5.

Nous avons également procédé à une lecture de la littérature spécialisée dans ce domaine. La principale impression qui ressort de cette lecture est le manque de définition précise de ce qu'on entend par le vocable "évaluation de la recherche".

L'évaluation de la recherche est un domaine nouveau qui est constitué de trois ensembles de travaux, que nous avons regroupés par chapitre :

1. des travaux sur l'évaluation et la sélection des projets de recherche, notamment dans des organismes de recherche privé et en milieu industriel. Ces travaux ont démarré dans les années soixante et font partie du management de la R et D (chapitre 2).
2. des travaux d'histoire et de sociologie des sciences

qui ont mis au point des méthodes quantitatives. Ces travaux furent surtout des recherches sur les publications scientifiques (la bibliométrie) et des études sociologiques sur des groupes scientifiques et leur productivité à partir d'enquêtes sur le terrain (chapitre 3).

- enfin, des analyses sur l'état de la science au niveau national (Chapitre 4) et au niveau des programmes des agences fédérales (Chapitre 5). Ces travaux ont souvent été impulsés par le pouvoir politique (le Congrès et la Maison Blanche) et par les administrations des agences de recherche (NSF, NIH, Department of Defense, ...) et des agences économiques et financières (Office of Management and Budget, Department of commerce, etc...)

2. Evaluation des projets et évaluation des programmes

La présence de travaux avec des objectifs très différents rend difficile une définition de ce que doit être l'évaluation. La plupart des chercheurs ont utilisé des définitions pragmatiques qui "collaient" bien avec leur objet d'analyse. Parmi les principaux résultats méthodologiques de ces travaux nous devons en mentionner un qui est fondamental : l'évaluation de la recherche dépend de l'objectif de cette évaluation.

Ainsi nous pouvons établir une première distinction entre l'évaluation des projets de recherche a priori, c'est-à-dire en vue de l'allocation de ressources pour ce projet, et l'évaluation des programmes de recherche a posteriori dont les objectifs peuvent être différents.

3.1. L'évaluation des projets de recherche

L'évaluation des projets de recherche au NIH et à la

NSF reposent essentiellement sur le mécanisme de l'évaluation par les pairs. Il s'agit d'une procédure qui ressemble fort à celle des actions concertées de la DGRST ou des appels d'offres au CNRS. Toutes les agences de recherche aux Etats Unis n'utilisent pas cette procédure et comme le rappellent certains scientifiques, beaucoup de recherches ont été financées sans l'aide de cette procédure (cas de l'office of Naval Research ou de la NASA par exemple).

L'évaluation par les pairs a d'ailleurs été principalement utilisée par la NSF pour financer des travaux de recherche fondamentale. Sa mise en application et son acceptation par l'administration et le pouvoir politique paraissent, historiquement, comme des coups de force. En effet, l'évaluation par les pairs est la reproduction exacte du mode de fonctionnement traditionnel de la communauté scientifique universitaire (par exemple dans le cas de l'acceptation d'articles dans les revues scientifiques). Il s'agit d'une procédure qui ne fait intervenir de manière déterminante que les scientifiques, à l'exclusion des administrateurs. Ainsi l'administration se voit dans le mauvais rôle du censeur par l'argent, car ce sont les contraintes budgétaires qui déterminent les limites des choix que peuvent effectuer les scientifiques.

Dans d'autres agences fédérales, il existe bien d'autres systèmes d'évaluation des projets. Ils peuvent tous être regroupés sous l'appellation d'évaluation hiérarchique qui repose non pas sur les pairs scientifiques mais sur les supérieurs hiérarchiques dans un centre de recherche. Bien sûr, les supérieures hiérarchiques ne sont pas seuls à décider de ce qui devra être fait, et dans la plupart des cas on assiste à un mélange d'évaluations dites internes (les collègues) et externes (les experts et pairs scientifiques les plus compétents dans le domaine en question). Ces évaluations hiérarchiques sont en apparence plus

bureaucratiques que l'évaluation par les pairs qui s'apparente plus à un système de marché financier. D'une façon générale on observe que plus les projets de recherche ont lieu dans des grands centres et plus les décideurs font appel à des évaluations externes. Par contre, le fait d'être en présence de grands projets n'implique pas nécessairement des évaluations externes comme le prouve abondamment la recherche spatiale et militaire.

Entre ces deux extrêmes que sont l'évaluation par les pairs et l'évaluation hiérarchique, il existe une infinité de configurations qui sont déterminées par plusieurs facteurs, en particulier : la discipline scientifique ou domaine de recherche , l'environnement de recherche, les modalités de l'évaluation et certains facteurs culturels comme par exemple le fonctionnement traditionnel de la recherche universitaire ou celui des centres de recherche industriels.

Les tentatives pour introduire des méthodes d'évaluations quantitatives dans l'évaluation de projets de recherche ont été des échecs ou furent rapidement abandonnés, tant dans le milieu industriel que dans la recherche publique, tant dans l'évaluation par les pairs que l'évaluation hiérarchique. Cependant si par évaluations quantitatives on entend les méthodes de notation des projets, en ce cas les recherches sont évaluées quantitativement aux USA un peu partout. Mais c'est là de la métrologie élémentaire, pas de l'évaluation quantitative. On a aussi quelque peu facilement opposé l'évaluation qualitative et subjective à l'évaluation quantitative et objective. Cette opposition n'a de sens que si on s'arrête aux mots ; si on regarde d'un peu plus près comment sont élaborés les indicateurs quantitatifs et sur quoi s'exerce le jugement subjectif on se rend compte que la limite entre ces mots

est fluctuante et imprécise. Les évaluations quantitatives reposent à des degrés divers sur des jugements et à l'inverse les jugements reposent aussi sur des variables quantifiables ou dénombrables.

3.2. Enseignements issus de la pratique américaine de l'évaluation des projets

Concluons sur quelques enseignements issus de ces pratiques et de l'échec relatif de la quantification des évaluations a priori :

- l'évaluation a priori fait partie de la stratégie de recherche des scientifiques. Quand un chercheur soumet un projet à l'évaluation, il sait pertinemment quels sont les points faibles et les points forts de son projet. Il sait aussi déguiser une recherche dans des termes adéquats à l'appel d'offre (ce que Philippe Roqueplo appelle "l'habillage rhétorique"). Il sait enfin placer son projet dans le cadre de tout un ensemble de recherche.

- l'évaluation a priori aux Etats Unis est avant tout une évaluation des recherches elles-mêmes. En particulier, les américains dissocient totalement l'évaluation des chercheurs (des hommes) de l'évaluation des recherches. Ceci est vrai dans le système d'évaluation par les pairs pour l'attribution des "grants", mais c'est aussi vrai dans l'évaluation hiérarchique comme le prouvent les témoignages des chercheurs chez Dupont, IBM, Ford ou encore à la NASA. Les responsables d'une recherche peuvent éventuellement faire l'objet d'une critique au regard de leur passé professionnel. Mais l'équipe qui exécutera un projet ne sera pas évaluée. L'évaluation des hommes et des équipes est donc totalement dissociée de l'évaluation des recherches. Il nous semble important de dire que très probablement l'une

des forces du système américain est cette dissociation entre les problèmes concernant le contenu des recherches et les problèmes de personnel scientifique. La sélection des personnes pour travailler dans une équipe se fait ailleurs, autrement et selon d'autres critères (1).

● l'élément crucial dans l'évaluation a priori est souvent une personne : le responsable du projet, le responsable de l'agence payeuse ou du groupe d'experts. Cette personne n'a pas toujours - et même rarement - le pouvoir de dire oui ou non ; elle n'a pas le pouvoir du censeur. Elle a par contre, un pouvoir bien plus fort qui est lié à sa capacité de définir des objectifs, choisir des domaines de recherche, choisir les autres personnes. A la NSF, le "programm manager" joue ce rôle clef qui consiste à négocier avec les postulants des objectifs, du budget, du matériel des propositions en même temps que c'est lui qui choisi les examinateurs. Au NIH, "l'exécutive secretary" a ce même pouvoir atténué par le fait que les examinateurs sont en poste pour une période donnée et n'évaluent pas de projet au coup par coup comme à la NSF (cas des "ad hoc reviewers"). Dans des centres de recherche comme celui d'IBM, de Ford, de Bell Laboratories, ou de DuPont, ce rôle revient au responsable du laboratoire ou du centre de recherche. Il est très évident, que les choix de cette personne, son attitude ses jugements, ne peuvent pas faire l'objet d'une quantification.

● La forme de l'évaluation a priori dépendra pour beaucoup de l'organisme dans lequel elle se déroule, de

- (1) Pour les méthodes d'évaluation des hommes et des équipes, R. Arvanitis, L'évaluation des chercheurs aux Etats Unis, présenté au colloque "Méthodologies Evaluatives de la Recherche", C.P.E., Avril 1983.

l'histoire de cet organisme. Ce n'est pas un hasard si la NSF a choisi un mode d'évaluation par les pairs impersonnel, où les évalués et les évaluateurs ne se rencontrent pas et où les évaluateurs entre eux ne se connaissent pas. Cela assure aux yeux des scientifiques qui gèrent les programmes de la NSF une plus grande objectivité car ce mode d'évaluation ressemble à une bourse où la valeur des offres et des demandes se déterminent les unes par rapport aux autres. La métaphore économique de la concurrence n'est pas innocente et sa prééminence au sein de la NSF ne s'explique pas autrement que par l'histoire même de la NSF. Celle-ci en effet a été dès le départ une représentation organisée du fonctionnement de la communauté scientifique universitaire.

Tout autre est le cas du NIH qui est plus soumis aux pressions politiques et à la demande du public en matière de santé. Son historique est celui d'une tentative d'intégrer la recherche dans la politique de la santé. Le fait que les examinateurs se connaissent et discutent ensemble des projets de recherches vient de ce que les groupes d'évaluations ont aussi été des groupes de programmation de la recherche.

Comme le montrent bien des travaux de sociologie dans les firmes qui effectuent de la recherche, le même type de détermination peuvent être trouvées quant aux modalités du système de sélection des projets.

● Les méthodes quantitatives qui furent conçues pour effectuer le travail des évaluateurs ont toutes échouées. Mais elles ont permis de montrer quels devraient être les critères d'évaluation. En effet, le principal effet des méthodes quantitatives de sélection des projets fut un effet de révélateur. Face à un "tableau de bord" détaillé, à une présentation précise des moyens et des objectifs, à une liste qui se veut exhaustive des méthodes, du matériel

des données qui existent etc..., les chercheurs ont eu à faire le point. Plus question de dire que untel est un bon scientifique et qu'on lui fait confiance. Cette confiance doit reposer sur des garanties précises. Les seuls systèmes quantitatifs qui ont survécu sont ceux qui précisément permettaient d'effectuer des négociations entre chercheurs et gestionnaires, ceux qui plaçaient, les uns et les autres sur un terrain commun, sur un objet commun (listes d'objectifs, de matériel, de moyens, etc...). C'est donc plus une fonction d'information qu'ont rempli les outils quantitatifs et aussi de plaider ou de pièce à conviction dans une négociation. La morale peut se résumer dans une maxime : informez-vous plus précisément, plus justement, ce ne sera que justice.

4.1. L'évaluation des programmes de recherche

L'évaluation des programmes de recherche ou évaluation a posteriori (1), n'a pas pour objectif l'allocation des ressources. Elle peut avoir plusieurs objectifs, et l'expérience américaine nous permet de faire une première typologie des évaluations de programmes en fonction de leur objectif.

1) Evaluations des moyens et du fonctionnement d'un ensemble de projets d'un programme.

2) Evaluations des performances ou résultats d'un centre de recherche ou d'un programme.

3) Etudes sur la génération d'une recherche, sur la filière d'un produit (études dites rétrospectives).

(1) Nous employons les termes évaluation a posteriori et évaluation de programmes comme des synonymes car les évaluations a posteriori sont toujours - ou presque - des évaluations de programmes.

4) Etudes d'impact économique et social d'une recherche, qui sortent du cadre de notre travail (1).

5) Etude de cas sur la structure d'un groupe de recherche, de sa productivité scientifique, etc... (voir les travaux de sociologie des sciences au chapitre 3).

6) Etudes sur les pratiques évaluatives dans des organismes de recherche et recherches sur la méthodologie de l'évaluation (dont ce travail fait partie).

—

Dans ce rapport , nous avons voulu montrer en quoi ces travaux peuvent être utiles à l'évaluation de la recherche. Ce n'est pas là une tâche superflue pour deux raisons :

- d'une part, ces études ne sont pas toutes des évaluations au sens strict. Seules les évaluations de performances peuvent prétendre à ce statut. Cependant, dès que l'on commence à rentrer dans une telle étude on s'aperçoit qu'il faut passer par un ensemble de données et de constructions théoriques qui n'ont pas de lien direct avec les performances du programme;

- d'autre part, et c'est peut être le point le plus important, les résultats de chacune de ces recherches dépend de manière cruciale de la définition du "programme" et de la construction des indices en cas d'études quantitatives. Plus que tous autres, ce furent les travaux de sociologie qui ont permis de montrer ces deux problèmes méthodologiques.

(1) Une bonne présentation de ces travaux est celle de Eleanor Thomas, Recent Research on R & D and productivity growth : A changing Relationship between input and Impact Measures, présenté à l'OCDE, Conférence sur les indicateurs de science et de technologie, Septembre 1980.

4.2. Origine des évaluations de programmes et des travaux quantitatifs

Ces travaux ont été pour la plupart impulsés par la demande accrue de rapports sur l'utilisation des fonds publics en période de crise économique et de restrictions budgétaires.

"Accountability, productivity, and evaluation are becoming watchwords in government sponsored programs"
(1)

Les années soixante dix ont été aux Etats Unis les années d'or de l'évaluation. Le General Accounting Office, audit et cour des comptes du Congrès, publie chaque année un répertoire des travaux d'évaluations dont le dernier, publié en 1982 décrit environ 1770 études effectuées en 1980-1981 par des agences fédérales. Le GAO a répertorié de 1972 à 1980, environ 7000 rapports d'évaluation. L'évaluation est d'ailleurs une discipline à part entière dans les Universités américaines (2).

Les mêmes années furent celles d'un intense débat sur la politique scientifique. Le rôle de la NSF dans ces débats fut déterminant. Tout d'abord, la NSF a rempli un rôle de coordinateur de la recherche scientifique. Mais ce rôle, que les directeurs de la NSF ne voulaient pas remplir, a pris une tournure différente de celle prévue par des hommes comme Bush ou Steelman. Ce fut un rôle de coordination par

(1) The Evaluation of Federal Research Programs, J. Salasin, L. Hattery, T. Ramsay, The MITRE corporation, June 1980, p.5.

(2) L'évaluation est maintenant devenue une discipline universitaire à part entière aux Etats Unis. Une large part de ces études ont abouti à des analyses méthodologiques assez élaborées, qui sont présentées maintenant sous forme de manuels. Voir par exemple Struening, Elmer L., Marcir Guttenberg (eds) Handbook of Evaluation Research, 2 vol., Beverly Hills, Sage Publications, 1975.

la statistique. La NSF a publié en 1973 le premier rapport statistique sur l'état de la science et de la technologie : les "Science Indicators" (SI).

Le rapport en question est une présentation de statistiques en fonction des thèmes de politique scientifique. Ceci permet donc d'illustrer les différents thèmes comme la baisse de la productivité scientifique, le vieillissement du matériel et de recherche, l'état du marché de l'emploi scientifique ; ces thèmes d'une façon ou d'une autre reviennent également dans les études d'évaluation. Ceci permet d'avoir une vision globale des problèmes habituellement examinés dans ces études.

Le deuxième avantage que présente la publication de SI est qu'il favorise les débats de politique scientifique en permettant aux parties prenantes d'éviter de se retrancher sur des principes indiscutables. Le fait que ce soit plutôt les statistiques, mais pas elles seulement, (1) qui ont joué ce rôle d'intermédiaire n'est pas étonnant si on se souvient qu'ils sont issus des débats sur la responsabilité des scientifiques dans le nucléaire, l'armement et autres thèmes de l'écologie politique. Les rédacteurs de Toward a Metric of Science (2) plaçaient les racines du "mouvement des indicateurs de science" dans la réflexion qu'illustraient le livre de Théodore Roszak, The Making of a Counter Culture (1969) et le numéro de Daedalus consacré à la science (Automne 1974).

Ainsi le "mouvement quantitatif" et l'évaluation des programmes scientifiques ont-ils convergé dans un climat -----

(1) Voir un chapitre 3 les documents "participatifs" de prospection scientifique comme le Five Year outlook, par exemple.

(2) Elkana et alli Toward a Metric of Science : The Advent of science Indicators, Wiley, New York , 1977.

d'effervescence politique.

4.3. Les composantes du "mouvement quantitatif"

Initialement, deux types de travaux se trouvent à la source de ce "mouvement pour les indicateurs de la science" comme l'appellent les auteurs de Toward a Metric of Science : les travaux de sociologie et d'histoire des sciences et les études rétrospectives sur les filières technologiques.

a) Sociologie et histoire des sciences

Ces deux disciplines ont cherché à identifier la croissance de l'activité scientifique et son fonctionnement au moyen d'outils "bibliométriques".

Le terme bibliométrie fut forgé par Pritchard en 1969 pour désigner "l'utilisation des méthodes statistiques dans le domaine des publications, livres ou autre moyens de communication" (1).

La bibliométrie s'appuie sur des décomptes soit de publications (articles, livres) soit de citations reçues et émises dans ces publications. L'utilisation de ses dernières a été grandement facilitée par la base de citations du Science Citation Index mis au point par Garfield et l'Institute for Scientific Information (ISI). Sa popularisation est due pour beaucoup aux travaux de Derek de Solla Price et des sociologues américains de l'école mertonienne ou fonctionnaliste.

(1) Thackray
in Toward a Metric of Science, pp.

L'utilisation d'indicateurs bibliométriques est sujette à plusieurs controverses. Elles n'est en effet pas innocente, ni neutre, puisque les résultats statistiques dépendent du contenu des bases de citations utilisées et des indices utilisés. Chaque fichier a ses propres critères de sélection des travaux et on peut constater des différences importantes dans des résultats comparatifs à partir de bases de données différentes.

Par exemple, si l'on compare le pourcentage d'articles en français dans la production des labos français (à partir du fichier de l'ISI et celui du CNRS-(PASCAL) on obtient les résultats suivants (1)

% ARTICLES EN FRANÇAIS			
		1976	1980
CNRS			
28.579 articles	82 %		67 %
ISI			
25 027 articles	60 %		49 %

De plus les critères évoluent dans le temps. Enfin, l'utilisation du fichier de l'ISI présente des défauts techniques souvent importants. D'où la nécessité d'établir des fichiers "propres" comme ceux de l'équipe de Francis Marin à Computer Horizons.

Koester, Sullivan et White (2) assignaient un rôle spécifique aux études de citations : l'étude diachronique de l'évolution d'une discipline ou d'une communauté scientifique à un niveau agrégé. Les analyses de citations sont de peu d'utilité pour examiner la constitution d'un champ scientifique, ce qui nécessite d'aller voir de près ce

-
- (1) Cet exemple m'a été remis par Ginette GABLOT, chargée de mission à la MIDIST.
 - (2) Koester, Sullivan, White, A quantitative case study of the Evolution of weak Electromagnetic Unification Theory, Social Studies of Science, vol. 12 (1), Février 1982, pp. 73-100.

que les acteurs ont fait. Mais elles peuvent être très utiles pour tracer l'évolution d'une discipline dans son ensemble. L'analyse des citations a connu son heure de gloire quand on a cru pouvoir identifier la qualité d'une recherche au nombre de citations qu'elle recevait. Depuis, même les bibliométriciens les plus acharnés s'accordent pour dire que le nombre de citations reçues ne mesure rien de plus que l'impact d'une recherche ou encore l'utilisation de l'article par les pairs ; les comptages de citations sanctionnent en quelque sorte le fonctionnement de la communauté scientifique, et en tous cas de cette portion qui s'exprime à travers les revues scientifiques.

Les travaux dont nous examinons la teneur au chapitre 3, sont essentiellement des travaux qui cherchent à comprendre le fonctionnement d'une équipe de recherche ou d'un labo à un niveau fin. En dehors des travaux bibliométriques, il y a aussi toute une série d'études qui puise son origine dans la sociologie des organisations et de la psychosociologie. Ces dernières pourtant quand il s'agit d'étudier une organisation de recherche se rapprochent beaucoup des travaux plus traditionnels de la sociologie des sciences. Cela est peut-être aussi le fait d'une remise en cause des données quantitatives dans la sociologie des organisations. "On assiste à une scepticisme grandissant sur les possibilités qu'offrent ces outils quantitatifs pour produire des données sans altérer et déformer le phénomène que cet outil est censé décrire"(1) Il faut dire que même le travail d'un historien de la pensée scientifique comme Kuhn n'est pas exempt de tendances psychologisantes. Dans "la tension essentielle" (1969), Kuhn disait à propos

 (1) John Van Mannem (éd.), Qualitative Methodology, Administrative Science Quarterly, 24 (4), Novembre 1979, p.522.

de Tesla et Gramme que "les traits de personnalité requis pour être un chercheur en science pure sont très différents de ceux requis pour être un inventeur, ceux requis pour être un chercheur en science appliquée étant peut être entre les deux" (1), p.292). Ce type de raisonnements à base de psychologie deviennent de moins en moins fréquents. Les travaux sur la créativité de la source de l'inventivité n'expliquent en effet pas grand chose. Il y a maintenant un consensus parmi les "spécialistes de l'évaluation pour affirmer l'importance des "variables contextuelles", c'est-à-dire de l'environnement de la recherche (institutionnel, culturel, etc...) (2)

Toujours est-il que la sociologie se détache de ses origines de plus en plus, pour se rapprocher de méthodes dites plus qualitatives, ce qui veut dire dans beaucoup de cas de méthodes moins simplistes. et qui mélangent parfois avec bonheur le quantitatif au qualitatif.

b) Les études rétrospectives

Les études rétrospectives dont les plus connues sont celles du Department of Defense (HINDSIGHT) et de la NSF (TRACES), ont été beaucoup critiquées. Elles ont cependant permis de mettre à jour les notions de filiations scientifiques, de délais dans l'application des résultats des recherches et de problèmes méthodologiques importants comme celui de la composition sociale des panels d'experts qui évaluent une recherche. (Voir chapitre 5, II.3).

-
- (1) THOMAS KUHN, La tension essentielle in Jacob , De Vienne à Cambridge, Gallimard, 1980, pp. 277-292.
 - (2) Salasin, Hattery, Ramsey, The Evaluation of Federal Research Programs, The MITRE Coop. 1980, p. A-2.

Ces idées ont été reprises par des bibliométriciens et des sociologues. Notamment Francis Narin a pu montrer que les publications pouvaient être utilisées dans ces études de filiations scientifiques. Il a aussi récemment montré que les brevets peuvent être exploités de la même manière.

De son côté, le problème de la composition sociale des panels d'experts est très précisément celui qu'abordent les travaux de sociologie par exemple dans l'analyse du fonctionnement de l'évaluation par les pairs.

Il nous semble que les études rétrospectives ont été souvent critiquées à tort. Souvent les critiques se sont adressées à l'administration qui finançait ces travaux en leur opposant l'intégrité de scientifiques tels que Weinberg ou de politologues comme Harvey Brooks qui ne se sont pas "mouillés" dans ces entreprises. Car en effet, les études rétrospectives ont souvent été des plaidoyers pour le Department of Defense ou la NSF ou tout autre organisme cherchant à justifier leurs dépenses. Cette attitude de critique contre tout ce qui se fait et qui risque d'être récupérée est sclérosante et véhicule une vision de l'administration très éloignée de la réalité. Il est important que les scientifiques et les administrateurs aient une base de discussion commune. Les études rétrospectives ont permis de montrer que des études quantitatives pourraient jouer ce rôle d'intermédiaire dans les discussions. Elles ont eu le mérite de montrer qu'une discussion d'ensemble sur la politique scientifique était faisable.

L'histoire de la politique scientifique aux Etats Unis est révélatrice à cet égard car entre les rapports de Bush et Steelman dans les années 40 et les rapports de la NSF dans les années 70 les conseillers scientifiques auprès de la Maison Blanche ont souvent purement et simplement nié la

possibilité de mener une réflexion d'ensemble sur la politique scientifique.

c) Les études d'évaluation a posteriori

Avec le soutien financier de la NSF, des indicateurs bibliométriques fiables ont été mis au point. Ces indicateurs peuvent concerner un pays dans son ensemble ou bien un organisme de recherche. En ce cas nous sommes en présence de travaux de bibliométrie évaluatrice.

Ce type de travaux a été mis au point au NIH dans le cadre de son programme d'évaluation. Nous l'examinerons en détail dans le chapitre 5.

Ces travaux se sont développés dans un autre cadre que celui des études rétrospectives et souvent ce furent des recherches faisant partie de tout un ensemble de travaux pour amorcer des discussions non seulement entre l'administration et les chercheurs mais aussi entre les chercheurs eux-mêmes et entre les agences fédérales et le pouvoir politique. Initialement conçues comme des audits, les évaluations a posteriori ont dû poser de plus en plus des questions comme celle de la définition des conditions du succès ou de l'échec d'un programme, de la communication entre les chercheurs et de la définition de la qualité scientifique. Autant de questions qui sont du ressort de la sociologie des sciences mais que les gestionnaires et les administrateurs de la recherche cherchent aussi à résoudre.

En ce sens nous pouvons affirmer que les études d'évaluations quantitatives ont eu essentiellement un effet d'information et de communication.

4.4. Quelques enseignements des études évaluatives américaines

L'expérience américaine est riche d'enseignements. Nous nous bornerons à en rappeler certains.

• Les indicateurs scientifiques de même que les études d'évaluation ont puisé beaucoup de leur force dans les débats sur la politique scientifique et sur l'organisation - ou ré-organisation - des centres de recherche.

Dans ces débats, études d'évaluation et travaux statistiques ont pu être des intermédiaires, souvent des modérateurs et en tout cas des informations utiles à la discussion. Ils ont permis aux différents acteurs de se comprendre mutuellement. Il y avait d'un côté des scientifiques savants et de l'autre des administrateurs ignorants. Les travaux statistiques et d'évaluation ont mis à égalité les uns et les autres en formant un nouveau terrain commun de discussions. D'aucuns ont cru que cela affaiblissait les scientifiques dans un débat où les détenteurs d'argent - et donc de moyens - sont de toute façon supérieurs. C'est là oublier que les scientifiques disposent aussi de moyens - leur savoir, leur connaissance du matériel, leur capacité politique aussi - et surtout qu'une politique scientifique sans les scientifiques est une vue de l'esprit.

Ainsi nous pouvons affirmer que dans le cas des travaux quantitatifs d'évaluation de programmes, comme dans le cas des travaux quantitatifs d'évaluation des projets que leur principale fonction fut une fonction d'information et de communication (1)

(1) Mais les indicateurs statistiques n'ont pas été les seuls moyens d'information et de communication, voir chapitre 4.

● Les études d'évaluation de la recherche ont eu tout intérêt à se rapprocher de disciplines comme la sociologie des sciences et l'histoire des sciences (1). Ceci a considérablement enrichi méthodologiquement les études en question, en même temps que les autres disciplines ont vu apparaître un domaine de recherche nouveau - le développement des indicateurs quantitatifs dans des études d'évaluation a posteriori n'a été rendu possible que par des objectifs précis qui furent assignés à ces études. Leur validité dépend de la définition précise de la base de recherche (du programme) et elle-même dépend de l'objectif défini.

Ceci explique les difficultés de comparaison d'études effectuées dans des cadres organisationnels différents. C'est pour cette raison que des statistiques plus agrégées - au niveau d'une région, d'un pays, d'un ensemble de pays - sont nécessaires.

Il est nécessaire que statistiques de la science et études d'évaluations se développent parallèlement afin de renforcer les unes et les autres.

● l'impact des études d'évaluation sur les pratiques est parfois difficile à cerner.

Dans le cas de l'étude du système d'évaluation par les pairs en NIH nous pouvons voir assez clairement que la procédure a été modifiée pour tenir compte de certaines observations sur son fonctionnement issues de l'étude. Mais là aussi cela n'a été possible qu'en encourageant la discussion et la communication entre les personnes

(1) Nous avons eu l'occasion de voir des centres de recherches financer des travaux d'histoire des sciences dans le cadre de leurs programmes d'évaluation.

concernées (1).

Dans d'autres cas, l'impact a été moins direct - ou bien il est moins bien perçu. Certaines études comme celles des Laboratoires des matériaux de la NSF furent exploitées par la NSF auprès du Congrès. Mais la direction de la NSF prétend avoir exploité les résultats de l'étude à des fins de planification et de budgétisation. En ce cas, leur utilisation a dû se faire par les discussions qui ont lieu au sein du processus de planification de la NSF, à plusieurs niveaux différents de l'organisme.

4.5. Les perspectives d'avenir en matière d'études d'évaluation

Les travaux d'évaluation aux Etats Unis vont très probablement se développer dans trois directions :

- une topographie de la recherche
- une réflexion sur la notion de qualité de la recherche et de critères d'évaluation
- une liaison plus étroite entre la sociologie des sciences, les statistiques sur la science et l'économie de la RSD (ou de l'innovation).

a) Topographie de la recherche

Les techniques issues de l'analyse des citations ont permis de mettre au point des méthodes qui non seulement décrivent l'évolution d'une discipline mais permettent de positionner des domaines de recherche les uns par rapport aux autres. L'intérêt de ces méthodes, les co citations et

(1) Actuellement le NIH cherche à évaluer l'impact de ses procédures de planification et d'évaluation. Il nous semble que cette expérience doit être suivie de près en prenant un contact avec ma direction de l'évaluation du NIH.

l'analyse de co-occurrence des mots, est qu'elles donnent la possibilité de tracer les positions relatives de domaines relativement "pointus" et non pas seulement de grandes disciplines, physique, chimie, etc... qui sont les seules que connaissent les institutions et par extension les bases de données bibliographiques.

Les cartographies des domaines de recherche ont été développées par l'ISI ("Atlas de la Science") et le Centre de sociologie de l'Ecole des Mines ("Cartindex"). L'interprétation de ces cartes et leur extension est certainement l'une des voies de recherche les plus prometteuses(1).

b) Qualité et critères d'évaluation

La deuxième voie que les études sur les activités scientifiques doivent explorer est la meilleure compréhension des critères souvent implicites d'évaluation de la recherche. Jusqu'à maintenant les auteurs d'études d'évaluations se sont contenté de constater et parfois mettre à jour les critères selon lesquels un projet de recherche doit être évalué. Il s'agit maintenant de comprendre les liens entre ces divers critères et leur définition. Il faut en quelque sorte mettre en évidence l'activité, souvent appelée rhétorique, d'intéressement, de canalisation des intérêts, de problématisation, que mettent en oeuvre les scientifiques (2).

Ceci doit se faire conjointement à l'exploration de la notion de qualité de la recherche, afin d'explicitier ce -----

- (1) Une bonne présentation de ces cartes et de leur méthode est celle de A. Guggenheim, Quand les sociologues étudient la littérature scientifique, La Recherche, n° 137 vol.13, octobre 1982, pp. 1200-1202.
- (2) Ces notions ont été développées par Michel Callon et John Law, On Interests and their transformation : Enrollment and counter Enrollment, Social studies of Science, vol. 12, 1982, pp. 615-625.

qui fait qu'une recherche est dite "bonne" et réussie là où une autre a échoué.

c) Lier la sociologie, l'évaluation de la recherche et l'économie de la R et .D.

La troisième voie de recherches futures est l'exploration des liens entre technologie et science. Nous n'en sommes plus à prétendre que la technologie dépend nécessairement de la science. Il n'y a pas de déterminisme de la science, pas plus que de la technologie. Mais bien que ce soit là un point acquis, cette indétermination pose plus de problèmes qu'elle n'en résoud (1).

Nous savons peu de choses sur la relation qui existe entre la littérature scientifique et les brevets, bien qu'il commence à y avoir des travaux sur ce thème (1). Nous avons peu de choses sur l'organisation des activités des recherches dans des laboratoires mixtes, Universités - industrie (2). Nous savons assez peu de choses sur les processus d'innovation et leur gestion (3). Enfin nous savons assez mal quantifier une technologie (4).

-
- (1) F. Narin a mis au point une liaison entre les indicateurs bibliométriques et les analyses de citations issues de brevets américains. En France, l'équipe du Centre de sociologie de l'innovation et celle du CDST ont amorcé une collaboration conjointe avec le Patent and Trademark Office aux USA et l'INPI en France.
 - (2) Michel Callon et Jean-Pierre Vignolle, Mythes et Réalités de la recherche collective, les Centres Techniques Industriels, les Cahiers de l'Anrt, n°5, juin 1983, voir aussi le brevetabilité de la recherche fondamentale, étude menée par la NSF sur le programme Chimie, janvier 1983 ; ce domaine commence à faire l'objet de recherches très poussées à cause de l'émergence des biotechnologies et des technologies informatiques.
 - (3) F. Engels, et F. Pellet, L'administration de l'innovation : l'ANVAR, Centre de gestion Scientifique, Ecole des Mines, Paris, janvier 1982.
 - (4) Mais les travaux de Battelle et ceux sur les indicateurs de technologie dans Science Indicators (voir chapitre 4) sont des recherches de pionniers.

Tout cela indique une voie prometteuse : la liaison entre l'économie de la R et D et la sociologie des sciences.

Jusqu'à maintenant les économistes ont cherché à évaluer l'impact d'une technologie, les pourcentages de réussite d'une technologie, les pourcentages de réussite d'une innovation de la recherche à la commercialisation et éventuellement la gestion de la recherche et du développement. Ces tâches ardues ont donné des résultats très précieux mais dont l'interprétation nécessite de plus en plus de regarder les stratégies des groupes industriels, des PME, et des chercheurs dans ces instances. Or c'est bien là le sujet des travaux de sociologie et des études d'évaluation.

Enfin un effort doit être consacré pour mener de front des travaux d'évaluation de programmes scientifiques et des évaluations techniques, voir de "technology assessment".

o
o o

Ces perspectives d'avenir sont une occasion à saisir : afin de développer une réflexion nouvelle sur la politique scientifique au niveau national et au niveau d'organismes de recherche ou de programmes de recherche. L'enseignement majeur que nous pouvons tirer des études citées dans ces pages, est qu'elles furent un moyen de réflexion nouveau, un véhicule de discussions entre des personnes jusque là fort distantes. Ainsi on a pu voir aux Etats-Unis, des administrateurs discuter de sociologie et des sociologues discuter d'évaluation. Ce n'était pas là de simples révérences des uns aux autres. C'était des débats en profondeur

sur le fonctionnement de la science (1). Mais la seule discussion ne suffit pas. Il faut la nourrir de travaux empiriques, d'études de cas, de recherches sur le terrain. Ce n'est qu'à ce prix que peut naître un débat fructueux.

(1) En témoignent les débats du Congrès d'octobre 1982 à Philadelphie de la 4S Society.

CHAPITRE 2

L'EVALUATION ET LA SELECTION DE PROJETS DANS LES

ENTREPRISES

Résumé

L'évaluation de la recherche sera différente selon qu'il s'agit de recherches fondamentales ou de recherches appliquées. Mais la limite entre ces deux termes varie, notamment en fonction des rapports de force et de la position des personnes qui effectuent et qui gèrent la recherche.

Dans la recherche appliquée, et plus encore le développement de produits et de processus nouveaux dans la production, les critères d'évaluation et de sélection des projets sont assez précis et font l'unanimité.

Par contre dans la recherche fondamentale, les critères sont plus flous, notamment à cause de l'incertitude plus grande quant à l'issue de la recherche. Ces critères dépendent beaucoup plus des responsables de la recherche et de leurs liens avec la direction de l'entreprise, dans le cas du secteur privé. Mais il est certain que dans un milieu industriel la liaison entre les critères "d'affaire" et les critères de la science est difficile à établir.

L'évaluation non formelle et non-quantitative est la forme d'évaluation la plus courante dans les agences

fédérales et les entreprises qui font de la recherche. Son organisation cependant ne vas pas sans mal, et l'utilisation de procédures plus formalisées de sélection des projets et de prise de décision n'est pas une panacée. Enfin, on peut apercevoir tant dans la littérature spécialisée que dans l'historique des centres de recherche publics et privés une opposition entre l'évaluation hiérarchique - par les collègues et les supérieurs hiérarchiques - et l'évaluation par les pairs scientifiques.

Diverses méthodes formelles de sélection des projets ont été présentées dans la littérature manageriale (méthodes de classement, méthodes de notations, méthodes économiques, méthodes d'optimisation et d'analyse du risque), que nous rappelons ici.

Mais l'utilisation de ces méthodes est rare comme le montrent les quelques travaux empiriques sur la question que nous examinons dans le détail. En conclusion nous rappelons les principales causes d'échecs de la mise en application de ces méthodes quantitatives de sélection des projets. Nous relevons aussi une certaine confusion sur ce qui doit être et ne doit pas être évalué.

Dans ce chapitre nous examinerons les principales conclusions de la littérature américaine consacrée au management de la recherche en milieu industriel. Nous examinerons les méthodes proposées aux managers pour évaluer et sélectionner les projets de R et D et les problèmes que ces méthodes soulèvent.

La gestion de la R et D peut être vue sous deux angles différents. Celui de l'action, et celui de l'analyse.

La littérature professionnelle dans les deux cas est considérable. Il ne s'agira donc pas ici de faire une "survey" de la littérature mais plutôt une synthèse des éléments les plus importants.

Comment effectuer le meilleur choix parmi plusieurs solutions proposées ? Comment exécuter la recherche de la manière la plus efficace ? Ce sont là les deux questions auxquelles s'efforcent de répondre les gestionnaires. Les méthodes de choix sont aussi des méthodes d'évaluation. Ce seront donc les seules examinées dans les pages suivantes (1).

Toutefois, avant de les aborder il faut préciser ce que l'on entend par recherche fondamentale ou appliquée. Ceci permettra de montrer les critères proposés pour l'évaluation de la recherche en milieu industriel. Ce débroussaillage nous permettra de mieux comprendre les conflits que suscite l'évaluation de la recherche.

1. Distinction entre la recherche fondamentale et la Recherche Développement

Les décisions à prendre concernant la R et D seront différentes selon qu'il s'agit de recherche fondamentale, orientée, appliquée, de développement, etc...

Différentes typologies des milieux, laboratoires,

(1) Pour un classement des méthodes analytiques voir OCDE Politique scientifique gouvernementale et Méthodes analytiques. Evaluation, 1972, p. 46. (Conférences OCDE 24-25, 3.69 et 19-20.11.70).

et institutions de recherches ont été proposées (1).
Il existe maintenant une littérature assez abondante sur ces distinctions.

Tous les auteurs s'accordent pour souligner l'arbitraire d'une distinction entre la recherche fondamentale et la recherche appliquée. A-t-elle néanmoins un sens ?

Il semble que oui, mais qu'il ne s'agit pas d'une distinction absolue comme on aurait tendance à le croire. Elle a un sens parce qu'elle est issue de la pratique et non d'une conceptualisation a posteriori. En effet, la distinction entre recherche fondamentale et recherche-développement (2) indique des différences non pas tant dans la nature du travail qui est effectué mais dans le contexte dans lequel se trouve le laboratoire ou l'unité de recherche. Le Directeur d'IBM-Recherche déclare que :

"La même activité technique aura une finalité différente selon qu'elle se trouve dans un contexte de recherche fondamentale ou appliquée ou bien dans un contexte d'applications industrielles" (Branscomb, 1982) (3)

-
- (1) WIRT J.G., LIEBERMAN A.J., LEVIEN R.E., R.D. Management in Federal Agencies, Lexington Books, 1975.
 COLE, G.A. (1979), Classifying Research units by patterns of performance and influence, un F.M. Andrews (ed) Scientific productivity of research groups in six countries UNESCO and Cambridge, 1979.
 VISART N. (1979). The utility of using different typologies of research units to understand their functioning and management, R & D Management, 9, pp. 193-204.
 GEE R.E., A Survey of current Project Selection practice" Research Management, vol.14, n° 5, septembre 1971, pp.38-45
 Gee distingue la recherche exploratoire, la recherche à haut risque commercial et la recherche de soutien des activités commerciales en cours.
- (2) C'est cette distinction qu'emploie par exemple Harvey BROOKS.
- (3) BRANSCOMB, Lewis, Industry Evaluation of Research Quality Science, Technology and Human Values, 7 (89), Spring 82, pp. 15-20.

L'arbitraire de la distinction a été soulignée par Salomon (1970) "les recherches qui ont précédé la découverte du transistor dans les "Bell Laboratories" ont exigés des travaux qui dans un contexte universitaire auraient été qualifiés de fondamentaux ; seul l'objectif de la firme qui les a soutenus (...) justifie de les tenir pour appliqués" (1).

Ce qu'on attend de l'unité de recherche fondamentale c'est des connaissances ... ce qu'on attend de l'unité "développement" c'est des produits ou des procédés nouveaux commercialisables. Ainsi, pour continuer à utiliser l'exemple d'IBM, le budget total de la R et D représente pour cette firme 6,5 % du chiffre d'affaires. 10 % de ce budget est alloué à la recherche (c'est-à-dire à l'unité "Recherche" d'IBM) et 90 % au développement. Mais "on peut dire que la recherche fondamentale représente 15 à 20 % du budget R et D car dans l'unité "développement" on effectue nécessairement de la recherche fondamentale" (Branscomb).

La firme postule qu'il existe d'un côté la recherche et de l'autre le développement ; et ce postulat repose sur un critère très simple. La Recherche c'est l'activité de recherche d'où n'émanent pas des "produits" commercialisables. Le Développement, c'est l'activité de recherche qui permet d'obtenir des "produits" commercialisables (2)

Les critères d'évaluation seront différents dans le cas de la recherche-développement et dans celui de la recherche fondamentale.

(1) J.J. SALOMON, Science et Politique, Le Seuil, Paris, 1970, pp. 153.

(2) Cette distinction et de même que la discussion sur les critères de la recherche chez IBM coïncide assez bien avec celle de A.M. Clogston sur les Bell Laboratories, voir NAS, Quality in Science, 1982).

2. La Recherche - Développement

La principale différence entre recherche-développement et recherche fondamentale tient donc au risque de la commercialisation. Mansfield (1971) a remarqué que parmi les projets de R et D dans 5 branches (industries électroniques, pharmaceutiques, chimiques et industrie du pétrole) 60 % étaient techniquement achevés, 30 % étaient commercialisés mais seuls 12 % étaient rentables.

Dans ces conditions la perception des pressions qui s'exercent sur le marché qui devraient également s'exercer sur le personnel de recherche des firmes sera un élément important dans l'évaluation de la recherche. Cette perception est, bien sûr, subjective "du moins en ce qui concerne une technologie qui n'a pas encore permis de produire".

" Nous devons deviner ce que les concurrents seront capables de faire quelques années plus tard par rapport à ce que nous espérons pouvoir faire (en ne sachant rien pour le moment, puisque nous n'avons encore rien fait)" (Branscomb, 1982, p. 16).

Dans ces conditions ceux qui arrivent à "deviner" seront aussi ceux avec le plus de chances de succès commerciaux.

E. MANSFIELD. How Economists see Rand D., Research Management, jul. 82, pp. 23-29.

Comment deviner ce qui se passera sur le marché, comment repérer un besoin ? C'est en se posant cette question que les managers eurent l'idée d'associer des spécialistes de la commercialisation aux évaluateurs. Déjà en 1971, le projet Sappho de Sussex avait démontré que les innovateurs qui enregistrent des succès commerciaux sont aussi ceux qui effectuent des études de marché pour leurs produits (cité par Clarke, 1974).

On peut observer dans la littérature spécialisée un débat jusqu'aux années 73-74 sur la nécessité d'associer les services marketing aux services de R et D. A quel moment faut-il effectuer des études de marketing ? Ne risquent-elles pas d'inhiber la recherche ? Les travaux économétriques ont semble-t-il permis de clore cette controverse. Leur verdict énonçait que plus la R et D est proche du marketing, plus la probabilité de succès commerciaux pour les produits nouveaux était élevée (Mansfield, et al. 1977).

Ces travaux ont aussi montré qu'il faut tenter de maintenir un équilibre entre projets R et D à long terme et à court terme. Jusqu'à un certain seuil dans le temps, il est intéressant pour la firme d'investir dans des recherches ambitieuses et à long terme à fortes probabilités de succès (ou d'échec) commerciaux, et simultanément d'investir dans des projets à plus court terme, moins ambitieux mais plus sûrs, dont les bénéfices financeront les projets à plus long terme. Passé ce seuil, le succès commercial n'est plus garanti.

CLARKE T.E. (1974). Decision - Making in Technologically Based organizations : a literature survey of Present Practice, IEEE transactions on Engineering Management EM-21 (1) Feb.1974.

MANSFIELD Edwin, RAPOPORT, John, SCHNEE, Jerome, WAGNER Samuel and HAMBURGER, Michael. research and innovation in the Modern corporation. New York, W.W. Norton, 1971.

Le débat sur le succès commercial de la R et D s'est déplacé avec le verdict des économistes. Là où auparavant les gestionnaires avaient à affronter un problème de gestion du personnel et d'organisation, on trouve un problème de gestion de "portefeuilles de R et D" (Mansfield, 1982, op.cit.).

Dans un portefeuille de R et D, chaque élément comporte une série de propriétés : un horizon à court, moyen ou long terme, un domaine d'activités, une orientation plutôt vers la conception d'un produit ou/et d'un procédé nouveau, un type d'avantages potentiels attendus (bénéfices, adoption à la réplémentation en vigueur, etc...), un type de marchés "ciblés", un degré de crédibilité technique, etc.

On ne décidera donc plus sur un seul projet ou même sur la base de la comparaison des valeurs intrinsèques définies pour chaque produit, mais plutôt sur le mélange optimal des projets. Il s'agit bien là de l'application des concepts financiers à la R et D (EIRMA, 1979).

3. Critères d'évaluation de la Recherche-Développement

Les critères d'évaluation concernent donc autant la recherche, que la production et la commercialisation des produits et procédés nouveaux.

Dean (1968) a ainsi repéré 10 critères concernant la R et D, 5 la production et 7 la commercialisation.

EIRMAN, Systems and Methods for Planning R and D in industry, 1979.

DEAN (1968). Evaluating, selecting, controlling R and D Projects, AMA. Research Study 89, New York, 1968.

A. RECHERCHE-DEVELOPPEMENT

1. Temps et coût du projet
2. Possibilité d'acheter des produits conçus ailleurs
3. Capacité en terme d'emploi, de matériel et de fonds
4. Chance de succès (évaluation des projets)
5. Unicité de l'approche
6. Avantages techniques
7. Equilibre de l'effort de recherche
8. Formation du personnel technique
9. Recherche dite fondamentale, définitions du savoir nouveau
10. Brevetabilité.

B. PRODUCTION (1)

1. Montant. Capital investi
2. Compatibilité du produit / procédé avec les capacités de production
3. Besoins nouveaux induits dans la production
4. Amélioration du processus de fabrication
5. Evaluation de la qualité des produits.

C. COMMERCIALISATION

1. Part du marché éventuelle et potentiel de développement du marché
2. Temps et taille (de recherche et de production) nécessaires à la commercialisation
3. Risques de commercialisation
4. Domaine de la firme et objectifs de croissance
5. Position stratégique de la firme, propriété et protection des produits (brevets, licences et royalties)

- (1) Jacques LACHNITT. L'analyse de la valeur. Que sais-je ? n° 1815, PUF, 1980. Les contrôles de qualité forment un ensemble essentiel à la compréhension de la stratégie de recherche d'une firme. Ceci est d'ailleurs aussi vrai de la recherche publique.

6. Compatibilité de la commercialisation du produit avec les structures de commercialisation existantes
7. Image de la firme / du produit.

D'après une survey sur les méthodes de décisions dans les organisations de R et D, ces critères sont universellement acceptés par les managers (Clarke, 1974). A quelques variations près tout le monde en effet utilise l'un ou l'autre de ces critères.

Une étude effectuée auprès de laboratoires de l'industrie pharmaceutique a permis de se rendre compte que les critères énumérés ci-dessus sont employés implicitement par les dirigeants quand ils doivent effectuer des choix (Faust, 1971).

4. Critères d'évaluation de la recherche fondamentale dans les entreprises

Concernant la recherche fondamentale la question de l'évaluation se pose en des termes différents. Kochan (1978, p. 111) rappelle deux impératifs de la recherche fondamentale en milieu industriel. Tout d'abord il faut pouvoir assurer la "transférabilité" des "produits" de la recherche vers le "développement" ; ensuite il faut assurer un "contrôle stimulant" du laboratoire de recherche pour que celui-ci soit capable de répondre aux besoins des autres divisions de la firme.

Dans le cas de recherche fondamentale, Branscomb, directeur du centre de recherche d'IBM, a indiqué très clairement "ce que la direction (the management) de la

CLARKE T.E. 1974 , Decision Making in Technologically Based Organizations : A Literature Survey of Present Practice, IEEE Transactions on Engineering Management vol. EM.21 (1), Février 1974.

FAUST R.E. Project selection in Pharmaceutical Industry Research Management, vol.14 (5), septembre 1971 : 46-55.

firme veut évaluer (assess) et quelquefois mesurer (...):

Premièrement, bien sûr, nous voulons connaître la production de connaissances de nos activités de recherches et nous voulons savoir si ces connaissances correspondent à celles que nous attendions.

Deuxièmement, nous voulons savoir si les produits de nos recherches sont effectivement utilisés, c'est-à-dire si les outputs de l'unité "Recherche" deviennent des inputs technologiques stratégiques et tactiques de l'unité "Développement". Ainsi quand nous demandons quelles idées cette unité reçoit et produit, nous effectuons une évaluation implicite de la division "Recherche", même si cette dernière n'a pas de moyens pour obliger les unités "développement" d'utiliser leurs "bonnes idées" (...)

Troisièmement, nous voulons connaître l'évolution du personnel en fonction des besoins de la firme (...)
Quatrièmement, la gestion impose des contraintes financières et ceci est aussi vrai de la division de recherche que du reste de la firme.

Enfin, nous cherchons à connaître une catégorie d'attributs que j'appellerais "l'image scientifique et technologique" (...) Nous n'engagerons pas un projet de recherche uniquement pour atteindre un but fixé ; certaines recherches sont pébiscitées et inspirent la confiance face à certaines questions. Ceci permet de comprendre pourquoi des projets de recherches très coûteux sont engagés alors qu'aucun bénéfice n'est prévu avant 10 ou 15 ans". (Branscomb, 1982 op. cit).

Les critères d'évaluation de la recherche fondamentale font moins l'unanimité, notamment en raison de l'incertitude plus grande qui existe dans le déroulement de la recherche.

"L'incertitude inhérente à toute recherche fondamentale rend difficile sinon impossible une évaluation des résultats attendus" (NAS, 1982).

Tous les articles s'accordent pour évoquer la nécessité d'une évaluation subjective : ou "qualitative", comme l'évaluation par les pairs, les collègues ou les responsables hiérarchiques. Ainsi, pour la recherche fondamentale plus encore que pour la recherche-développement, les résultats d'une évaluation ex-ante pour sélectionner des projets, dépendra de l'évaluateur.

Cela veut dire que l'évaluation dépendra non seulement de la personnalité et des connaissances de l'évaluateur, mais aussi de sa position dans l'organigramme de l'entreprise, de sa fonction et de la confiance dont il jouit auprès de la direction.

Ainsi, plusieurs études ont reconnu un clivage très net entre les techniciens occupant des postes subalternes et les managers occupant des positions élevées.

L'EIRMA dans une étude sur les méthodes de gestion de la R et D s'est aperçu que les critères étaient différents selon que l'évaluateur était plus ou moins proche des centres de recherche des entreprises (EIRMA, 1979, op.cit.)

Pour surmonter cette difficulté l'EIRMA recommandait aux entreprises d'avoir un spécialiste de la R et D chargé de

N.A.S. The Quality of Research in Science, Methods for Post performance Evaluation in the NSF, Washington D.C., 1982.

défendre les projets auprès des "business managers" sans que cela nuise à sa carrière. On remarquera que de cette manière les auteurs de ce rapport préconisent l'institutionnalisation d'une fonction celle "d'avocat du diable" (ibid.) la transformation d'un "rôle" en "statut", pour employer le vocabulaire de la sociologie américaine. (1)

Les responsables de la Recherche dans un milieu industriel insistent beaucoup sur cet aspect d'intermédiaire du responsable de la recherche entre les chercheurs et la direction du travail. A ce sujet l'incompréhension entre Branscomb et Holton, entre le scientifique de l'industrie et l'universitaire, illustre bien notre propos.

Dans un séminaire sur "la qualité de la science" Holton reprochait à Branscomb d'avoir une vision quelque peu "paternaliste" : le PDG d'IBM imposerait à l'ensemble des activités sa propre vision des choses, les chercheurs n'étant qu'un "feedback" du bas vers le haut. Finalement le chercheur responsable du laboratoire ne serait qu'un signal sur la santé de l'entreprise, au même titre que le chiffre d'affaire est un signal sur les ventes.

A ce reproche, Branscomb répondait qu'il n'y a une vision "paternaliste" dans cette présentation que si on fait abstraction du rôle d'intermédiaire du directeur de la recherche, son rôle d'avocat auprès de la direction. En effet, n'importe quel demande d'investissement doit être justifiée. Or les justifications qui accompagnent la recherche ne sont finalement que peu convaincantes. La direction

"Les statuts sont les relations de rôle les plus complètement institutionnalisées, ou qui comportent le plus grand nombre d'éléments institutionnels" - GOODE, Norm commitment and conformity to Role-status obligations, American Journal of Sociology, Vol 66, 1960, p. 250.

n'engagera les dépenses que parce qu'elle sait qu'il lui faut investir dans la Recherche : elle n'accordera le montant des investissements dans les projets proposés que parce qu'elle a confiance dans ses chercheurs, et son directeur de recherche.

Ce qu'il y a de plus difficile dans l'évaluation de la recherche en milieu industriel est de faire concorder les objectifs et critères de la programmation des activités "d'affaires" des entreprises avec la planification des activités de recherche (EIRMA, 1980).

Terleckyj (1972) a montré dans le cas de la recherche médicale sur l'immunisation pour la fabrication de vaccins, que les objectifs du management (mesurés par les coûts unitaires des vaccins) peuvent être contraires à ceux des programmes (mesurés par des taux de maladie). Par exemple, pour maintenir un faible coût unitaire on peut ne pas vacciner une partie de la population qui est difficile à atteindre (1).

5. Méthodes d'évaluation

5.1. Méthodes non formelles d'évaluation

Nous pouvons distinguer deux façons de concevoir l'évaluation. L'évaluation où aucun critère numérique n'est utilisé et celle où les critères sont quantifiés.

La méthode de sélection / évaluation des projets sans critères numériques peut être illustrée par ce que les mana-

TERLECKYJ N.E. Productivity Analysis Tempered with Judgment Improves Efficiency. Defense Management Journal vol. 8 (3), oct. 1972, pp. 25-28.

- (1) il s'agit d'un problème plus général connu des économistes dans les modèles d'optimisation qui doivent respecter les "conditions aux limites", les objectifs d'un plan d'ensemble peuvent être concurrents voir, De Boissieu, Ch., Principes de Politique économique, Economica, Paris, 1978, pp. 36-42.

gers appellent le Management-by-objectives ou MBO.

Le "Management by Objectives" comporte deux idées. L'une est que les objectifs de l'entreprise doivent faire l'objet d'une décomposition de façon à ce que chaque "manager" se voit attribuer la responsabilité d'un objectif, l'ensemble des objectifs ainsi attribués devant contribuer à la réalisation des objectifs globaux de l'entreprise.

La deuxième idée est que la fixation des objectifs se fait dans la discussion, sinon l'accord, de l'ensemble des personnes responsables d'un objectif. C'est le côté participatif, participation restreinte tout de même aux seuls cadres des entreprises. La décomposition des objectifs devrait permettre une représentation chiffrée des moyens servant chaque objectif (les coûts ne doivent pas dépasser 5 % du budget).

Chiffré ou pas le MBO est considéré par ses partisans comme une "méthode d'évaluation interne" (voir Manuel de Karger et Murdick par exemple, pp. 416-418).

Leur argument peut se résumer ainsi : Il vaut mieux des objectifs négociés où l'accord s'effectue après discussion, que des objectifs issus d'un instrument formalisé qui seront toujours perçus comme des ordres imposés (Branscomb, 1982).

Le MBO a été énormément prisé à l'Office of Management and Budget. Il fut introduit par Malek et Ash sous la présidence de Nixon. Chaque agence fédérale soumet tous les ans son budget. Ce budget est alors discuté, c'est-à-dire que chacun des objectifs énoncés par l'agence est passé en revue. Une fois ses objectifs fixés et son budget déter-

miné, l'agence est responsable de l'accomplissement de chacun des objectifs. Des services intermédiaires ont lieu où les objectifs et l'état d'avancement des programmes sont examinés. (1)

D'après Malek en 1974, la NSF avait été l'une des agences ayant essayé d'être la plus précise dans la formulation des objectifs. Mais il ne s'agit là que du budget de la Fondation, pas de la sélection des projets (voir chapitre 4).

Cette vision de la fixation des objectifs est en fait une façon de dire aux agences fédérales qu'il faut fixer des objectifs précis et s'y tenir. C'était au moment de sa création une façon de décomposer de trop grosses agences (comme l'Office of Management and Budget) en de petites unités indépendantes, avec des délégations de pouvoir successives.

Comparant le Management by Objectives au PPBS, Malek disait qu'il voulait éviter ce qui a coulé le PPBS : le "S", c'est à dire le fait qu'il s'agissait d'un "system" (2) "nous ne voulons pas établir une technique de gestion, nous voulons établir un état d'esprit, celui qui permet de faire des affaires. Certaines agences comme la NASA, font cela très bien et ce n'est pas une nouveauté. Ce n'est pas le cas de toutes les agences" (op.cit., p. 284).

Comment "faire des affaires" c'était aussi, dans la tête des gestionnaires et budgétistes de Nixon, introduire

(1) John Welsh, The accent on Management, Science, v.183, 25 janvier 1974; p. 286-290. in T.J. KUEHN et A.C. PORTER, Science Technology, and National Policy, Cornell University Press, 1981, pp. 281-287.

(2) il y a un malentendu ici. "System" signifie plus technique que système en américain.

l'esprit du monde des affaires dans les agences fédérales, "une pression pour que le gouvernement soit plus "business-like" (Walsh op.cit., p. 281).

Une critique de cette méthode de management, notamment en ce qui concerne la science fut publiée dans Science (Culliton, 1974).

"Vu les relations qui existent entre la communauté scientifique et l'Office of Management and Budget, la forme que prennent les rapports des scientifiques ne satisfont ni les budgétistes de l'OMB, ni eux-mêmes, bien que l'OMB donne des consignes précises sur la forme que doivent prendre les rapports budgétaires (...). A chaque requête budgétaire sont joints les documents justificatifs qui sont censés expliquer pourquoi ce programme doit être financé. Ces documents qui sont loin d'être amusants à lire, (...) sont rarement convaincants et probablement ils ne sont même pas lus" (Culliton, 1974, pp. 290-291).

Le PPBS a été critiqué pour avoir provoqué une inflation de rapports. Le MBO est critiqué pour la même raison, A la NSF en décrivant le travail des bureaux des budgets il m'a été dit plus d'une fois que "les budgets sont discutés en détails, trop en détails".

En bref on peut dire que le MBO n'est pas une technique mais l'orchestration des négociations concernant les budgets.

Comme pour le "strategic planning" qui est une tentative de planification intégrée, l'un des objectifs avoués du

Barbara CULLITON, A skeptical view on Scientific Advice, in KUEHN et PORTER (1981), op.cit., pp. 287-295.

MBO est d'expliciter les étapes du processus de décision depuis la conception jusqu'à la production et la distribution. Un des sous produit de cette tentative est ce qu'on peut appeler un "tableau de bord". Ce sont là des documents récapitulatifs qui permettent de rassembler pour chaque fonction les diverses tâches qu'il faut accomplir (voir tableau ci après).

Le travail de chaque fonction est décomposé en étapes successives. Chaque étape peut bien sûr être quantifiée, si on le désire. C'est ce que prônent les manuels de management afin, disent ils, d'augmenter la productivité de la firme. La question qui se pose alors est : quel sera l'effet d'une telle quantification ?

L'une des choses que ne disent pas les articles sur le management de la R et D c'est que de telles méthodes ne peuvent fonctionner que dans des organismes très hiérarchisés. Pour cette raison on peut les qualifier de méthodes hiérarchiques d'évaluation.

L'évaluation hiérarchique par sa composante bureaucratique était quelquefois perçue comme un danger, même en milieu industriel. Déjà en 1912, Willis Writting le directeur du Laboratoire de la "General Electric" mettait ne garde ses collaborateurs sur les dangers d'une bureaucratisation de la recherche. Ce qui est étonnant c'est que ces avertissements avaient lieu au moment de l'apogée des laboratoires de la General Electric. "Mais le temps où de telles attitudes anti-hiérarchiques et individualistes pouvait avoir quelques effets, n'est plus. La première génération d'entrepreneurs de la R et D a été remplacée par une nouvelle race de

PHASE I - STUDY	Manager	Market- ing	Engineer- ing	Manufac- turing
1. Establish Business Objectives, Timing and Strategy	R			
2. Define Responsibilities and Relationships	R			
3. Establish Customer Requirements		R		
4. Establish Specific Product Requirements and Schedules	R			
5. Establish Product Design Philosophy			R	
6. Establish Manufacturing Philosophy				R
7. Establish Marketing Philosophy		R		
8. Select Preliminary Product Design Approach			R	
9. Identify Critical Problems and Alternate Approaches			R	R
10. Plan Feasibility Program Including Resource Requirements		R	R	R
11. Estimate Design and Production Program Resource Requirements		R	R	R
12. Review and Authorize Feasibility Program	R			
PHASE II - FEASIBILITY PROGRAM				
13. Establish Subsystem and Component Functional Design Requirements			R	
14. Establish Component Design Approaches			R	
15. Make Electrical Schematics and Mechanical Layouts			R	
16. Preliminary Product Design Review			R	
17. Propose Possible Manufacturing Methods for Critical Components			R	R
18. Determine Design Alternatives for Critical Components			R	
19. Producibility Review for Critical Components				R
20. Establish Design Specifications for Critical Components			R	
21. Determine Design Alternatives for Critical Manufacturing Processes			R	
22. Prove Feasibility of Critical Manufacturing Processes				R
23. Establish Design of Critical Manufacturing Processes				R
24. Plan Design Program Including Resource Requirements			R	R
25. Re-estimate Production Program Requirements		R	R	R
26. Review and Authorize Design Program	R			
PHASE III - DESIGN PROGRAM (PRODUCT AND PROCESS)				
27. Review Product Design Approach			R	
28. Revise Electrical Schematics and Mechanical Layouts			R	
29. Product Design Review			R	
30. Propose Possible Manufacturing Methods			R	R
31. Determine Product Design Alternatives			R	
32. Producibility Review				R
33. Complete Product Design Evaluation			R	
34. Finalize Product Design Specifications for Initial Production			R	
35. Complete Production Process Design				R
36. Establish Market Plan for Distribution and Service		R		
37. Prepare Production Program Plan Including Resource Requirements		R	R	R
38. Review and Authorize Production Program Plan	R			

Integrated program plan.

PHASE IV - INITIAL PRODUCTION	Manager	Market- ing	Engineer- ing	Manufac- turing
39. Release Product Design to Manufacturing			R	
40. Interpret Product Design Specifications			R	
41. Establish Production Equipment Methods and Procedures				R
42. Build Products to Established Product Design Specifications				R
43. Establish Product Installation, Maintenance and Operating Procedures		R		
44. Distribute and Service Products		R		
45. Evaluate Adequacy of Product Design Specifications			R	
46. Evaluate Production Samples			R	
47. Evaluate Manufacturing Process from Product Viewpoint			R	
48. Evaluate Manufacturing Process from Capability and Productivity Viewpoint				R
49. Evaluate Distribution and Service System		R		
50. Evaluate Customer Satisfaction		R		
51. Evaluate Business Performance of Initial Production	R			
PHASE V - CONTINUING PRODUCTION				
52. Propose Design Changes to Improve "Quality" or Reduce Costs			R	R
53. Propose Process Changes to Improve "Quality"			R	R
54. Propose Process Changes to Reduce Cost			R	R
55. Evaluate Proposed Design and Process Changes from Product Viewpoint			R	
56. Evaluate Required Process Changes from Productivity Viewpoint				R
57. Evaluate Proposed Design and Process Changes from Marketing Viewpoint		R		
58. Propose Design Change Program		R	R	R
59. Review and Authorize Design Change Program	R			
60. Establish Revised Product Design Specifications			R	
61. Establish Necessary Process Changes				R
62. Produce Redesigned Products				R
63. Evaluate Redesigned Products			R	
64. Distribute Redesigned Products, Spare Parts, and Service Information		R		
65. Evaluate Productivity of Redesigned Manufacturing Process				R
66. Evaluate Business Performance	R			

R - Primary Responsibility for Result

Integrated program plan. Source: "An Integrated Approach to Product Quality and Producibility," *Engineering Services* (Schenectady, N. Y.: General Electric Company).

Karger and Murdick (1980) Research and Engineering Management, New York, Industrial Press, pp.364-365

managers de la R et D" (Birrr, 1979, p. 205). (1)

Plus encore dans les organismes qui financent la recherche fondamentale, on peut voir à quel point ces évaluations hiérarchiques et budgétaires sont difficiles à concilier avec la hiérarchie implicite de la communauté scientifique.

Glaser (1964) (2) rappelait des travaux effectués au sein d'organismes de recherche en milieu industriel (3) où les auteurs avaient remarqué que la promotion des scientifiques ne se faisait pas en fonction de la réputation des chercheurs au sein de la communauté scientifique. Par contre, la promotion au sein d'organismes de recherche publics dépend de la réputation du chercheur dans sa discipline ou dans la communauté scientifique. (4)

Les problèmes qu'a eu à affronter la direction de la NSF sont très révélateurs à cet égard. D'une part les exigences de l'Office of Management and Budget poussait la direction à fournir un effort de quantification des objectifs de la Fondation. D'autre part la très forte tradition de l'évaluation par les pairs qui tient plus du jeu de la bourse que de l'organisation hiérarchisée s'opposait à

- (1) K. BIRRR industrial Research laboratories in Nathan Reingold (ed) The Sciences in the American Context: new Perspective Smithsonian, Washington DC, 1979, pp.193-207.
- (2) GLASER B.G. Organizational scientists: their professional careers. New York: Bobbs-Merrill, 1964, p.5.
- (3) William KORNHAUSER, Scientists in Industry (Berkeley: University of California Press, 1962),
 Leo MELTZER, "Scientific Productivity in Organizational Settings", Journal of Social Issues, XII, n° 2, (1956).
 Simon MARCSON, The Scientist in American Industry (New York: Harper & Brothers, 1960),
 Hebert A. SHEPARD, "Nine Dilemmas in Industrial Research", in Barber and Hirsch, p. 347. Science and Social order, New York, The Free Press, 1952.
- (4) Cette question a été examinée dans le détail dans R. ARVANITIS, L'évaluation des chercheurs aux Etats Unis: l'exemple du National Institute of Health, du National Bureau of Standards et de l'Université, CPE, février 1983.

de telles quantifications.

Les "program reviews" (voir chapitre 4), par exemple qui avaient lieu au sein de la NSF et où chaque division de la NSF devait plaider sa cause en présentant l'intérêt scientifique des recherches en cours et justifier ses demandes de budget devinrent vite une place où les scientifiques marchandaient pour obtenir le plus de crédits possibles. (les program Reviews furent d'ailleurs introduits à la NSF par un ancien de la NASA, dont les recherches plus orientées que celles de la NSF, pouvaient faire l'objet d'une discussion plus précise sur la définition des objectifs, tant scientifiques que financiers).

Cette situation devint intenable au point que récemment les program reviews ont été scindés en deux types de séances: Une séance de présentation orale sur l'intérêt scientifique et technique des programmes de la NSF et une séance de négociations budgétaires entre la direction centrale de la NSF (2 ou 3 personnes) et 2 ou 3, directeur et assistant de la division dont il est question.

On touche ici au deuxième élément occulté par les partisans du MBO. C'est le petit nombre de personnes qui, à chaque niveau de l'organigramme de la firme ou de l'organisme, participent à la discussion. Là encore, il y a un abîme qui sépare, le monde de la communauté scientifique plein de colloques, de discussions, de tables rondes et de cours prononcés devant des auditoires diversifiés et celui du management où les discussions se font par petits groupes de responsables.

Ici encore, il faut rappeler que malgré tout les efforts fournis par l'Office of Management and Budget une analyse qui tiendrait plus compte des réalités politiques et moins de phantasmes administratifs du management, révélerait une image du processus budgétaire où les rapports sont aussi invisibles que les collègues du même nom de la commu-

nauté scientifique.

Introduire "l'esprit des affaires" en science est probablement une utopie, les scientifiques sachant très bien devenir des commissaires priseurs ou des entrepreneurs (1) quand il le faut. Introduire l'esprit du management est par contre une ambition de contrôle qui déplaît fortement aux scientifiques car cela suppose d'implanter une hiérarchie en plus de celle qui existe implicitement déjà dans la communauté scientifique.

D'une manière générale, les auteurs qui étudient l'attitude des scientifiques et des managers dans un environnement industriel s'accordent pour opposer deux types d'attitudes : celle de l'entrepreneur (on peut même ajouter Schumpeterien) à celle du gestionnaire ou manager professionnel . Le premier aura tendance à voir l'entreprise comme un moyen de compétition, l'autre à y voir un organisme mécanique gouverné par un ensemble de lois et de règlements (Dailey, 1971). (2).

-
- (1) Voir par exemple dans le domaine des biotechnologies : G. MARKIE et S. ROBIN, The Scientific Ethic and the Spirit of Capitalism : Recombinant DNA Research for profit, papier présenté au Congrès de sociologie des sciences, Science and Technology, Retrospectives and perspectives, Philadelphia, 28-31 oct. 1982. Une orientation plus générale a été fournie par H. ETZKOWITZ Entrepreneurial scientists and the university.
- (2) DAILEY, G.A. Entrepreneurial Management, Mc Graw Hill N.Y., 1971, p. 208.

5.2. Méthodes formelles et quantitatives

Pour la plupart élaborées dans les années soixante, les méthodes dites quantitatives ont surtout été développées dans des industries ou des laboratoires fédéraux et des agences principalement tournés vers la R et D (NASA, Department of Defense, etc...). Elles concernaient plutôt les projets technologiques à grands budgets que la recherche dans des petites unités. De plus, les méthodes quantitatives étaient plutôt orientées vers le contrôle de projets déjà commencés que l'évaluation/sélection a priori des projets. Les méthodes dites de "network" font plutôt partie de ce type de méthodes de contrôle (1).

En ce qui concerne cependant l'évaluation/sélection a priori des projets de recherche, Moore et Baker (2) (1969) ont répertorié 4 types de méthodes de sélection

- les méthodes de notation
 - les méthodes économiques et financières
 - les méthodes d'optimisation
 - les méthodes d'analyse du risque, ou méthodes probabilistes,
- D'autres classifications des méthodes de sélection des projets ont été proposées dans la littérature managériale

-
- (1) PERT (Program Evaluation Review Technique), CPM (Critical Path Method), PEP (Program Evaluation Procedure), VERT (Venture Evaluation Review Techniques) sont des exemples de ces méthodes. Leur nom "network methods" vient de ce qu'elles permettent de dessiner des relations complexes entre divers éléments, des réseaux (network) (CZAYAKA cité plus loin, p. 53).
- (2) MOORE Jr J.R. and BAKER N.R. "Computational Analysis of Scoring Models for R and D Project Selection", Management Science, vol.16, n° 4, December 1969, pp. B212-B232. MOTTLEY C.M. and NEWTON R.D. "The Selection of Projects for Industrial Research", Operations Research, vol.7, nov.-december, 1959, pp. 740-751.

qui sont moins exhaustives (1). Aux quatre catégories énoncées ci-dessus il faut ajouter les méthodes de classement (ranking) (Clarke, 1974, op.cit.)

Les méthodes de classement

Les méthodes de classement ont des appellations très diversifiées. Le principe est de sélectionner les projets deux à deux. On compare donc soit deux projets, soit deux groupes de projets et l'on retient l'un des deux qui est ensuite comparé à un autre etc... jusqu'à ce qu'un des projets soit sélectionné. Ce type de sélection de projets n'implique pas la notation des projets.

Les méthodes de notation (2)

Les méthodes sont fondées sur une notation de chaque projet proposé. Certains critères sont définis pour chaque projet, puis noté par un groupe de pairs ou de collègues, C'est la méthode qui est habituellement choisie dans la sélection de projets de recherche fondamentale. La méthode de l'évaluation par les pairs utilisée à la NSF et aux NIH sont des méthodes de notations.

- (1) COCHRAN et al. (1971), COCHRAN M.A., M.A. Pyle III, E.B., GREENE, L.C., CLYMER H.A. and BENDER A.D. "Investment Model for R and D Project Evaluation and selection", IEEE Transactions of Engineering Management, vol.EM-18,n°3, August? 1971, pp.89-100.
BREENBLOT B.J. (1970) Astructure for Management. Decision Making IEEE TRansactions on Engineering Management, EM 17 (4), nov.1970, pp.145-158.
- (2) On peut trouver une bonne présentation de ces méthodes d'évaluations formelles dans Research Policy, 1 (1), Nov. 1971 due à L. CZAYKA, "The importance of graph theory in research planning, pp. 60-87. On peut y lire notamment une critique des méthodes arborescentes (68-73) du type PATTERN, PROFILE ou PPBS. L'auteur prône l'utilisation des graphes pour les mêmes raisons que celles invoquées par les ingénieurs qui proposent les méthodes quantitatives : cela permet d'explicitier, clarifier les notions et les critères d'évaluation.

Cependant si les méthodes de sélection par notation proposées dans les articles de gestion de la R.D. sont semblables à l'évaluation des projets du NIH et de la NSF car toutes deux impliquent une notation, ce n'est là qu'une ressemblance superficielle. En effet les méthodes de notation comme celles proposées par Mottley et Newton (1959) sont des notations hiérarchiques. La méthode du "project grant" de la NSF est une méthode qui s'apparente plus au modèle du marché avec commissaire-priseur, comme le marché des valeurs boursières. Il y a donc lieu de distinguer entre les méthodes hiérarchiques et les méthodes que l'on peut qualifier de "concurrentielles" (*)

Concernant ces méthodes de notations hiérarchiques, Moore et Baker (1969) ont remarqué que la structure de la méthode peut avoir une très forte influence sur la décision finale. La note finale attribuée à chaque projet peut ainsi être totalement différente selon que l'on additionne les notes de chaque critère ou selon que l'on multiplie ces notes intermédiaires. Ce n'est bien sûr qu'un rappel des complexes problèmes qu'étudie la métrologie.

L'une des méthodes de notation qui est le plus souvent citée dans la littérature est la méthode proposée par Newton et Mottley (1959). Il s'agit d'une notation de quelques critères, formulés sous forme de questions, chaque

(*) qui, de plus, s'apparente fort au modèle de concurrence pure et parfaite introduit en théorie économique par Walras, puisque toutes les informations concernant chaque projet sont supposées connues et puisque aux yeux des offreurs et demandeurs le choix effectué est optimal car réalisent "l'optimalité de situations des micro-sujets", J.P. FITOUSSI Le fondement micro-économique de la théorie Dynamique, Cujas, Paris, 1974, p.30. Il ne faut cependant pas pousser l'analogie plus loin car la NSF est simultanément le secrétaire du marché et le seul offreur de fonds.

réponse reçoit une note (voir tableau ci-après). Ces notes sont multipliées pour obtenir la note finale pour chaque projet.

Une autre méthode assez réputée est la méthode PATTERN . Honeywell prétend l'utiliser couramment . Il s'agit d'un arbre de décision. A chaque niveau de l'arbre on effectue une notation en fonction de critères définis à l'évidence et qui sont applicables à tous les niveaux de l'arbre. Puis on multiplie ces notes le long des branches de l'arbre. La méthode PATTERN fut proposée en 1960.

Les méthodes économiques

Les méthodes économiques ont été les premières conçues. Elles s'appliquent bien à l'évaluation de produits nouveaux, ou à la R.D. notamment quand les projets sont proches de l'aboutissement. Ces méthodes sont largement utilisées par les entreprises (*)

On peut distinguer deux types de méthodes économiques d'évaluation:

a) Rendement escompté par unité d'investissement

Il s'agit du calcul d'un indice de rendement.

* EIRMA (1979).

Exemple de notation d'après Mottley et Newton

RATING SYSTEM

Criterion	Question	Range of answers	Numerical rating
Promise of success (P)	What is the best estimate of the promise of technical success consistent with known economies and the state of the art?	Unforeseeable	1
		Fair	2
		High	3
Time to completion (T)	How long will it take to complete the research effort from this time forward?	Greater than 3 years . .	1
		1 to 3 years	2
		Less than 1 year.	3
Cost of project (C)	How much will it cost to complete the research effort from this time forward?	Greater than \$1 million .	1
		\$100,000 to \$1 million. .	2
		Less than \$100,000 . . .	3
Strategic need (N)	To what extent is successful research needed from a market standpoint?	No apparent market application; must be developed.	1
		Desirable to maintain, reinforce, or expand position within market applications currently served	2
		Essential in relation to current or projected markets within market applications not currently served.	3
Market gain (M)	What is the net market gain potential for the company after taking into account losses through product replacement?	Less than \$1 million/yr.	1
		\$1 to \$10 million/yr. . .	2
		Greater than \$10 million/yr.	3

Rating system for selection of industrial research projects. Source: C. M. Mottley and R. D. Newton, "The Selection of Projects for Industrial Research," *Operations Research* (November-December 1959).

$$\frac{\begin{matrix} \text{(valeurs estimées)}^* \\ \text{(au cas de succès)} \end{matrix}}{\text{Coût estimé de la recherche}} \frac{\begin{matrix} \text{(Probabilité de)} \\ \text{(succès)} \end{matrix}}{1} = R$$

C'est un indice fondé sur la comparaison des recettes et des dépenses, flux auxquels on associe une probabilité de réussite (F. Olsen et O. Mathinson cités par Karger et Murdick 1980 et Rubinstein, 1963).

Sur ce modèle, plusieurs variantes ont été proposées. On peut grossièrement en distinguer deux qui concernent l'estimation du succès du projet.

Succès technique (le projet de R et D a x % de chances d'aboutir à une innovation) et celles qui insistent sur la publicité ou succès commercial (le produit ou procédé nouveau a y % de chances d'être rentable une fois commercialisé). Là encore on voit deux conceptions différentes, les "deux bouts" que doit faire se rejoindre au sein d'une entreprise : l'aspect technico-scientifique et l'aspect économique-commercial.

b) Indices d'actualisation

Ceux-ci sont plus connus en France, puisqu'ils ont fait l'objet de plusieurs discussions pendant la fabrication du VII^e Plan (CGE 1969)(1). Issus de la "théorie pure de l'investissement" élaboré par les économistes, les indices d'actualisation sont tous fondés sur un principe très simple et connus des financiers et comptables du monde entier.

Il n'y a pas lieu d'insister sur ces méthodes d'actualisation. Elles servent très souvent à déterminer la part

-
- (*) Ce peut être aussi par exemple : la valeur des économies réalisées grâce à ce procédé nouveau, sur 3 % du C.A. dû à la vente du produit nouveau sur 5 ans, ou 2 % des ventes du produit, etc... (Rubinstein, 1963)
- (1) Commissariat général au Plan, l'Entreprise face à la décision d'investir, 1969, et Calcul Economique et planification, Economie et Planification, mars 1973.

des dépenses de R.D. allouée à chaque projet et aussi pour déterminer la rentabilité intrinsèque d'un investissement. C'est une méthode qui, formellement, ressemble aux méthodes de détermination des taux d'intérêts et des échéanciers bien connues des analystes financiers. Ceux-ci sont, il faut le rappeler, la fonction la plus ancienne et la mieux établie au sein des entreprises. De plus, les méthodes d'actualisation évitent de porter tout jugement sur les individus.

Mieux comprises des gestionnaires, les méthodes économiques sont très largement répandues. Elles sont moins comprises des scientifiques et techniciens qui sont plus méfiants sur leur utilité.

Les méthodes dites du coût-bénéfice sont proches des méthodes économiques. Mais la littérature s'accorde pour rappeler la faible utilité des méthodes de coût-bénéfice appliquées à la recherche (1).

Méthodes d'optimisation

Ce sont des méthodes où il faut maximiser une fonction-objectif sans contraintes. La procédure de calcul est complexe et s'appuie le plus souvent sur des mathématiques de la programmation linéaire (2).

L'une des plus répandue est la méthode TORQUE. Torque est une méthode pour déterminer le niveau des fonds que l'on peut allouer à un projet. (Voir NUTT A.B. Testing

(1) Voit notamment ANDREWS (1979) Scientific Productivity Cambridge U. Press et UNESCO, ch. 2.

(2) SOUDER, W.E. "Comparative Analysis of R & D Investment Models", AIIE Transactions, vol.4, n° 1, march, 1972, pp.57-64.

Torque, IEEE/TEM, vol. EM-12, March 62 : 9-13). On établit une note d'ensemble selon la formule (f. d'utilité du projet)

$$U = \sum_{i=1}^N C_i W_i C_f t_i$$

N = nb de "systèmes" pour un niveau donné de difficulté

C_i = "Criticality" du système i (noté de 0 à 1)

W_i = pondération du système i

C_f = fonds alloué la première année sur total des fonds requis.

T_i = fonction du temps, non-définie par Nutt mais qui peut être définie par chaque équipe utilisant la méthode. Cette fonction d'utilité est ensuite maximisée sous contraintes.

Enfin, les méthodes d'analyse du risque sont des simulations des données d'input et permettent de tracer une distribution des avantages attendus, comme par la part de marché ou le chiffre d'affaire (Ayers, 1969). Mais l'utilisateur d'un calcul de probabilité a été remis en cause plus d'une fois "car en principe, on n'observe pas de distributions de fréquences concernant le succès ou l'échec des projets de R et D" (Czayka, 1971, op.cit., p. 85). Cependant une méthode de simulation du management des projets est issu des méthodes de network, il s'agit de VERT (voir plus haut p. 47)

5.3. L'utilisation des méthodes formelles d'évaluation hiérarchique

Toutes ces méthodes d'évaluation peuvent être appelées hiérarchiques, D'une part car ce sont les supérieurs hiérarchiques qui effectuent les évaluations, c'est-à-dire

AYERS, R.U. "Technological Forecasting and Long-Range Planning", New York, McGraw-Hill Book Co, 1969.

les responsables des laboratoires et équipes de recherches
La direction décide en fonction de la confiance qu'elle
accorde aux directeurs de recherche.

D'autre part, car le processus de l'innovation est
conçu de manière séquentielle, des évaluations devant être
réalisées à chaque niveau (voir le schéma de Gloskey page
suivante qui a fait fortune dans la littérature spécialisée).

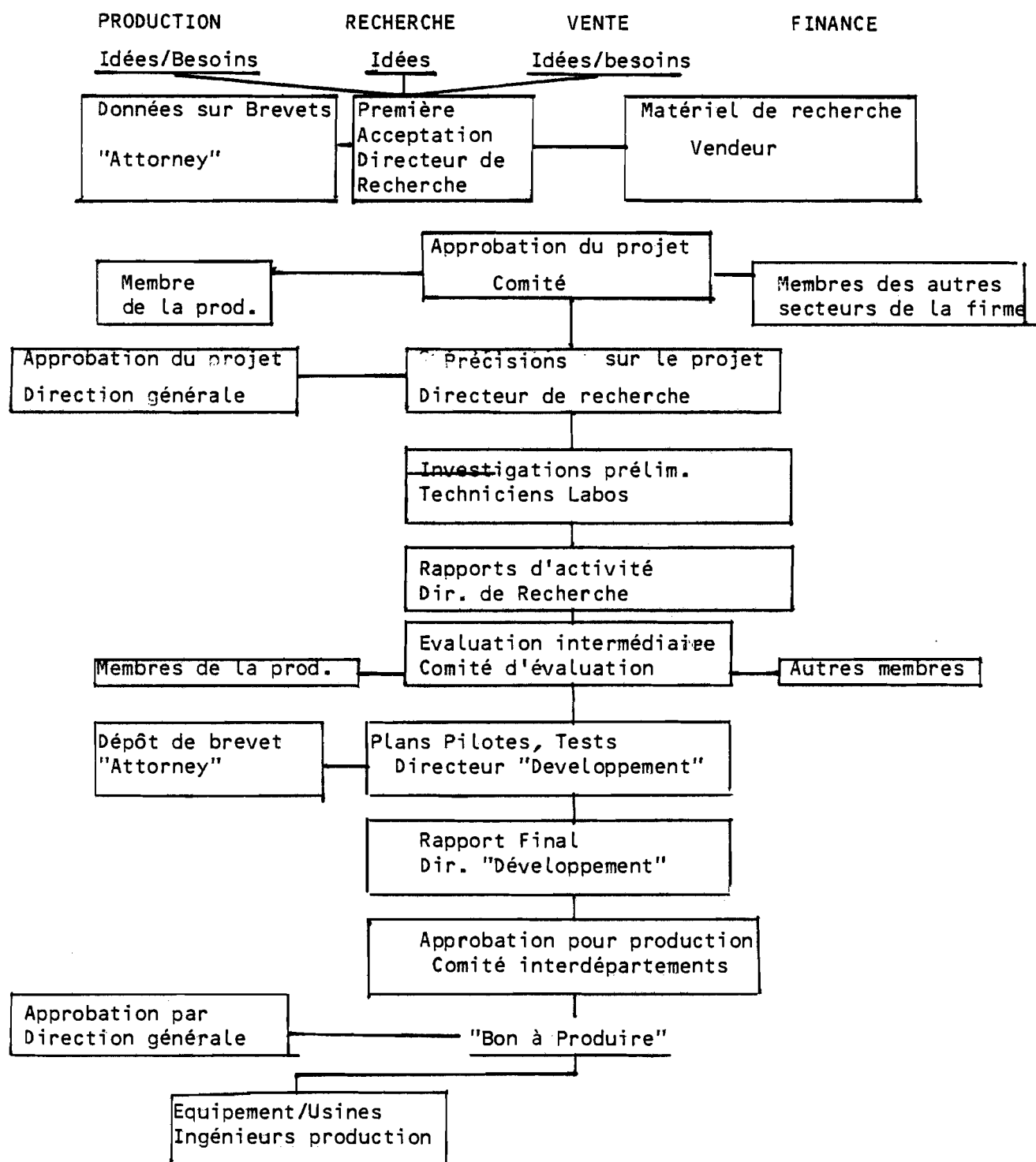
Plusieurs des méthodes ci-dessus sont issues
de la Recherche opérationnelle où "l'ingéniererie était
considérée comme une entreprise héroïque conduite à grande
échelle, dans des grands organismes pour produire des tech-
nologies de pointe" (Rettig et Wirt, 1978, p.33) conception
qui peut être résumée dans ce précepte: "il faut trouver une
solution technique quel qu'en soit le prix" (ibid.)

Ces modèles ont été critiqués à plusieurs reprises,
même par les ingénieurs gestionnaires. Ainsi Samuel Eilon,
rédacteur en chef de la très sérieuse revue Omega dit
que les méthodes d'optimisation reposent sur une conception
fallacieuse de la maximisation qui suppose qu'un change-
ment dans les données, n'implique pas un changement dans
la nature des relations entre ces données (S. Eilon, 1977,
p. 623).

De plus ces modèles supposent des méthodes mathémati-
ques trop compliquées et souvent pèchent par la mauvaise
définition des problèmes.

Ces méthodes sont coûteuses, car elles nécessitent
Samuel Eilon, More Against optimization.
OMEGA, The International Journal of Management Science, 5(6)
1977, 627-633.

Modèle de décision séquentiel
de Gloskey



l'emploi d'ordinateurs et elles sont longues à mettre sur pied. Le coût de PATTERN à Honeywell a été estimé à 250.000 \$ par an (Mansfield, 1982, op.cit.)

Toutes les méthodes signalées plus haut nécessitent des estimations des coûts, du temps nécessaire pour les projets, des prix, du succès technique, etc...

Les estimations sont souvent fausses d'après plusieurs études mentionnées par Clarke (1974) (1), surtout en ce qui concerne les estimations de temps et de coût. Généralement les coûts sont largement sous estimés (2)(3) Ceci est souvent dû à la volonté des chercheurs qui veulent que leur projet soit "acheté" par la direction (Mansfield, 1982). Par contre dans des laboratoires R.D. importants, certains auteurs ont remarqué que les prévisions concernant des aspects techniques des projets sont correctes. (Souder, (4) Mansfield et Brandenburg) (5)

La plus grande difficulté concernant les estimations sur le temps nécessaire et les coûts et donc les rendements escomptés, vient de ce que les distributions des fréquences

- (1) Thomas H. "Some evidence on the accuracy of Forecasts in R.&D Projects", R&D Management, vol.1,n° 2,Fébruary 1971, pp.55-69 qui cite lui même plusieurs études avec des données quantitatives.
- (2) Allen D.H. and NOVEMBER P.J. "A Practical study of the Accuracy of Forecasts in Novel Projects", Chemical Engineer, n° 229, June 1969,
- (3) MEADOWS, D. 'Estimate Accuracy and Project selection Models in Industrial Research', Industrial Management Review, vol.8, n°3, Spring, 1968, pp.105-121.
- (4) SOUDER W.E. "The validity of subjective Probability of success Forecasts by R & D Project Managers", IEEE Transactions on Engineering Management, vol.EM-16, n°1, February 1969, pp. 35-49).
- (5) MANSFIELD Edwin and BRANDENBURG Richard. The Allocation Characteristics and Outcome of the firm's Research and Development Portfolio : A case study", The journal of Business, vol.39, n° 4, october, 1966, pp.447-464.

des divers paramètres sont inconnues. Un manuel à l'usage des managers rapportait une anecdote évocatrice à ce sujet. Un manager après avoir vu les résultats issus d'une simulation par ordinateur, de l'exécution d'un des projets déclarait : "personne ne m'achètera ce truc, il va falloir changer quelques variables" (Karger et Murdik, 1980, p. 294).

Les méthodes formelles sont peu utilisées. C'est là un constat généralisé. Dean a étudié 34 firmes qui sans utiliser des méthodes formelles, prétendent employer des "critères quantitatifs". Son étude a montré que ces firmes utilisaient de simples méthodes de notation. Moore et Baker ont également remarqué que seule la méthode de notation du type de Newton-Motlley est utilisée. L'étude menée par Gee a révélé que très peu de méthodes informatisées étaient employées. Les industries pharmaceutiques que Faust a étudié ont cherché de temps en temps à utiliser plusieurs méthodes parmi celles citées ci-dessus mais "on a toujours douté de leur utilité" (Faust, 1971) (1)

Cetron et Ralph (1971) (2) notaient que 20 % des firmes qu'ils avaient interrogées avaient testé des méthodes d'optimisations ou de simulation et que seules 10 % de ces firmes utilisaient l'une de ces méthodes couramment.

Dans un article récent, Mansfield (1982) lui aussi rappelait que l'on ne sait toujours pas à quel point les méthodes quantitatives sont ou non utilisées aux Etats Unis (p.25)

 (1) FAUST R.E., "Project selection in the Pharmaceutical Industry", Research Management, vol.14, n° 5, september 1971, pp.

(2) cités dans Clarke, 1974.

Enfin, l'étude de l'UNESCO (Andrews, 1979, op.cit.) confirme également que la planification de la science et des activités de laboratoires est rarement effectuée au moyen de modèles formels"... les méthodes formelles de planification - ex. : analyse d'opportunité, modèles dynamiques, analyse des systèmes analyse du risque, etc. - et les méthodes de programmation - ex. : PPBS, PERT, etc. - ne sont pas utilisées fréquemment dans les 6 pays étudiés. (les méthodes de programmation sont encore moins utilisées que les méthodes de planification), et aucune relation n'a été remarquée entre la fréquence d'utilisation des méthodes de planification et celle des méthodes de programmation". (Haraszthy, Szanto, in Andrews, 1979 : 158). Les auteurs de cette étude ne fournissent pas de renseignements plus précis sur les causes invoquées par les directeurs de recherches (les renseignements proviennent d'eux pour l'essentiel).

Clarke (1974) cite 6 enquêtes menées auprès de laboratoires de R et D ou de RF en milieu industriel ou dans des laboratoires fédéraux qui montrent que les méthodes formelles sont rarement utilisées. Wirt et al. (1975) rappellent les conclusions des travaux de Backer et Pound, Rubinstein et Souder. Ils disent que les mêmes conclusions s'imposent toujours. "Aucun article dans la littérature ne fait état de l'utilisation de ces méthodes, ni des raisons pour lesquelles elles ne sont pas utilisées ... En suivant de près le domaine de l'évaluation de la R et D depuis 1950, on se rend compte que les méthodes quantitatives utilisées dans les organismes de R et D sont à peu près les mêmes et toutes aussi peu utilisées".

Rubinstein - et depuis les autres auteurs de ce domai-

ne - parlent des causes probables. L'une est la complexité des formulations mathématiques, l'autre est l'éventuel manque de données numériques (voir plus haut p. 30-32)

Plus récemment deux études ont été effectuées qui illustrent et confirment les différents aspects des méthodes d'évaluations des centres de R et D ; principalement en milieu industriel.

La première enquête a été menée en Europe et aux USA par l'EIRMA (1979), la deuxième a été dirigée par A. Schainblatt (1982) et concerne les centres de R et D américains.

D'après l'EIRMA ce qui prime dans la programmation de la R et D, c'est l'organisation des centres de recherches et l'établissement d'un bon système relationnel et de communication ; les méthodes de sélection et d'évaluation des projets sont tout à fait secondaires (p.38). Les méthodes les plus sophistiquées sont les moins employées (voir tableau ci-après). La prévision (ou prospective), l'identification des priorités et les problèmes d'allocation des ressources sont des problèmes où se fait sentir le besoin de techniques nouvelles de planification (p.89). 1/3 des compagnies de l'EIRMA emploient pour se faire des panels d'experts (p.39).

EIRMA, Systems and Methods for Planning R and D in Industry, 1979.

SCHAINBLATT Alfred H. How companies measure the production of engineers and scientists, Research Management, vol.25, (3), May 1982, p. 10-18.

Tableau issu de l'étude de l'EIRMA

QUESTIONNAIRE REPLIES ON THE USE OF SPECIAL TECHNIQUES IN R & D PLANNING.

Those methods and techniques frequently named as 'often in use' by the R & D function as an aid to preparing and/or controlling the execution of R & D plans are :

	Percentage and/or number of the 38 organizations naming the tech- nique as 'often in use'	
	%	Number
- profitability methods	50	19
- cost/benefit analysis	50 50	19
- project form/checklist	50	18
- market research	45	16
- resource/capacity allocation	40	16
- bar-charts	35	13
- estimating credibility of technical success *	35	13
- statistics, surveys and fore- casts of national and inter- national organisations	35	13
- expert panels	30	12
- estimating credibility of commercial success *	25	10
- portfolio concepts	25	10

Some techniques named by only few organisations are :

- network techniques (PERT CPM, Etc.)	6
- multicriteria decision-making techniques	5
- mathematical methods (expo- nential smoothing, curve-fitting etc.)	3
- scenario writing	2
- modelling, simulation techni- ques	1
- decision analysis, decision tree concepts	1
- Delphi techniques	1
- relevance trees	0

As might be expected, the use of these methods and techniques is not evenly distributed between all Companies, but tends to be concentrated in relatively few, mainly those with the larger R & D organisations. Further-more, these figures do not necessarily represent the relative significance to the user.

* See also EIRMA Working Group Reports n° 6 & 16 - Methods of Evaluation of R & D Projects, vol.V, "Credibility of Technical Success : its definition, estimation and utilisation", 1973. Vol. VII, "Credibility of Commercial Success : its definition, estimation and integration, with credibility of technical success". 1975.

Les méthodes financières sont les plus usitées (p.45). Les méthodes arborescentes considérées les meilleures sont les moins utilisées (p.46). Les méthodes qualitatives (descriptives ou méthodes de notation) peuvent être utiles, dit l'EIRMA, mais pas un modèle de décision fondé sur une telle méthode (49). Les notations doivent être comparatives

Enfin toutes les méthodes économique-financières sont souvent fausses et inapplicables car les problèmes de R et D ne peuvent être décrits qu'imparfaitement par des méthodes financières (49).

La deuxième étude est une enquête auprès des firmes industrielles pour recenser le type d'instrument quantitatif utilisé pour mesurer la productivité des laboratoires de R et D. Là aussi les résultats ne montrent guère une large utilisation de méthodes quantitatives.

En effet, l'enquête dirigée par A. Schainblatt, portant sur 84 firmes US a montré que seules 7 firmes utilisaient un système quantitatif (dont Détroit - Edison Philips-Pétroleum et Borg-Warner).

Détroit-Edison emploie deux indicateurs quantitatifs : le nombre de "conceptions" (drawings) par ingénieur par homme/heure et le nombre de projets, pondéré par un "facteur de complexité" par ingénieur par employé/heure.

A Borg-Warner on effectue deux types d'évaluations. Une évaluation interne fondée sur le jugement d'experts de la compagnie de la qualité du travail d'une équipe qui travaille pour la compagnie (note de 0 à 3) ne évaluation du potentiel du marché du produit, du coût total du produit pour que le consommateur puisse le faire fonctionner (il s'agit d'éléments ou de procédés de production); du chiffre d'affaire si le prix du produit ou du procédé était égal à ce coût de fonctionnement; de la valeur de l'unité de

recherche qui est le coût de fonctionnement de tous les produits inventés Par l'unité de recherche divisé par les charges de cette unité de recherche.

Philips-Petroleum calcule tous les deux ans des indicateurs d'outputs qui sont essentiellement financiers, issus de la comptabilité analytique. En plus, un indice de productivité est calculé (nombre d'heures travaillées par produits inventés). Mais un responsable de cette compagnie ajoutait que "ces indices de productivité sont utilisés pour satisfaire une demande de la direction concernant la création d'un système de mesures de productivité intégrées au système de management" (...) "mais le personnel technique (et scientifique) et l'ensemble des employés résistent à ce type d'évaluation, et se sentent très mal à l'aise quand on les "mesure" de la sorte "(ibid., p. 14).

Certaines compagnies comptent le nombre de brevets par scientifiques et ingénieurs et par employés, combiné avec un jugement interne d'experts permettant de pondérer les brevets leur attribuant une note de 1 à 5.

Bell a effectué une évaluation rétrospective de ses laboratoires. Ils ont repris 683 innovations, commercialisées ou non, par leurs soins de 1927 à 1977, qui étaient le résultat positif ou négatif de recherches dans leurs laboratoires. Pour chaque innovation cela leur a permis d'établir le coût du programme de recherche par le nombre de bonnes innovations. Ce type d'indice cependant reste très global (*)

(*) McKay, J.G. The measure of Innovation Bell Laboratories Record July/August, 1980.

Schainblatt a aussi remarqué que certaines firmes distinguent la gestion de la recherche fondamentale, la gestion de la R et D et la gestion des activités d'engineering (p. 15). Sans dévoiler les noms de ces firmes, Schainblatt remarque que ce ne sont pas les mêmes types d'évaluations qui sont alors menées. Dans la perspective adoptée par l'auteur, celle d'un ingénieur-manager, ces évaluations quantitatives sont "disparates". Mais c'est précisément cela qui les rend intéressantes.

Les évaluations quantitatives sont

. en Recherche Fondamentale :

- jugement de panels d'experts sur les ressources et les impacts des recherches
- citations dans la littérature scientifique
- brevets
- notation des projets, qui "doivent également, tenir compte de la probabilité d'une découverte et de sa valeur potentielle" (dixit sur des managers interrogés).

. en Recherche-Développement :

- indices d'allocation des ressources
- nombre de tests expérimentaux effectués par les scientifiques / ingénieurs,
- nombre de techniciens par scientifiques/ingénieurs.
- nombre de techniciens et autres employés/ouvriers par prototypes par ingénieurs/scientifiques.

. en Gestion des activités engineering

- charges
- besoins en capital humain

Enfin, l'auteur propose de distinguer les "algorithmes quantitatifs" des mesures de productivité. Il évoque notamment la confusion systématiquement entretenue entre ces deux types d'évaluations, formelles et quantitative. ce que nous remarquons ici même (p. 48) en distinguant les les méthodes de sélection, des méthodes de contrôle.

Il en conclut qu'il n'existe actuellement aucune mesure de la productivité des scientifiques/ingénieurs et des laboratoires qui soit satisfaisante et rappelle que la plupart des gestionnaires interrogés sont très sceptiques sur la possibilité même de mesurer la productivité en matière de recherche. Plusieurs gestionnaires considèrent que ces mesures de productivité sont insignifiantes et les qualifient de "bean counting".

Pour la plupart des gestionnaires interrogés, c'est la comparaison entre les budgets réalisés et les budgets prévus qui est la plus significative. Schainblatt qualifie cette méthode de gestion par étape (management - by - milestone) et cite quatre études de ce type (*), dont le principal défaut, comme nous l'avons déjà dit, est la difficulté d'estimer les inputs nécessaires et la probabilité de succès ou d'échec des recherches.

- (*) BAUMGARTEL G.P. and JOHNSON T.D. Productivity Measurement in a Base Level USAF Civil Engineering Organization; Master's thesis presented to Air Force Institute of Technology. Air University, School of Systems and Logistics. Wright-Patterson AFB, OH. June 1979.
 ERICKSON J.M. A firm Model for productivity Assessment and White Collar Productivity Measurement, Masters Thesis, University of Minnesota, october 1980.
 TAUSS K.H. "A Pragmatic Approach to Evaluating R & D Programs". Research Management, September, 1975, 13-15.
 ZINK D.A., "Monitoring the Adequacy of the Amount and Productivity of Engineering and Construction Manpower" 1980 Transactions of the American Association of Civil Engineers, Morgantown W.V. American Association Cost Engineers, 1980.

Des méthodes quantitatives que nous venons d'évoquer il faut retenir la distinction très nette qui existe entre :

- variables temps (ou temps de travail)
- variables monétaires
- variables de performance.

Comme nous le verrons les études sur l'évaluation des performances des unités de recherches *ne distinguent pas les variables de performances des variables temporelles et monétaires.*

Trop souvent, les variables de temps et les variables monétaires sont utilisées comme des variables de performance. Ainsi l'étude de Baumgartel et Johnson (1979, op.cit) mesure le nombre d'heures productives par rapport au nombre d'heures improductives (tâches administratives). La distinction entre temps productif et improductif n'a cependant pas beaucoup de sens car, comme Pelz et Andrews (1966) l'ont montré, les scientifiques les plus performants sont aussi ceux qui passent 1/4 à 1/3 de leur temps à des tâches administratives. Ceci peut être aussi le résultat de la position qu'occupent les scientifiques au sein de l'organisme de recherche. Il est bien évident qu'un "undergraduate" ne remplira pas les formulaires de contrôle, les demandes de financement etc... Il exécutera plutôt les tâches dites "productives", car nécessaire au projet de recherche mais qui seront aussi les moins intéressantes.

Pour que des mesures du temps productif aient un sens il faudrait pouvoir distinguer le temps "intéressant" et innovatif et le temps où l'on n'effectue que des tâches de routine (disséquer une souris, vérifier la pureté d'un matériau, etc...) ce qui semble impossible.

PELZ D.C. ANDREWS F.M. Scientists in Organizations, John Wiley and Sons, 1966.

De même les variables monétaires ont moins de sens pour la recherche, son contenu, sa qualité que pour les financiers et la comptabilité de l'entreprise - on peut, à ce propos rappeler que l'étude de l'UNESCO, n'a remarqué aucune corrélation significative entre les fonds attribués aux unités de recherche et les indices subjectifs de qualité (voir plus loin).

Pour finir, rappelons que de très grandes firmes innovatrices, telles IBM, EXXON ou DU Pont qui financent une part considérable de la recherche fondamentale dans les entreprises, font reposer l'essentiel de leur capacité d'évaluation sur "la compréhension des individus" et non pas sur l'évaluation des recherches déjà accomplies. Cette confiance dans les capacités des chercheurs diffère de personne en personne, d'entreprise à entreprise. Ce flou même s'il n'est pas aussi présent dans la recherche la plus appliquée, demeure la cause essentielle de toutes les confusions et tous les échecs des tentatives de quantification qu'ont essayé les managers.

Conclusion

L'évaluation et la sélection des projets de recherche en milieu industriel se fait pratiquement de la même façon que l'évaluation des projets en milieu universitaire : elle repose plus sur le jugement d'experts que sur des évaluations quantitatives. Les experts en question sont des scientifiques des firmes plutôt que des scientifiques externes. Seules les plus grandes entreprises ont recours à des évaluateurs externes (IBM, General Electric, Ford, etc...). Ceci semble lié au fait que ce sont ces entreprises qui effectuent des recherches plus fondamentales, mais c'est aussi un phénomène probablement lié au coût plus élevé de l'expertise externe, coût que seules les plus grandes entreprises semblent prêtes à financer.

Plusieurs firmes ont tenté de mettre au point des méthodes quantitatives, mais dans la plupart des cas ces méthodes ont échoué. Parmi les causes de ces échecs nous pouvons rappeler :

- l'incertitude élevée du succès ou de l'échec d'un projet, qui rend difficile sinon impossible l'estimation des coûts et des bénéfices futurs pour la firme.
- l'utilisation de critères de sélection qui ont plus de sens pour les gestionnaires et les administrateurs que pour les scientifiques. Ceci explique en grande partie le refus des scientifiques d'être évalués par des personnes qui n'ont pas de savoir technique et scientifique. La plupart des méthodes quantitatives sont dérivées des méthodes de contrôle financier et comptable, et de ce fait elles peuvent difficilement distinguer les variables monétaires des variables de performance scientifique.
- les méthodes quantitatives sont souvent difficiles et longues à concevoir ; leur coût est souvent élevé, trop

élevé, aux yeux des gestionnaires.

- Enfin, on reproche aux méthodes quantitatives d'empêcher le jeu des subjectivités, c'est-à-dire de remplacer la subjectivité et la difficulté de l'évaluation par des hommes par une évaluation austère, mécanique, voire inhumaine. Cette objection n'est pas fondée, car l'utilisation d'indicateurs quantitatifs dans des domaines tels que la comptabilité et les contrôles financiers montre à quel point les personnes qui sont "mésurées" et celles qui mesurent continuent à discuter, controverser et modifier la mise au point et le fonctionnement de ces indicateurs (1).

Il existe cependant une raison plus importante qui permet d'expliquer l'échec des méthodes quantitatives. Il est très courant de voir les termes évaluations, sélection, contrôle apparaître pour désigner aussi bien l'évaluation de projets de recherche que l'évaluation des équipes de recherche. Ceci est d'autant plus vrai que les équipes de recherche sont associées à un travail de recherche précis. Cette confusion a lieu dans des environnements de recherche hiérarchisés, où les chercheurs sont des salariés ou des fonctionnaires, c'est-à-dire dans les laboratoires qui appartiennent à l'organisme qui finance les recherches. Dans la recherche publique aux USA c'est le cas de la recherche intra-muros du NIH, ou des laboratoires du National Bureau of Standard (1). Cette confusion entre l'évaluation des projets et l'évaluation des équipes sera d'autant plus forte que les évaluations seront proches des évalués (collègues), comme c'est souvent le cas dans les organismes où il y a coïncidence entre l'organisme finan-

-
- (1) Voir à ce sujet, C. Riveline, Evaluation des coûts, Ecole des Mines de Paris, 1980, ch.VI.
 - (2) Pour une description détaillée de l'évaluation des chercheurs et des équipes dans les organismes voir A. Arvanitis, 1983, op.cit. ci-dessus, p. 45.

ceur et l'organisme qui exécute la recherche.

En résumé notons donc deux confusions à éviter :

- confondre l'évaluation de projets et l'évaluation des équipes de chercheurs
- confondre les indicateurs temporels et monétaires avec les indicateurs des performances.

CHAPITRE 3

LA PRODUCTIVITE DES CHERCHEURS : LES RESULTATS DES TRAVAUX SOCIOLOGIQUES

Résumé

Les travaux de sociologie des sciences utilisant des méthodes quantitatives ont permis de faire le point sur les problèmes méthodologiques et d'interprétation des données. Essentiellement axés au début sur l'étude de la stratification sociale de la science, ils se sont petit à petit déplacés jusqu'à rejoindre les interrogations d'études de cas n'utilisant pas de données quantitatives.

Dans l'ensemble des modèles sociologiques présentés, c'est la production de publications qui constitue la principale variable explicative du prestige acquis par les scientifiques au cours de leur carrière. Mais l'interprétation des données concernant les publications reposent de manière cruciale sur la définition des unités d'analyse (équipe, individu, laboratoire, etc...) et des instruments de mesure (choix des indicateurs : brevets, publications, citations, personnel, financement ; choix des indices construits).

Les travaux ont permis de montrer deux types différents d'indices : indices "objectifs" ou quantitatifs;

et indices "subjectifs" ou qualitatifs. Les indices subjectifs sont issus directement du jugement des pairs ou collègues ou autres personnes interrogées alors que les indices objectifs reposent plus sur des unités dénombrables.

Les travaux les plus significatifs sont ceux qui intègrent des données sur l'organisation des centres de recherches. Nous examinerons donc plus dans le détail ce type de travaux, notamment ceux de Pelz et Andrews, ceux qui reposent sur les données de l'enquête de l'UNESCO sur plusieurs unités de recherche dans six pays différents et des travaux qui examinent des cas d'organisation de la recherche.

Enfin, nous mettons en garde les lecteurs sur les prétendus résultats généraux de ces enquêtes. En effet, plusieurs travaux se contredisent et la cause de ces contradictions vient des définitions des unités de mesure et d'analyse. Un des principaux résultats qui cependant semble être confirmé à peu près toujours, est que la productivité scientifique dépend largement de la position des chercheurs dans les équipes, et que la productivité des équipes dépend elle-même du rôle et des activités des responsables de recherche.

Les études quantitatives sur la science se sont initialement penchées sur les thèmes de la croissance de la science et de la communauté scientifique. Très rapidement ces études ont permis d'observer une très inégale distribution de la productivité des chercheurs, de l'attribution de titres honorifiques, gratifications et prix scientifiques, et de l'allocation des ressources pour la recherche. Les sociologues des sciences ont alors cherché à comprendre ces

phénomènes qui se trouvent à la base des processus de stratification sociale de la science. Il existe maintenant plusieurs études quantitatives sur ce thème. La plupart de ces travaux utilisent la notion de productivité des chercheurs ou de productivité de la recherche.

1. Les problèmes de méthode

Les travaux de mesure de la croissance de la science ont posé plus de problèmes qu'ils n'en ont résolu. Ils ont permis de bien cerner les deux obstacles majeurs dans la mesure de la science :

- la définition d'une unité pertinente d'analyse
- la définition d'un instrument de mesure, des indicateurs et des indices qu'il faut construire.

Généralement les indicateurs utilisés sont les publications, citations, participations à des colloques et des réunions professionnelles, brevets, etc... A partir de ces indicateurs on construit des indices.

Ces indicateurs, et par extension les indices, sont appelés objectifs car ils concernent des unités dénombrables. Mais à côté de ces indices objectifs sont construits des indices subjectifs qui reposent sur des jugements traduits sous forme de notes, issus d'enquêtes auprès des chercheurs.

1.1. La "loi" de Lotka

Les travaux sur la croissance de la science utilisent essentiellement des décomptes de publications scientifiques. L'article de Alfred Lotka (1) fut le premier à étudier la

(1) A.J. Lotka, The Frequency Distribution of Scientific Productivity, Journal of The Washington Academy of Science, vol. 16, 1926, 317-323.

fréquence de la distribution de la productivité scientifique. Il a formulé une "loi" empirique sur la productivité des chercheurs qui fut reprise par Dereck Price (1) :

"Le nombre de personnes produisant n papiers est proportionnel à $1/n^2$. Ainsi, pour chaque 100 auteurs produisant un papier unique pour une période donnée, on trouve 25 auteurs produisant 2 papiers, 11 auteurs produisant 3 papiers, et ainsi de suite" (p.43).

Price a tiré d'autres généralisations à partir de ces travaux : 50 % des publications scientifiques seraient ainsi produites par 6 % de la population scientifique (phénomène de concentration) ; le scientifique "moyen" produirait environ 3 papiers tout au long de sa carrière.

Ces généralisations concernant des individus sont douteuses, sinon purement et simplement fausses. Elles sont effectuées à partir de calculs sur des données bibliographiques qui n'enregistrent pas l'ensemble de la production des chercheurs individuels (2). On ne peut que souligner que ces outils quantitatifs sont par nature des outils statistiques et ne s'appliquent pas aux individus. De plus, J. Hastings (3) a montré que les indices de publications ne sont pas aussi efficace qu'on le prétend pour évaluer l'activité scientifique dans une discipline, comme par exemple l'économie de l'agriculture. Cela tient entre autres au fait que le nombre de revues disponibles par pays influence énormément sur les résultats.

En effet, il faut remarquer que la productivité dépend de plusieurs facteurs, en particulier de la discipline

(1) D.J. de Solla Price, Little Science, Big Science, New York, Columbia U. Press, 1963.

(2) Une critique récente de la loi de Lotka est : M. H. et B.R. MacRoberts, A Re-evaluation of Lotka's Law for Scientific Productivity, Social Studies of Science, vol. 12, 1982, 443-450.

envisagée et de l'institution ou organisme dans lequel se déroule la recherche. C'est pour cela que les travaux les plus significatifs sur la productivité des chercheurs ne sont pas des travaux qui mettent l'accent sur l'ensemble de la science mais plutôt sur des domaines particuliers, des organisations particulières.

1.2. Reconnaissance et productivité scientifique

Les sociologues des sciences et ceux notamment qui se sont rattachés à la tradition mertonienne ont beaucoup approfondi la question de la distribution inégale des prix et gratifications scientifiques que les scientifiques attribuent à leur pairs en récompense d'un travail remarquable ou remarqué. Les citations ont aussi été envisagées comme une forme de reconnaissance, de dette du scientifique envers son collègue qu'il cite, ou plus exactement qu'il se donne la peine de citer. D'où l'idée d'utiliser les citations comme indicateurs du prestige sinon de la qualité des chercheurs. Les frères Cole sont même allés jusqu'à définir la qualité d'un article scientifique par le nombre de citations qu'il reçoit (1, p.21). Les auteurs qui utilisent les citations comme indicateurs de qualité prétendent que la corrélation entre le nombre de citations et le jugement par des experts est élevée (Garfield, 1979, p.23). Certains auteurs ont pourtant fait remarquer que des taux de corrélation de 0,214 ne peuvent pas être qualifiés de taux élevés (3).

- X (1) J. Cole et S. Cole, Social Stratification in Science, The University of Chicago Press, 1973.
- X (2) E. Gardfield, Citation Indexing, It's Theory and Application in Sciences, Technology and Humanities, New York, Wiley, 1979.
- (3) R. McGinnis, compte-rendu de Citation Indexing de Garfield, Science, vol.207, 29 février 1980, p.972.

Garfield, qui est le principal promoteur des études de citations et producteur du Science Citation Index, a rectifié ce jugement en disant que les citations ne permettent pas de mesurer la qualité mais plutôt l'utilité et l'impact d'un article (p.246).

L'utilisation de citations ne va pas sans de sérieux problèmes de définition (1). Même si l'on accepte que les citations reçues par un scientifique sont de bonnes approximations de la qualité de son travail, la relation entre cette forme de reconnaissance de la qualité et la productivité est loin d'être évidente.

Les Cole affirment avoir donné des preuves empiriques qui démontrent que "le bon travail est reconnu quel qu'en soit la provenance" (ce qui correspond à la norme d'Universalité qui serait propre à la science, p.95) et que "c'est la qualité et non la quantité de travail qui est récompensée, la plupart des fois, au moyen de gratifications honorifiques" (p. 94). Pour aboutir à de telles conclusions les Cole examinent le nombre de citations reçues par les chercheurs reconnus et s'aperçoivent qu'ils ont un nombre de citations plus important que la moyenne. Mulkay (1980) affirme que ceci ne prouve rien ou plus exactement que les deux affirmations des Coles sont triviales. Elles veulent dire que "les physiciens les plus cités reçoivent des gratifications quelle que soit leur université d'origine". et que "les physiciens les plus cités ont une probabilité plus forte de recevoir des gratifications que leurs collègues avec la même production mais qui sont moins cités" (1, p.33). La trivialité des résultats empiriques vient de ce

 (1) En français il n'y a pas eu véritablement de débat, sur ces questions mais voir J. Ninio, L'objectivité des critères objectifs, Le Progrès Scientifique, Avril 78; PP. 26-28.%

X (2) Michael Mulkay, Sociology of Science in the West, Current Sociology, vol. 28 (3), 1980, pp. 1-116.

que la relation qui unit la reconnaissance, le nombre de citations, la productivité et la qualité est postulée dès le départ. La seule chose qui est montrée est que ce sont les mêmes scientifiques qui reçoivent les citations et les gratifications honorifiques. Mulkay formule sa critique en ces termes :

"Le succès d'un scientifique par le passé (dans ces études) est le seul indicateur de ses capacités et, partant le manque de succès est supposé être une bonne indication du manque de capacités. (...) Il est simplement postulé que ceux qui connaissent le succès sont les plus aptes" (p. 34).

Ceci est d'autant plus vrai que les Cole ont cherché à corréler le QI et le succès mesuré en termes de citations requises ; ils n'ont pas trouvé de relation significative entre ces deux grandeurs. Gaston (1) a passé en revue les tentatives d'explication de la reconnaissance et de la productivité en termes de capacité individuelle des chercheurs. Il en a conclu que le pouvoir explicatif de variables telles que la motivation, le choix des recherches, l'habileté etc... est quasiment nul.

Le seul résultat certain de ces travaux est le caractère cumulatif de la relation qui unit la reconnaissance par les pairs et la productivité. Plus la productivité est élevée, plus élevée sera la reconnaissance par les pairs. A l'inverse, plus la reconnaissance est élevée et plus la productivité augmentera. Ces deux propositions ont donné lieu à la formulation de deux hypothèses.

La première hypothèse consiste à dire que la productivité des scientifiques s'accroît cumulativement au fur et

X (1) J. Gaston, The reward System in British and American Science, New York, Wiley, 1978.

à mesure de la carrière du chercheur ; cet accroissement cumulatif sera d'autant plus élevé que la productivité en début de carrière sera forte. Ainsi, les chercheurs ayant une forte productivité dès le début de leur carrière accroîtront leur productivité plus vite que ceux n'ayant qu'une faible productivité en début de carrière.

La deuxième hypothèse lie la productivité à la reconnaissance. Cette hypothèse édicte un accroissement de la productivité quand les travaux en début de carrière sont souvent cités (1). Les Cole ont observé la productivité des scientifiques les plus éminents et remarquent que leurs premiers travaux avaient été beaucoup cités (2). Lightfield (3) a observé le même phénomène parmi les sociologues.

Cette hypothèse a été critiquée par Reskin (4). Elle a montré que "l'effet des citations sur la productivité est plus faible (elle souligne) dans les contextes où l'accent est mis sur la recherche". Ce résultat, apparemment paradoxal, tient d'après Reskin au fait que dans les environnements de recherche "la reconnaissance immédiate et informelle de la part des collègues-chercheurs est probablement plus importante que la reconnaissance plus formelle mais aussi plus lente que ~~est~~ sont les citations" (p.502). C'est pour cela

- (1) Voir Cole-Cole, 1973. p.114 où se trouve la première définition de cette hypothèse ; voir aussi Gaston, 1978, p.152-157. Cet auteur fait un très utile survol de la littérature sur ces questions.
- (2) Cole-Cole, 1973, p.110-116.
- (3) T.E. Lightfield, Output and Recognition of Sociologists, The American Sociologists, vol.6, 1971, pp. 128-131.
- (4) B. Reskin, Scientific Productivity and the Reward Structure of Science, American Sociological Review, vol.42, 1977, pp. 491-504.

que Gaston notait qu'il faut repérer "les contextes où la reconnaissance en termes de prix et gratifications formelles et/ou de citations n'est pas nécessaire (pour la reconnaissance des scientifiques) "(p.157). Lui aussi indiquait que la reconnaissance ne semblait pas avoir d'effet d'encouragement auprès des chercheurs. C'est donc l'hypothèse de l'avantage cumulatif qui semble la plus probable car un chercheur ayant eu une forte productivité en début de carrière aura plus de facilité à mobiliser des ressources, ce qui lui permettra de consolider sa position et ainsi de suite.

La recherche de Zuckerman et Merton (1) vient corroborer cette priorité donnée à la productivité par rapport aux gratifications. Ils ont mis en évidence le fait que la productivité des chercheurs les plus éminents est plus élevée que pour les chercheurs moins réputés, mais que leur productivité est très élevée avant qu'ils ne soient devenus éminents. Ainsi, parmi les physiciens ayant reçu le prix Nobel, le prix Fermi ou étant membres de la National Academy of Science (U.S.A.) ou de la Royal Society (G.B.), 28 % avaient envoyé 15 manuscrits à la Physical Review de 1948 à 1956. Le même pourcentage parmi les physiciens sans distinction autre que leur appartenance à l'American Institute of Physics n'était que de 18 %. Parmi ceux n'ayant aucune de ces distinctions il est de 3 %. Or tout les physiciens en question ont acquis leurs titres de gloire après ces 15 envois à la Physical Review, qui est la plus prestigieuse revue de physique.

- (1) H. Zuckermann, R.K. Merton, Age, Aging, And the Age Structure of Science, in R.K. Merton, N.W. Storer, The Sociology of Science, Chicago, The University of Chicago Press 1973.

1.3. Remarques sur la sociologie des sciences quantitatives : les prémisses d'un modèle

L'exemple cité-ci-dessus montre la force et la principale faiblesse des travaux de sociologie des sciences. Ceux-ci examinent presque exclusivement les plus éminents chercheurs, ou les plus éminentes et prestigieuses revues, ou les institutions les plus prestigieuses. Ceci permet sur des échantillons relativement faibles, censés représenter la "crème de la crème" de la science, de tirer des conclusions fracassantes sur le fonctionnement de la communauté scientifique cosmopolite. Le défaut de ces travaux est qu'ils évacuent entièrement les conditions qui doivent être réunies pour qu'un chercheur devienne éminent. Cet "oubli" est voulu car cette conception de la science repose sur un modèle d'une science autonome et autoréglée, donc qui a ses règles propres nécessairement très différentes des règles sociales qui s'appliquent dans le reste de la société (1). Comment ne pas voir dans cette description l'idéal que les chercheurs veulent mettre en avant. Il est tout à fait naturel pour eux d'éliminer les liens qui les rattachent à la société (2). Ceci explique leur rejet de travaux de sociologie des sciences (3).

-
- (1) Une comparaison des divers modèles de la science est : Kochan, Models of Scientific Output, in Elkana et alii, Toward a Metric of Science, New York, Wiley, 1977.
 - (2) Pour les causes de cette élimination des liens qui rattachent les scientifiques à la société, voir la magistrale démonstration de David Bloor, Sociologie de la logique, Pandore, Paris, 1983, plus spécialement le ch.3.
 - (3) Ces débats sur l'utilité ou non de la sociologie de la science prennent en France des tournures pathétiques et agressives, voir Pandore, n° 9, juin 1980, pp. 2-4, et les débats dans les autres numéros de Pandore. Le rejet de la sociologie et en tout cas de la sociologie Mertonienne est assez étonnant puisque le principal résultat de ces travaux est de renforcer la spécificité des scientifiques, et donc de les conforter dans leur image d'eux-mêmes. Mais peut-être aussi qu'en bons scientifiques ils savent que quand commence une recherche, on ne sait jamais où elle va aboutir.

Pourtant il commence à y avoir une littérature assez conséquente sur les organismes scientifiques et leur fonctionnement. Avant de les présenter, il faut remarquer que les travaux des sociologues des sciences appartenant à l'école mertonienne (ou fonctionnaliste) ont permis d'aboutir à une vision dynamique du fonctionnement de la communauté scientifique. L'hypothèse de l'avantage cumulatif me semble être la pierre de touche de cette conception dynamique. En effet, l'avantage cumulatif fait reposer la productivité présente sur la productivité passée et la reconnaissance présente sur la productivité passée. Cette dynamisation risque de poser énormément de problèmes (comme d'ailleurs en économie).

Mais cette dynamisation doit reposer sur certains éléments qui doivent être examinés simultanément. Ces éléments sont les suivants :

1. les conditions de production : ressources humaines (les conditions de production des diplômés, les conditions d'embauche, etc.) ; ressources matérielles (coût des projets, du matériel, ressources financières).
2. les modes de fonctionnement de la recherche : structure des groupes de recherche (âge, taille, organisation, relations entre les groupes et à l'intérieur des groupes, communications, évaluation des recherches, etc...)
3. les marchés scientifiques, ou si on n'aime pas la métaphore économique, les lieux où sont élaborés et les lieux où sont écoulés les produits scientifiques (publications, littérature grise, brevets, produits technologiques, mais aussi les étudiants, les chercheurs et les techniciens) ; ces lieux peuvent être des disciplines scientifiques, des communautés, des marchés économiques, des organismes de recherche, etc.

Bien sûr la construction d'un tel modèle pose le pro-

blème du type d'indices qui doivent être construit. Les travaux des sociologues nous ont permis de voir que les citations et les publications ne sont pas des instruments neutres. En particulier les études qui prennent comme unité d'analyse les communautés scientifiques, comme celles de Merton, les Cole, Gaston, etc..., sont particulièrement exposées aux problèmes de définition des limites de l'instrument quantitatif.

Nous allons rappeler ici les méthodes et principaux résultats des études qui portent sur les organisations de recherche, comme celles de Pelz et Andrews et celle de l'Unesco.

1.4. La méthode de Pelz et Andrews (1)

Deux types de données sont recueillies :

a) des données subjectives : il s'agit des réponses à 2 questionnaires envoyés respectivement aux responsables hiérarchiques des laboratoires ou centre de recherche, et aux scientifiques et ingénieurs.

Dans ces questionnaires deux classements des scientifiques sont demandés aux collègues.

Les juges devaient classer la performance de leurs collègues selon deux critères :

- la contribution de la personne jugée à la connaissance scientifique dans son domaine.

(1) Pelz D.C., Andrews F.M. (1976). Scientists in Organizations : Productive Climates for Research and Development (rev. ed. of 1966) Ann Arbor, U. Michigan (Institute for Social Research).

- l'utilité de la recherche : en quoi les personnes jugées ont-elles permis à l'organisme auquel elles appartenaient de remplir sa fonction ou ses objectifs.

b) des données quantitatives sont recueillies auprès des scientifiques (production d'articles, de rapports, dépôts de brevets (1), durant les 5 années précédant l'enquête).

Les auteurs disposaient ainsi de quatre indicateurs de "production", par chercheur (1311 scientifiques).

1. jugements sur la contribution scientifique,
2. jugements sur l'utilité,
3. production de rapports
4. production d'articles et de brevets

Ces indicateurs sont traités statistiquement pour tenir compte de biais dûs au temps ou à l'âge des scientifiques, mais aussi pour disposer de données comparables à celles des "pairs". Les pairs sont ici définis de manière plus stricte qu'on ne l'entend habituellement : il s'agit des scientifiques qui disposent du même diplôme (Ph.D. Bachelor, etc...), du même temps d'expérience et qui travaillent dans des environnements similaires.

Afin de tenir compte des différents environnements de recherche, la population totale des scientifiques étudiés a été divisée en 5 groupes selon 3 dimensions :

- (1) Notons que, mis à part le cas des ingénieurs dans des organismes de R-D, le dépôt de brevet est difficilement interprétable car un brevet peut faire partie de la stratégie commerciale d'une firme. C'est pour cette raison que nous ne retenons pas ici les statistiques de brevets. Par contre, une étude sur l'innovation doit absolument se fonder sur une analyse des statistiques de brevets.

- 1) Orientation principale de l'unité de recherche (recherche fondamentale ou recherche-développement)
- 2) Possession d'un Ph D,
- 3) Majorité de chercheurs possédant le PhD dans l'unité de recherche (ceci semblait énormément influencer le statut des chercheurs sans PhD. Ils étaient en permanence des assistants dans les unités de recherche dominées par les PhD ou pouvaient devenir des directeurs de recherche dans les unités de recherche non dominées par le PhD.

Ainsi, les 5 sous-populations de l'étude en question étaient :

1. les PhD dans des laboratoires principalement orientés vers la recherche fondamentale (par commodité nous les appelons "docteurs de RF" (recherche fondamentale)
2. Les PhD dans des laboratoires de R-D ("docteurs de R-D")
3. Les non-PhD dans des labos R-D dominés par des PhD (appelés "Ingénieurs" dans cette étude)
4. les non-PhD dans des labos de recherche fondamentale ou de R-D dominés par les PhD ("assistants").
5. les Non-PhD dans des labos de recherche fondamentale non dominés par des PhD ("chargés de recherche").

1.5. La méthode de l'enquête de l'Unesco (1)

L'enquête de l'UNESCO porte sur les équipes de recher-

- (1) Andrews (ed.), Scientific Productivity, The Effectiveness Of Research Groups in Six Countries, Unesco, Cambridge U. press, 1979.

che et non sur les chercheurs individuels comme c'est le cas de Pelz et Andrews. Elle permet de comparer des résultats pour environ 1200 équipes de recherche dans six pays européens (1).

a) Définition de l'équipe de recherche

Les critères qui permettent d'identifier une équipe de recherche sont les suivants :

- il existe au moins un chef d'équipe, travaillant au moins 8 heures par semaine dans l'équipe ;
- la durée de vie du groupe est au moins de un an ;
- l'équipe est composée d'au moins trois membres travaillant à temps plein.

L'enquête a été menée)à partir de questionnaires envoyés auprès des chefs de groupes de recherche, des chercheurs, des techniciens, des administrateurs et des évaluateurs externes.

Les différentes équipes sont classées selon une typologie qui a été effectuée d'après les performances de l'unité de recherche et non d'après les seules caractéristiques institutionnelles (université, industrie, etc...). La typologie que les auteurs ont obtenu est la suivante :

1. équipes universitaires effectuant des recherches dans les sciences naturelles et exactes,
2. équipes universitaires effectuant des recherches médicales et dans les sciences sociales,

(1) L'enquête a été effectuée trois fois de suite en portant chaque fois sur des équipes différentes, Etude comparative internationale sur l'organisation et la productivité des unités de recherche, Unesco, sept. 1981, document SC.TECH./SEM.9/R.2+Add.16.

3. équipes universitaires effectuant des recherches appliquées et technologiques,
4. équipes en R-D "coopérative", c'est-à-dire dont les institutions sont au service d'un secteur industriel public
5. équipes de recherche en milieu industriel (privé).

Ces cinq catégories ont été choisies parmi 35 combinaisons possibles entre le type d'institutions et les domaines de recherche?

b) Les indices "objectifs"

Les indices mis au point par les enquêteurs sont de deux types : des indices de performance objectifs et subjectifs. Les trois indices de production sont

- les publications,
- les brevets,
- les rapports internes, documents de travail, etc...

Les auteurs mentionnent une caractéristique intéressante de ces indices objectifs : que ce soit au niveau individuel ou au niveau des équipes les indices présentent les mêmes caractéristiques (1). Etant donné les difficultés rencontrées par les sociologues qui ont cherché à expliquer les variations de la productivité individuelle en fonction de caractéristiques individuelles, nous pouvons inférer qu'en ce qui concerne les produits rendus publics c'est la productivité du groupe qui détermine la productivité individuelle.

(1) Identification of RU output Clusters (document STP/524), Unesco, 1975 ; Construction and Testing of RU Output Composite Measures (Document STP/1713), Unesco 1975.

La bonne piste en ce qui concerne la productivité est donc celle qui permet d'expliquer la productivité de l'équipe de recherche.

c) Les indices "subjectifs"

Par la suite sept indices sont construits à partir de 41 variables identifiées dans les divers questionnaires. Il s'agit d'indices effectués d'après les jugements des personnes interrogées. Les indices sont les suivants :

1. Contribution générale de l'équipe à la science et la Technologie,
2. Reconnaissance internationale du groupe de recherche (réputation internationale et demandes d'articles produits par l'équipe),
3. Efficacité sociale (valeur sociale des travaux et utilité),
4. Efficacité pédagogique ou d'enseignement,
5. Efficacité administrative (plan de travail respectés, budgets prévus respectés, concordance de ce qui est prévu et de ce qui a été accompli),
6. Efficacité de R & D (productivité et degré d'innovation),
7. Efficacité des applications ou extension des applications (résultats et utilisation des activités "développement"-).

La construction de ces indices pose beaucoup de problèmes d'interprétation. En particulier, les mesures subjectives dépendent de l'observateur, c'est à dire des personnes interrogées. Les auteurs ont examiné la validité de ces mesures en termes d'analyse de la variance. Andrews étudie les sources d'erreurs de ces mesures subjectives (1). En

 (1) F.M. Andrews, Estimating the Construct Validity and Correlated Error Components of the Rated Effectiveness Measure in Andrews (1979), pp. 405-422.

particulier le biais issu de "l'effet de halo" ou de contagion sont examinés. Ces effets sont particulièrement élevés parmi les évaluateurs externes auprès desquels les enquêteurs ont demandé des jugements sur le travail des équipes. Ce résultat nous impose beaucoup de prudence dans l'utilisation des jugements des évaluateurs externes : les pairs scientifiques ont les yeux moins bien ouverts sur leurs pairs que les chercheurs sur leur propre travail. De plus, ajouter le jugement d'évaluateurs externes à celui des chefs d'équipe et membres de l'équipe n'altère pas significativement les résultats quantitatifs des indices. L'introduction du jugement externe ne doit pas être effectué pour améliorer les indices mais sur la base d'autres considérations (par exemple avoir l'opinion de "la communauté scientifique").

1.6. Rapport entre mesures objectives et subjectives

Les auteurs ayant travaillé sur les données de l'enquête de l'Unesco ont tous remarqué que les indices subjectifs sont relativement indépendants des indices objectifs. Même quand ils sont liés, les corrélations sont assez faibles. Ceci est vrai de toutes les études portant sur la productivité des chercheurs. De plus, nous l'avons déjà mentionné, les chercheurs sont plus critiques vis-à-vis de leur propre travail que vis-à-vis du travail de leurs pairs. Enfin, on remarque une absence de relations entre les indicateurs, tout à fait objectifs, de ressources (ou inputs) et les indices de productivité.

Le problème des jugements subjectifs a été récemment abordé par Keller et Holland (1982). Ils proposent d'utiliser conjointement aux mesures subjectives des collègues et

supérieurs hiérarchiques, le jugement des scientifiques sur leur propre travail (1). Une autre enquête avait déjà permis de voir que les physiciens qui ont une haute estime d'eux-même étaient très satisfaits par leur travail et très concernés par leurs recherches (2). Keller et Holland ont remarqué que les scientifiques ayant une bonne image d'eux-même perçoivent correctement le jugement que leur supérieur hiérarchique porte sur eux. Par contre, ils se trompent sur le jugement que leurs collègues émettent sur eux. Cette perception plus aiguë des opinions et jugement des supérieurs hiérarchiques de la part des chercheurs n'implique pas nécessairement que ce supérieur hiérarchique aura à son tour un bon jugement sur les chercheurs qui lui sont subordonnés. Mais d'après les résultats de l'étude de Andrews, on peut considérer que c'est le cas la plupart des fois. Par contre, *l'activité et le jugement du chef d'équipe est essentielle car il oriente le niveau, l'orientation et la quantité de recherche effectuées par les membres de son équipe.*

La question de la correspondance entre les mesures de ressources et la performance des équipes de recherche est abordée par Veronica Stolte-Heiskanen (1979). (3)

-
- (1) Keller R.T., Holland W.E. The Measurement of Performance among R & D professional employees : A longitudinal Analysis. IEEE Transactions of Engineering Management, Vol. 29 (2), Mai 1982.
 - (2) Procédure proposée par R.L. Diphoye, W.H. Zultowski, H.D. Dewhirst and R.D. Arvey. "Self-esteem as a moderator of the relationship between scientific interests and the job satisfaction of physicists and engineers". Journal of Applied Psychology, vol. 63, 289-294, 1978.
 - (3) V. Stölte-Heiskanen
in Andrews, (1979).

Cet auteur a remarqué que les ressources en matériel, en homme et en informations sont relativement faiblement liées à la perception subjective qu'en ont les chercheurs. Ainsi, la satisfaction des membres des équipes de recherche quant à leurs ressources et à l'adéquation de ces ressources pour leurs recherches n'est pas liée au niveau réel de ces ressources. De plus, les chercheurs expriment une attitude différente en fonction du type de ressources. Ainsi, les ressources les moins traditionnellement envisagées (présence d'un spécialiste en information chargé de faire circuler l'information, présence d'un personnel technique important et compétent) sont perçues comme bénéfiques plus facilement que des ressources plus traditionnelles comme par exemple le nombre de scientifiques. Enfin, les performances des équipes de recherche identifiées dans l'étude de l'Unesco ne sont pas liées à l'interaction entre ressources réelles et perception subjective de ces ressources. Les contacts des unités de recherche avec l'extérieur, la compétence des chefs d'équipe interviennent plus dans les mesures subjectives des performances que le niveau réel des ressources.

Ces remarques sur les rapports entre mesures subjectives et objectives doivent attirer notre attention sur les relations complexes qui existent entre les performances et les ressources. Il semble bien que ce soit la structure des groupes de recherche qui détermine la plupart des variables de production.

2. Les résultats de ces travaux

2.1. Les performances individuelles et la liberté des chercheurs

Une des caractéristiques aux yeux des scientifiques

qui rend attrayante la recherche en milieu universitaire, est la liberté dont ils jouissent dans les labos universitaires Pelz et Andrews ont examiné ce problème en détail au niveau des chercheurs individuels. Ils ont mesuré la pression exercée sur et par les chercheurs de deux façons. D'une part en comptant le nombre de sources de décision (le chercheur lui-même, le responsable hiérarchique, les collègues, les directeurs et clients, ces sources de décision peuvent être combinées). D'autre part en mesurant l'effet de ces sources de décisions combinées sur l'orientation des recherches.

Dans l'ensemble les chercheurs présentent de meilleures performances quand un subtil mélange de ces sources de décisions et de l'influence qu'exercent les chercheurs a lieu.

Les scientifiques les plus influents sont plus performants quand ils combinent plusieurs sources de décisions (3 ou plus), Ceci est encore plus vrai des docteurs de Recherche Fondamentale qui n'exercent pas une forte influence sur les décisions. Par contre, cette constatation s'avère fausse dans le cas des non-PhD de Recherche Fondamentale (chargés de recherche) : quand plus de trois "sources" décident, leurs performances décroissent considérablement. Enfin, les ingénieurs et assistants sont plus performants quand 2 ou plus sources de décisions sont impliquées dans les décisions, et leurs performances n'augmentent que faiblement avec le nombre de sources de décisions.

Les auteurs rappellent d'autres travaux, notamment une étude sur des sociologues spécialisés dans la sociologie médicale, qui montrent que les travaux les plus innovatifs étaient dûs à des sociologues dont les projets étaient dis-

cutés par des supérieurs hiérarchiques qui ne déterminaient pas le contenu de la recherche, ou, rarement, par un supérieur hiérarchique qui détermine assez substantiellement le contenu des recherche. Ce dilemme entre l'existence d'une source supérieure de décision et son absence est interprété de la sorte : *les chercheurs dont les résultats de recherche sont visibles sont en même temps ceux qui jouissent du plus de liberté dans la conduite de leur recherche.*

Les résultats de Pelz et Andrews peuvent être interprétés de la même manière : les scientifiques dont les résultats sont les plus visibles sont aussi ceux où les décisions concernant l'orientation de la recherche incluent plusieurs personnes. Les plus performants sont donc ceux qui combinent des résultats visibles et une influence plus forte, qui les sauvegardent de la main-mise d'un supérieur hiérarchique.

De toute manière, ceux qui disent décider eux-mêmes et entièrement du contenu de leur recherche, ne sont pas les plus performants. Dans ces conditions, l'autonomie du chercheur est-elle une mauvaise chose ? Pas nécessairement, répondent les auteurs, car cela dépend de l'intérêt que portent les chercheurs au contenu des recherches.

Ceux qui s'intéressent aux caractères généraux des nouveaux domaines sont aussi ceux pour qui l'autonomie bénéficie le plus. Ceux par contre, qui sont plus tournés vers les recherches de détail, bénéficient d'avoir moins d'autonomie. Ceci est vrai autant dans la recherche Fondamentale que dans la recherche-développement. En bref, on peut dire que la liberté profite à qui ne se spécialise pas trop.

2.2. Les performances des équipes et la structure du pouvoir

Cette notion de l'influence des sources de décision sur les performances a été reprise dans les enquêtes ultérieures, notamment dans l'enquête de l'Unesco (1). L'influence exercée par les chefs d'équipe et les chercheurs semble être liée aux aspects les plus universitaires de la production scientifique (publications, reconnaissance, efficacité pédagogique). Par contre l'influence exercée par une source extérieure à l'équipe est liée aux performances atypiques de la recherche appliquée (efficacité des applications, brevets et prototypes) mais pas aux performances de type universitaire. Il s'agit ici des performances des équipes de recherche et pas des performances individuelles. De plus ces résultats empiriques ne montrent pas de relations de causalité. Il manque à ces travaux un minimum de réflexion théorique sur la notion d'influence et de pouvoir. De toute manière on ne peut qu'être frappé par la faiblesse des corrélations qui ont été trouvées entre des indices d'influence et des indices de performances.

Plus significative nous semble être la contribution de Knorr et alii (2)

Ces auteurs ont essayé de comprendre les différences de productivité des scientifiques de haut statut ou de grand prestige par rapport aux scientifiques de statut moins élevé. Ils font remarquer que les travaux sur la sociologie des sciences ne font pas suffisamment attention

-
- (1) Salomea Kowalewska, Patterns of Influence and the Performance of research units, in Andrews (1979), pp.169-189.
 - (2) K.D. Knorr, R. Mittermeir, G. Aichholzer, G. Waller Individual publication productivity as a social position effect in academic and industrial research units, in Andrews (1979), pp. 55-94 ; et les mêmes auteurs, Leadership and group performance : a positive relationship in academic research units, in Andrews (1979), pp. 95-120.

aux structures d'organisation et la recherche. Les principaux résultats de ces auteurs sont les suivants :

- Les différences de productivité en termes de publications (1) dépendent de la position des chercheurs au sein des équipes. Ainsi, les scientifiques qui détiennent une position dominante dans l'équipe sont ceux qui ont la productivité la plus élevée. Ceci s'explique par le fait que les scientifiques dominants décident des objectifs et des orientations de la recherche, ce qui leur permet de dépenser moins de temps à effectuer des recherches pour publier des articles. Comme ils se préoccupent plus de la définition des recherches que de leur exécution, ils ont "la main haute" sur un grand nombre de travaux que s'ils se consacraient exclusivement à l'exécution des recherches. De plus, leur position dominante leur confère un pouvoir plus grand à mobiliser des ressources ; les ressources jouent alors un rôle d'amplification de la productivité, de l'ensemble des chercheurs du groupe, mais plus encore des chercheurs en position dominante. C'est bien le phénomène de l'avantage cumulatif. Rappelons que l'avantage cumulatif permet d'expliquer le fait que les effets d'amplification sont plus importants pour ceux qui se trouvent en position dominante que ceux qui se trouvent en position dominée.

- En comparant la productivité individuelle des chercheurs universitaires dans les sciences naturelles à la productivité des chercheurs en milieu industriel, on s'aperçoit que la position dominante des chefs d'équipe ne joue pas de la même façon : les chefs d'équipes de chercheurs industriels doivent détenir des fonctions plus diversifiées à l'extérieur des équipes. Mais ceci est dû au fait que la production des chercheurs industriels est différente de

(1) La production individuelle d'articles et de livres = nombre d'articles et livres que signe un chercheur individuel.

celle des chercheurs universitaires. Ces derniers produisent essentiellement des publications alors que les premiers produisent des brevets et des rapports techniques. Mais quel que soit le contexte, c'est toujours la position du chercheur qui permet d'expliquer les variations de la productivité.

Pelz et Andrews ont aussi cherché à comprendre les différences de productivité en fonction de la structure du groupe. Parmi les groupes de recherche qu'ils ont étudié, certains ont une structure hiérarchique très stricte, alors que d'autres ont une structure plus lâche. Cela ne concorde pas nécessairement au type d'institution.

Ce ne sont pas nécessairement les universités qui présentent des groupes à structure hiérarchique lâche, bien que ce soit souvent le cas. Les auteurs ont distingué cinq ensembles de groupes selon leur structure hiérarchique, du plus lâche au plus strict.

Le résultat essentiel est que plus une structure est lâche, plus fortes doivent être les motivations pour effectuer les recherches. Les chercheurs autonomes dans une structure lâche sont moins productifs que ceux dans une structure stricte. A l'inverse une structure hiérarchique trop stricte, une dépendance trop étroite entre la direction du centre de recherche et l'équipe de recherche semble avoir un effet contre-productif. On peut rappeler ici cette observation de Lewis Branscomb à propos des contacts entre les chercheurs et la direction. C'est la confiance de la direction de l'entreprise envers le chef de labo qui détermine le travail de recherche. Si la direction se met à douter de ce que fait le chef de labo et si elle cherche à contrôler directement les chercheurs, les travaux du labo risquent de ne plus être fructueux.

2.3. Les performances individuelles et caractéristiques individuelles des chercheurs

Les travaux de sociologie des sciences purement sociologiques sont relativement récents. Initialement les recherches portant sur les scientifiques étaient des recherches psycho-sociologiques. Elles tentaient d'isoler des caractéristiques individuelles censées influencer les performances des chercheurs.

Ainsi, Pelz et Andrews ont examiné les motivations des chercheurs et les rapports entre ces motivations et les performances. Ils ont définis trois types de motivations :

(1) motivations scientifiques : elles sont mesurées par les réponses aux questions suivantes : "la recherche me permet :

- a) d'utiliser mes connaissances et mes capacités au maximum,
- b) de jouir de liberté dans l'élaboration de mes idées,
- c) de contribuer au savoir dans mon domaine de recherche ou ma spécialité ou ma discipline.

(2) motivations professionnelles : les trois questions ci-dessus, plus :

- d) apprendre ou acquérir plus de connaissances,
- e) affirmer ma réputation professionnelle,
- f) travailler sur des sujets difficiles.

(3) motivations quant aux statuts

Je préfère les problèmes qui me permettent :

- a) d'affirmer ma réputation professionnelle,
- b) un avancement au sein de l'organisme,
- c) un salaire plus élevé,
- d) d'affirmer mon autorité
- e) de m'associer aux membres de la direction-décideurs (top executives").

Les docteurs dans la Recherche Fondamentale ont des motivations principalement scientifiques et ont des motivations professionnelles légèrement plus fortes que les autres groupes. Ils sont suivis de près par les docteurs dans le R & D. Les assistants de recherche dans l'industrie montrent un désintérêt concernant ces motivations. Enfin, les ingénieurs sont essentiellement motivés quant à l'obtention d'un statut plus élevé.

Quand on relie ces réponses sur les motifs déclarés et les mesures de performance, on obtient des résultats étonnants.

Les motivations scientifiques sont très faiblement corrélées aux indices de performance, même en ce qui concerne les jugements sur l'avancement de la science. Les motivations quant au statut n'ont pratiquement pas de corrélations avec les indices de performance. Seul le nombre de rapports techniques internes, non publiés, semble lié aux motivations scientifiques (faiblement). Les auteurs en concluent que les rapports techniques sont peut-être la mesure d'un facteur d'énergie associée au travail scientifique

Les docteurs en R & D et ingénieurs, qui travaillent dans les laboratoires fédéraux, ont des corrélations plus élevées entre performances et motivations scientifiques et professionnelles. Les docteurs en Recherche fondamentale ne présentent aucune corrélation entre les performances et toutes les mesures des motivations. Les assistants dans l'industrie, mais aussi dans les laboratoires fédéraux, en R & D et en Recherche fondamentale sont le seul groupe où l'on observe une corrélation entre motivations de statuts et la production de rapports, brevets, et articles. La corrélation est aussi vérifiée entre la production de rapports des assistants et les motivations scientifiques et professionnelles, et leur production d'article dans les laboratoires fédéraux.

Il faut de plus rappeler que d'autres travaux de psycho-sociologie ont été fait sur cette notion de motivation. Dans aucun de ces travaux on n'a pu établir de relations concluantes entre les motivations déclarées et les performances ((1), (pp.255-257)).

Andrews (1) a repris cette notion de la motivation, mais au lieu de s'intéresser aux motivations individuelles et aux performances individuelles, il a comparé les motivations de l'ensemble d'une équipe de recherche et les performances de l'équipe, en utilisant les données de l'étude de l'Unesco. Il semble bien que motivations et performances soient deux choses liées, même en ce qui concerne la production d'articles et de livres. Par contre, la production de brevets, de prototypes, etc... n'est pas liée aux motivations. Ces observations sont vraies dans toutes les sortes d'équipes et pour l'ensemble des pays étudiés. Mais il n'existe pas de liens de causalité entre les motivations exprimées par le groupe et ses performances. La corrélation entre ces deux dimensions est une simple concomittance. L'impression qui prédomine à la lecture de ces travaux est que les performances et les motivations résultent de la formation d'une équipe de recherche et ne sont pas une condition à sa formation.

2.4. La nature des recherches et les performances

Pelz et Andrews ont remarqué que les chercheurs performants sont ceux qui ont une approche globale plutôt que spécialisée des problèmes, large plutôt qu'approfondie. De plus, une plus grande attention aux concepts abstraits donne une corrélation négative ou nulle, des indicateurs de

(1) Voir Andrews, F.M., Motivation, diversity and the performance of Research Units, in Andrews (1979), pp.253-289.

performance, excepté l'indicateur du jugement sur la contribution scientifique dans le seul cas des docteurs en R & D dans des laboratoires fédéraux. On observe aussi avec étonnement que seuls les assistants bénéficient, en termes de jugement sur l'utilité de la recherche, à développer des recherches concernant des problèmes immédiats, du moment. Travailler sur des sujets pratiques et immédiats ne bénéficie à aucun autre ensemble de chercheurs quel que soit l'indicateur de performances choisi.

On peut rattacher ces résultats à ceux sur la position des chercheurs. En effet, les chercheurs ayant une position dominante sont souvent ceux qui ont une approche globale du sujet des recherches.

2.5. Les performances sont liées à d'autres activités que la recherche

Comme nous l'avons déjà remarqué pour les chefs de labo ou d'équipe la productivité individuelle est liée à la position du chercheur au sein de son équipe. Ceci est dû au fait que selon sa position, le chercheur aura des activités différentes. En particulier, les activités autres que la recherche, comme la coordination des tâches, la détermination des objectifs de recherche, leur justification auprès de ceux qui financent les travaux (en particulier dans le domaine des recherches appliquées en milieu industriel); les diverses tâches d'administration, les activités d'enseignement, sont des activités qui influencent le contenu même des recherches.

Pelz et Andrews ont examiné le temps consacré à des fonctions d'administration, de gestion, d'enseignement et de recherche. Les docteurs de R & D sont plus performants quand ils ne consacrent que $3/4$ de leur temps à la recherche. En Recherche fondamentale, apparaît le même

résultat. Mais, en ce qui concerne les ingénieurs, les meilleures performances sont obtenues quand ils se consacrent que la moitié de leur temps à la recherche. Par contre, le nombre de projets auxquels sont confrontés les chercheurs n'influe pas leurs performances.

Comme Knorr et alii, ces auteurs ont remarqué que c'est la position hiérarchique des scientifiques qui explique cette plus forte productivité des chercheurs quand ils ne s'occupent pas exclusivement de l'exécution des recherches.

2.6. Les performances individuelles en fonction de l'âge

Une des questions qui revient en permanence dans ces travaux est celle de la variation de la productivité en fonction de l'âge des scientifiques. Lehman (1) avait remarqué que les découvertes importantes avaient lieu quand les chercheurs avaient entre 35 et 45 ans. Cela d'ailleurs correspond à peu près avec l'âge où les chercheurs ont des postes de responsabilité au sein des équipes, comme l'ont montré Knorr et alii. D'après Lehman, la productivité décline régulièrement après cette tranche d'âge. Le déclin serait d'ailleurs plus rapide dans les disciplines les plus abstraites (mathématique, physique théorique) que dans les disciplines empiriques (géologie, physique appliquée).

Les travaux de Lehman ont été critiqués notamment sur la méthode. D'une part, il ne tient pas compte de la longévité des chercheurs ; les recherches d'un scientifique mort à trente ans sont toujours les travaux d'un "jeune" chercheur

(1) Lehman, Age and Achievement, Princeton U. Press, 1953.

Si tous les chercheurs avaient la même durée de vie, Lehman aurait probablement eu des résultats différents. D'autre part, l'auteur n'a pas tenu compte du nombre de chercheurs dans chaque classe d'âge. Il a supposé, à tort, que la population était la même dans chaque classe d'âge. Ceci est d'autant plus grave que l'importante croissance démographique des chercheurs a eu pour conséquence une très forte augmentation du nombre de jeunes chercheurs. Les conditions difficiles du marché de l'emploi scientifique sont là pour le montrer.

Pelz et Andrews furent plus prudents et leurs données ont été calculées pour tenir compte de ces différents problèmes. Ils ont remarqué que les courbes de productivité des chercheurs ne suivaient pas des fonctions linéaires décroissantes à partir d'un certain âge. Elles sont marquées par un premier maximum aux environs de la quarantaine (35 à 45 ans) puis par un autre maximum aux environs des 50 à 55 ans.

Les Docteurs en R & D présentent la même configuration mais légèrement plus tard que les chercheurs en recherche fondamentale. Enfin, les ingénieurs dans l'enquête de Pelz et Andrews produisent beaucoup de brevets régulièrement jusqu'en fin de carrière ; par contre, les autres indicateurs (publications et indices subjectifs) chutent très fortement à partir de 49 à 50 ans. Il faut voir dans leur cas que le contexte de la R et D est très différent de celui de la recherche fondamentale en milieu universitaire, ce qui semble expliquer ces résultats.

Reskin (1) rappelle plusieurs études sur la productivité individuelle en fonction de l'âge. L'une d'elles a permis de constater que seuls les chercheurs affiliés aux institutions les plus prestigieuses avaient une courbe à deux sommets, comme celle décrite par Pelz et Andrews. Zucker-

(1) B. Reskin, A Review of the Literature on the Relationship between Age and Scientific Productivity, in National

man avait aussi remarqué la même courbe en ce qui concerne les Nobélistes ; ces derniers avaient une productivité qui plafonnait un peu plus tardivement que pour les non-Nobelistes. Les deux maxima des prix Nobels surviennent environ dix ans plus tard que dans le cas des autres scientifiques, mais ces maxima sont plus élevés que ceux de leurs confrères.

Cependant, il est dangereux de généraliser cette courbe à deux sommets contrairement à ce que prétendent des auteurs comme Pelz et Andrews. En effet, une étude de l'American Council on Education a montré très clairement que la courbe de la productivité individuelle, même agrégée, varie en fonction de la discipline. Ainsi dans au moins deux disciplines, le génie chimique et la biochimie, les courbes ne présentent qu'un seul maxima et décroissent régulièrement après. Dans ces deux cas cependant l'âge explique très peu la variation de la productivité (2 % dans la variance de la productivité en biochimie, 4 % dans le génie chimique. En biologie, cette courbe à deux sommets est moins marquante qu'en physique, chimie et mathématique. Il semble bien que cette productivité à deux sommets ait lieu dans les disciplines anciennes ou traditionnelles où l'âge des chercheurs ont un facteur plus important.

Ces résultats montrent à quel point le contenu des recherches détermine la productivité. Ils indiquent aussi que les variables les plus "objectives", comme la production d'articles, sont fortement dépendantes du mode d'organisation non seulement des lieux de recherche mais des disciplines ou des communautés scientifiques.

... Academy of Science, National Research Council : Research Excellence through the Year 2000, Washington, D.C., 1979 pp. 189-207.

La courbe à deux sommets a ainsi lieu parmi les scientifiques en psychologie expérimentale mais pas parmi l'ensemble des psychologues. La psychologie expérimentale est une discipline ancienne aux Etats-Unis, ce qui n'est pas le cas de bien d'autres sous-disciplines de la psychologie.

Une autre étude portant sur les économistes américains et qui isole la productivité des présidents de l'American Economic Association a aussi permis d'observer cette courbe à deux sommets. Ce qui confirme que c'est dans les disciplines les plus anciennes et auprès des scientifiques les plus réputés qu'a lieu cette productivité à deux maxima (1).

2.7. Les performances des groupes de recherche

Nous avons dit que les performances individuelles sont liées aux structures des groupes de recherche. Il est loisible de penser qu'elles sont aussi liées aux performances des groupes de recherche. Quelques travaux ont exploré cette voie et ont mis en avant les variables des groupes de recherche qui influencent la productivité non plus individuelle mais des groupes.

La productivité des groupes peut être mesurée, en ce qui concerne les publications, par le nombre d'articles publiés par le groupe (Voir Knorr et alii dans Andrews, 1979, op.cit., p. 81). Elle diffère de la productivité individuelle car on décompte le nombre d'articles, pas le nombre d'auteurs.

Les déterminants de la productivité des groupes sont :

(1) R. Eagly, Contemporary Profile of Conventional Economists History of Political Economy, vol.6, (1), 1974, pp.76-91.

- la productivité individuelle des chefs d'équipe
- la productivité individuelle des membres de l'équipe
- la taille de l'équipe
- les échanges d'informations du groupe
- l'âge (ou la longévité) du groupe.

Parmi ces déterminants, les trois derniers sont des variables structurelles. Knorr et alii ont remarqué que plus on va vers des domaines de recherche appliquée et technologique plus leur importance grandit. La taille de l'équipe devient, en particulier une variable cruciale. De plus on observe une relation inversement proportionnelle entre l'âge du groupe et la production d'article par tête (qui est différente de la productivité individuelle).

Ce résultat intéressant appelle des précautions, d'une part en ce qui concerne la définition de l'équipe, et d'autre part en ce qui concerne la variable "âge du groupe". Cette variable dépend très certainement de la communication et de l'activité du chef d'équipe. Enfin, toutes les variables structurelles des groupes sont variables selon le pays et selon la discipline.

Pelz et Andrews ont pu remarquer que les notes issues du jugement sur la contribution scientifique (voir plus haut la définition), baissent régulièrement avec l'âge du groupe, c'est-à-dire le nombre d'années de fonctionnement de l'unité de recherche, avec un léger accroissement quand le groupe atteint les 4 ou 5 années de fonctionnement.

Par contre, les notes issues du jugement sur l'utilité des recherches augmentent quand le groupe a 4 ou 5 ans de fonctionnement, baissent par la suite et augmentent légèrement après 8 ou 9 ans de fonctionnement. La courbe d'utilité a donc la même forme que la courbe de contribution mais elle est décalée d'environ 4 à 5 ans.

Les groupes les plus anciens sont moins actifs que les groupes récents ; moins de contacts ont lieu, moins fréquemment entre leurs membres. Les chercheurs perçoivent moins l'esprit de compétition, sont moins secrets et plus spécialisés dans leurs recherches.

Parmi ces groupes anciens, les plus performants sont ceux qui résistent à ces différents changements dans le temps. Mais ils ne sont pas pour autant identiques aux groupes récents ; dans le temps les chercheurs transforment les relations de rivalité intellectuelle en liens plus chaleureux, plus amicaux. Les auteurs remarquent à ce sujet que les chercheurs productifs dans des groupes récents sont inhibés par la critique alors que les chercheurs productifs dans les groupes anciens sont au contraire stimulés.

Une autre recherche menée aux Etats-Unis parmi 50 laboratoires de recherche et de développement a permis de voir que les groupes de recherches deviennent de plus en plus autonomes dans le temps et ont tendance à s'isoler. Cela semble avoir un effet contre-productif (1). Pelz et Andrews eux aussi ont bien remarqué que les docteurs et ingénieurs accroissent leurs performances en fonction du nombre de contacts qu'ils ont avec leurs collègues (2). Seul le nombre d'articles suit des courbes différentes. Pour les docteurs en R & D plus de 2 contacts avec les collègues par semaines fait baisser la production d'articles. Par contre, pour les docteurs en Recherche Fondamentale, la production d'articles est croissante avec des contacts mensuels ou moins

-
- (1) R. Katz, The Effects of Group Longevity on Project Communication and Performance, Administrative Science Quarterly, vol.27, 1982, pp. 81-104.
 - (2) Les collègues sont les personnes qui travaillent dans un même laboratoire, en excluant les directeurs de labs et les subordonnés.

et avec des contacts bi-hebdomadaires et plus souvent ; pour eux la production d'articles est faible quand ils ne rencontrent leurs collègues qu'une fois par semaine, et seulement dans ce cas. Les assistants et non-docteurs ont trop peu de contacts pour obtenir dans cette analyse des résultats significatifs.

D'autres mesures des contacts avec les collègues sont effectuées par les auteurs. Toutes confirment ce résultat. Par contre, le nombre de contacts n'est pas lié au nombre de collègues.

Les contacts avec les collègues peuvent être un résultat plutôt qu'une cause de bonnes performances. Mais les auteurs ont pu observer que le nombre de contacts était fortement corrélé aux indicateurs de performances, même quand on ne retenait que les contacts que les scientifiques provoquaient eux-mêmes, en excluant tous les contacts dûs au hasard ou provoqués par un autre scientifique.

La similarité de point de vue entre les chercheurs a aussi fait l'objet d'une analyse approfondie. Les auteurs ont pu observer que les performances étaient légèrement meilleures pour les chercheurs qui maintenaient des relations avec des personnes n'ayant pas la même façon d'attaquer un problème, n'ayant pas le même type d'approche.

Nicole Visart (1) a mesuré la communication entre les groupes de recherches à partir d'un certain nombre d'indices objectifs : nombre de visites effectuées, publications reçues et envoyées par le groupe, participation à des réunions, délais dans la réception de l'information, nombre de membres du

(1) Nicole Visart, Communication between and within research units, in Andrews (1979), pp. 223-252.

groupe émettant des informations. La communication à l'intérieur du groupe peut en revanche être mesurée par des indices subjectifs. L'auteur a remarqué que les indices subjectifs (représentant la communication interne du groupe) ne sont pas dépendant des indices objectifs (p.249). Seule la "reconnaissance de l'équipe" à l'extérieur est liée à une mesure objective (le nombre de publication), mais cela n'a rien d'étonnant puisque la reconnaissance se fait principalement au moyen des publications et de plus cette relation était postulée dans la construction de l'indice de "Reconnaissance" exploité par l'étude de l'Unesco.

Rien de très concluant n'a été montré dans ces études. En effet, comme Andrews le remarque à propos de l'étude de l'Unesco, il est nécessaire de faire plus de travaux théoriques sur ces questions pour construire un cadre cohérent d'analyse. Pour le moment ce sont les études de cas qui restent plus significatives.

Nous exposerons ci-dessous une étude quantitative menée dans un seul laboratoire de recherche, et des travaux non-quantitatifs sur des études de cas, avant de conclure, ce chapitre.

3. LES ETUDES DE CAS

3.1. Productivité et perception dans un laboratoire militaire

Stahl et Koser (1978) ont effectué une étude empirique de la performance d'un grand laboratoire de R & D de l'Armée de l'Air américaine. Des questionnaires furent adressés à des

responsables/dirigeants du laboratoire et à des scientifiques et ingénieurs travaillant dans le laboratoire.

L'output du laboratoire était composé d'après cette étude de :

- rapports techniques,
- documents de propositions de recherche adressés à la direction du labo,
- articles publiés,
- inventaire de matériel,
- rapports d'évaluation des propositions
- programmes, logiciels, etc...
- présentations techniques d'une activité de recherche,
- rapports de gestion.

A chaque output, une pondération était calculée à partir des notes attribuées par les responsables du labo, en fonction de l'importance qu'ils accordent à chacun des éléments de l'output. Le degré de corrélation entre les outputs sans pondération et les out-puts pondérés est de 0,96, ce qui signifie que tous les documents appelés "output" sont aussi utiles aux yeux de la direction du laboratoire qu'aux yeux des chercheurs.

Les scientifiques devaient dire quel est le nombre de chaque type de documents qu'ils ont rédigés. Un numéro de classement est attribué à chaque output.

Le classement des responsables est approximativement le même que celui issu des réponses des scientifiques, excepté pour les rapports techniques.

Afin d'expliquer les variations dans les classements un certain nombre de variables ont été construites et testées par un questionnaire envoyé aux scientifiques et ingénieurs. Ces variables sont censées expliquer les variations dans la production des divers output.

Comme dans toutes ces études, les corrélations sont extrêmement faibles (1). Les corrélations positives entre ces mêmes variables ont été trouvées dans d'autres études (2).

Les auteurs proposent également de regrouper leurs 8 mesures d'output en 2 groupes : output des ingénieurs / scientifiques (nombre de rapports techniques, articles publiés, présentations techniques et logicielles d'ordinateurs), et output de direction/gestion (les 4 autres mesures).

Les résultats des corrélations entre ces deux groupes d'output et les variables sont les suivants :

- le niveau de diplôme est associé positivement aux outputs des scientifiques, mais ne l'est pas avec ceux des gestionnaires/dirigeants.

- La communication entre scientifiques est liée positivement à la production totale, la production d'articles et les présentations techniques.

- La fréquence de participation à des congrès, colloques, réunions professionnelles est aussi liée positivement avec la production totale, la production d'articles, de logiciels et surtout de présentations techniques.

- La participation à la décision des objectifs du labo est liée positivement avec ces mêmes variables.

(1) 0,487 à un degré de confiance de 0,99 est la corrélation la plus élevée et la plus exceptionnelle.

(2) M.J. Stahl and M.A. Steger, "Innovation and productivity in R & D : Associated individual and organizational variables", R & D Manag., vol.7, n° 2, pp.71-76, Feb. H.F. Vincent and A.Miraktor, "Relationship between productivity, satisfaction, ability, age, and salary in a military R & D organization", IEEE Trans.Eng. Manag. vol. EM-19, pp.45-52, May 1972.

- Le sentiment des scientifiques d'être récompensé pour leur travail (c'est-à-dire leur production) est liée à la production d'articles, et de présentations techniques. Mais les corrélations sont plus élevées pour la production de documents des gestionnaires/responsables. Les chercheurs se sentent moins bien récompensés que les administrateurs. Ceci est confirmé par l'étude de Pelz et Andrews.

- L'évaluation des projets de recherche est faiblement liée aux productions des gestionnaires, insignifiante pour les scientifiques. Les jugements par les collègues (évaluations) ne sont donc pas corrélés aux indicateurs quantitatifs objectifs. Par contre 'l'évaluation par les pairs (évaluation des projets) corrèle avec la production de brevets. Ceci est dû au fait que les labos étudiés sont des laboratoires principalement de R & D et de conception de produits nouveaux. La corrélation entre le nombre de publications et l'évaluation par les collègues est deux fois plus faible que dans le cas du nombre de brevets déposés.

Comme on peut le voir, ce type de travaux complète et confirme les résultats déjà remarqués dans les autres recherches sociologiques. La très forte relation entre la structure, la hiérarchie des groupes et leur productivité nous semble être un argument suffisant pour présenter certains travaux non quantitatifs.

3.2. Etudes de cas non quantitatives

La plupart des évaluations de l'efficacité et de la productivité des unités de recherche se limite à étudier des unités de recherche industrielle ou des grands laboratoires qui regroupent plusieurs disciplines. Leur sujet est rare-

ment un programme de "big science" (1).

L'étude de Allan Mazur et Elma Boyko (1981) (2) est donc une exception. Ils se sont penchés sur cinq grands projets de recherche océanographique financés par la NSF. Ils cherchent les causes qui permettent d'expliquer leurs succès ou leur échec.

Les projets étudiés étaient coûteux (au moins 3 millions de dollars) avaient plus de dix chercheurs principaux (1), et au moins deux institutions étaient rattachées à ces projets.

Les cinq projets avaient été sélectionnés pour cette étude à partir d'un classement initial de 19 projets. Celui-ci était établi par 4.000 chercheurs auxquels on avait demandé d'effectuer le classement en fonction du succès qu'ils attribuaient à chacun des projets.

Parmi les cinq projets sélectionnés, 3 projets faisaient partie du haut du classement et les deux autres faisaient partie du bas du classement (4)

-
- (1) A l'exception d'une recherche menée en France par Gérard Darmon, Psychosociologie d'une décision en Science lourde, Thèse, 3e cycle, EHESS, 1981;
 - (2) A. Mazur, E. Boyko, Large Scale Ocean Research Projects what makes them succeed or fail ? Social Studies of Science, 11, 1981, pp. 425-449.
 - (3) Les chercheurs principaux sont ceux qui reçoivent le montant de la bourse de recherche de la NSF (principal investigator).
 - (4) Ces projets sont : CLIMAP qui étudiait les variations du climat à très long terme (18 000 ans) ; MODE-1, sur les tourbillons dans les courants des océans ; GEOSECS sur les méthodes de mesure des caractéristiques physiques et chimiques dans l'océan Pacifique, Atlantique et Indien au moyen d'une nouvelle instrumentation ("radioisotope tracers") ; les deux autres projets qui ont été des échecs étaient : CEPEX, sur l'effet des polluants sur la faune et la flore marine, en utilisant de très larges containers, ce qui était une méthode nouvelle à cette époque ; MANGANESE NODULES, sur la formation des nodules de manganèse dans les fonds marins.

Les auteurs ont remarqué que les 5 projets étaient organisés selon la même structure administrative : une hiérarchie à 3 niveaux, mais assez souple typique des universités américaines. De plus, il ne semblait pas avoir de problèmes de communications entre les chercheurs dans aucun des projets étudiés. Les frictions, malentendus, dissensions entre les chercheurs de l'équipe ne semblait pas non plus affecter les résultats de la recherche (1).

Par contre, pour les auteurs de cette étude, quatre éléments permettent de distinguer les projets réussis des échecs.

Les 3 projets ayant abouti ont été initiés par les scientifiques eux-mêmes, qui cherchaient des solutions à des problèmes scientifiques. Les deux autres projets avaient été initiés dans le premier cas pour répondre à un objectif social (la pollution) et dans le second cas parce qu'il y avait des facilités de financement immédiates pour le projet avant la définition des questions scientifiques auxquelles un projet pourrait contribuer. L'indépendance des chercheurs vis-à-vis de la NSF en ce qui concerne les objectifs de la recherche semble être un élément important dans le succès des projets de recherche. Dans ces deux cas bien que certains problèmes étaient identifiés par les scientifiques comme étant importants, les scientifiques engagés dans ces projets ne s'y intéressaient que de manière secondaire. L'intérêt que portent les scientifiques au projet est donc essentiel pour son succès.

- (1) Le sens des termes organisation, communication et frictions sociales et personnelles n'est pas suffisamment clair pour être généralisables à d'autres cas. La question de la communication entre chercheurs a été examinée par J.J. Allen, Managing the Flow of Technology. Cambridge, Mass., MIT Press 1978, Allen cependant n'examine que des unités de recherches en milieu industriel.

Par ailleurs, les scientifiques engagés dans les deux "ratés" recevaient des appréciations moins chaleureuses de la part des pairs, en ce qui concerne leur éminence, ou leur qualité.

Enfin, l'existence d'un noyau stable de chercheurs qui dirigent les projets semble être une caractéristique des projets réussis, bien que la composition de ce noyau puisse varier.

Les auteurs de cette étude remarquent que l'ordre de causalité peut éventuellement être différent de celui qu'ils attribuent au succès ou à l'échec des projets. Certains traits identifiés comme étant des causes de succès peuvent être, en fait, des résultats de ce succès. Ils notent qu'aucun moyen ne leur était disponible pour vérifier ces causes invoquées.

L'indépendance des chercheurs par rapport à l'organisme qui finance la recherche dans la formulation du sujet de recherche, l'intérêt que portent les chercheurs à cette recherche, la qualité des scientifiques qui y participent et l'existence d'un noyau solide qui dirige la recherche, sont donc autant d'éléments qui permettent de comprendre le succès ou l'échec d'un projet.

Une autre étude menée par Gilbert permet de cerner les problèmes d'organisation et leur influence sur la productivité des chercheurs (1).

(1) G.N. Gilbert, Competition, Differentiation, and Careers in Science, Social Science Information, vol.16, 1977, pp.113-122.

L'étude de Gilbert porte sur 25 scientifiques travaillant sur la détection radar des météores entre 1940 et les années cinquante. Au sein de ce groupe 4 thèmes de recherche différents ont apparus, définis par les scientifiques eux-mêmes. La productivité (publications) de ces quatre groupes avait atteint un maximum en 1948 pour le premier groupe, en 1949 et 1951 pour le deuxième, en 1953 et 1956 pour le troisième et en 1959 pour le quatrième groupe. A ce niveau fin de l'analyse on s'aperçoit que la productivité à deux maxima n'a pas nécessairement lieu. Chaque groupe était dirigé par un responsable de recherche (2 chefs d'équipes pour l'un des groupes). Ce furent ces chercheurs qui avaient défini les problèmes, et qui avaient un droit de propriété sur chacun de ces domaines. Ces chefs d'équipes avaient une productivité environ deux fois plus importante que les autres chercheurs. Les scientifiques qui rentraient dans les domaines de recherche après leur définition par les chefs d'équipes avaient toujours une position inférieure à ces derniers, surtout s'ils venaient à un moment de croissance (avant que ne soit atteint un maximum de production d'articles pour chaque problème) ; par contre, les scientifiques qui rentraient à une période de déclin d'un problème étaient dirigés vers de nouvelles recherches. Cette remarque est, d'après Gilbert, assez générale : elle se produit dans tous les domaines.

Il existe bien d'autres études de cas qui rentrent non seulement dans le détail de l'organisation de la recherche mais aussi dans le contenu des recherches. L'objectif des recherches de ce genre est de montrer comment sont construits les faits scientifiques. (1)

-
- (1) Une discussion et compilation de ces travaux : M. Mulkey Science and the Sociology of Science, Londres, Allen and Unwin, 1979. Une série de traductions de ces travaux La Science Telle qu'elle se fait, Paris, Pandore, 1982. Enfin il existe une très bonne étude historique qui aboutit aux mêmes grandes conclusions que celle de Mazur et Boyko et celle de Gilbert : J.B. Morell, The Chemist breeders : the research schools of Liebig and Thomas Thomson, AMBIX, vol. 19 (1), 1972), pp. 1-46 ; voir aussi D. Bloor, Sociologie de la logique, éditions Pandore, Paris, 1983, pp. 38-41 qui replace ce type de travaux dans le contexte des débats épistémologiques.

4. CONCLUSION

A ce stade de notre survol de la littérature nous pouvons faire un certain nombre de remarques générales :

- C'est la position d'un chercheur dans une équipe qui permet de comprendre sa productivité, qu'il soit chef d'équipe (auquel cas il aura un avantage certain sur ses collègues) ou simple membre de l'équipe.

- La force d'un chercheur vient de sa capacité à formuler des problèmes de recherche, à enrôler des chercheurs sur ces problèmes, à convaincre les autres, scientifiques et non-scientifiques, de la pertinence de cette voie de recherche.(1)

- La productivité du groupe et sa structure seront donc déterminés par cet ensemble d'opérations, que l'on peut appeler *opérations de traductions*.

- Quant à l'évaluation des recherches, on ne peut donc la concevoir que comme un résultat de ces opérations de définition des problèmes et de traduction. Un processus de recherche scientifique n'est jamais pur en ce qu'il mélange des considérations scientifiques avec des considérations "basement matérielles", comme les questions de financement de carrière, de promotion, etc. Si, et seulement si les diverses considérations non scientifiques étaient exclues ou effacées de l'esprit des chercheurs eux-mêmes, nous pourrions dire que l'évaluation se fait avant la recherche. Mais

 (1) Problématique développée par Michel Callon, De problèmes en problèmes, Itinéraire d'un laboratoire universitaire pris dans l'aventure technologique, rapport CORDES, 1978.

cette vision est mécaniste et fait abstraction des chercheurs (d'où peut être leur méfiance vis-à-vis des débats sur l'évaluation, qu'ils qualifient de "trop politiques"). La participation massive des scientifiques dans les procédures d'évaluation aux Etats-Unis est là pour rappeler cette réalité.

Les travaux quantitatifs dont nous avons examiné la teneur dans ces pages sont dans un certain sens des évaluations a posteriori. Leurs résultats sont solides uniquement dans la mesure où ils portent non pas sur des recherches en train de se faire mais sur des recherches déjà faites. Ils peuvent être lus, et ne manqueront pas d'être lus comme des sanctions. C'est ce qui les rend d'un usage délicat. Si ces travaux n'ont montré qu'une chose c'est celle-ci : les instruments utilisés pour évaluer ne sont pas neutres et le résultat d'une évaluation en dépendra de manière cruciale.

CHAPITRE 4

L'EVALUATION DE LA SCIENCE ET LA POLITIQUE SCIENTIFIQUE

RESUME :

Par sa position particulière au sein de l'ensemble des institutions qui financent la recherche aux Etats-Unis, la NSF a acquis un rôle central dans l'évaluation de la science au niveau national, malgré des réticences exprimées de manière vigoureuse jusqu'au milieu des années soixante dix. Depuis lors, la NSF a joué un rôle actif dans la définition de la politique scientifique en cherchant à faire participer les scientifiques à la définition des objectifs de la politique scientifique. Elle a aussi permis un rapprochement ou plus exactement une collaboration des administrateurs et des scientifiques.

Trois documents indiquent cette nouvelle orientation : les rapports annuels auprès du Congrès, le "Five Year Outlook" et les "Sciences Indicators".

Nous examinerons dans ce chapitre en détail l'origine et le rôle des Sciences Indicators dont l'originalité réside dans la présentation de données statistiques concernant les grands débats scientifiques du moment, ainsi que le rôle de centre d'information sur la science et la technologie et centre de recherche qu'est la "Science Indicator Unit" au sein de la NSF.

Nous rappelons l'essentiel du contenu des Sciences Indicators en insistant tout particulièrement sur deux types d'indicateurs ; les indicateurs bibliométriques et les indicateurs composites de technologie.

Enfin, nous rappelons que les Science Indicators ont stimulé plutôt qu'ils n'ont empêché la discussion entre scientifiques et administration car il est devenu le terrain commun sur lequel ont pu se dérouler ces discussions.

L'évaluation de la science et de la technologie au niveau national est un processus complexe. Bien que le processus de fixation des objectifs à long terme de la politique scientifique aux Etats-Unis soit extrêmement décentralisé. Cependant "la détermination des objectifs globaux, l'évaluation des programmes ("program assessment and integration") a été pendant longtemps l'une des prérogatives du Président des Etats-Unis, principalement à travers le processus budgétaire et les propositions de loi des finances que l'exécutif soumet au Congrès (Beckler, 1974, p. 130⁽¹⁾).

La cohérence de ce processus est assurée par le processus budgétaire qui fait l'unité de ce système. Mais les objectifs budgétaires sont des objectifs à court terme. Comment alors fixer les objectifs à long terme ? On ne répondra pas ici à cette question qui fait l'objet de

(1) David Z Beckler, *The Precarious Life of Science in the White House, Daedalus*, Eté 1974, pp. 115-134.

plusieurs études de politique scientifique⁽¹⁾. Mais on se bordera à examiner le rôle de la NSF dans ce processus d'évaluation au niveau national des activités scientifiques.

1. Evaluer et coordonner

Dès son origine, il était prévu que la NSF devait, en plus de son soutien financier des projets de recherche, "évaluer les programmes de recherche des agences fédérales et coordonner les programmes de recherche de la Fondation avec ceux entrepris par des chercheurs individuels et des groupes de recherche publics ou privés"⁽²⁾.

Cependant, tout au long de son histoire, la direction de la NSF a été très réticente à exercer ce pouvoir d'évaluation et de coordination⁽³⁾.

Waterman, le premier directeur de la NSF, avait déjà dit, quand il travaillait à l'Office of Naval Research, que l'évaluation de l'effort de recherche de l'ensemble du pays est une tâche impossible ; une fondation sage et intelligente ne devrait pas s'y essayer"⁽⁴⁾.

(1) Une excellente étude comparative des appareils de politique scientifique aux Etats-Unis, en France et en Hollande, a été présentée par Ronald Brickman et Arie Rip (1979), Science Policy Advisory Councils in France, The Netherlands and the United States, 1957-77, Social Studies of Science, vol. 9, pp. 167-198.

(2) Loi Publique constitutive de NSF, PL 507

(3) Voir M. Lomask, A. Minor Miracle, and Informal History of the NSF, NSF, 197_, chapitre 6 "To coordinate or not to coordinate", pp. 91-110

(4) ibid. p. 96

Mais cette fonction qu'ont refusé les premiers directeurs de la NSF, a été de fait imposée à la NSF. La loi de réorganisation de la NSF en 1976 stipule en effet que la Fondation "doit évaluer l'état et les besoins des diverses sciences comme ceux-ci apparaissent à travers les programmes et études effectuées par les agences fédérales, par des particuliers ou des groupes privés ou publics de chercheurs financés par bourses, contrats (la NSF) peut s'adresser à des groupes de consultants pour effectuer de telles évaluations ; la (NSF) doit tenir compte du résultat de ces évaluations pour coordonner ses programmes de recherche et d'enseignement avec les programmes, projets et études des autres agences fédérales, des autres chercheurs particuliers ou groupes de recherches publics ou privés"⁽¹⁾.

Il faut noter que ce travail d'évaluation et de coordination nationale devait faire l'objet de plus d'attention de la part du Policy Science Advisory Council auprès du Président des Etats-Unis et plus tard de l'Office of Science and Technology Policy (OSTP) constitué en 1976.

La NSF a acquis une autorité incontestable en matière de politique scientifique tant vis-à-vis de la communauté scientifique que vis-à-vis du pouvoir exécutif et législatif ⁽²⁾. Même depuis la constitution de l'OSTP en 1976, la NSF a continué à avoir un rôle important dans la politique scientifique et l'OSTP a continué à déléguer ses fonctions d'analyse à la NSF.

(1) P.L. 94-282, The National Science and Technology Policy, Organization and Priorities Act of 1976.

(2) Voir, Etats-Unis d'Amérique, dans le document de l'UNESCO, Politiques scientifiques et technologies nationales en Europe et Amérique du Nord, 1978, Etat actuel et Perspectives, Etudes et Documents de politique scientifique n° 43, pp. 129-148.

L'OSTP a d'ailleurs été très durement critiqué au sujet de cette délégation des fonctions d'analyse et d'évaluation auprès de la NSF.

2. L'expertise de la NSF

Etablir les priorités de la politique scientifique a long terme comme à court terme (c'est-à-dire en termes budgétaires) ne peut être effectué que par l'évaluation de l'impact de la science et technologie sur l'économie et la société, l'évaluation de la productivité interne de l'activité scientifique afin l'évaluation du soutien de l'Etat par les dépenses publiques, ses institutions, son influence juridique (réglementation) et fiscale.

Quelle doit être la nature de ces évaluations ?

Peuvent-elles être quantitatives ? Doivent-elles l'être ? C'est poser là la question de l'information nécessaire à la politique scientifique. Le rôle du "Conseil scientifique", du type d'expertise nécessaire, pour établir une politique, sont des sujets très abondamment étudiés aux Etats-Unis ⁽¹⁾.

Les personnes qui établissent la politique scientifique veulent des avis d'experts, objectifs, "positifs, quantitatifs et d'autorité"⁽²⁾. L'information fournie doit être utile.

(1) Un inventaire des propositions concernant les conseils et conseillers de la politique scientifique, a été établi par Susan G. Hadden, Technical Advice in Policy Making : a propositionnal inventory, dans Haberer (1977) op. cit. pp. 81-98.

(2) Subcommittee on Science, Research, and Development, Committee on Science and Astronomy, U.S. House, 92nd Congress Session, Technical Information for Congress, Washington D.C., G.P.O. 1972.

Le contenu et le type d'informations sur la science et la technologie dépend très largement du contexte politique⁽¹⁾. Le Rôle croissant de la NSF vient en partie de son activité, sa place, et sa réputation au moment de la période de crise politique du milieu des années 70. La réputation de la science, la réputation du scientifique lui-même, celle enfin de l'institution scientifique sont nécessaires pour rendre crédible une analyse de la politique scientifique⁽²⁾.

Au moment où fut aboli le PSAC par Nixon, la NSF était l'institution scientifique jouissant de la plus forte réputation et qui était la plus crédible⁽³⁾. De plus, en reprenant un rôle actif dans la définition de la politique scientifique, elle ne faisait que reprendre une fonction qui lui avait été attribuée légalement dès sa constitution en 1950. C'est plus par la force des situations que par sa volonté que la NSF s'est trouvée au centre de ces activités d'analyse de la Science et Technologie. Enfin, la NSF disposait d'un important appareil de statistiques sur la science, ce qui faisait d'elle l'institution toute indiquée pour effectuer des études documentaires sur les ressources scientifiques du pays⁽⁴⁾.

-
- (1) C'est ce que montrent toutes les études sur la politique et la science (voir Salomon J.J., Science et Politique, Le Seuil, Paris 1970), et ce que savent tous les administrateurs par leur pratique quotidienne.
 - (2) Subcommittee, 1972, op. cit. p. 415
 - (3) Par le fait qu'elle était indépendante, les pressions politiques au Congrès, que les scientifiques chérissaient son mode de fonctionnement et qu'elle entretenait l'excellents rapports avec le pouvoir exécutif, notamment le bureau du Budget, cf. J. Walsh (1974) The view from the Executive Office repris, dans Kuehn et Porter (1979) Science, Technology and National Policy, pp. 274-281
 - (4) Des documents d'analyse avaient été publiés bien avant 1976 : "Graduate education in sciences" en 1969, "State of the Physical Sciences" en 1970, "Environmental Science" en 1971, "Role of engineers and Scientists in shaping a national Policy for technology" en 1972.

3. Les documents d'analyse issus de la NSF

La NSF établit trois documents publics d'analyse de la politique scientifique :

- le "Five Year Outlook on Science and Technology" qui est un document de prospection de la science et de la technologie aux Etats-Unis,
- le rapport annuel sur l'état de la science et de la technologie, qui identifie les principales tendances actuelles de la recherche, rappelle les questions les plus actuelles de la politique de la science,
- le "Science Indicators" qui fournit des informations quantitatives sur la science et la technologie. Science Indicator est publié tous les deux ans.

Ces trois documents d'analyse tracent les objectifs prioritaires de la politique scientifique américaine. Ces objectifs sont les suivants :

- (1) Garantir la paix mondiale,
- (2) Accroître l'utilisation efficace des produits et matériaux nouveaux,
- (3) Assurer une offre alimentaire, énergétique et matérielle suffisante,
- (4) Contribuer à la sécurité nationale,
- (5) Améliorer les soins médicaux,
- (6) Sauvegarder, favoriser et restaurer l'environnement,
- (7) Protéger l'océan et les régions polaires et utiliser leurs ressources efficacement,
- (8) Soutenir l'économie et assurer le plein emploi au moyen des innovations scientifiques et techniques,

- (9) Accroître la qualité de l'enseignement,
- (10) Promouvoir la conservation des ressources humaines et naturelles nationales,
- (11) Améliorer le logement, les transports et les systèmes de communication et fournir les services publics efficaces,
- (12) Eliminer la pollution de l'eau et de l'air ainsi que les médicaments inefficaces et les additifs alimentaires dangereux,
- (13) Promouvoir l'exploration et l'utilisation pacifique de l'espace.

Ces 13 priorités, celles de la loi de 1976, sont périodiquement discutées dans les documents d'analyse de la science et de la technologie. Les maîtres mots en sont l'efficacité et la productivité par l'innovation, la science et la technologie.

Les efforts de l'appareil de décision de la Maison-Blanche depuis cette époque ont été de plus en plus tournés vers la coordination des activités scientifiques à long terme et l'analyse des budgets par fonctions plutôt que par institutions. Cet effort, relativement récent, marque un tournant dans la politique scientifique américaine⁽¹⁾.

En effet, après le célèbre rapport Bush (1945) et le rapport Steelman (1947) et pendant 30 ans, l'administration américaine n'a produit aucun document global sur la science. Seuls des travaux ponctuels ont été publiés. Les conseillers scientifiques du président des Etats-Unis se sont assez systématiquement opposés à la publication de documents sur

- (1) Pour une bonne description historique de la notion de politique scientifique aux Etats-Unis on peut lire J. Schmandt (1977), Science Policy : one step forward, two steps back, dans Haberer (ed.) Science and Technology Policy, Lexington Books 1977, pp. 9-24 ; voir aussi Beckler (1974) op.cit.

l'ensemble des activités scientifiques américaines. Kistiakowski qui fut le second conseiller scientifique de Eisenhower illustre mieux que tout autre ce refus quand il disait que tout rapport sur l'ensemble de la science est une pure perte de temps (Schmandt 1981)⁽¹⁾.

Peut-être faut-il voir dans ce changement d'attitude vis-à-vis de l'analyse de la science, un changement du personnel qui peuple les administrations à Washington. Beaucoup plus de chercheurs en sciences sociales constituent le corps des administrations. L'un des premiers rapports d'ensemble sur la politique de l'administration américaine fut le rapport présenté par Henry Kissinger sur la politique internationale des Etats-Unis. Ce rapport devait exprimer "la philosophie générale du pays en matière de politique étrangère". "Kissinger avait une formation en sciences sociales et de ce fait était plus tourné vers la tradition des généralistes qui soulignent la nécessité d'une analyse politique globale ("comprehensive")" (Schmandt 1981, op. cit., p. 121). Le même phénomène s'est produit dans d'autres domaines, comme par exemple la protection de l'environnement⁽²⁾.

A ce sujet il faut souligner que ce changement de perspective a été stimulé par le pouvoir législatif (le Congrès). Ainsi, le General Accounting Office (GAO) qui est l'organisme de contrôle et l'audit (accounting) du Congrès, a instauré un institut d'évaluation chargé de contrôler, d'examiner

(1) Schmandt J. (1981). The five year outlook as a policy document : historical and comparative perspectives, Technological Forecasting and Social Change, 20 (2), septembre 1981, pp. 113-138.

(2) Perl (1979). The scientific advisory system : some observations, in Kuehn et Porter (eds.) Science Technology a National Policy, pp. 262-274, pense que le système de décision de l'exécutif, même en associant la communauté, n'est pas efficace.

et éventuellement soutenir des études d'évaluation dans tous les domaines ⁽¹⁾.

Notons cependant que l'évaluation n'a pas été coordonnée par une instance particulière. A part les efforts du NIH (voir ch. 5) et du GAO, l'évaluation de la recherche fut toujours menée au coup par coup selon les besoins et les demandes politiques du moment.

4. Le rapport annuel au Congrès

Les documents de politique scientifique de la NSF sont des documents qui émanent de la communauté scientifique. Ils sont des produits de l'autogouvernement des scientifiques ⁽²⁾ tant dans leur contenu que dans leur forme.

Dans le contenu des rapports annuels de l'OSTP on peut voir s'exprimer cet autogouvernement : "les critères essentiels pour décider du financement d'une discipline ou d'un

(1) Voir la publication annuelle du répertoire de toutes les études d'évaluation des programmes fédéraux, GAO, Evaluation Studies, 1982.

(2) M. Lomask (1976) A minor miracle, NSF, p. 277, rapporte ces paroles de Stever, directeur de la NSF au moment de la création de l'Office of Science and Technology Policy (OSTP) en 1976 : "l'arrangement institutionnel actuel (avant la création de l'OSTP où le rôle de la NSF est prédominant) a beaucoup d'avantages. Voyez par exemple certains de nos problèmes nationaux les plus durs. Grâce aux moyens de la NSF, nous pouvons examiner ces questions plus profondément que n'importe quel bureau de la Maison-Blanche (...). Voyez par exemple l'analyse fonctionnelle du budget que nous avons effectué en juin 1975, le rapport IMOS (rapport sur les fluorocarbones et l'environnement et les effets inattendus des aérosols sur la stratosphère)".

projet sur fonds publics doit être l'excellence des chercheurs, l'excellence du sujet de recherche et l'excellence des résultats attendus (...). En outre, un autre critère important (...) doit être la pertinence de la recherche par rapport aux besoins économiques et sociaux de la Nation" (Annual Report to the Congress : 1981, p. 5). L'adéquation de la recherche aux objectifs sociaux et économiques est toujours désignée en dernier.

Mais dans leur forme également, ces documents sont bien des produits de la communauté scientifique. En effet, sur chaque sujet, chaque discipline ou chaque domaine de recherche examiné, ces rapports comportent une argumentation scientifique précise sur ce qui s'y fait et ce qui doit être entrepris.

La première analyse de cette sorte fut produite en 1975, avant la création de l'OSTP⁽¹⁾. C'était là le premier document portant sur des disciplines ou des fonctions budgétaires (santé, énergie, climat) et concernant les activités et les budgets de plusieurs agences fédérales. Stever (alors directeur de la NSF) disait à son sujet : "Je suppose que de tels documents budgétaires seront un jour préparés pour le congrès annuellement"⁽²⁾. La prédiction de Stever s'est bien réalisée : le rapport annuel de l'OSTP au Congrès, auquel participe actuellement la NSF est bien une analyse fonctionnelle du budget et des activités des agences fédérales de R et D.

Les rapports annuels de l'OSTP, élaborés conjointement par l'OSTP et la Direction de la Planification, la Direction de Recherches sur la politique Scientifique, Technologique

(1) "Report on the Federal R & D Program : F.Y. 1976"

(2) M. Lomask (1976) op. cit. p.273

et des Affaires Internationales de la NSF, comme le Five-Year Outlook, permettent de cerner les principales questions de politique scientifique. En 1980, la commission de la science et de la technologie de la Chambre des Représentants reprenait les divers thèmes problématiques que le Congrès devait examiner.

Ces thèmes étaient (et sont encore) les suivants :

- Audit des centres de R & D,
- Extension des relations entre les universités et la recherche industrielle,
- Contrôle du matériel afin de s'assurer de leur utilisation optimale, et afin d'empêcher l'obsolescence du matériel,
- Examen de la qualité de l'enseignement supérieur scientifique et technique,
- Examen des diminutions budgétaires et de leur influence sur le fonctionnement des universités et centres de recherche publics,
- Réévaluation des besoins des institutions de recherche ; réévaluation des Centres Nationaux de Recherches voir plus loin l'évaluation des programmes MRL - Laboratoires de recherches sur les matériaux),
- Examiner les processus de fixation des priorités budgétaires et des priorités à long terme de la recherche.

"Les avantages et les inconvénients du système actuel de fixation des objectifs scientifiques, doit faire l'objet d'une analyse approfondie ; le financement de ce processus pourrait être envisagé sous la forme d'un financement commun ("joint venture") entre le gouvernement, les universités et les industries, et en incluant dans ce processus les représentants du management et des travailleurs ("management and labour"). Ce

mécanisme devrait être indépendant du pouvoir législatif, exécutif ou juridique et devrait être uniquement un mécanisme consultatif". Cette proposition de la commission ne faisait finalement que décrire l'état actuel des choses.

5. Le Five Year outlook : document "participatif".

La loi de réorganisation de l'appareil scientifique fédéral de 1976 marque donc un tournant que traduisent les publications d'ensemble sur la science et la technologie. Elles sont probablement un premier pas vers ce que Nelkin a appelé le contrôle participatif de la science et de la technologie⁽¹⁾.

Le texte final du "Five Year Outlook" a été rédigé à partir de documents fournis et discutés par plusieurs "partenaires" : la NAS, le SSRC, l'AAAS, des experts en matière de politique scientifique, des scientifiques, des universitaires, et des représentants des différentes agences fédérales⁽²⁾. Le mouvement de va et vient entre les rédacteurs de la NSF et ces partenaires était assuré essentiellement par quatre personnes du "Office of Spécial Projects" de la NSF⁽³⁾.

Les textes d'analyse sur des sujets précis rédigés par les experts portaient sur certains domaines scientifiques et

- (1) Voir notamment D. Nelkin (1977) Trends in Science Policy : The Search for Controls, dans J. Haberer (ed), Science and Technology Policy, Lexington Books, 1977, pp. 69-80.
- (2) Voir : The Five Year Outlook, numéro special de technological Forecasting and Social Change, 20 (2) septembre 1981, notamment W.A. Blanpied et A.I. Leshner, Origins and Aims of the Five-Year Outlook, pp. 145-154.
- (3) L'office of Special projects était composé de 4 membres, mais d'après D. Chubin et F.A. Rossini, ce furent essentiellement William Blanpied et Alan Leshner qui jouèrent ce rôle d'intermédiaire. Voir The Five Year Outlook, op. cit.

technologiques qui demandent à être mis en priorité⁽¹⁾. Ces domaines étaient la commercialisation des produits spatiaux, le génie génétique et la robotique. Quatre autres documents devaient porter leur attention sur les problèmes plus généraux de politique scientifique : la R & D industrielle et le processus de l'innovation, la contribution des sciences sociales à l'innovation et à la productivité, l'obsolescence du matériel scientifique et l'enseignement scientifique et technique, la coopération internationale en science et en technologie.

Le thème dominant de ces textes est l'accroissement de la productivité : productivité industrielle, économique, productivité de la recherche, des chercheurs, des agences fédérales. Ce thème dominant assure la cohérence de ces documents.

Ainsi le Five Year Outlook apparaît-il comme une évaluation prospective de la science et de la technologie américaine en même temps qu'il fut un forum de discussion sur les priorités nationales de la politique scientifique.

6. Les "Science Indicators" (SI)

Ce document fournit une synthèse des données quantitatives disponibles sur la science et la technologie. Contrairement aux deux autres documents de politique scientifique examinés ci-dessus, le "Science Indicators" ne fournit pas de recommandations politiques.

Très largement distribué⁽²⁾, SI est un pôle autour duquel gravitent universitaires, gestionnaires, administrateurs,

(1) Ils furent d'ailleurs publiés par la NSF : *Emerging Issues in Science and Technology*, 1981, A compendium of working Papers for the NSF, NSF 1982;

(2) Science Indicators est distribué gratuitement. L'édition de 1980 a été diffusée à plus de 9 000 exemplaires dès sa parution ; actuellement, sa diffusion est estimée à plus de 15 000 exemplaires

et hommes politiques. La position de ce document est stratégique. De plus, les données contenues sont accompagnées d'un commentaire interprétant les chiffres présentés. Il est élaboré par un bureau spécial de la NSF, la "Science Indicator Unit" qui fait partie du bureau statistique de la Fondation. Malgré ce rattachement institutionnel à la NSF, le SI est un document crédible aux yeux de ses différents utilisateurs. Cela est probablement dû en partie au "capital de confiance" dont dispose la NSF, qui ne peut être soupçonnée de déformer la réalité ; mais c'est aussi dû à son origine et au fonctionnement de l'unité qui a conçu et qui élabore tous les deux ans ce document, depuis 1970 (la première édition de SI date de 1972).

a) Naissance du SI.

Le SI initialement a été voulu par le Président des Etats-Unis et ses conseillers scientifiques. L'idée était de faire un document semblable aux rapports économiques pour le Président des Etats-Unis. Ce document de politique scientifique devait contenir une description de "l'état de santé de la science" et des informations statistiques concernant la Science et la Technologie, à l'image des rapports économiques des conseillers économiques.

Il semble qu'il y ait eu un débat politique qui a opposé la Maison-Blanche au NSB, l'organe directeur de la NSF.

Finalement, la situation, semble-t-il, assez épineuse, fut dénouée par une décision abrupte du NSB. La décision était de créer un document informatif, sans recommandations de politique scientifique, ce qui évitait d'entrer dans le débat politique.

Le caractère moteur du SI-72 venait d'une part de la masse de données compilées, qui étaient auparavant difficilement accessibles ; d'autre part, les statistiques concernant la science recevaient leur baptême. C'était là une façon d'annoncer un domaine nouveau⁽¹⁾.

L'impact de ce document est considérable. SI non seulement a permis de disposer de données quantitatives sur les questions de politique scientifique. Il a également été l'un des principaux catalyseurs dans la formation de la "scientométrie"⁽²⁾.

b. Le rôle des S.I.

Le coût de fabrication des SI-80 représente environ un million de \$, (soit 0,0014 % du budget 1979 de la NSF). Ceci est peu si on examine le type de travaux qu'effectue l'unité des Science Indicators et l'impact qu'ils ont eu.

- (1) L'importance du titre du document "Science Indicators" m'a été signalée par plusieurs personnes.
- (2) S.I. a donné lieu à beaucoup de discussions sur l'analyse quantitative de la science. Trois conférences ont eu lieu jusqu'à maintenant sur ce sujet. Elles ont donné lieu à la publication de plusieurs articles d'une part dans le livre Y. Elkana, J. Lederberg, R.K. Merton, A. Thacray, H. Zuckerman (eds), "Toward a Metric of Science : The Advent of Science Indicators", Wiley and Sons, 1978 ; d'autre part, un numéro spécial de Scientometrics est consacré au Science Indicators (1980). Enfin, plus récemment, 2 numéros de la revue Science, Technology and Human Values réunissent des contributions sur les indicateurs de qualité en Science. Vol. 7, n° 38 et Vol. 7 n° 39, 1982. Voir chapitre 1 de ce rapport.

L'Unité des "Science Indicators" effectue deux sortes de travaux :

- Premièrement, l'unité donne des informations détaillées à quiconque les demande. Ce travail de "clearing house" en information est un service très prestant. Les demandeurs sont pour la plupart des personnes engagées dans les processus d'élaboration de la politique scientifique. La plupart des demandes concernent des données qui sont trop agrégées ; les demandeurs veulent des données plus détaillées.

- Deuxièmement, l'unité s'intéresse à l'amélioration des travaux quantitatifs sur la science. Tous les ans, l'unité finance 5 projets de recherche qui travaillent sur la création d'indicateurs scientifiques. De plus, le personnel de l'unité des S.I. effectue des recherches sur ces domaines.

C'est dans cette double caractéristique des S.I., bourse d'informations et centre de recherche, que réside son originalité.

c) Les qualités de l'unité des S.I.

Une unité d'informations scientifiques doit avoir certaines qualités qui lui permettent de ne pas voir son activité périlcliter. Parmi ces qualités, la plus importante est sa crédibilité. Les informations sur la science sont nécessaires à l'élaboration de la politique scientifique. L'unité de "Science Indicators" ne doit pas rentrer dans ce processus politique, sinon les données du S.I. risquent de ne plus être crédibles, disent les responsables de l'unité.

Une autre qualité doit être sa capacité d'identifier les principales questions de politique scientifique et de présenter les données, en fonction de ces questions. Ceci lui permet de sélectionner l'information et de la présenter sous une forme utile.

Le choix qu'il faut effectuer dans l'information dépend du public auquel le document est destiné. C'est cette capacité d'effectuer des choix en fonction du public qui distingue le S.I. des documents de statistiques "brutes", et lui confère son dynamisme.

Une autre qualité que possède l'unité est la facilité d'accès aux données. Elle peut sembler contradictoire avec l'exigence de probité qui nécessite de maintenir une distance autant vis-à-vis de la NSF que des autres demandeurs. Ce qui est vital pour les S.I., c'est de maintenir toujours la même distance entre l'unité de SI et les utilisateurs.

d) Les S.I. et l'évaluation

L'unité qui conçoit les S.I. n'est pas directement concernée par l'évaluation des institutions de recherche. Cependant les indicateurs élaborés peuvent être utilisés dans des études d'évaluation. C'est le cas des indicateurs bibliométriques élaborés par l'équipe de Narin pour la NSF (voir plus loin), des indicateurs composites de technologie et des statistiques de brevets.

Certains auteurs ont remarqué que le "Science Indicators" ne peut pas être évaluatif parce qu'il n'existe pas de mesure d'output satisfaisante, et que le S.I. ne répondait pas aux exigences d'évaluation de l'activité scientifique, exprimées par les agences fédérales, notamment l'Office of Management and Budget (Y. Ezrahi, in "Toward a Metric of Science", 1978, p. 323).

Il faut aussi rappeler que les S. I., dans leur forme actuelle, ont été conçus contre la volonté du pouvoir exécutif.

Enfin, même si les mesures d'output ne sont pas exhaustives et ne représentent pas la véritable production de la science et de la technologie, il n'en demeure pas moins que les séries statistiques que fournit le S.I. sont suffisamment développées pour permettre des comparaisons dans le temps, entre institutions et pays. La démarche d'une étude d'évaluation est principalement comparative. En ce sens, le S.I. fournit une base de données suffisamment valable pour effectuer des études d'évaluation.

Les critiques adressées aux S.I. émanent habituellement du secteur universitaire. Or, "les universitaires ont une vision du gouvernement et de l'administration où toute activité est le résultat d'une conspiration" dit-on à la NSF. Sans aller jusque là, il est certain que les universitaires ont tendance à voir l'administration et l'Etat comme un groupe homogène. Or, tout membre de l'administration ne peut à aucun moment avoir cette vision car toute activité est le résultat de négociations parfois laborieuses. L'équipe du S.I. qui par la nature de son travail se trouve à la frontière de l'administration et du monde "privé et académique" (les industries et les universités). Cette position les rend évidemment doublement vulnérables.

Enfin, outre les critiques qui lui sont adressées par les représentants du secteur universitaire, le S.I. est souvent mis à l'index par tel ou tel service de l'administration.

Dans un tel contexte, la raison pour laquelle S.I. a pu être publié pendant 10 ans réside dans cette double caractéristique de l'unité que nous avons souligné ci-dessus : être une source d'informations, c'est-à-dire où tous peuvent prendre des informations et en faire ce qu'ils veulent, et être une unité de recherche, ce qui garantit le sérieux des statistiques qui sont élaborées.

e) Les mesures de la science contenues dans le S.I.

Les chapitres du S.I. montrent les différents niveaux d'intérêt dans la construction d'indicateurs de la science. Les indicateurs de ressources pour la science, de productivité de la R & D, d'output de la R & D (littérature ou brevets), sont étudiés au niveau international et au niveau national.

Au niveau international s'ajoutent le commerce international de produits à haute intensité technologique ; les brevets U.S. dans le monde comparés aux brevets étrangers déposés aux Etats-Unis ; la R & D étrangère menée aux USA ; les canaux que suivent les transferts de technologie, la coopération scientifique internationale.

Au niveau national s'ajoutent les études plus détaillées des dépenses de R & D par sources et par utilisateur, ainsi que par des fonctions (défense, espace, R & D civile, etc...) les moyens indirects pour soutenir les activités de R & D (politique fiscale, financement des centres de recherche indépendants contractés principalement par le DOD et la NASA) ; la participation de chaque agence fédérale à la Recherche fondamentale nationale ; le temps de travail de chaque ingénieur et scientifique consacré à la recherche et celui consacré à des tâches d'enseignement et d'administration ; des statistiques sur l'emploi des scientifiques et des ingénieurs et leur formation.

Le SI-80 a la particularité de désagréger la recherche fondamentale de la Recherche appliquée ou encore du "development". Confondre la recherche "pure" et la recherche "appliquée" étaient, d'après plusieurs observateurs, un défaut du S.I. Holton, de manière quelque peu provoquante, disait que le modèle implicite du S.I. était un modèle "baconien" qui met l'accent sur les sciences "dures" et la technologie

et qui souligne plutôt l'aspect concurrentiel que l'aspect coopératif de la recherche. Dereck Price assignait à la bonne scientométrie, la tâche de distinguer la "R" du "D".

Avec l'édition de 1980 du S.I., cette désagrégation tant voulue est maintenant chose faite. Deux chapitres différents couvrent respectivement la science fondamentale et la science et la technologie industrielle. De la sorte, le S.I. met l'accent sur l'aspect coopératif de la science, notamment sur la coopération entre l'industrie et l'université, qui fait actuellement l'objet d'un très large débat de politique scientifique.

S.I. aborde la question de la définition de la science fondamentale et de la science appliquée en ces termes : "une activité particulière sera dénommée fondamentale en fonction de l'organisme qui finance cette recherche, de l'organisme ou de l'individu qui effectue la recherche. Mais les définitions dont nous disposons sont comparables dans le temps (...). Ainsi, les tendances sont mesurées comparativement avec un degré de confiance suffisant" (SI-80, p. 73).

Les dépenses en Recherche Fondamentale sont examinées en dollars courants et constants. Pour ces derniers, la NSF a élaboré un déflateur implicite sur le PNB (voir Manuel de Frascati).

La provenance des fonds pour la recherche est par la suite examinée à partir des documents du budget (équivalent à la présentation fonctionnelle de la "Loi des Finances" en France). Ainsi on s'aperçoit par exemple que la masse de laboratoires industriels financés par des fonds fédéraux a énormément augmenté.

Les dépenses en Recherche Fondamentale sont ensuite examinées par utilisateurs. les universités représentent 50 % de l'effort en matière de Recherche Fondamentale ; les laboratoires fédéraux 15 % ; les laboratoires privés également 15 %. S.I. estime cependant que sur les 15 % de l'industrie, seuls 5 % sont véritablement de la Recherche Fondamentale.

Dans S.I. 80, les tendances de financement de la Recherche industrielle et celles de la Recherche universitaire sont comparées avec plus de précisions qu'auparavant ; de même, le nombre de chercheurs et de diplômés dans la Recherche Fondamentale est comparé aux dépenses de R & D. On voit ici apparaître de nouveau le thème dominant de discussions sur la politique scientifique entre 1978 et maintenant les rapports industries-universités.

Enfin, S.I.-80 consacre plusieurs pages à ce problème, particulièrement en évaluant les coûts administratifs en temps et en argent qu'engendre la gestion des programmes de recherche.

f) Les indicateurs bibliométriques de S.I.

Le moyen le plus efficace dont on dispose pour évaluer la productivité de la recherche fondamentale c'est le dénombrement des articles produits et éventuellement des citations que ces articles résolvent. Narin appelle cette discipline qui s'intéresse à la production écrite des chercheurs la bibliométrie⁽¹⁾.

Etant donné l'importance croissante des indicateurs bibliométriques, nous allons ci-dessous examiner les indicateurs qui ont été formulés par Narin et son équipe pour les Science Indicators.

(1) F. Narin, M. Carpenter - Bibliometric Indicator Series in the U.S. Science Indicators Data Base, Science and Technology indicators Conference, 15-19 sept. 1980.

S.I. comporte 13 indicateurs bibliométriques. Dans le tableau ci-joint ces 13 indicateurs ont été rassemblés sous forme synthétique.

(1) Nombre de publications au niveau national.

Le nombre d'articles, notes et comptes rendus écrits par des scientifiques qui appartiennent à une institution donnée dans un pays donné.

A partir de cet indicateur statistique, 3 indices peuvent être calculés :

- Part d'un pays dans un domaine (en %),
- Part d'un domaine dans un pays (en %)
- Indice d'activité nationale qui est défini :
 - $\frac{\% \text{ des articles d'un pays dans un domaine}}{\% \text{ du nombre total d'articles dans ce domaine}}$ (1)

(2) Indicateurs de citations au niveau national, Probabilités

nombre de citations allant d'un pays A à un pays B, estimé par la contribution de A au journal de référence, et par la contribution de B au journal cité, puis par la sommation des deux. Figurent également des ratios de citation relatifs montrant l'influence de chaque pays.

Cet indicateur, par son caractère probabiliste, sous-estime les citations internes à un pays et surestime les citations entre pays. Notons qu'il est calculé pour les Etats-Unis, la Grande-Bretagne, l'Allemagne Fédérale, la France, l'U.R.S.S., le Japon, le Canada et le reste du monde.

- (1) L'article de J.D. Frame F. Narin, M. Carpenter (1977). "The Distribution of World Science" Social Studies of Science, 7 (4), 501-516, novembre 1977, fournit une analyse à partir de ces indices.

TABLE 6

AVAILABLE INDICATOR SERIES

<u>Indicator</u>	<u>Countries</u>	<u>Fields/Subfields</u>	<u>Years</u>	<u>Comments</u>
1. National Publication Counts	"190 country aet"	9 major fields and 106 subfields	1973-1977 year by year	Also available in copy form Activity Index % of country in field % of field in country
2.A. National Citation Counts- Probability Assignment (old method)	7 major countries + all other + outside of self	9 major fields	Citations from 1973- 1977, year-by-year, to all previous years	Underestimates within country citation
2.B. Relative Citation Index	"	"	"	"
3. U.S. Sector Publication Counts-by Field, by Level	U.S. only, by 6 sectors	9 major fields, and by level withinfield	1973-1977 year by year	Also provided in hard copy form % of sector in field % of field in sector
4. Publication Counts for Active Institutions within Sector	U.S. only, by 6 sectors	All fields combined	"	100 most prolific publishing institutions are listed for each year, with count and % of sector
5.A. National Citation Counts- Exact Assignment	7 major countries + all others + outside of self	9 major fields	Cited years 1973-1977 year by year	Citations to these years are summed for that year, and all subsequent years. 1974 is the first citing year.
5.B. Normalized National Citation Counts-Exact Assignment	"	"	"	Above counts recalculated to account for speed of citation normalization
5.C. Relative Citation Index	"	"	"	% of cites/% of papers
5.D. Normalized Relative Citation Index	"	"	"	Above with speed of citation normalization

TABLE 6
(Continued)

AVAILABLE INDICATOR SERIES

	<u>Indicator</u>	<u>Countries</u>	<u>Fields/Subfields</u>	<u>Years</u>	<u>Comments</u>
6.	International Institutional Co-authorships	"21 country set"	9 major fields and all fields combined	Cited years 1973-1977 year by year	Also provided in hard copy form % of country coauthorship index patterns of coauthorship field degree of coauthorship
7.	Country Distribution among Fields	"21 country set"	Calculated across 106 subfields	Cited years 1973-1977 year by year	Two Indicators: Gini Index relative concentration index also, in hard copy
8.	Country Influence	8 major countries	9 major field	Cited years 1973-1977 year by year	Five Indicators: influence weight references per paper influence per paper number of papers total influence
9.	Country Isolation/Integration	3 major countries	9 major fields and All fields combined	1973-1977 year by year	Four Indicators: reference isolation index citation isolation index reference interactivity index citation interactivity index
10.	Field Internationality/Interactivity	8 major countries	9 major fields	1973-1977 year by year	Two Indicators: field internationality index field interactivity index
11.	Journal Internationality	"21 country set"	9 major fields and all fields combined	1973-1977 year by year year by year	Basic data is number of papers from each country in each countries' journals: also provided in hard copy % of country of publication by country of author % of country of author in country of publication

TABLE 6
(Continued)

AVAILABLE INDICATOR SERIES

<u>Indicator</u>	<u>Countries</u>	<u>Fields/Subfields</u>	<u>Years</u>	<u>Comments</u>
12. U.S. Sector Coauthorship	U.S. only, by 6 sectors	9 major fields	1973-1977 year by year	A measure of cooperation between sectors of the U.S. scientific community
13.A. U.S. Sector Cross Citations	U.S. only, by 6 sectors	9 major fields	Cited years 1973-1977, year by year	Citations to these years are summed for that year and all subsequent years (Exact assignment)
13.B. U.S. Sector Relative Citation Index	U.S. only, by 6 sectors	9 major fields	Cited years 1973-1977, year by year	% of cites/% of papers

SOURCE: NARIN F., CARPENTER H. (1980), op.cit. p. 139

- (3) Nombre de publications américaines par secteur, discipline et niveau: le nombre de papiers publiés par les auteurs affiliés à des institutions américaines dans six secteurs et quatre niveaux de recherche.

Les secteurs de recherche sont les laboratoires gouvernementaux fédéraux, ou d'Etat, Centre de recherche développement financés par le secteur public, universités, firmes privées, et firmes privées sans but lucratif. Cette typologie est, comme on peut le voir, essentiellement administrative.

Les niveaux de recherche sont au nombre de quatre du plus appliqué au plus fondamental.

- Niveau 1 : Technologie appliquée
- Niveau 2 : Science technologique - Science des Ingénieurs
- Niveau 3 : Recherche appliquée
- Niveau 4 : Recherche fondamentale

La définition de ces 4 niveaux n'est pas entièrement conforme à la définition de S.I. qui ne repère que deux niveaux Appliqué (niveau 1 + 2 + 3) et Fondamental. (Niveau 4).

- On remarquera toutefois que délimiter d'un côté la science fondamentale et de l'autre toute recherche plus ou moins appliquée, donne en termes financiers et économiques beaucoup plus de poids à la technologie qu'à la recherche fondamentale. Nous reviendrons sur ce point plus loin.

- (4) Institutions actives américaines : le nombre d'articles publiés par des auteurs affiliés aux 100 institutions américaines les plus prolifiques. Un classement de ces 100 institutions est effectué en fonction du nombre d'articles.

(5) Indicateurs de citations au niveau national : décompte exact.

Le même que l'indicateur (2) sauf que dans ce cas les calculs sont exacts et non des estimations.

Cet indicateur permet de produire des indicateurs de citations qui tiennent compte du temps nécessaire pour qu'un article soit cité.

(6) Coauteurs

Le nombre d'articles cosignés par des scientifiques du pays A et du pays B. 3 indices sont liés à cet indicateur :
 1) le nombre de coauteurs exprimé en pourcentage de publications du pays A - 2) le nombre de coauteurs exprimé en pourcentage des publications cosignées du pays A -
 3) les indices de coopération entre les pays A et B⁽¹⁾.

(7) Participation de chaque pays par discipline

Deux indicateurs définissent la distribution relative d'articles nationaux par discipline.

(8) Influence d'un pays : c'est un décompte du nombre de citations nationales par discipline, combiné dans un indice de spécialisation (voir formule page suivante).

(9) Degré d'autonomie/intégration : deux indicateurs mesurent l'autonomie ou l'intégration de l'activité nationale dans la production mondiale par discipline (indicateur (5)).

(1) Un exemple d'utilisation de cet indicateur :

J.D. Frame et M. Carpenter, International Research Collaboration, Social Studies of Science, 9 (3), 1979, p. 481-497.

INDICATOR 8: COUNTRY INFLUENCE

DERIVE "INFLUENCE WEIGHT" W FOR EACH COUNTRY

(9 FIELDS, 8 COUNTRIES, CITED YEARS 1973-1977)

$$W_i = \frac{\sum_{k=1}^n W_k C_{ki}}{S_i}$$

WEIGHT OF i th COUNTRY = SUM OVER ALL COUNTRIES, k WEIGHT OF k th COUNTRY $\times$ NO. OF CITES FROM k th TO i th COUNTRY

THE 8 EQUATIONS, FOR A GIVEN FIELD AND CITED YEAR, PROVIDE A SELF-CONSISTENT 'BOOTSTRAP' SET OF RELATIONS. EACH COUNTRY PLAYS A ROLE IN DETERMINING THE WEIGHT OF EVERY OTHER COUNTRY.

MOST USEFUL MEASURE IS INFLUENCE/PAPER

THE WEIGHTED AVERAGE NUMBER OF TIMES EACH ARTICLE FROM A COUNTRY IS CITED

EXAMPLE: INFLUENCE MEASURES FOR 1974 CITING 1973 IN BIOMEDICAL RESEARCH

COUNTRY	INFLUENCE/PAPER
U.S.	27.7
U.K.	25.0
CANADA	19.1
JAPAN	16.3
OTHER	14.3
FRG	13.7
FRANCE	10.8
U.S.S.R.	2.8

Un indicateur d'interaction est aussi calculé qui varie de - 1 (minimum d'interaction entre ce pays et le reste du monde) à + 1 (maximum).

Manuellement ces indices, ainsi que les indicateurs 10 et 11 suivants, ont été calculés dans le cas de l'analyse économique⁽¹⁾.

(10) Internationalisation des disciplines.

Deux indicateurs permettent d'examiner la participation de plusieurs pays dans chaque discipline.

Même procédure statistique que pour l'indicateur (9).

(11) Internationalité d'un journal : le nombre de papiers écrits par un scientifique du pays A qui sont publiés dans un journal du pays B.

(12) Coauteurs américains par secteur : le nombre d'articles cosignés par des américains dans les secteurs A et B de la science américaine.

(13) Indicateur des citations entre secteurs américains : décompte exact : le nombre de citations d'un auteur affilié au secteur A à des articles écrits par des auteurs du secteur B.

g) Les mesures composites de technologie

Le chapitre consacré au développement et à la technologie est asymétrique par rapport au chapitre consacré à la Recherche Fondamentale. Il ne traite pas du "développement" (entendu comme le fonctionnement d'une technique à l'échelle industrielle, ex. un prototype ou une usine pilote), ni de la science appliquée.

(1) R. EAGLY, Economics Journals as a Communication Network, Journal of Economic Literature, 13 (3), 1975, pp. 878-888.

Le chapitre est en effet consacré à un secteur : l'industrie. Alors que l'unité d'analyse du chapitre précédent était un type de recherche : la Recherche Fondamentale. Cette asymétrie est due en grande partie aux types d'indicateurs dont on dispose.

D'une part, en ce qui concerne les output de l'activité d'innovations. L'indicateur quantitatif le plus évident sont les brevets. Cependant, bien que l'on puisse inférer les tendances d'un secteur à partir des brevets, les relations qui existent entre plusieurs technologies et les rapports entre la production d'articles et la "production" de brevets, plusieurs difficultés subsistent dans l'utilisation des brevets; ces difficultés sont liées à l'utilité qu'une firme accorde à un brevet. Les brevets font partie de la stratégie même des firmes, stratégie industrielle et commerciale. Le nombre de brevets, est lié aux décisions microéconomiques des firmes . Les statistiques de brevets indiquent donc certains aspects de la croissance de la technologie mais on ne sait pas encore très bien lesquels.

D'autre part, concernant les inputs de la R & D, les statistiques d'emploi et les statistiques financières sont liées à plusieurs problèmes qui ne trouvent pas forcément leur source dans les activités de R & D. L'environnement économique et la politique fiscale par exemple influent sur ces données.

La collecte de ces statistiques connaît plusieurs limitations et de plus les données collectées sont souvent inexploitable car elles ne permettent pas d'identifier ce qui ne concerne que la R & D.

La dernière limitation, et non la moindre, est le fait que le financement et le personnel de la R & D ne représentent

pas le véritable input du processus d'innovation. En effet, il n'existe pas d'effet mécanique par lequel augmenter le financement permettrait d'augmenter la productivité. Cette limitation très grave et mal connue reflète l'impossibilité d'imputer un input à un output.

La plupart des travaux sur l'innovation tentent d'élucider le rapport entre input et output en effectuant des calculs économétriques assez sophistiqués.

Un type relativement récent d'indicateurs commence à être exploité. Il s'agit des "indices composites de technologie". Construire des indices composites consiste à tracer dans le temps le niveau de certaines variables qui sont censées représenter une bonne ou une mauvaise caractéristique d'une technologie. Par exemple, dans le cas des ordinateurs, on peut mesurer la capacité des ordinateurs en nombre d'opérations effectuées à la minute ; ou bien la vitesse de calcul, liée au coût par opération et à la capacité de la mémoire de l'ordinateur. Une pondération est calculée pour chacune des variables qui rentrent dans la composition de l'indice. Cet indice dont le choix des variables et des pondérations est établi à partir des jugements d'experts de la technologie en question, permettent de rendre compte de l'évolution d'une technologie en termes quantitatifs⁽¹⁾. En établissant le même indice pour les USA et d'autres pays, on peut avoir une idée assez globale de l'état des technologies aux U.S.A. et ailleurs.

- (1) Pour les ordinateurs, voir Montgomery Phister, Jr. Data Processing : Technology and economics, Bedford Mass. Digital Press 1979. L'indice est :

$$100 \times \left(P_1 \frac{(X_1)}{(X_1^*)} + P_2 \frac{(X_2)}{(X_2^*)} + P_3 \frac{(X_3)}{(X_3^*)} \right)$$

où $P_1 = 0,5$, $P_2 = 0,3$ et $P_3 = 0,2$ sont les pondérations de X_1 la vitesse X_2 le coût et X_3 la capacité, in S.I. 80

Tableau 4-19 p. 294

Le FUTURES GROUP a été financé par la NSF pour calculer des indices sur les biotechnologies.

CONCLUSION

Dans le cas des indices composés de technologie, comme dans le cas des indicateurs bibliométriques, on se rend compte du caractère évaluatif de toute tentative de mesure. Les indices composés reposent sur le jugement d'experts. Les indicateurs bibliométriques reposent également sur le jugement des spécialistes : soit quand ils sont directement sollicités, comme c'est le cas des "reviewers" dans les évaluations de performances ; soit quand il s'agit des referees des périodiques où sont publiés les articles. En ce sens, on peut voir que les indicateurs de la science sont une conjonction de jugements évaluatifs, sur la science, sur une discipline ou sur une technologie.

On peut d'ailleurs rappeler que les "Sciences Indicators" ont été conçus pour "évaluer la santé de l'effort scientifique national, le statut et la santé de la science"⁽¹⁾.

Et curieusement, c'était le directeur du GAO (Comptroller General) qui précisait que "la plupart des termes invoqués pour développer des mesures de la science (l'état de santé et la vitalité de la science) sont vagues et évaluatifs" (GAO, 1978) p. 1. Le rapport finissait sur cette remarque : "Il serait bon de n'utiliser que des termes non évaluatifs comme état de la science ou condition et orientation de la science et de la technologie" (p. 17).

Les auteurs de ce rapport associent le terme évaluatif à subjectif et indicateurs à objectifs ; comme on vient de le

(1) D. Hoenig : The health of scientific and technical enterprise, an Advisory Panel Report to the Office of Technology Assessment, Washington DC, oct. 1978.

rappeler les indicateurs reposent sur des jugements. Ces jugements (de valeur, bien entendu) sont plus ou moins proches de la construction et conception de l'indicateur. Et c'est cette distance qui permet de parler d'objectivité.

Il faut souligner l'apport bénéfique essentiel des indicateurs quantitatifs. Ils ont permis d'alimenter un véritable débat entre les scientifiques et les administrateurs en mettant les uns et les autres sur un terrain commun. Dans l'ensemble des controverses sur la politique scientifique aux USA, l'un des principaux points d'accord entre les scientifiques, les administrateurs, les législateurs et les décideurs était la nécessité de disposer de données statistiques, de bons indicateurs.

Puisque nous venons de dire que l'objectivité réside dans la distance qui existe entre celui qui émet le jugement et celui qui conçoit l'indicateur, il nous faut ajouter que la force de l'objectivité vient de l'intégration de plusieurs autres jugements subjectifs dans cet intervalle. L'indicateur est en quelque sorte une borne, un point de repère que tous peuvent apercevoir. Scientifiques et Administrateurs ont tout intérêt à se mettre d'accord sur cette borne quitte à la déplacer et la transformer. D'ailleurs les statistiques sur la Science n'ont rien d'absolu, seuls ceux n'ayant jamais participé à une négociation l'ignorent.

Ainsi les scientifiques et les organisateurs ont du discuter ensemble plutôt que de se renvoyer la balle au nom d'une expertise que chaque partie aurait du mener de son côté.

CHAPITRE 5

L'EVALUATION DE LA RECHERCHE A LA NSF ET AU NIH

RESUME :

Les agences fédérales examinées (la NSF et le NIH) financent essentiellement des bourses de recherche (grants) en allouant les fonds selon le mécanisme de l'évaluation par les pairs.

A la NSF, nous avons repéré quatre types d'évaluations : l'évaluation par les pairs des projets de recherche, les études prospectives par disciplines scientifiques -c'est-à-dire par programme NSF- par écrit (Status of Science Reviews) et par oral (Program Reviews) et les évaluations des performances des programmes de recherche NSF. Ces différentes évaluations reposent toutes sur le jugement qualitatif des pairs à des degrés divers. Seules les évaluations des performances des programmes utilisent des indicateurs quantitatifs (montant des financements, matériel, coûts de la recherche, publications, citations, brevets, etc...).

Au NIH, l'évaluation par les pairs se fait selon une procédure différente qu'à la NSF, procédure dite "Dual Review" car ayant lieu à deux niveaux différents.

En plus de cette évaluation des projets de recherche, le NIH a mis au point un ensemble d'études d'évaluations des programmes qu'il finance. Ces évaluations de programmes portent sur les aspects scientifiques des recherches, sur les processus de diffusion des produits de la recherche et des technologies médicales, sur l'enseignement et la formation scientifique, en particulier la formation de chercheurs et médecins dans les hôpitaux et centres de recherche de l'Institut. Le NIH a développé des indicateurs quantitatifs d'output, essentiellement bibliométriques qui sont ici présentés dans le détail.

Nous examinons par ailleurs deux autres études sur la productivité de laboratoires et centres de recherches : l'étude menée par Francis Narin et son équipe portant sur les laboratoires de la marine, et l'étude menée par Irvine et Martin portant sur plusieurs centres d'astronomie. De plus, nous examinons rapidement certaines études de type rétrospectif.

Les agences fédérales de recherche utilisent 3 mécanismes de financement pour la recherche fondamentale menée dans les universités : les bourses individuelles, les contrats de recherche et les accords de coopération (1).

Les bourses de recherche individuelles sont attribuées sans que l'agence qui finance la recherche contrôle la gestion des activités de recherche

Les contrats sont établis pour des recherches orientées et l'agence qui finance la recherche contrôle les activités de plus près. Les accords de coopération sont un arrangement entre les agences, les institutions dans lesquelles la recherche est effectuée et les chercheurs. La responsabilité de chacune des parties en présence est négociée pour chaque accord par les contractants.

Les "block-grants" ou "core-grants" sont les accords de coopération. Ce ne sont pas des chercheurs qui reçoivent alors les fonds de recherche mais un laboratoire qui regroupe plusieurs chercheurs;

La gestion des ressources est en général entièrement sous la responsabilité d'un comité directeur du laboratoire. L'agence qui finance le laboratoire effectue des contrôles occasionnels, on peut faire pression sur le laboratoire pour que les recherches se fassent dans un domaine particulier ("thrust areas").

(1) "Research grants", "research contracts", "cooperative agreement", voir GAO (1981) Better Accountability Procedures needed in NSF and NIH Research grant system, 30 septembre 1981, p. 3. Ces modes de financement ne concernent pas les équipes internes à un organisme de recherche qui ne financerait que ses propres laboratoires (comme c'est le cas du National Bureau of Standards par exemple).

TABLEAU 5.1

RELATIONS FINANCIÈRES-ADMINISTRATIVES ENTRE POUVOIRS PUBLICS ET CHERCHEUR

	"PROCUREMENT CONTRACT"	"COOPERATIVE AGREEMENT" (1)	"GRANT" (2)
	CONTRAT ADMINISTRATIF	ACCORD DE COOPERATION	BOURSE DE RECHERCHE
RÔLE des Pouvoirs Publics	« Acheteur »	« Partenaire » ou « Participation Active »	« Tuteur » ou « Participation Passive »
RESPONSABILITÉ	Pouvoirs Publics	Responsabilité Partagée	Boursier (« Principal Investigator »)
DÉCISIONS CONCERNANT FIXATION OBJECTIFS DE RECHERCHE	Pouvoirs Publics uniquement	Récipiendaire avec accord préalable des pouvoirs publics	Récipiendaire sans droit de regard de la part des pouvoirs p.
ENGAGEMENT DES POUVOIRS PUBLICS	Engagement nécessaire tel qu'il est défini par la loi.	Engagement concernant la gestion du programme	Délégation de pouvoirs aux instances gérant le projet de recherche

* Tableau d'après N.S.B., Basic Research in Mission-Oriented Agencies, 1978 p.276, Table 5.1.

- (1) dit aussi "Block-Funding" ou "Core-Funding" : ce mode de financement permet de financer des laboratoires, ou centres de recherche, ou institutions dans le cadre d'un programme
- (2) ce type de financement ne concerne que des projets et des chercheurs individuels

Près de 50 % de la recherche fondamentale a lieu dans les universités ; elles reçoivent du gouvernement, en 1980, plus de 2,9 milliards de dollars, et environ 2,2 milliards de dollars provenaient de la NSF et du NIH.

Près de 2000 institutions reçoivent des fonds de la NSF. En 1980, quelques 26.000 projets de recherche individuels ont été évalués et 11.500 d'entre eux (44 %) ont été financés, ce qui correspond à un montant de 653 millions de dollars.

Le NIH a consacré 1,6 milliard de dollars en 1980, et 1,9 milliard en 1981 pour financer des bourses individuelles dans les universités, soit plus de 1/3 du budget des NIH. 26.600 projets ont été évalués et 16.500 ont été financés.

Les bourses du NIH sont en moyenne plus importantes que celles de la NSF (97.000 \$ pour les bourses NIH, contre 56.000 \$ Pour les bourses NSF) et la durée moyenne pour les bourses NIH est plus longue (3,2 ans) que celles de la NSF (2,2 ans).

On entend parfois dire que ceci est anormal étant donné que le NIH ne finance que la recherche biomédicale alors que la NSF finance des travaux dans tous les domaines (excepté certains domaines spécifiques pris en charge par d'autres agences). Il est vrai que la recherche médicale a toujours bénéficié d'une plus grande indulgence en ce qui concerne son budget de la part du congrès que la NSF.

I. L'évaluation de la recherche à la NSF

La National Science Foundation en dehors de son activité de politique de la science, effectue quatre types d'évaluations.

1. Evaluation ex-ante des projets de recherches qui demandent des bourses de recherches. C'est le processus d'évaluation par les pairs.

2. Evaluation des programmes de la Fondation. C'est une présentation orale des programmes de la NSF, plus exactement de certains travaux qui sont financés dans le cadre de ces programmes, des perspectives d'avenir qu'offrent ces programmes et des problèmes rencontrés par les chercheurs.

3. La rédaction de "Status of Science Reviews". C'est une évaluation prospective par sous-discipline scientifique (qui correspondent aux programmes de la NSF).

4. Evaluation ex-post, des résultats des recherches financées par la NSF, (sélection d'un échantillon de projets dans le cadre d'une division, ou évaluation d'un échantillon de projets selon le mécanisme de financement).

Les 3 premières formes d'évaluation sont issues des traditions mêmes de la communauté scientifique. Ces traditions, comme le rappelle Peter Healey, privilégient l'évaluation par les pairs "où la science n'est évaluée que dans ses propres termes", ce que l'on pourrait appeler "une vision purement internaliste de la science". L'évaluation par les pairs est une étiquette qui recouvre une variété d'arrangements institutionnels minutieux. Ils sont supposé tous aboutir au même résultat : effectuer le choix optimal pour le développement, l'avancement de la science." Dans cette perspective internaliste

les scientifiques éminents seraient censés condenser dans leurs jugements toute la sagesse accumulée dans chacune des disciplines" (P. Healey, 1982, p. 3-4). Seule l'évaluation a posteriori n'est pas à proprement parler une évaluation par les pairs. Cependant, comme nous le verrons dans le détail plus loin, là aussi une certaine forme d'évaluation par les pairs a lieu ; les auteurs des études d'évaluation ex-post d'ailleurs appellent leur travail en abrégé "post-peer evaluation".

Avant de décrire ces quatre procédures nous devons rappeler que la NSF n'effectue pas d'évaluation du personnel puisqu'elle n'emploie pas les chercheurs mais finance soit des projets individuels (bourses) soit des laboratoires de recherches (core-grants).

Que ce soit pour financer des chercheurs individuels, ou pour financer des laboratoires, la NSF ne répond qu'à des demandes qui lui sont volontairement adressées. Ceci est vrai en théorie, quoique dans la pratique on ait pu observer des entorses à ce système (1).

1. L'évaluation a priori des projets

L'évaluation par les pairs des projets peut prendre plusieurs formes (Mail Review, Dual Review). Elle est couramment utilisée dans les agences qui financent la recherche fondamentale (NSF, NIH). Les principes qui la

P. HEALEY, (1982) The Research funding organization as a focus for science studies, sciences conference, New College, Oxford, 27-9-1982.

(1) Voir par exemple le projet MANGANESE NODULES en océanographie, décrit par MAZUR et BOYKO (1981)

régissent sont toujours les mêmes :

- . l'allocation se fait pour un projet précis et pour une durée précise limitée.
- . les projets ne sont pas sollicités mais sont présentés volontairement pour les scientifiques demandeurs.
- . l'avis de la communauté scientifique, qui est supposée être représentée par les "reviewers" (les pairs) est prépondérante dans les décisions de financement.

Il faut toutefois préciser que les reviewers représentent mais ne sont pas représentatifs de la composition de la communauté scientifique. En effet ils sont plus âgés, plus connus, ont publié plus d'articles et se trouvent dans des institutions plus prestigieuses que le scientifique moyen.(1) Ceci concorde assez bien avec l'opinion d'un scientifique, directeur de division de la NSF qui m'a dit : "ici nous ne gérons pas n'importe quelle population comme dans d'autres programmes gouvernementaux ; nous sommes la crème de la crème du pays."

(1) Voir C. MULLINS (1972).

The Structure of an Elite : The Advisory Structure of the Public Health Service, Science Studies, vol. 2, pp. 3-29.

L. GOENEVELD, N. KOLLER et N. MULLINS, The Advisers of the US National Science Foundation, Social Studies of Science, vol. 5, août 1975, pp. 343-354 ; M.J. Mulkay, The Mediating role of an Elite, Social Studies of Science vol. 5 (1976), pp. 445-470.

- . une fois le projet en cours de financement, les chercheurs peuvent en altérer les perspectives et le contenu s'ils le considèrent nécessaire. Cela n'aura aucune influence sur la continuation du financement.
- . des demandes de renouvellement de financement sont généralement présentées ; à ce moment théoriquement les "reviewers" examinent les résultats des recherches passées en même temps que les propositions de renouvellement. Cependant ce n'est pas là une obligation et très souvent cette évaluation des recherches passées n'a pas toujours lieu. D'après plusieurs directeurs de programmes ce n'est que récemment que cela est rendu possible, car les propositions décrivent avec plus de précisions leurs objectifs de recherche, ce qui permet de comparer les objectifs accomplis avec les objectifs initialement prévus, ainsi que les raisons pour lesquelles ces objectifs ont été ou n'ont pas été atteints. (voir annexe 1 du chapitre).

Deux procédures différentes sont employées par la NSF pour évaluer et sélectionner les projets.

- le "Mail Review" : les projets sont envoyés aux juges par correspondance
- le "Panel Review" : un groupe de juges est réuni, et ils doivent classer les projets. Cette procédure, qui est plus coûteuse est utilisée dans le cas de grands projets seulement.

Pour certains projets un mélange des deux procédures peut avoir lieu. Mais la plupart des projets sont évalués et sélectionnés par correspondance.

1.1. Le Mail Review

Les propositions envoyées aux pairs doivent être no-

tées globalement (excellent, très bon, bon, acceptable, mauvais). Mais on demande aux juges de commenter la note globale. Des commentaires sont demandés selon 4 types de critères (11 critères au total, (annexe 1)

- A. Critères concernant les capacités du chercheur.
- B. Critères concernant la valeur scientifique des projets
- C. Critères concernant l'utilité des projets
- D. Critères concernant l'impact à long-terme de la recherche

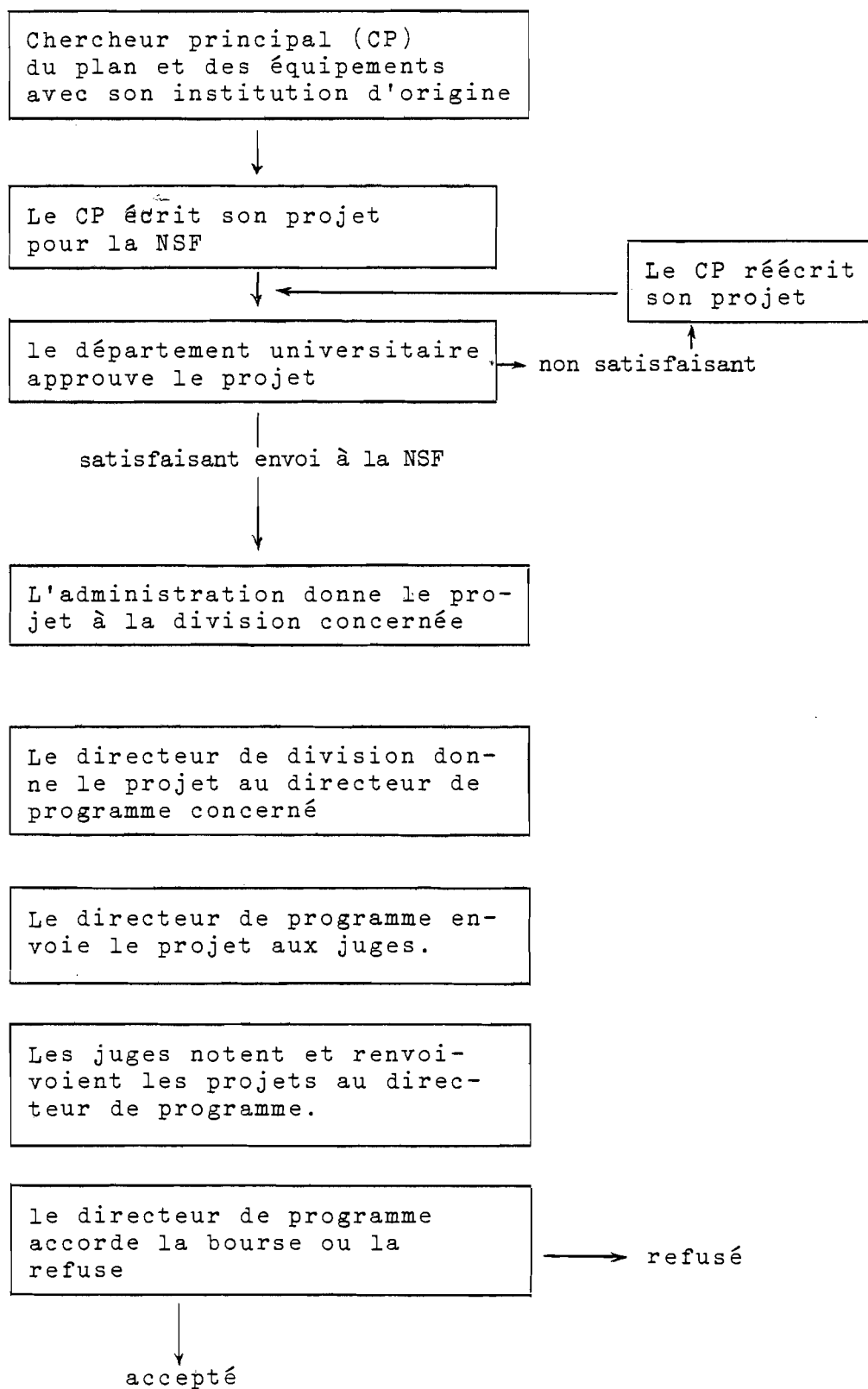
Une fois les projets notés, le directeur de programme les rassemble, décide quels projets doivent être financés. Le directeur de division doit finalement approuver cette décision. En cas de désaccord entre les directeurs de programme et le directeur de division, on peut éventuellement faire appel à un comité qui doit trancher. Ce cas est assez rare.

Le directeur de division joue en quelque sorte le rôle du secrétaire de marché à la bourse. C'est lui qui rassemble toutes les décisions de financement et qui les rend définitives. Mais ce n'est pas lui qui établit ni notes des projets, ni la hiérarchie entre les projets.

L'étude de Cole, Rubin et Cole (1978) a montré que dans l'ensemble ce système d'évaluation/sélection fonctionne correctement, du moins si on se place au niveau des divisions (disciplines) de la NSF. Mais il peut y avoir de nombreux biais au niveau (les divisions rassemblent plusieurs projets).

On peut schématiser l'ensemble des opérations de l'évaluation a priori des projets sous la forme d'un organigramme.

COLE S., RUBIN L., COLE J.R. Peer Review in the NSF, NAS Washington 1978.



Une fois la bourse accordée, le budget doit être également approuvé. Si l'administration financière de la NSF ou le directeur de division s'oppose au budget, le budget peut être renégocié entre le directeur de programme et le chercheur principal.

Comme on peut le voir le directeur de programme est le personnage clé dans ce processus. C'est sur son avis que sont acceptées ou refusées les bourses de recherche. C'est lui qui discute directement avec ceux qui soumettent des demandes de bourses, qui les conseille, les oriente, leur propose des modifications dans la proposition. Le directeur de division a le pouvoir du censeur tandis que le directeur de programme a le pouvoir d'infléchir, de modifier et de modeler le contenu des propositions contenu d'abord scientifique mais aussi budgétaire. Il faut souligner un aspect important de cette activité. Le directeur de programme se trouve devant plusieurs projets qui sont donc en situation de concurrence. Les négociations qu'engagera éventuellement le directeur de programme avec les chercheurs dépendre grandement du montant total du budget qui a été alloué à son programme. S'il est totalement maître de sa décision en ce qui concerne la "valeur intrinsèque" du projet, il n'en est pas moins contraint de refuser certains bons projets mais qui comparativement sont inférieurs aux autres. Le travail du directeur de programme est donc de traduire les termes budgétaires en termes scientifiques. La NSF lui donne l'occasion d'effectuer la traduction inverse dans les "status of science reviews" et au moment des négociations budgétaires avec la direction de la NSF ("Directors Managing Review"). Mais alors que la traduction des budgets des programmes en "science" se fait projet-par-projet, au fur et à mesure des demandes de financement, la traduction de la science en budget se fait plutôt à partir d'une évalua-

tion du potentiel de la sous-discipline dans laquelle se trouve le programme en question.

Ainsi c'est l'image que se fait le directeur de programme de sa discipline, de sa sous-discipline, qui va en grande partie orienter le type de recherches qui auront lieu dans le futur et qui seront financées par la NSF (1). Ce rôle d'intermédiaire, de traducteur, du directeur de programme n'est pas abordé dans l'étude des Cole, Rubin et Cole (1978). Et pourtant bien plus que les manipulations du mécanisme de sélection, c'est bien leur perception globale de l' "état de la science" qui façonne les grandes tendances des recherches financées par la NSF.

1.2. Le débat sur l'évaluation par les pairs

Le système d'évaluation par les pairs fait l'unanimité au sein de la communauté scientifique. Le directeur des NIH, James B. Wyngaarden s'exprimait ainsi à son propos : "I think peer review is one of the great creations ... ans its really the envy of the world. Many other countries are copying it, in a way or another ..."

-
- (1) Les scientifiques ne sont pas tous d'accord avec cette interprétation. Ils prétendent en effet que la NSF, ne fait que respecter la volonté de la communauté scientifique. Leur argument repose sur une vision de la communauté scientifique, toute entière vouée au développement d'un savoir universel . La recherche est fondamentale pour eux précisément en ce qu'elle est universelle, désintéressée, communiste pour employer la terminologie de Merton.

Mais leur aptitude à négocier les budgets de leurs projets, à populariser leurs recherches, et les structures même de la NSF, démontrent abondamment que le système de la communauté scientifique est loin d'être aussi uniforme que veulent bien le dire les scientifiques.

James B. WYNGAARDEN, directeur au NIH in Science and Government Report, 15, Mai 82, p.5.

D'après les documents des hearings(1) au sujet du système d'évaluation par les pairs on peut identifier 5 points de consensus :

1. "Il n'existe aucune méthode meilleure que le peer-review pour juger de la compétence scientifique des propositions de recherche. Les pairs sont les seuls aptes à juger de la conception de la recherche, de son importance, et des performances des recherches passées. Les procédures les plus appropriées du peer-review ont pour résultat de financer les propositions de meilleure qualité. En utilisant ces procédures, la NSF a pu faciliter de grandes découvertes." (P.2.).

2. "Les témoins (des hearings) s'accordent dans leur majorité pour vouloir maintenir le système d'évaluation par les pairs pour l'allocation des ressources" (p.25)

3. "Malgré le fait que des projets de grande qualité sont financés, dont certains véritablement innovatifs, la plupart des témoins s'accordent pour dire que les travaux innovatifs qui sont à contre courant de la pensée dominante ont peu de chance d'être financés" (p.27).

4. Le sous comité (du Congrès) a examiné la proposition selon laquelle le congrès devrait examiner les projets financés par la NSF en plus des décisions de financement de la NSF... Les opinions exprimées étaient majoritairement opposées à cette proposition" p.41. Il faut aussi rappeler que c'était là l'une des propositions de réorganisation de la NSF, en 1975,(2) connue sous le nom d'amendement Bauman (du nom -----

(1) "NSF Peer Review" vol.1 "A report of the Subcommittee on science, Research and Technology, 94th Congress, Janv. 1976.

(2) Congressional report 1 Août 1975, pp. J. 8063-8077 ; 30/7/75, pp. H 7826-7829.

de Robert E. Bauman représentant républicain du Maryland à la Chambre) qui fut rejeté à cette époque.

5. "Personne ne soutint la position extrême selon laquelle les proposants ne devraient rien connaître sur qui sont les reviewers et sur ce qu'ils ont dit au sujet des propositions" p. 43.

Il faut remarquer que la période où furent formulées des critiques les plus sérieuses vis-à-vis du système d'évaluation pour les pairs correspond à une période de hausse du chômage des diplômés. Les totaux de chômage les plus élevés apparaissent en 1976 : 8,2 % de chômeurs (100 = total des diplômés chômeurs et non chômeurs de cette année) en 1976, contre 4,7 % en 1974 et 3,9 % en 1978 (1). C'est aussi une période de ralentissement de l'accroissement des dépenses en R et D. De 1969, à 1972, les dépenses en R et D s'accroissent légèrement moins vite qu'avant et après cette période? Ce ralentissement se traduit par une stagnation des dépenses de R et D en dollars constants pour la période 1968-1976 (Graphique 2.1, p. 52 SI-80). Mais surtout ce fut la recherche fondamentale qui fut touchée (Graphique 2.-9, p.60). et la R et D financée par l'état (graphique 2-3 p.53). Enfin, le début des années 70, est une période de crise économique et politique (le "Watergate"). C'est dans ce contexte le processus de sélection des deux grandes agences fédérales qui financent principalement la recherche fondamentale (la NSF et les NIH) a été, remis en cause. Dans les deux cas comme il s'agit du système d'évaluation par les pairs, une grande controverse fut engagée sur l'efficacité de ce système d'évaluation?

(1) Science Indicators, 1980, pp. 50-53.

C'est à la suite de cette controverse que la NSF dans les années 1976-78 a engagé des sociologues de la science pour effectuer des études sur le système du "peer-review". Une étude a été effectuée en 1975-1976 par Deborah Hensler (1) ; l'autre fut menée par les frères Cole (Cole, Rubin, Cole, 1978). Plus récemment le Congrès (1979) a de nouveau examiné cette question mais l'impression que l'on tire de cette nouvelle enquête est qu'elle est beaucoup moins sévère que les hearings précédents (3). Enfin le GAO a effectué une étude comparative du système de la NSF et du NIH (4).

Ces deux derniers documents, celui du GAO et celui du sous comité du Congrès, sont bien différents des rapports issus pendant les années 75-76 sur l'évaluation par les pairs. Auparavant, il s'agissait de savoir si le système d'évaluation par les pairs était équitable. Aujourd'hui, il s'agit de savoir si les agences fédérales, et les universités sont "accountable" (5), c'est-à-dire transparentes du point de vue de leur système de gestion et de comptabilité (6)

-
- (1) HENSLER D., Perceptions of the National Science Foundation Peer Review Process : A report on a survey of NSF Reviewers and Applicants. Prepared for the Committee on Peer Review, National Science Board, and the Committee on Science and Technology, US House of Representative (Washington, DC ; NSF, 77-33).
 - (2) COLE S., RUBIN L., COLE J.R. (1978). Peer Review in the NSF, NAS Washington, DC, 1978.
 - (3) Interviews of National Science Foundation Program Officers, Final Report, Subcommittee on Science, Research and Technology, Committee on science and technology, House of Representatives, July, 1979.
 - (4) Better accountability procedures needed in NSF and NIH Research grant systems, by the comptable general, General accounting office, PAD 81-29, september 1981.
 - (5) Accountable : que l'on peut compter et conter.
 - (6) Voir National commission on Research, Accountability, Restoring the Quality of Partnership, Science, v.207, 14 mars 1980, pp. 1177-82, voir également, Science Indicators, 1980, pp. 83-85.

Ces documents sur la transparence des méthodes d'administration ont deux origines.

L'une est la réaction des universitaires à une circulaire de l'Office of Management and Budget, dite "A-21", qui préconisait une comptabilisation plus stricte des temps de recherche, d'administration et d'enseignement des universitaires.

L'autre est le besoin pressant de mieux spécifier les liens administratifs entre l'université et l'industrie car de plus en plus de recherches se font dans des laboratoires universitaires financés par l'industrie ou dans le cadre de contrats entre l'université et l'industrie.

Actuellement, la NSF elle-même effectue les études d'évaluation sur l'injonction du Congrès. Dans les années 75-76 ceci aurait été impensable. En effet puisqu'il s'agissait d'examiner le rôle de la NSF elle-même, une étude sur son propre rôle n'aurait pas été crédible. D'ailleurs l'étude que les Cole ont effectuée a été critiquée précisément pour avoir été financée par la NSF (1).

2. Les Program Reviews :

Les Program Reviews font partie du processus interne de la NSF pour planifier son activité. Il s'agit d'une présentation orale de l'état des recherches dans chaque direction de la NSF. Pendant une demi-journée, les présentateurs d'un programme doivent illustrer de manière convaincante les travaux qu'ils ont financés. Ils doivent aussi évaluer les

- (1) "We feel compelled to remark that despite Cole's claim of complete autonomy from NSF in conducting the research, the research they have reported betrays a commitment to show that even where peer review is not equitable within NSF the inequity is for the good of science" (P.216). Mitroff et Chubin, Peer-Review at the NSF : A dialectical Policy Analysis, Social Studies of Science, 9(2), 1973, pp. 199-232.

effets et la qualité de ces travaux. C'est en quelque sorte un exercice en persuasion, puisqu'il faut persuader les administrateurs que le travail effectué est suffisamment important pour que son financement ne soit ni diminué, ni interrompu.

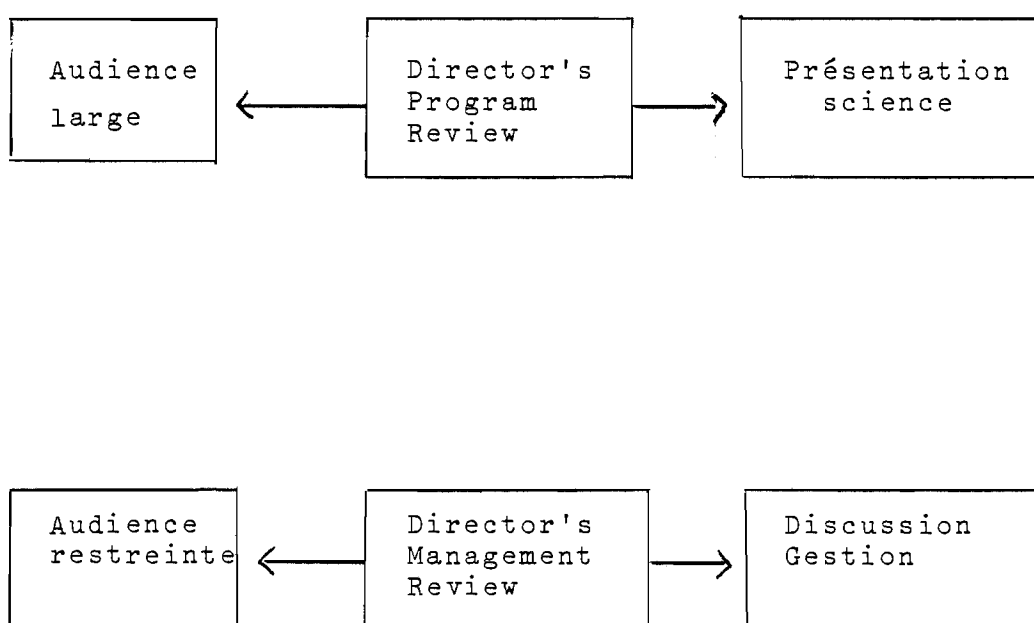
Cette année (juin 82) la forme des revues de programmes a été modifiée. En effet la direction de la NSF considérait que les présentations des revues des programmes était devenu trop ouvertement une discussion pour justifier les dépenses des divisions de la NSF. "Trop de gestion, pas assez de science". Les sessions ont donc été séparées en 2 séances distinctes. L'une concerne uniquement la science, l'autre uniquement sa gestion. Cette décision met à jour l'une des principales fonctions de ces revues de programmes.

Elles sont effectuées devant un large public comprenant des membres de l'administration de la NSF, des représentants de l'OMB, du personnel du congrès et de l'OSTP et des journalistes.

C'est donc une façon pour la NSF non pas tant pour savoir ce qui se passe dans les divisions de la fondation, (1) mais pour disséminer des informations courantes à l'extérieur de la NSF. La présentation orale des travaux les plus intéressants a en effet l'avantage de ne pas être aussi lente à diffuser que l'information écrite. D'autre part ces "program reviews" sont un plaidoyer en faveur de la NSF puisqu'on y présente ce que la NSF a permis de faire.

(1) Objectif bien mieux atteint par les "STATUS OF Science Reviews" voir Infra.

SCHEMA 5.3.

DIRECTOR PROGRAM REVIEW

Pour se rendre compte de l'importance d'une "program review" orale, il faut se souvenir que la mobilité des professeurs, assistants et chercheurs est plus grande aux USA qu'en France(1). Ainsi, la qualité de la recherche dans un domaine dépend énormément de sa capacité à attirer de bons chercheurs. Faire miroiter des possibilités nouvelles peut donc inciter des chercheurs à d'intéresser et finalement venir dans un nouveau domaine de recherche.

Les program reviews sont un moyen pour la NSF de montrer son rôle aux scientifiques et aux non scientifiques. Ce sont donc autant des opérations de prestige "pour la science" que pour la NSF.

Dans les séances des "Management Reviews" les directeurs de programme et les directeurs de division discutent avec la direction de la NSF des possibilités de financement de la recherche dans chaque division, chaque programme. C'est à cette occasion que les directeurs de programmes peuvent transmettre à la direction leurs sentiments concernant la gestion, la budgétisation, les actions à entreprendre, etc... Certains travaux peuvent être alors protégés, d'autres mis en compétition pour l'allocation des ressources. Certaines recherches peuvent être privilégiées, en augmentant les crédits dans un secteur, etc. Toutes ces décisions peuvent avoir lieu dans les "director's management reviews".

En résumé on peut qualifier les program reviews de plaidoyer sur le rôle de la NSF, justification de ses dépenses, exercice de persuasion et de séduction auprès des scientifiques et des personnes proches de la politique scien-

(1) Notamment en ce qui concerne les jeunes chercheurs.

tifique. Les management reviews sont le lieu de négociations budgétaires et plus généralement de gestion des programmes de la NSF.

Du point de vue qui nous intéresse ici il faut remarquer que les program reviews et management reviews sont essentiellement des évaluations prospectives à court et moyen termes, intuitives, effectuées par des scientifiques, (puisque le personnel de la NSF est constitué de scientifiques). Ici, l'évaluation s'associe à la gestion et à la prospective et non plus à la sélection comme dans le cas de l'évaluation des projets de recherche.

3. Les status of Science Reviews

Les status of science reviews, sont des documents d'évaluation prospective rédigés par les directeurs de programme. Par ces documents la direction de la NSF dispose d'une évaluation, de l'état de la discipline des besoins en personnel et en financement. Comme les divisions de la NSF correspondent assez bien à des disciplines scientifiques, la direction dispose en réalité d'une évaluation des activités de la science, dans chaque discipline. Ces évaluations sont prospectives car elles tentent d'identifier les problèmes qu'auront à affronter les administrateurs et les scientifiques, problèmes de gestion mais aussi problèmes scientifiques. Tous les documents par discipline sont regroupés dans un document unique intitulé "Status of Science Reviews" qui comporte deux parties distinctes. Les rapports des directeurs de programme font l'objet de la première partie ; des données statistiques sont rassemblées dans la deuxième partie.

Les "Status of Science Reviews combinent ainsi une évaluation subjective (ou qualitative) des disciplines scientifiques émanant de l'intérieur de la fondation, et des données statistiques sur chaque division ou discipline. Les rapports des directeurs de programmes sont présentés selon une formule fixe comportant 4 sections. Tout d'abord il y a une présentation générale de la discipline. Cette section est généralement peu contestée. La deuxième section, intitulée "perspectives nouvelles" est plus sujette à controverses. Elle identifie ce qu'aux yeux du directeur de programme constitue les problèmes les plus importants dans le domaine, tant du point de vue du contenu de la science que des ressources pour la science. Cette section reflète assez bien l'opinion qu'ont les directeurs de programme de leur discipline, des événements importants qui s'y déroulent, des besoins qui s'y manifestent, des tendances qui s'y expriment. Il est probable que le caractère contestable du diagnostic qu'établit le directeur du programme est voulu, ou du moins attendu, puisque les SSR sont des documents internes discutés pratiquement dans toutes les instances administratives de la NSF.

La troisième section est une évaluation des moyens matériels, financiers et en personnel dans le domaine, tels que ceux-ci sont perçus à la NSF. D'une manière générale il s'agit de la présentation des données quantitatives par discipline ou par programme et de leur éventuelle interprétation.

Enfin, la quatrième section s'intitule "NSF Role". Dans cette section le rôle que devrait jouer ou que joue la NSF dans la discipline en question est défini. On y trouve également des descriptions des relations entre la NSF et d'autres agences fédérales, ou internationales, toujours

au niveau de la discipline. D'une certaine façon il s'agit d'une désagrégation du paragraphe de science Indicators intitulé : "Basic Research. The role of Federal Agencies".

Dans la seconde partie du document annuel se trouvent réunies des données statistiques. Sur les programmes de la Fondation, sur les tendances de la R & D fédérale, et sur les moyens en ressources pour la recherche (statistique de diplômés, d'emploi des scientifiques, de fonds alloués à la recherche etc...)

On ne peut pas discerner avec précision quelle est l'influence exacte du SSR. Mais on peut affirmer que c'est un document assez largement distribué.

Théoriquement, le SSR est un document à usage interne. Il est adressé principalement au National Science Board puisque les informations qui s'y trouvent sont l'opinion des directeurs de programme et de division. Le SSR fait "monter" l'information ; les besoins des scientifiques trouvent ainsi une tribune auprès des instances consultatives et décisionnelles supérieures (NSB, OSTP, Congrès, etc...) Car, en effet, le SSR bien qu'étant un document interne est assez largement lu dans toutes ces instances de l'appareil de la politique scientifique, auprès d'organes consultatifs et d'associations professionnelles.

Le SSR démontre que la frontière entre le dedans et le dehors de la NSF est fictive, et que ce qui se passe dans la NSF détermine largement la politique scientifique.

En effet le SSR sert à désigner quelles doivent être les axes futurs de la recherche. Il est particulièrement

utilisé par le NSB et le Congrès, au même titre que les listes de priorités qu'établissent les associations professionnelles (par ex. AAAS ou National Academy of Sciences)!

Plusieurs préoccupations des directeurs de programme et de division apparaissent dans les SSR et dans les discussions Program Review (PR) et Director's Management Reviews ; elles sont très souvent reprises dans le rapport annuel du NSB (1). Mais contrairement à ces rapports annuels du NSB qui s'intéressent à l'ensemble de la science américaine, les PR et SSR concernent exclusivement les disciplines financées par la NSF (2).

4. L'évaluation a posteriori à la NSF

Il s'agit d'évaluer les résultats des recherches en fonction de certains critères (qualité, efficacité, efficience, productivité).

Les évaluations a posteriori ne sont voulues que par le Congrès. En effet, pour le Congrès qui cherche depuis longtemps à contrôler le financement de la science (3),

- (1) Annual Science and Technology Report to the Congress, Office of Science and Technology Policy in cooperation with the National Science Foundation.
- (2) Par exemple l'astronomie / NSF ne concernant que l'astronomie au sol, alors que la NASA finance et effectue essentiellement des travaux à partir d'observation dans l'espace.
- (3) Voir l'amendement Bauman (infra p.). La première année fiscale de la NSF avait failli d'ailleurs être sa dernière année. En effet, le 15 août 1951, la Chambre des Représentants avait voté une baisse du budget de la NSF de 98 % pour l'année fiscale 1952 par rapport à l'année précédente. La petite guerre entre le Congrès et la NSF n'allait plus jamais s'arrêter depuis ce jour là et à ce qu'il semble toujours à l'avantage de la NSF (voir M. Lomask, A. Minor Miracle, p. 80-82).

l'évaluation apparaît comme un outil de contrôle inestimable.

Le Comité des appropriations du Sénat qui s'occupe du budget de la NSF, du NIH et de la NASA, entre autres, en 1979 invoquait la possibilité pour la NSF "d'établir un programme pilote ("prototype") d'évaluations a posteriori" (1). En 1982, la NSF fournissait au Congrès les résultats de ce programme pilote. Le comité avait alors demandé à la NSF de faire un rapport sur l'état de la question de l'évaluation qui est actuellement en cours (2).

Trois commentaires s'imposent ici :

1) Il n'est pas étonnant que ce soit le congrès et uniquement lui qui demande des études d'évaluation a posteriori; ni l'office of Management and Budget, ni l'office of Science and Technology Policy, ni les associations professionnelles comme la National Academy of Science n'ont demandé des études d'évaluation. L'OMB a suffisamment de moyens de pression, notamment le budget, pour ne pas avoir besoin d'un outil de contrôle supplémentaire. Ceci est aussi vrai de l'OSTP, quoique dans ce cas on puisse dire que l'OSTP est "inféodé" à la NSF. Enfin, les associations professionnelles rejettent bien évidemment tout moyen de contrôle des activités de la communauté scientifique. Seul le pouvoir législatif entend contrôler/évaluer.

2) L'évaluation est en passe de devenir - ou est déjà, une discipline universitaire en science sociale. "Nous abordons les années où les principaux thèmes des sciences

(1) Rapport 92-258, 24 juillet 1979.

(2) Hearings du Comité des Appropriations du Sénat, 27 avril 1982.

sociales seront la performance, la productivité et l'unité" disaient les sociologues et humanistes de la NSF (1).

Des programmes d'évaluation existent dans des domaines comme la justice, la santé, l'éducation, la politique fiscale. Or cela déplaît profondément aux scientifiques de se savoir évalués comme le sont les malades mentaux, les cancéreux, les jeunes délinquants ou les contribuables du New Jersey. Contrairement à ce que l'on peut lire à propos des sciences sociales dans des documents officiels, les techniques ne sont pas "content-free" (2) ; l'évaluation des programmes publics a, en effet, des connotations qui déplaisent aux scientifiques; Demander des études d'évaluation, cela veut dire s'opposer à une résistance très spontanée des scientifiques.

3) L'évaluation rentre dans le cadre des discussions sur l'accountability (voir ci-dessus, p.) . L'évaluation apparaît comme un moyen de contrôle de gestion.

En effet une lettre d'un lecteur de Science remarquait : "qu'on ne connaît pas le poids exact en termes financiers des lois fédérales et des exigences réglementaires, si ce n'est par anecdotes". (3)

Milton Friedmann d'ailleurs s'appuie sur cette constatation pour attaquer le principe même du financement de la

(1) Status of Science Reviews, 1982, NSB, p.52.

(2) Emerging issues in Science and Technology, 1982, NSF, p.38.

(3) D.R. CORSON, Science, v.208, p. 1212, 13 juin 1980.

science au moyen de fonds publics.

L'ère des réductions budgétaires, qui s'annonce depuis la venue au pouvoir de Reagan, ne font que réactiver ce type d'exigences et de débats.

L'évaluation n'est pas un outil neutre de gestion. L'exemple des Etats Unis montre à quel point la situation politique et les nécessités financières prédéterminent entièrement les débats sur la productivité de la science et donc son évaluation?

Le service d'évaluation de la NSF a effectué, ou commandité des études définies selon 3 axes :

1) établir la qualité, l'efficacité, ou l'efficience des équipes de recherche ou des programmes de recherche.

2) Déterminer les effets économiques éventuels de la recherche, c'est-à-dire la capacité d'une recherche à produire des travaux brevetables. (2)

3) Vérifier si les décisions des pairs au moment de l'évaluation a priori sont équitables: c'est à dire vérifier les décisions ne sont pas injustes par exemple vis-à-vis des femmes, des noirs, des universités les moins prestigieuses, etc... Cette dernière forme d'études d'évaluation

-
- (1) "De quelle justification éthique dispose-t-on pour prélever de l'argent des contribuables pour l'allouer à des activités. sans rendement ou à faible rendement ? Il faut pouvoir prouver que chaque dollar supplémentaire dépensé en recherche aura un rendement supérieur à celui que cette personne attribue à ce dollar s'il l'avait gardé" "Je défie quiconque de trouver une seule étude justifiant l'appui financier gouvernemental à la recherche" M. Friedman, Science, 210, p. 33, 3.10.80.
 - (2) Une étude de la brevetabilité de la recherche fondamentale en chimie est actuellement en cours à la NSF. Des résultats provisoires seront probablement disponibles à la fin de l'année 1982, en plus d'une étude d'évaluation de la qualité de la recherche.

intéresse le Congrès pour des raisons évidentes. L' "evaluation staff" effectue une étude de ce type chaque année (Depuis, 1980-81).

Il faut remarquer que toutes les études d'évaluation sont effectuées dans le cadre de ce que la NSF définit comme étant un programme. Cela peut être un programme de financement de laboratoires ou centres de recherche. Mais cela peut aussi être une évaluation des résultats d'un échantillon de projets individuels dans le cadre d'un programme/division de la NSF. Ce ne sont donc pas les projets individuels qui sont évalués ou, plus exactement, le résultat des évaluations concerne un programme.

Nous allons ci-dessous examiner les études d'évaluation a posteriori dont nous disposons. Il s'agit de l'évaluation des laboratoires de recherche des matériaux, (programme MRL de la NSF), l'évaluation du programme d'océanographie. Une évaluation du programme de Chimie est actuellement en cours. Enfin, nous allons également étudier le dernier rapport sur l'équité des décisions de financement des projets individuels de la NSF.

4.1. Evaluation du programme MRL de la NSF.

L'étude d'évaluation que la MITRE (1) a menée de 1976 à 1978 est l'une des plus importantes étude d'évaluation. Son but était d'examiner l'efficacité du système de financement des laboratoires des Matériaux, dit programme MRL (Material Research Laboratories) de la NSF. Les labora-

(1) Evaluative Study of the Materials Research Laboratory Program, The MITRE Corporation.

toires du programme MRL sont rattachés à des universités, de caractère interdisciplinaires, Le financement permet à ces laboratoires de remplacer leurs matériels.. Mais aucun chercheur particulier n'est financé ; c'est le laboratoire qui est financé et les chercheurs peuvent demander des bourses individuelles à la NSF en plus.

Ce type de financement (le "block-funding" ou "core-funding") est très en vogue en ce moment car il est complémentaire du système de financement par bourse individuelle (1).

Les MRL étaient des laboratoires du programme interdisciplinaire du Department of Defense (ARPA) créé en 1960. En 1972, le programme interdisciplinaire du DOD fut transféré à la NSF (c'est à cette époque que fut forgé le terme de core-funding, évoquant la possibilité d'acquérir des moyens financiers par les bourses individuelles autour d'un financement du labo). Avant l'étude de Lin et al. du MITRE, les laboratoires MRL et les programmes interdisciplinaires de recherche sur les matériaux avaient été évalués 5 fois.

Comme l'évaluation de la MITRE, les précédentes évaluations étaient surtout concernées par l'efficacité de ce mécanisme de financement qui est relativement marginal aux USA. La plupart des recherches sont financées soit par des bourses, soit par des contrats où les objectifs de la recherche sont déterminés à l'avance, ce qui n'est pas le cas du "core-funding".

(1) H.J. SANDERS, Peer Review, How Well is it working ? Chemical and Engineering News, 15 mars, 1982, p. 38.

De nombreuses données ont été rassemblées concernant les inputs, les processus et les output de la recherche. Elles furent rassemblées par : des "reviews" des papiers publiés par les membres des laboratoires évalués et un choix effectué à partir d'une évaluation d'experts (Assessment by experts) des principales découvertes de ces laboratoires ; d'une collecte de citations effectuée par l'ISI ; des interviews avec les chercheurs et administrateurs dans les laboratoires et les universités ; des propositions de financement et les rapports de fin ou de renouvellement de financement ; et, des études d'évaluation des programmes de recherche des matériaux qui avaient été effectués auparavant.

Les données en questions ont été collectées sur une période de 10 ans. Elles sont désagrégées par universités, type de recherches, agences, etc... Le nombre de données fut considérable. A tel point que les auteurs notaient : "Si une nouvelle évaluation des programmes MRL devait avoir lieu, il faudrait trouver une méthode standard pour affecter les indicateurs d'outputs aux programmes (...) sinon les données ne seront jamais comparables" (p. 167). Les données étaient douvent difficilement accessibles. L'équipe de MITRE, considère que l'on ne doit remonter que 5 ans en arrière dans le temps pour la collecte de données à des fins d'évaluation (1)

Les résultats de cette étude concernent :

- la capacité des universités à effectuer de la recherche scientifique.
- les coûts associés à la gestion des activités de recherche.
- le type de recherche effectuée dans les programmes
- la qualité des recherches effectuées
- le "mérite" (ou impact) des recherches effectuées.

(1) Cette observation ne concerne pas les données statistiques macro.

LISTE DES VARIABLES DE L'ETUDE D'EVALUATION DU PROGRAMME
MRL EFFECTUEE PAR LA MITREVARIABLES D'OUTPUT

- Nombre de publications et de rapports par discipline ou champ de recherches soumis pour approbation de bourse de recherche MRL.
- Nombre d'auteurs
- Nombre de citations par publications
- Nombre de PhD - domaine de R. - emploi actuel
- Nombre de jeunes membres de l'université recevant une bourse NSF individuelle ou autre aide à la recherche.
- Nombre de jeunes membres de l'université ayant un emploi.
- Autres fonds que le fonds NSF reçus par les labos MRL.
- Résultats des recherches
- Découvertes importantes.

VARIABLES DE PROCESSUS

- Importance respective donnée par les directeurs des labos à la conception et manipulation (design and operation) dans la recherche sur les matériaux (matériels engineering Research).
- Degré d'intégration de chercheurs émanants de champs de recherche différents (formation d'équipes multidisciplinaires cohérentes)
- Rôle de l'instrumentation des labos (financée par le "core funding") dans les recherches
- Ratio heures/dollars investis pour la gestion des "block ou core funding" par rapport au même ratio dans la gestion des bourses individuelles.
- Type de gestion pour les labos de recherche comparés (NSF-DOD NASA et AEC)
- Nombre d'étudiants de niveau maîtrise ou post PhD qui furent aidés financièrement.
- Coût, Sophistication, Unicité, originalité de l'équipement.
- disponibilité, temps d'utilisation du matériel par type et coût du matériel.

VARIABLES INPUTS

- Liste (nombre) des chercheurs, leurs publications et leurs recherches
- domaines de recherche / disciplines inclues dans les recherches effectuées dans les labos MRL.
- d° de concentration sur un problème de recherche, même interdisciplinaire (THRUST AREA)
- Continuité / ancienneté du personnel MRL.

- la productivité des laboratoires et institutions financés par le même programme.

Ces diverses conclusions que nous allons examiner ici peuvent être considérées comme étant les objectifs intermédiaires de l'étude d'évaluation elle-même dont l'objectif initial était d'examiner l'efficacité d'un mode de financement.

(Voir tableau page suivante)

A. Capacité de recherche des laboratoires MRL

La capacité de recherche se résume dans la capacité

- a) d'attirer les chercheurs de meilleure réputation ; b)
- b) d'attirer plus de fonds ; c) d'avoir un équipement moderne, spécialisé et coûteux .

a) Le matériel

Les inventaires permettent d'établir le coût d'achat du matériel, mais pas la valeur actualisée et une étude d'évaluation ne doit pas effectuer ce calcul qui implique des choix souvent arbitraires et qui peuvent éventuellement s'avérer être faux.

La recherche sur les coûts et l'âge du matériel, permettent aux évaluateurs d'établir les utilisations du financement NSF. Cependant il faut noter que la seule distribution statistiquement significative concerne les labos MRL constitués récemment par la NSF, qui détiennent le matériel le plus récent et le plus coûteux.

Les universités financées par les fonds de la NSF et qui disposent de laboratoires, accessibles à plusieurs chercheurs (science centers) ont pu être comparés aux universités sans laboratoires centraux ; cette

TABLEAU 5.5.

EVALUATION DU PROGRAMME MRL DE LA NSF

<u>DONNEES</u>	Eléments concernés par éva- luation	OBJECTIFS INTERMEDIAIRES (Catégories utilisées par la MITRE)	OBJECTIF FINAL
INPUTS	Gestion des laboratoires	CAPACITE DE RECHERCHE COUTS D'ADMINISTRATION	EFFICIENCE DU FINANCEMENT dit "CORE-FUNDING"
PROCESSUS	Caractéristique des travaux éva- lués	TYPE/COHERENCE DE LA R. QUALITE MERITE : IMPACT	
OUTPUT	PRODUCTION des laboratoires	PRODUCTIVITE	

L'efficience est décelable compa-
rativement et concerne la gestion
des travaux et la production com-
parés aux mêmes éléments dans un
autre système de financement
(Bourses individuelles)

comparaison donne l'avantage aux universités avec des laboratoires centraux.

Les évaluateurs n'ont pas pu déceler à quel point ce matériel était spécialisé pour un certain type de recherche ou au contraire s'il était général, c'est-à-dire utilisable dans plusieurs types de recherches (unicité du matériel).

La complexité du matériel est aussi impossible à évaluer - dans les inventaires on ne dispose que de pièces détachées ! - De plus, la "complexité, est une notion qui dépend du domaine de recherche. Les études d'évaluation par définition sont inadaptée pour effectuer ce type d'évaluation (de la complexité ou unicité du matériel) puisqu'elles ne se préoccupent pas du contenu des recherches.

Enfin, rien n'est dit sur la stratégie d'acquisition du matériel, ce qui est probablement le point important. En effet les négociations concernant le matériel comme d'ailleurs les notions de complexité ou unicité du matériel, reflètent "le caractère conflictuel de ce que les scientifiques et technologues appellent l'état de la question" (Callon, 1981, p. 385) (1). C'est là une des limitations les plus importante des analyses d'évaluation.

b) Attirer des chercheurs éminents[^]

Concernant l'attraction de chercheurs éminents, les évaluateurs ont examiné les listes de professeurs ayant travaillé dans les laboratoires MRL selon deux critères :

(*) M. CALLON, 1981, Pour une sociologie des controverses technologiques, *Fundamenta Scientiae*, 2(3/4), 1981, pp. 381-399.

jugement par les pairs et nombre de distinctions honorifiques (Nobel, etc...). Un classement fut effectué à partir de ces données. Les laboratoires MRL attirent les chercheurs avec des notes supérieures. Il faut ici rappeler que les laboratoires retenus sont les 19 laboratoires les plus grands des Etats Unis. Les Universités en dehors du programme MRL qui ont été retenus à des fins de comparaison sont les 15 universités recevant le plus d'argent de la part de la NSF pour la physique, chimie et technologie des matériaux.

Un des enseignements de ce type d'étude est qu'elles reproduisent les situations en place et même peuvent éventuellement consolider la position de ceux qui se trouvent dans des situations avantageuses. Cela rend très délicate l'utilisation d'études d'évaluations.

De plus la capacité de recherche ne se résume pas seulement dans des quantités d'argent, de matériel et de personnel disponibles ; c'est ce qui explique la faiblesse des relations issues des calculs statistiques dans l'étude en question.

c) Les coûts administratifs. Il s'agit du temps passé et de l'argent dépensé dans la gestion des projets. C'est là une question de politique scientifique et probablement celle qui concerne le plus directement la question très actuelle aux USA de "l'accountability". A titre d'exemple nous reproduisons ici le type d'activités qui sont répertoriées et évaluées en heures/dollars, homme-heure/dollar, ou en effort comparé (tableau 5.6)

TABLEAU 5.6

COMPARAISON "CORE " et "PROJECT" : EVALUATION ADMINISTRATIVE DES COÛTS.

ACTIVITE

	Type de financement c = Core P = Project	Variations dépendent du Nb (N) ou du type de financement (F)	" CORE FUNDING"	"PROJECT" FUNDING"
1. Le chercheur discute du plan de recherche avec le directeur du MRL ou l'Administration du programme NSF	C, P	F	60-90 % de taux de succès	30-70 % de taux de succès
2. Le chercheur prépare la proposition	C, P	F	2 h à 20 h 0,4-0,7 heure par \$	2 semaines à 2 mois 1,7-6,8 h/\$
3. Le directeur et le comité directeur du MRL décident à partir de la proposition	C		60 à 100 hommes par heures 0,07-0,11 h / \$	
4. Le directeur du MRL prépare la proposition de financement auprès de la NSF à partir du contenu des projets individuels et des domaines de recherche spécifique dans lesquelles le laboratoire se spécialisera	C		80-100h/homme 0,09-0,2h/ \$	
5. Les services de l'Université passent en revue la proposition	C, P	N	≈ 1 pour 19 (voir notes)	
6. La NSF examine les propositions(éval. par les pairs)	C, P	F, N	≈ 1 pour 2	
7. La NSF prépare la note d'acceptation ou de refus	C, P	N	≈ 1 pour 19	
8. Le budget du projet est établi par l'université	C, P	N	≈ 1 pour 19	
9. Les Bourses non renouvelées sont "soldées" par l'université	C, P	N	≈ 1 pour 19	
10. Plus de temps nécessaires pour les bourses accordées	C, P	N	≈ 1 pour 10	
11. Reexamen éventuel du budget de la proposition	C, P	N	≈ 1 pour 19	

12. Définition du domaine de recherche	C		20-110 hommes 0,03-0,12 h/ \$
13. Directeurs de laboratoires se rencontrent 2 fois par an	C		27 heures 0,03 heures/ \$
14. Visites des laboratoires 1 fois par an au moins	C		<u>Visites prévues</u> 180-240 heures 0,20-0,26 h/ \$ <u>Visites informelles</u> tous les 3 ans 420-480heures 0,46-0,53 h/ \$
15. Rapports techniques	C, P	F	60-100 heures 4-16 heures 0,07-0,11h/ \$ 0,09-0,34 h/\$
16. Rapports budgétaires	C, P	N	1 pour 19

TOTAL COUTS ESTIMES EN HOMMES/HEURES
(moyenne)

CORE = 1 à 1,7 heure par millier de dollars
PROJECT = 1,8 à 7,1 heures par millier de dollars

Note : 1 pour 12 se lit : 1 heure dans le cas du "core" contre 19 heures dans le cas du "project"

L'analyse a montré effectivement que pour le "core-funding" les coûts d'administration et le nombre d'heures est inférieur au "grant-funding" des projets. (Voir tableau) Mais l'analyse de cette étude d'évaluation ne permet pas de comparer le temps consacré à la recherche par rapport au temps consacré à des tâches administratives. Une telle comparaison a été menée par d'autres auteurs (1).

Elle montre, au niveau fédéral que 11 % du temps total professionnel (46 heures par semaine en moyenne) est consacré aux tâches administratives, 24 % à la recherche, 39 % à l'enseignement.

A partir d'études de cas, on a pu évaluer la part du temps de recherche (25 heures par semaine) consacré à des tâches administratives : environ 25 % à 33 % du temps de recherche (2).

B. Type des recherches MRL

Les domaines de recherche sont déterminés pour l'étude d'évaluation à partir des publications, des journaux dans lesquels paraissent les articles émanant de ces laboratoires. Le libellé du domaine de recherche est un croisement de la classification des sujets concernant la recherche sur les matériaux et une liste fournie par la National Academy of Science.

L'étude d'évaluation regardait seulement si les domaines de recherches du MRL étaient les mêmes que ceux des autres laboratoires américains des matériaux (30 % de ces domaines sont les mêmes).

(1) Professional Activities of Science and Engineering Faculty in Universities and Four Year Colleges. Rockville, Maryland Information and Communications Applications, Inc. 1980.

(2) Science Indicators, 1980, p.83.

Il faut remarquer que ce type de manipulations n'est valable que dans le cas d'études comparatives.

C. Cohérence des recherches MRL

Ce vocable désigne la durabilité et stabilité des recherches menées aux MRL. Les auteurs ont examiné la concentration des fonds par projet et par domaine de recherche, cette fois tels que les chercheurs les définissent durant l'enquête ; ils ont examiné le taux de changement de domaine de recherches des projets, par an ($\approx 30\%$) ; ils ont aussi examiné la longueur des périodes de recherche par domaine de recherche par projets (les résultats ne sont pas donnés), enfin ils ont examiné la permanence du personnel employé dans les programmes MRL, sur 13 années ($\approx$ de 62 à 64 % du personnel employé en 1965 était encore présent en 1978).

Les différences entre les programmes MRL et les programmes non-MRL ne sont pas significatives.

D. Jugements d'experts à partir des publications

L'analyse de publications a été menée pour identifier et mesurer les critères :

- 1) de qualité de recherche
- 2) du caractère innovatif
- 3) l'impact sur le progrès scientifique
- 4) le niveau d'interdisciplinarité des recherches

"L' évaluation est effectuée par 215 "reviewers" qui passent en revue un échantillon de 690 publications. Des comparaisons sont effectuées entre les publications issues des laboratoires avant d'être financés par la NSF et après avoir été financés par la NSF, entre publications issues des laboratoires

MRL financés par le mécanismes du "core-funding" et celles qui sont issues de projets financés par les bourses individuelles. Enfin, les papiers issus des MRL mais de recherches financées par des bourses individuelles sont comparés à ceux de recherches financées par le "core funding".

Par ailleurs les publications étaient regroupées en articles théoriques et articles expérimentaux ou encore articles mélangeant les deux genres (ce sont les reviewers qui établissent la distinction) ; en articles avec une orientation scientifique ou avec une orientation d'ingénieur. Toutes les caractéristiques permettant d'identifier les différents critères sont désagrégés suivant l'orientation des articles pour savoir si l'orientation influence par exemple, la qualité, l'interdisciplinarité, etc...

Les comparaisons sont établies sur la base des réponses des reviewers à un questionnaire. Ainsi le degré d'interdisciplinarité est "mesuré" par les réponses aux 3 questions suivantes :

- les auteurs représentent ils plus d'un département d'université ?
- l'input provient-il de plusieurs disciplines ?
- les résultats sont-ils utiles à plus d'une discipline ?

Mais les résultats concernant ce degré d'interdisciplinarité ne sont pas significatifs pour comparer les programmes MRL aux non-MRL.

Trois éléments permettaient de mesurer le degré d'innovation :

- les outils de la recherche, les techniques et l'approche analytique,

Seule la moitié des articles cumulaient l'emploi d'outils nouveaux, de techniques et d'approches analytiques

nouvelles. Mais 80 à 90 % des articles recensés avaient qu'au moins une de ces caractéristiques.

Là aussi les différences dans les comparaisons ne sont pas significatives. Enfin la qualité des processus de recherche est testée au moyen de quatre éléments :

- . L'adéquation de la méthode de recherche, c'est-à-dire sa solidité, sa justesse et sa précision.
- . La précision des outils de recherche.
- . le degré selon lequel les objectifs fixés ont été atteints
- . La justesse - ou solidité - de l'interprétation des résultats.

Par la suite, les contributions qu'apportent ces publications au progrès scientifique sont identifiées par les reviewers. L'article présente :

- un nouveau matériau
- une nouvelle technique
- la création d'un nouveau champ de recherche
- la résolution d'une controverse
- s'attaquer à un problème difficile.
- la meilleure compréhension d'un phénomène physique.

A ce classement, les reviewers ont ajouté les possibilités suivantes :

- "review" d'un champ de recherche
- applications supplémentaires (d'un nouveau matériel, théorie ou technique) à celles prévues par les auteurs.
- encouragement à des recherches futures.
- amélioration de notre connaissance dans un domaine par l'apport de nouvelles données ou des informations supplémentaires, etc...
- confirmation de notre connaissance dans un domaine (state-of-the-art).

- nouvel instrument
- nouvelle application,
- réfutation (d'une méthode, théorie, données ...)

Les catégories qui reçoivent le plus grand nombre de réponses sont :

- la recherche fournit une meilleure compréhension d'un phénomène physique
- le chercheur s'attaque à un problème difficile
- l'article montre des applications supplémentaires à celles prévues par les auteurs.

Les auteurs de l'étude ont pu d'ailleurs relever les articles sur lesquels les reviewers faisaient beaucoup plus de commentaires. Il s'agit des articles les plus importants et des articles les plus insignifiants (aux laboratoires MRL cela donne environ 7 % d'articles insignifiants et 2 % d'articles très importants).

Il faut toujours garder à l'esprit, que cette étude d'évaluation compare des unités qui se distinguent par le type de financement qu'elles reçoivent. Il est donc normal que les auteurs ne cherchent pas trop à approfondir les résultats et à identifier les "causes" qui seraient susceptibles d'expliquer ces résultats. Ils se contentent de dire que les résultats ne comportent pas de biais. Malgré cette réserve, on ne peut pas adhérer entièrement à la façon d'interpréter et d'utiliser les questionnaires comme le font les chercheurs du MITRE. Le seul fait que la plupart des articles soient classés principalement comme "fournissant une meilleure compréhension d'un phénomène physique" rend ces résultats suspects. Car, en effet, quel est le but de la recherche scientifique si ce n'est précisément de fournir une meilleure compréhension du monde physique ?

Ensuite, tant qu'on n'a aucun moyen pour savoir quelle est la perception des diverses catégories de "contributions au progrès scientifiques" par les reviewers, il est très difficile de considérer significatives leurs réponses. Par exemple, les scientifiques repugnent peut être à dire qu'un article arrête ou résoud une controverse et préfèrent dire d'un article qu'il améliore notre connaissance d'un phénomène.

Mais pour faire des comparaisons entre diverses institutions pour les mêmes reviewers on peut se contenter des formulations ci-dessus où ne sont retenues que les catégories avec le plus grand nombre de réponses ; c'est ce que font les auteurs de cette étude.

E. Analyses bibliométriques

L'étude d'évaluation des laboratoires MRL comporte une étude bibliométrique qui porte sur les publications et les citations.

Tout d'abord, le nombre d'auteurs a été utilisé comme mesure de la recherche en collaboration ($\approx 2,5$ auteurs par articles, jamais plus de six auteurs).

Ensuite les citations des auteurs à des travaux de membres des mêmes universités, les auto-citations et l'absence de citations à des travaux de la même université ont été recensés à partir des 690 publications passées en revue. Il n'y avait pas de différences significatives entre les différentes populations examinées. On constate que environ 21 à 22 % de ces articles citent des travaux effectués dans la même université ; environ 65 à 70 % des auteurs citent leurs propres travaux et 9 à 15 % ne citent aucune recherche effectuée dans la même université.

Par ailleurs, les citations des papiers des MRL publiés en 1973 ont été analysées par l'ISI. Le nombre de citations atteint assez régulièrement son maximum 2 à 3 ans après la publication de l'article (donc ici maximum de citations entre 1975 et 1976). Les auteurs de l'étude constatent que le jugement des reviewers corrèle assez bien avec le nombre de citations. Cependant ils sont assez laconiques sur leurs résultats et la façon de les obtenir.

F. Production des laboratoires

La production des laboratoires est mesurée par le nombre de publications, les résultats scientifiques et le nombre de jeunes chercheurs ayant reçu leur PhD en travaillant dans ces laboratoires.

Les publications sont analysées en fonction du montant du financement. Le montant moyen du financement par projet divisé par le nombre moyen des publications par projet, est d'environ 14 à 15, quelque soit le laboratoire et le montant du financement. Cet indice aurait donc tendance à montrer que la production d'article est fonction du montant du financement. Mais ce point est loin de faire l'unanimité, comme le montre l'étude de l'UNESCO examinée au chapitre 3.

403 résultats des recherches MRL ont été identifiées par un groupe d'experts. 102 parmi ces 403 résultats sont considérés importants, et ceux-ci furent produits en majorité par les laboratoires MRL ; c'étaient aussi les travaux expérimentaux qui produisirent le plus de résultats et les MRL effectuent essentiellement des travaux expérimentaux. Les critères selon lesquels un résultat est dit "important" ne sont pas spécifiés.

Enfin, le nombre de diplômés dans le cadre des MRL fut

difficile à dénombrer, bien qu'étant un indicateur "exclusivement quantitatif". La proportion de diplômés se dirigeant vers le secteur public, industriel ou universitaire est le seul chiffre significatif qui a pu être établi. Les auteurs ont ainsi comparé les chercheurs MRL et non-MRL. Ces derniers sont plus attirés par l'industrie que les premiers. Les chercheurs MRL auraient plus tendance à demeurer dans un cadre universitaire. Ce qui tend à montrer qu'une bonne formation en terme d'environnement de recherche est nécessaire pour rester dans le domaine de la recherche fondamentale.

Encore une fois, comme nous ne connaissons pas la stratégie des chercheurs, nous ne pouvons pas savoir s'il s'agit là d'une tautologie ou bien d'un véritable résultat.

4.2. Evaluation a posteriori du programme d'Océanographie

L'étude du programme d'océanographie fut expérimentale dans la mesure où elle fut la première menée à l'intérieur de la Fondation. Son objectif était de déterminer la qualité de la recherche scientifique, à partir des publications issues de ce programme.

Le rapport de l'étude précise :

"Il n'existe pas de meilleure évaluation que le jugement par les pairs. Entreprendre une évaluation par les pairs (a posteriori) c'est affronter plusieurs problèmes, notamment ceux concernant la disponibilité et l'objectivité des juges, la sélection des projets, le choix des périodes de comparaisons, la collecte de données satisfaisantes, la validité des résultats statistiques, etc..." *

(*) Evaluation Staff NSF, Evaluation Study of NSF's oceanography program, executive summary, Aug. 20, 1980.

La procédure est la suivante (*) on sélectionne au hasard, 51 parmi 95 projets accomplis de recherche océanographique financés par la NSF, pendant au moins 2 ans. Par ailleurs, on constitue un groupe de reviewers (dans ce cas précis il s'est agit du oceanography Advisory Committee de la NSF) qui sont chargés de noter ces projets sur la base des résumés, et des articles entiers. Par ailleurs les directeurs de programmes de la division d'océanographie ont sélectionné des journaux océanographiques d'où furent extraits 25 articles écrits par des équipes non financées par la NSF.

Ces articles, ou ensembles d'articles, furent également examinés et notés par le même groupe de reviewers. De plus, les reviews de l'évaluation a priori par les pairs furent rassemblés et comparés aux reviews a posteriori. Ainsi, les données statistiques suivantes furent rassemblées :

- 1) les notations des reviewers établies sur la base des abstracts et sur la base des articles entiers.
- 2) le nombre total d'articles et le nombre d'articles indexés par Oceanic Abstracts.
- 3) Le nombre de citations aux publications et le nombre de citations de l'article le plus cité pour l'ensemble des publications et pour les publications indexées par Oceanic abstracts.
- 4) les informations sur le renouvellement ou non de la bourse par les mêmes chercheurs et les décisions de financement de ces propositions de renouvellement.
- 5) les notations individuelles et moyennes des reviews a priori (notations de reviewers postées, reviews en panel et notations des directeurs de programmes).

 (*) étant assez significative d'une étude d'évaluation, nous allons ici la reprendre dans le détail.

Les statistiques sont établies à l'intérieur de la Fondation (sauf les informations concernant l'index de Oceanic Abstract et les citations qui sont recueillies par l'ISI). On y voit deux mesures "qualitatives" issues du jugement des pairs en plus des jugements qui avaient été émis sur les projets au moment de leur évaluation a priori (jugements émis à partir des abstracts et jugements émis à partir des publications) des mesures bibliométriques (nombre de publication, nombre de citations reçues et nombre de publications indexées) et des informations concernant la "vie" des projets de recherche financé par la NSF (renouvellement ou pas des bourses, comparaison des jugements émis avant et après l'accomplissement de la recherche-.

Demander des évaluations par des pairs a posteriori est une opération longue. En plus, demander ce genre d'exercice à des scientifiques n'est pas aisé : Ainsi dans le cas de l'expérience du programme d'océanographie, deux sous groupes de reviewers (les chimistes et biologistes) ont refusé de faire des reviews sur la base des résumés. Ils ont également expliqué qu'à leur avis les publications ne sont pas suffisantes pour appréhender le contenu et la qualité d'une recherche. Cependant, une partie de l'Advisory Committee de la division océanographie a accepté d'effectuer les reviews.

Les résultats de cette étude d'évaluation montrent que les laboratoires d'océanographie financés par la NSF, dans l'ensemble avaient des notes moyennes supérieures à celles des groupes de contrôle (non-NSF). Mais les résultats sont statistiquement faiblement significatifs ou ne le sont pas du tout. Ceci est vrai autant des

 * Les sous groupes de Physique et géologie avaient accepté l'exercice.

notations des reviewers sur la base des abstracts que sur la base des articles entiers. L'analyse des citations reçues montre que les projets NSF reçoivent plus de citations que les projets non-NSF. Mais aucune cause de cette supériorité n'a pu être identifiée ni des variations du nombre de citations reçues entre les différentes disciplines. Ce qui montre que les déterminants ou facteurs de la production d'articles et de citations sont loin d'être spécifiés.

Le nombre de citations par article ne donne pas non plus de résultats significatifs. Le nombre de citations reçu par les articles ayant le plus grand nombre de citations donne des résultats faiblement significatifs (avantage aux projets NSF par rapport aux projets non-NSF). Ceci est également vrai si on s'en tient aux seuls projets dont les publications sont répertoriées, au moins une fois par l'Oceanic Abstract (*)

Des corrélations ont été calculées pour mesurer la validité comparative des notations avec les mesures bibliométriques. Les notes établies par les reviewers sur la base des abstracts corrélaient plutôt mieux que les notes sur la base des articles entiers avec les publications alors que les notes sur la base des articles entiers corrélaient mieux avec les citations. Cependant là encore les différences ne sont pas statistiquement significatives. Cependant la variance des corrélations entre les notations et les mesures bibliométriques sont attribuables à un facteur différent des citations et des publications, un "facteur résiduel" (**)

* ce qui ne prouve rien concernant les projets, mais montre qu'Océanic Abstract est une bonne base de données !

** on se souvient ici du facteur résiduel de DENISON qui est le "savoir-faire, la connaissance, etc...

Enfin, les mesures bibliométriques sont plus significatives à un niveau d'aggrégation élevé, ce qui est aussi vrai des mesures économétriques.

Une comparaison entre les évaluations a priori et a posteriori, par les pairs montre de faibles ou moyennes corrélations, sauf exception (non expliquées par les auteurs de l'étude. comme l'océanographie biologique). Les notes de l'évaluation a priori corrèlent moins bien avec les mesures bibliométriques que les notes a posteriori (*). Par contre, les évaluations des directeurs de programme (dernière note avant l'allocation d'une bourse) corrèlent mieux avec les mesures bibliométriques que les notes des évaluations a priori.

La concordance entre les notes a posteriori données par les reviewers a été calculée. 2/3 environ des notes sont les mêmes.

De plus les évaluations a priori concordent mieux avec les évaluations a posteriori effectuées sur la base des résumés que sur celles effectuées sur la base des publications.

Enfin une autre comparaison est effectuée entre les projets qui ont demandé et reçu un renouvellement du financement NSF et ceux qui n'ont pas été renouvelés, soit parce que le renouvellement n'a pas été demandé soit parce qu'il a été refusé après l'évaluation par les pairs (80 à 90 % des projets financés sont renouvelés) Cette comparaison montre que ce sont les "meilleurs" selon les critères

* Mais la corrélation ne tient pas compte des demandes de bourses refusées. A quel point cela fausse la corrélation n'est pas non plus spécifié.

retenus par l'étude qui sont renouvelés. Ceux-ci publient 2,5 fois plus que les projets qui n'ont pas été renouvelés. Ils reçoivent des citations, et le nombre de citations par publications est beaucoup plus élevé pour les projets renouvelés que pour ceux non renouvelés.

La méthode d'évaluation du programme d'océanographie est valide de la même façon qu'une évaluation par les pairs est valide. "Nous n'en sommes certainement pas encore au point où l'on peut se passer de l'évaluation par les pairs a posteriori et de la remplacer par l'utilisation des mesures bibliométriques".

Brièvement on peut dire que tout est corrélé mais rarement de manière significative et plus rarement encore de manière signifiante (puisque le sens des corrélations n'est pas établi).

4.3. Le projet d'évaluation a posteriori du programme de chimie

Un projet d'évaluation a posteriori des programmes de chimie est actuellement en cours. Il est plus axé sur une comparaison entre les projets renouvelés et non renouvelés. En plus des mesures bibliométriques des notes a posteriori seront établies par un seul chimiste. Mais il s'agit là d'un type de données différentes de celles obtenues dans l'étude d'océanographie. En effet ce ne sera pas le projet lui-même qui sera examiné, mais la concordance entre les objectifs et les résultats du projet ; de plus, la NSF prévoit d'examiner à quel point les résultats de l'évaluation a priori affectent les propositions de renouvellement. Il faut remarquer que dans cette étude les problèmes d'interprétations des résultats seront encore plus complexes. En effet, les résultats de la R

fondamentale sont incertains par nature. "On peut mesurer la qualité d'un travail scientifique par l'immensité de l'étonnement qu'il procure", disait Lewis Thomas (1974).

Kruystbush (1978) cité par le rapport du NAS (1) a pu remarquer que dans 65 projets de recherche financés par la NSF, le résultat scientifique de 59 % de ces projets (37 projets) n'était pas mentionné dans la proposition de bourse soumise à l'évaluation par les pairs.

Le rapport du GAO (1981) sur l'évaluation par les pairs a révélé que les reviewers et membres de l'administration de la NSF et des NIH ne faisait pas grief aux chercheurs de ne pas avoir atteint les résultats initialement prévus. En effet, le fait d'obtenir un résultat est jugé plus important que d'atteindre les objectifs initialement prévus.

Le NAS remarquait dans son rapport sur la qualité de la science (2) que : "l'exigence que toute recherche appuyée par des fonds publics atteigne des objectifs strictement définis n'est pas une bonne justification pour ces dépenses ; elle peut avoir un effet contre productif en freinant les recherches de bonne qualité" (p.11). On se rend compte dans l'étude sur la chimie que les comparaisons entre objectifs initiaux et finaux bien qu'elles soient possibles seraient très difficiles à interpréter. Cependant, l'évaluation staff de la NSF ne semble pas considérer qu'il s'agisse là de problèmes insurmontables. Mais le personnel de l'Evaluation Staff, de même que l'ensemble de la NSF justifie rarement le

LEWIS Thomas (1974). The Lives of a cell, NY Vicking, in NAS Quality of science, p.10.

(1) in NAS, op. cit.

(2) The Quality of Science, Methods for postperformance Evaluation in the NSF, NAS Press, 1982.

pourquoi des travaux qu'ils engagent (voir plus loin).

4.4. Etudes sur l'équité des décisions de financement de la NSF.

Les facteurs de variations des notations dans l'évaluation par les pairs sont examinées dans les études sur l'équité des décisions d'attribution des bourses. Ces variations sont étudiées selon les critères géographiques, (Etats et régions) de prestige des universités, âge des chercheurs, sexe ; la fidélité (*) des notations est aussi examinée comme ce fut le cas dans l'étude Cole-Cole (phase II).

La première étude, sur la justice des décisions d'attribution des bourses a été faite en 1980-81 ; la deuxième a été finie en août 1982 (1).

Bien que leur sujet soit la justice dans les décisions de financement, ces études commentent une définition de la qualité de la science. Cette définition repose sur les notations des pairs au moment de l'évaluation a priori (2).

Dans l'ensemble les attributions de bourses ne semblent pas être influencées par les caractéristiques énoncées plus haut pour des projets ayant reçu des notes équivalentes, excepté une légère préférence accordée aux projets émanant d'universités de la région ouest et Est contre le Midwest qui reçoit moins de bourses.

(*) Concordance des observations effectuées par des observateurs différents.

(1) O.A.O./ES : Geographic and Institutional Effects on the the Demand Decision, nov.81.
OAO/ES : A study of Fairness of the NSF Award Decision Process, Aug. 82.

(2) Il faut se méfier des définitions de "la qualité" car on peut trouver la qualité définie à partir de notations directes et quelques fois à partir de la quantité d'articles ou quantité de citations, ou encore une combinaison des deux.

Plus intéressantes sont les analyses concernant la fidélité des notations. La fidélité des notes attribuées par les pairs est de 91 % pour l'ensemble de la NSF. Ceci veut dire que 9 % des notes attribuées à un projet sont totalement différentes et que 91 % sont les mêmes.

De plus la probabilité pour qu'un projet noté deux fois ne soit finalement pas accepté est de 25 % pour l'ensemble de la NSF. Ceci signifie que si un projet accepté est noté deux fois de suite il a 75 % de chances d'être accepté à nouveau et 25 % de chances d'être refusé (1).

Ceci concorde parfaitement avec les résultats de la deuxième phase de l'étude des frères Coles (2) qui portait sur les projets en 1976.

4.5. Les évaluations a posteriori : un outil fort contesté :

Comme nous l'avons dit plus haut, les études d'évaluation a posteriori servent essentiellement à justifier des choix déjà effectués. Ceci est visible dans les dépositions effectuées au Congrès par les responsables de la NSF.

Les évaluations a posteriori servent également à comparer les programmes de la NSF aux programmes des autres agences fédérales.

Il faut remarquer que les études d'évaluation a posteriori sont essentiellement comparative et ne sont significative qu'à un niveau d'agrégation moyen (3). Dans le cas

(1) Si la décision de financement était entièrement due à la chance, alors un projet n'aurait eu que 50 % de chances d'être à nouveau financé, et 50 % de chances d'être refusé.

(2) J.R. COLE, S. COLE, Peer Review in the NSF. Phase II of a Study, NAS, Press, 1981.

(3) Moyen, car ni au niveau national, ni au niveau individuel.

de la NSF, le niveau d'agrégation significatif est le programme NSF. Un projet n'est jamais examiné pour lui même, mais toujours comparativement aux autres projets d'un même programme.

Ainsi ce n'est ni la qualité d'un scientifique, ni même d'une équipe de recherche qui est évalué. C'est le programme de la NSF et son fonctionnement institutionnel.

Ainsi, pour ceux qui n'ont accès ni au contenu des recherches scientifiques, ni au fonctionnement institutionnel de la science, l'évaluation peut paraître un excellent outil d'investigation. Il est donc normal que le Congrès soit demandeur de ce type d'études.

Mais ce qu'il faut surtout souligner c'est que l'évaluation a posteriori est contraire au principe même de l'évaluation a priori par les pairs.

Historiquement, l'attribution de bourses de recherche individuelles avait été un choix novateur. Pendant l'été 1951, la direction de la NSF a eu à choisir entre plusieurs mécanismes de financement. A cette époque à part le Public Health Service, aucune agence fédérale n'avait le droit d'attribuer des bourses individuelles pour des travaux extra-muros. Le seul mécanisme de financement autorisé était les contrats de recherche. Les contrats avaient l'inconvénient d'obliger les contractants à spécifier un objectif précis. Les bourses avaient l'avantage de permettre un financement sans indiquer un objectif précis. Les bourses semblaient être toutes indiquées pour des travaux de recherche "généraux par nature".

La NSF a porté sa préférence sur ce mécanisme aussi parce qu'il donne une plus grande liberté aux chercheurs. Les bourses "permettent d'énoncer moins de conditions de contrôle auxquelles seraient soumises les équipes (conditions and accountability requirements) (...) l'agence aura moins de contrôles à effectuer soit sous la forme d'audits financiers soit sous la forme d'évaluation des résultats" (NSF, 1951). (1)

Les évaluations a posteriori ont toujours paru comme un moyen de contrôle administratif qui entrave le bon fonctionnement de la recherche. Le soutien massif que les scientifiques ont apporté à la NSF pour une grande part a été certainement provoqué par cette perspective de liberté qu'offrait la NSF.

L'attitude de la section d'évaluation de la NSF est très révélatrice. Ses membres ne justifient jamais leur travail, les principaux intéressés ayant montré leur appréhension ou un total désintéressement. Et ceci est vrai non seulement des scientifiques mais aussi des hommes politiques ou des administrateurs et budgétistes qui seraient plus intéressés par une méthode de planification des dépenses à effectuer que par une méthode d'évaluation ex-post.

Une autre raison de ce manque d'entrain à se justifier vient de ce que la NSF repose sur les arguments que développent les scientifiques en faveur de la Recherche fondamentale dans d'autres agences ou dans des associations profession-

(1) "Working Paper on Fostering Research" NSF, July, 1951 ; ce document cité dans Lomask (1976) *Minor Miracle*, p. 84. semble avoir fait l'unanimité au sein du personnel de la NSF.

nelles (AAAS, NAS, etc...). La NAS ainsi joue un rôle de garant de la NSF auprès du gouvernement.

La NSF a pour stratégie de ne justifier son travail que quand elle ne peut pas faire autrement sinon, elle garde une prudente réserve à l'égard de ses critiques surtout si ceux-ci détiennent des postes clés *

Par contre, le bureau d'évaluation de la NSF tente de se démarquer du management : "ceci n'est pas un outil de gestion, on n'évalue pas le processus de la recherche ; on évalue uniquement les résultats". A Washington, les rapports entre le congrès et l'OMB sont particulièrement tendus, autant en ce qui concerne la R et D que en ce qui concerne les autres domaines de la société et de l'économie US. Dans ce contexte, la NSF n'a aucun avantage à chercher publiquement à s'identifier aux méthodes utilisées par une puissante administration de l'exécutif, mal vue du pouvoir législatif et du public. Mais, les rapports officieux entre l'OMB et la NSF sont monnaie courante ; l'OMB a une attitude protectrice vis-à-vis de la NSF qu'elle ne manifeste pas vis-à-vis d'agences mieux loties financièrement (Department of Defense, Department of Energy, etc...) ou trop soumises aux pressions politiques du moment (comme les NIH).

Les évaluateurs de la NSF, se démarquent des évaluations de programmes. Un membre de ce service m'a précisé "j'ai participé à des évaluation de programmes éducatifs. mais ces évaluations n'ont rien à voir avec l'évaluation des résultats scientifiques".

(*) voir les relations du Sénateur Proxmire du Comité Appropriations HUA du Sénat et la NSF, M. LOMASK, p. 257-259.

Enfin, ils se démarquent des méthodes de management. Le vocabulaire employé pour indiquer les activités administratives de gestion est différent de celui qui peut être entendu dans une entreprise. On parle d'oversight, pas de management.

Cette distinction qu'établissent les évaluateurs entre leur travail et celui de managers dans les entreprises privées est l'expression de la différence culturelle qui existe entre les scientifiques universitaires et industriels. Distinguer l'évaluation du management, c'est se démarquer par rapport à un système d'organisation où les scientifiques ne sont pas libres de leurs mouvements.

Si l'évaluation a posteriori est appelée "post-peer evaluation" c'est pour bien montrer que là comme dans l'ensemble de la NSF on continuera à faire confiance au mécanisme de l'évaluation par les pairs.

Nous allons examiner ci-dessous l'évaluation a priori des projets au NIH et les évaluations a posteriori qu'effectue cet organisme. Nous allons montrer deux autres évaluations cette fois non pas d'un programme mais d'un ensemble de laboratoires, ceux de l'office of Naval Research aux Etats Unis, et des observatoires d'astronomie étudiés par Irvine et Martin. Enfin, nous allons mentionner quelques études rétrospectives qui peuvent être considérées comme les ancêtres des travaux bibliométriques et des études quantitatives sur les technologies.

II. L'EVALUATION DE LA RECHERCHE AU NIH

1. L'évaluation a priori au NIH (1)

On peut trouver l'origine de l'évaluation duale par les pairs du NIH (Dual Review) dans la Biologics Control Act de 1902, qui institua un bureau d'experts scientifiques privés pour conseiller la direction du Laboratoire d'Hygiène. Lors de la réorganisation de 1930, ce bureau fut rebaptisé National Advisory Health Council (NAHC). Son rôle, inchangé depuis, est de conseiller le Surgeon General (médecin chef, directeur de la Santé).

Le service public de santé commença à attribuer des bourses en 1918. A l'origine, ni le Laboratoire d'Hygiène du NIH, ni le bureau de conseillers n'avaient de rôle dans la gestion de ces bourses.

En 1937, un premier institut de recherches est rattaché au NIH, L'Institut national du Cancer (NCI). Il était habilité à distribuer des bourses aussi bien au dedans qu'au dehors, et suivant l'exemple des relations entre la NAHC et le service public de santé, ce nouvel institut allait avoir un Conseil Consultatif pour le Cancer. Il devait conseiller le directeur et devait ratifier les bourses avant qu'elles ne soient accordées par l'institut. Pour la première

(1) Voir WIRT J.G., LIEBERMAN A.J., LEVIEN R.E.
R.D. Management in Federal Agencies, Lexington Books 1975.

fois, un corps de conseillers scientifiques avait autorité pour approuver, ou non, l'attribution des bourses.

D'autres Instituts furent rattachés au NIH au cours des années suivantes, et tous ont un conseil qui a autorité en matière d'attribution des bourses. Après le passage du Public Health Service Act en 1933, la NAHC avait autorité pour toutes les bourses de santé, même pour celles n'émanant pas du NIH.

En 1947, le NIH créa la division des bourses de recherche (DRG), dont la responsabilité est d'étudier toutes les demandes de bourses, avant leur approbation par les différents comités consultatifs. Les bourses sont examinées par les 25 sections que comprend la DRG. Chaque section comprend d'éminents chercheurs de la communauté scientifique médicale. Le rapport nouveau entre DRG et comités consultatifs marqua une seconde étape majeure dans le développement de l'évaluation duale. Jusqu'à une date récente, la DRG et ses sections ont eu plus d'influence que l'équipe de gestion des instituts et que les comités consultatifs pour décider quels projets seraient financés par le NIH. Lors de ses premières années, les directeurs du DRG assurèrent la programmation de la recherche, en animant ateliers et conférences, en stimulant la recherche dans les universités, et en offrant des bourses dans les domaines prometteurs.

Dans les années 1975-76, les instituts ont accru leur influence pour l'allocation du Fonds de la recherche médicale, en faisant plus de programmation de recherche, et en utilisant de nouvelles méthodes de gestion (centres R et D, contrats de recherche) peu employées par la DRG.

L'évaluation a priori au NIH s'appelle Dual Review car elle se fait en deux étapes.

Les propositions sont données à la Division of Research Group qui les distribue à un premier groupe d'examineurs, qui comporte une quinzaine de spécialistes. Les propositions retenues alors sont distribuées aux unités administratives qui gèrent le programme dans lequel ces propositions sont incluses. Là, les propositions sont à nouveau évaluées par un groupe d'experts. Les propositions retenues reçoivent un deuxième ensemble de jugements. La décision de financer un projet a lieu seulement si le projet a été deux fois recommandé.

Les scientifiques font plus confiance à ce système d'évaluation qu'à celui employé par la NSF, car les groupes sont physiquement rassemblés dans une même pièce. Mais l'étude que le NIH a effectué sur son propre système d'évaluation (1) montre que la plupart des scientifiques ignorent comment il fonctionne. De plus les scientifiques semblent s'accorder dans leur majorité pour dire que c'est là le meilleur système d'évaluation des propositions. Cependant beaucoup de scientifiques ont dit que le premier groupe d'évaluateurs, celui de la "Division of Research Grants" a trop de pouvoir de sélection. Enfin la plupart des scientifiques demandent à ce qu'un mécanisme d'appel soit mis en place.

Le NIH a depuis cette étude accepté d'envoyer le nom des évaluateurs, aux proposants, mais leurs commentaires - qu'ils soient bons ou mauvais - sont anonymes.

(1) Grants Peer Review Study Team, Report to the Director, NIH, Phase I, december 1976.

2. Comparaison du mécanisme NSF et NIH

Si dans l'ensemble le mécanisme "dual" paraît meilleur que le mécanisme de la NSF (1) cela est plutôt dû à la forme des deux mécanismes. Les juges du NIH peuvent discuter entre eux et ceci semble être essentiel dans l'opinion de la communauté scientifique. Alors que les juges de la NSF ne se voient pas et n'ont de rapport qu'avec le directeur de programme.

Le rapport du GAO (1981) prétend que les chercheurs sont plus influencés par les avis des évaluateurs au NIH qu'à la NSF. Cela serait en partie due en fait que le NIH demande aux proposants de spécifier le type de recherches qu'ils ont effectué précédemment et les résultats obtenus, ce que ne fait pas la NSF. La direction de la NSF a répondu au GAO sur ce point en disant qu'une description des travaux précédents des chercheurs est demandé aux proposants dans le cas des bourses qui demandent à être renouvelées. De toute manière juger à partir des recherches passées d'un scientifique ne garantit pas la qualité des recherches futures (Sanders, 1982) (2).

Roy (ibid, p.43) rappelle que l'Office of Naval Research et la Défense Advanced Research Project Agency n'ont pas utilisé une évaluation par les pairs mais n'en ont pas moins financé de très bonnes recherches. Le jugement des responsa-

 (1) Comme le prétend par exemple le rapport du GAO (1981) Better Accountability procedures Needed in NSF and NIH Research Grant Systems, 30 septembre 1981.

(2) H. Sanders, Peer Review How well is it working ? Chemical and Engineering News, 15 mars 1982, p.42.

bles de recherche sur la compétence des chercheurs était la seule évaluation en vigueur. (A l'ONR ce système s'appelait "No Review", la "non-évaluation", (Wirt et al, 1975) (1).

On se souviendra ici que c'est là l'argument des responsables scientifiques des grands laboratoires industriels (voir Branscomb (IBM) 1981, ch.3 ; voir aussi W.Dole Compton (FORD) 1982, (2)).

Si l'évaluation par les pairs jouit d'un tel prestige cela n'est pas dû à son efficacité. En fait, l'évaluation par les pairs quel que soit son mode de fonctionnement est l'institutionnalisation des rapports informels qu'établissent les scientifiques au sein même de la communauté scientifique.

L'évaluation par les pairs est bien l'expression de l'auto-gouvernement des scientifiques. "Le pouvoir investi dans ces mécanismes vient largement du pouvoir des scientifiques en tant que "referees", conseillers et gardiens (gatekeepers) de la science. Le système en vigueur ne fait que rationaliser ce pouvoir" (Mitroff et Chubin 1979) (3).

-
- (1) WIRT et al., R.D. Management methods used by Federal Agencies, Lexington Book 1975, chapitre 2.
 - (2) W. Dole Compton Leading Ph D. :Toward Greater Productivity, Research Management 25 (1) jan.82, pp.17-21.
 - (3) I.I. MITROFF; D.E. CHUBIN (1979) Peer Review at the NSF: A dialectical Policy analysis, social studies of Science, vol.9, 1979, p.201 ; voir aussi J.J. SALOMON, The Mating of Knowledge and Power, Impact of Science on Society, vol.22, 1972, pp.123-132.

TABLEAU 5.7

215.

Comparaison des systèmes d'évaluation a-priori du NIH et de la NSF *

	N.S.F.		H.I.H.
	Mail(1)	Panel(2)	
Conditions que doivent remplir les propositions de recherche (PR)			
- PR doivent indiquer les objectifs à long-terme, et objectifs de la recherche	oui	oui	oui
- prop. de renouvellement de bourse doivent indiquer objectifs du travail de recherche précédent	non	non	oui
- prop. de ren. doivent comporter résumé de la recherche précédente	non	non	oui
- prop. de ren. doivent comporter publications issues de la recherche précédente	oui	oui	oui
Evaluation par les pairs			
- évaluateurs doivent décider du financement	non	non	oui
- déterminer le budget	non	non	oui
- commenter le travail de recherche précédent	non	non	oui
- établir une hiérarchie entre les diverses propositions	non	quelques fois	oui
- commentaires sont automatiquement communiqués aux chercheurs	non	non	oui
Décision de financement peut avoir lieu contre l'opinion des évaluateurs	oui	oui	non
Déroulement de la recherche			
- rapports intermédiaires exigés	oui	oui	oui
- visites sur lieu de recherche	oui	oui	oui
Evaluation a posteriori			
- rapport final exigé	oui	oui	non
- rapport final sur un ensemble de recherches sur le même thème	non	non	oui
- Publications	oui	oui	oui

3. L'évaluation des programmes NIH

Le NIH plus que tout autre organisme a développé les études d'évaluation concernant la plupart de ses programmes et la plupart des programmes des Instituts qui sont rattachés au Service Public de Santé. Depuis la loi de 1970, un pour cent (1 %) des fonds alloués à la recherche médicale devait être attribué à des études d'évaluation. En 1981 cela correspondait à 3,2 millions de dollars et 6,3 millions en 1982. En 1982, 35 % de cette somme (2,2 millions) concernait des projets en cours et 38 % des projets non encore commencés (les 27 % restant concernent des projets acceptés provisoirement).

Les études d'évaluation couvrent 6 types de travaux :

A - Les évaluations des programmes

A proprement parler : "c'est la mesure des performances et des activités de recherche : l'efficacité, l'efficience (1), la capacité à répondre à des besoins (responsiveness)" (2). Elles peuvent être à court terme, ou concerner l'ensemble d'un programme (ex.: Evaluation du programme de vaccination du "National Institute of Allergy and Infections Diseases"), ou explorations afin de savoir si une évaluation est possible).

Le bureau d'évaluation du NIH demande aux différents instituts d'effectuer les évaluations d'ensemble des programmes systématiquement autour de 4 axes (le "paradigme SATT :

(1) Sur le sens de ces mots voir Ch.4, p. plus haut.

(2) Department of Health and Human Services (HHS) FY 1982 Evaluation Planning Procedure, dans "Evaluation Plan FY 1982, Part 1, pp.A.25.

Science Base/Applications/Technology Transfers/Training")(1):

1 Les "fondements" scientifiques du programme en question, c'est-à-dire la description des recherches liées à ce programme, déjà effectuées ou en cours, des disciplines concernées etc... (Science Base)

2 Les produits issus des recherches : médicaments, matériaux, procédés techniques, etc... (Application).

3 Les processus par lesquels se diffuse l'information scientifique concernant ces programmes de recherche, leurs résultats etc... (Technology Transfer).

4 L'enseignement scientifique, la formation des chercheurs, les formations universitaires etc... qui sont liés à ce programme (Training). (ex.: première étape d'une évaluation du programme de la santé infantile).

Ces études représentent 78 % du 1 % des crédits budgétaires 1982 alloués au NIH (4,916 millions).

B - Evaluation de l'efficacité des programmes

C'est ce qu'on pourrait appeler des études d'impact d'un programme sur une population, ou bien du fonctionnement d'un programme de recherche sur l'ensemble du fonctionnement d'un institut, etc... Par ex.: "Liaison entre le nombre de publications et le montant du financement du NIH".

(1) Evaluation Plan, FY 1982, Office of the Director, NIH, Bethesda Md. p.3.7.

Ces études représentent 13 % du budget 82 (832 millions de \$) et 16 % du budget 81.

C - Etat de la question (State of the art)

Il s'agit d'études sur un thème particulier, comme par exemple "les effets de l'utilisation de la Marihuana : recherches futures" ou encore "Evolution des besoins et priorités de la recherche dermatologique". Ces études sont généralement demandées par le Congrès au moment du vote du budget des Instituts.

En 1981, ces études représentaient 6 % du budget d'évaluation, soit 187 milliers de dollars.

D - Evaluation des programmes d'éducation, des mécanismes de financement des recherches des communications et de l'information S et T, et plus généralement des diverses procédures pour le soutien de la recherche biomédicale. Par ex.: "Evaluation du programme d'enseignement des Médecins" ou "Evaluation du système d'information MEDLAR pour la compréhension des troubles psychomoteurs".

Ces études représentent 6 % du budget d'évaluation en 1982 (soit 390 milliers de \$).

E - Des études méthodologiques pour évaluer les programmes.

F - Le développement de données statistiques et l'étude de ressources et du potentiel de recherche des Instituts (Economie de la Santé).

Par exemple : "le marché de l'emploi pour les chercheurs en recherche biomédicale" , ou encore, "les statistiques sur l'évolution des maladies".

°°°

Le bureau central d'évaluation du NIH effectue des réunions mensuelles où les différents problèmes concernant la planification, la gestion et l'évaluation de l'ensemble des programmes du NIH sont discutés. En 1981, un comité fut mis en place afin d'assurer la coordination des activités d'évaluation sur l'ensemble des NIH.

Tous les efforts déployés actuellement ont pour objectif d'intégrer les évaluations à la programmation et la planification.

Le deuxième objectif du "plan d'évaluation" du NIH concerne les indicateurs quantitatifs. Le Bureau de la direction du NIH veut développer des indicateurs quantitatifs, principalement bibliométriques, qui seraient fondés sur l'exploitation de toutes les données "on-line".

4. Indicateurs bibliométriques

La conception des indicateurs bibliométriques a été confiée à Francis Narin et à l'équipe de CHI (1) qui tra-

(1) Voir Ch.2, CHI a également élaboré les indicateurs bibliométriques de Science Indicators.

vaille depuis 1973 sur la production d'articles et documents de tous les instituts (excepté le National Institute of Aging et le National Institute of Environmental Health Service, pour lesquels on ne dispose pas suffisamment de données).

CHI a établi des indicateurs sur la production d'articles entre 1970 et 77 et sur les citations que reçoivent ces articles. Les citations sont celles que répertorie le Science Citation Index de l'ISI.

Les articles paraissent dans environ 260 revues et journaux. La liste des journaux est produite à partir des informations que fournissent les instituts. Les journaux et revues examinés permettent de couvrir environ 80 % de la production des articles issus de recherches financées par des bourses du NIH, et par des recherches effectuées dans les laboratoires et centres internes des Instituts.

La liste des journaux en question est réajustée à chaque nouvelle analyse bibliométrique. Cette liste est compilée selon deux classifications : type de recherche et sujet.

La classification par type de recherche se fait à 4 niveaux du plus appliqué au plus fondamental :

Niveau 1 : Observations cliniques (ex.: Journal of the American Medical Association).

Niveau 2 : Mélange clinique (clinical Mix) : c'est un niveau intermédiaire entre l'observation clinique et la recherche clinique (ex.: New England Journal of Medicine).

Niveau 3 : Recherches cliniques (ex.: Journal of Clinical Investigation).

Niveau 4 : Recherche biomédicale fondamentale (ex.: Journal of Biological Chemistry).

La classification par sujet est effectuée en attribuant une discipline à chaque journal, telles la physiologie, la chimie, la microbiologie, l'immunologie, la pathologie, etc...

Comme pour tous les indicateurs élaborés par CHI, il ne faut pas perdre de vue le fait que ce ne sont pas les articles qui sont classifiés par disciplines et sous-disciplines, mais les revues et journaux.

Le détail des tendances de la recherche est ainsi perdu au profit d'une procédure informatisée plus simple et au profit d'une analyse statistique agrégée.

Ainsi on peut savoir par exemple que les articles émanant du National Heart, Lung, and Blood Institute (NHLBI) a produit de plus en plus d'articles de recherche fondamentale (niveau 4) entre 1970-1977 jusqu'à dépasser la production d'articles descriptifs médicaux (niveaux 1 et 2) (1).

L'analyse des citations permet d'identifier les articles les plus cités (seuls les 10 articles les plus cités sont répertoriés).

(1) On remarquera là que la tendance à la "scientification" de la médecine a lieu en France. Voir Méthodes d'évaluation à l'INSERM, Séminaire sur l'évaluation de la recherche, CNRS, 1982, Compte-rendu n°3, 4 mai 1982, p.12.

Elle permet également d'identifier les institutions qui citent la production d'un institut. Ainsi pour prendre le même exemple qu'auparavant, le NHLBI reçoit 33 % de citations d'articles écrits par des scientifiques appartenant, ou qui sont financés par le NHLBI, 2,6 % par la NSF, etc... A l'inverse on peut identifier les institutions que le NHLBI cite le plus.

Ces analyses permettent de déterminer l'impact - en termes de publications et citation - d'un institut de recherche appartenant au NIH.

On peut finalement établir le "profil" d'une agence ou d'un institut pour une période donnée (voir ci-dessous le profil du NHLBI).

20-JAN-81

TABLEAU 5.8

PAGE 1

BID PUBLICATION PROFILES

HEART, LUNG AND BLOOD INSTITUTE - NHLBI

CITED YEAR : 1970 - 76

BASED ON CITES RECEIVED FROM ALL CITING YEARS 1973-77

CITED AND CITING ARTICLES FROM 293 BID JOURNAL SET

LEVEL/FIELD/SUBFIELD	(1) ⁺ NUMBER PAPERS	(2) ⁺ ACTIV INDEX	(3) ⁺ % INT EFFRT	(4) ⁺ % EXT EFFRT	(5) ⁺ TOTAL* CITES	(6) ⁺ CITES* PER PAPER	(7) ⁺ CITE T-SCR	(8) ⁺ % PAPERS AMONG TOP 10%
FIELDS [1-9]	15292.	1.00	100.00	3.06	149207.	9.8	53.39	18.58
CLINICAL-MIX [1-2]	5647.	1.35	36.93	4.14	45692.	8.1	57.32	26.80
CLIN INVESTIGATN [3]	4172.	1.24	27.28	3.79	37656.	9.0	53.94	17.70
BASIC RESEARCH [4]	5472.	0.71	35.78	2.16	66407.	12.1	53.64	19.37
CLIN MEDICINE [1]	9684.	1.43	63.33	4.36	82975.	8.6	54.20	19.29
BIOMED RESEARCH [2]	5048.	1.08	33.01	3.31	64498.	12.8	52.44	14.19
BIOLOGY [3]	196.	0.28	1.28	0.87	1593.	8.1	52.07	21.24
CHEMISTRY [4]	313.	0.16	2.04	0.48	2700.	8.6	56.78	34.72
PHYSICS [5]	18.	0.04	0.12	0.12	99.	5.6	51.88	10.38
EARTH & SPACE SC [6]	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
ENGRNG & TECHNOL [7]	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
PSYCHOLOGY [8]	33.	0.05	0.22	0.15	130.	3.9	53.88	18.59
MATHEMATICS [9]	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00

(+) Voir définitions des intitulés des colonnes p.

*Scaled to cited year 1972

20-JAN-81

PAGE 2

TABLEAU 5.8 (suite)

BID PUBLICATION PROFILES

HEART, LUNG AND BLOOD INSTITUTE - NHLBI

CITED YEAR : 1970 - 76

BASED ON CITES RECEIVED FROM ALL CITING YEARS 1973-77

CITED AND CITING ARTICLES FROM 293 BID JOURNAL SET

LEVEL/FIELD/SUBFIELD	NUMBER PAPERS	ACTIV INDEX	% INT EFFRT	% EXT EFFRT	TOTAL* CITES	CITES* PER PAPER	CITE T-SCR	% PAPERS AMONG TOP 10%
CLIN MEDICINE [11]								
GENRL & INTERNAL MED	2251.	2.20	14.72	6.72	25272.	11.2	55.02	18.29
ALLERGY	24.	1.34	0.15	4.10	229.	9.7	54.31	15.49
ANESTHESIOLOGY	127.	1.68	0.83	5.15	641.	5.1	53.72	20.69
CANCER	55.	0.15	0.36	0.45	428.	7.8	48.85	9.42
CARDIOVASCULAR SYSTM	2578.	9.46	16.86	28.92	27217.	10.6	54.50	16.11
DENTISTRY	7.	0.04	0.05	0.11	21.	3.0	54.25	19.05
DERMAT & VENERL DIS	15.	0.20	0.10	0.62	45.	3.0	50.09	15.56
ENDOCRINOLOGY	213.	0.64	1.39	1.97	2123.	10.0	50.87	11.65
FERTILITY	5.	0.05	0.03	0.16	34.	6.5	54.58	25.81
GASTROENTEROLOGY	33.	0.40	0.21	1.21	289.	8.9	53.72	16.84
GERIATRICS	37.	0.52	0.25	1.58	90.	2.4	56.69	30.22
HEMATOLOGY	453.	3.34	2.96	10.20	3137.	6.9	51.67	15.70
IMMUNOLOGY	273.	0.63	1.79	1.92	2791.	10.2	49.23	9.65
OBSTETRICS & GYNECOL	90.	0.52	0.59	1.58	462.	5.1	52.63	19.59
NEUROL & NEUROSURG	116.	0.19	0.76	0.58	745.	6.4	48.52	4.15
OPHTHALMOLOGY	41.	0.26	0.27	0.81	145.	3.5	49.03	4.05
ORTHOPEDICS	12.	0.12	0.08	0.38	54.	4.6	58.02	24.29
ARTHRITIS & RHEUMAT	7.	0.18	0.04	0.55	51.	7.6	56.77	25.00
OTORHINOLARYNGOLOGY	25.	0.22	0.16	0.66	51.	2.1	51.01	17.69
PATHOLOGY	463.	2.90	3.03	8.88	3123.	6.7	55.48	17.62
PEDIATRICS	217.	0.97	1.42	2.96	1291.	6.0	53.06	15.37
PHARMACOLOGY	561.	1.34	3.67	4.11	5056.	9.0	53.58	17.08
PHARMACY	99.	0.56	0.64	1.72	535.	5.4	55.31	26.57
PSYCHIATRY	23.	0.12	0.15	0.37	124.	5.4	54.65	21.01
RADIOLOGY & NUCL MED	320.	1.13	2.09	3.46	1466.	4.6	52.59	17.53
RESPIRATORY SYSTEM	384.	5.01	2.51	15.33	2066.	5.4	52.22	18.38
SURGERY	1019.	2.10	6.66	6.41	5018.	4.9	53.18	19.16
TROPICAL MEDICINE	9.	0.30	0.06	0.91	27.	3.0	49.19	5.45
UROLOGY	45.	0.39	0.30	1.21	109.	2.4	50.91	15.87
NEPHROLOGY	18.	1.34	0.12	4.10	39.	2.2	51.16	18.69
VETERINARY MEDICINE	82.	0.66	0.54	2.02	176.	2.1	54.09	21.30
ADDICTIVE DISEASES	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
HYGIENE & PUBL HLTH	83.	0.68	0.54	2.07	466.	5.6	62.66	41.68
MISC CLINICAL MED	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00

*Scaled to cited year 1972

20-JAN-81

TABLEAU 5.8 (suite)

PAGE 3

BID PUBLICATION PROFILES

HEART, LUNG AND BLOOD INSTITUTE - NHLBI

CITED YEAR : 1970 - 76

BASED ON CITES RECEIVED FROM ALL CITING YEARS 1973-77

CITED AND CITING ARTICLES FROM 293 BID JOURNAL SET

LEVEL/FIELD/SUBFIELD	NUMBER PAPERS	ACTIV INDEX	% INT EFFRT	% EXT EFFRT	TOTAL* CITES	CITES* PER PAPER	CITE T-SCR	% PAPERS AMONG TOP 10%
BIOMED RESEARCH [2]								
PHYSIOLOGY	1495.	4.83	9.78	14.77	13289.	8.9	49.40	10.26
ANATOMY & MORPHOLOGY	47.	1.06	0.31	3.24	158.	3.4	47.25	4.26
EMBRYOLOGY	28.	0.53	0.18	1.63	239.	8.5	52.05	19.05
GENETICS & HEREDITY	37.	0.24	0.24	0.75	185.	5.1	48.96	9.55
NUTRITION & DIETET	94.	1.15	0.62	3.52	496.	5.3	52.26	17.88
BIOCHEM & MOLEC BIOL	2202.	1.10	14.40	3.35	32993.	15.0	52.79	15.77
BIOPHYSICS	86.	2.59	0.56	7.92	574.	6.7	51.43	12.67
CELL BIOL CYT & HIST	235.	0.79	1.53	2.41	3296.	14.1	56.44	20.40
MICROBIOLOGY	28.	0.10	0.18	0.29	278.	10.1	60.45	32.73
VIROLOGY	10.	0.05	0.07	0.16	188.	18.8	57.38	18.33
PARASITOLOGY	4.	0.04	0.03	0.12	7.	1.7	48.71	0.00
BIOMEDICAL ENGINRNG	108.	3.86	0.70	11.79	198.	1.8	51.31	16.41
MICROSCOPY	0.	0.00	0.00	0.00	0.	0.0	0.00	0.00
MISC BIOMEDICAL RES	19.	0.49	0.12	1.51	20.	1.1	49.95	9.65
GENRL BIOMEDICAL RES	658.	0.63	4.30	1.92	13906.	21.1	55.47	18.95

*Scaled to cited year 1972

TABLEAU 5.8 (fin)

Intitulé des colonnes du profil de publication du NHLBI

1. Nombre d'articles du NHLBI publiés entre 1970 et 1976.
2. Indice d'activité : proportion d'articles du NHLBI par niveau/discipline/sous-discipline, divisé par la production totale d'articles inclus dans la liste de revue par niveau/discipline/sous-discipline (indice de spécialisation relative).
3. % d'effort interne : % des articles du NHLBI par niveau discipline/sous-discipline dans sa propre production d'articles (indice de spécialisation absolu).
4. % d'effort externe : % des articles financés par le NHLBI dans la production totale d'articles (part de marché).
5. Total des citations : Nombre de citations que reçoivent les articles NHLBI.
6. Citations par article : Nombre de citations reçues (indicateur 5) divisé par le nombre d'articles NHLBI (indicateur 1).
7. Nombre de citations par articles NHLBI (indicateur 6) moins le nombre moyen de citations totales par article dans chaque niveau/discipline/sous-discipline, divisé par la variance des citationstotales par article dans chaque niveau/discipline/sous-discipline. La moyenne de ce chiffre est = 50, la variance est = à 10.
8. % d'articles parmi les 10 % d'articles les plus cités : % d'articles NHLBI qui se trouvent parmi les 10 % d'articles les plus cités de l'ensemble des articles du niveau, de la discipline ou de la sous-discipline en question.

Les statistiques ci-dessus seront significatives avec un intervalle de confiance de 95 % avec un échantillon où $N = 20$.

Actuellement le NIH cherche à coordonner les bases et banques de données dont disposent les instituts avec la base MEDLINE de la Bibliothèque médicale, afin de pouvoir effectuer des études bibliométriques et des analyses statistiques du potentiel de recherche ponctuelles et à court terme.

Le bureau central d'évaluation du NIH a énormément donné d'importance aux indicateurs bibliométriques car ils permettent de faire un tableau de la recherche sans avoir recours aux jugements immédiats des personnes intéressées. Une des raisons qui expliquent cette décision de faire des études d'évaluation bibliométrique systématiquement, est que l'activité des NIH est soumise à de très fortes pressions des groupes d'intérêts. La bibliométrie permet d'éviter les arguments de circonstance qu'invoquent les chercheurs, comme par exemple : "donnez-nous de l'argent en immunologie car on ne peut plus effectuer nos recherches normalement".

Le bureau d'évaluation peut face à de telles requêtes montrer quels sont les domaines où de telles demandes sont justifiées ou pas.

Il est probable que les directions des instituts ont riposté à ce type d'instrument de contrôle en effectuant une profusion d'études d'évaluations. Une telle inflation de rapports a pour effet de "noyer" les résultats les plus significatifs dans une masse de considérations secondaires.

Mais il ne faut pas exagérer la puissance des études bibliométriques. En effet elles ne permettent pas de contrôler a priori des unités de recherche ou des chercheurs. Elles sont suffisamment agrégées pour que les scientifiques ne se

sentent pas "contrôlés". Brièvement on peut dire que les études de bibliométrie n'évaluent pas les recherches mais la réceptivité de la communauté scientifique à ces recherches.

La clé de voûte des études bibliométriques se trouve dans la liste de revues, donc dans le fonctionnement des revues scientifiques. L'évaluation bibliométrique repose sur les jugements des "referees" des revues scientifiques. Elle permet d'identifier les retards (lags) dans la réception des "messages" que livrent les articles scientifiques. Mais elles ne permettent pas d'évaluer la qualité de ces messages.

III. AUTRES EXEMPLES D'EVALUATION A POSTERIORI

Nous allons rapidement examiner deux autres études d'évaluations a posteriori ainsi que les études d'évaluations rétrospectives.

La première porte sur les laboratoires de recherche de la marine aux Etats-Unis (1). La deuxième est une étude, menée au Science Policy Research Unit de l'Université de Sussex, porte sur des observations d'astronomie en Angleterre, aux Pays-Bas et en Allemagne (2).

1. Analyse bibliométrique des laboratoires de la Marine (NRL)

Mc Allister et Narin ont effectué une étude des performances des laboratoires de l'ONR, et des comparaisons entre les NRL et d'autres institutions financées par le Department of Defense.

Les données statistiques concernent 500 institutions qui effectuent des recherches intra-muros. La liste des publications est établie à partir de la base de données de

(1) Mc ALLISTER P., NARIN F., Bibliometric indicators of Research Productivity and quality for Navy Research Laboratories, CHI, 22 février 1982.

(2) OCDE.

l'ISI. Les publications qui sont produites par des chercheurs externes à l'ONR, et financées par le Department of Defense, ne sont pas prises en compte.

La typologie du domaine de recherche (disciplines / sous-disciplines) est la même que celle élaborée pour les "Science Indicators". Le type des recherches est examiné au moyen d'une classification à 4 niveaux, comme celle établie pour l'analyse du NIH, du plus appliqué au plus fondamental :

- Niveau 1 : Technologie appliquée
- Niveau 2 : Sciences de l'ingénieur - technologie
- Niveau 3 : Recherche appliquée
- Niveau 4 : Recherche fondamentale.

De même que pour l'étude des instituts du NIH, les niveaux de recherche/discipline/sous-discipline sont assignés aux revues, pas aux articles pris individuellement. Par contre, dans cette étude, les articles paraissant dans des revues multidisciplinaires (Science ou Nature) sont assignés manuellement à un niveau/discipline/sous-discipline.

Enfin, un indice d'influence des citations est calculé pour chaque revue. Cet indice est le nombre de citations en tenant compte de l'influence de la revue, émettant la citation et de la taille de la revue recevant la citation. L'indice d'influence par article pour une revue est le nombre de fois qu'un article moyen paru dans cette revue est cité.

Les citations, comme les publications sont issues de la base du Science Citation Index de l'ISI. Les auteurs de l'étude prennent énormément de précautions pour éviter plu-

sieurs biais. Tout d'abord, le nombre absolu de citations n'est jamais utilisé dans les indices de citations. Seuls les nombres relatifs de citations sont utilisés. De plus des pondérations sont calculées pour tenir compte de la taille d'une discipline/sous-discipline.

Ce travail de "nettoyage" permet d'établir des "profils de publications" comme pour les instituts du NIH. L'analyse des citations - leur interprétation - se fait à partir de ces profils et en regroupant les citations par quartiles. Le quartile supérieur est plus particulièrement désagréé, par institutions, domaine, sous-domaine. Enfin, les auteurs font une analyse article par article, des 50 articles les plus cités.

L'interprétation des données pour chaque laboratoire de l'ONR se fait selon trois axes :

- leur nombre relatif de publications
- leur activité relative en termes de publications
- leur influence en terme de citations.

Le nombre relatif de publications permet de montrer quelle est l'institution la plus productive et de tracer des histogrammes de production. L'activité est étudiée au moyen d'un indice d'activité. C'est le nombre de publications d'une institution dans un domaine divisée par la part de ce domaine dans le nombre total de publications. Il permet de tenir compte des différences de taille des institutions et d'étudier ce qu'en économie on appellerait la part du marché d'un laboratoire, le marché étant dans ce cas le domaine de recherche.

Pour examiner l'influence des citations, un score des citations du quartile supérieur a été défini. C'est la différence entre le pourcentage d'articles d'un laboratoire inclus dans le quartile supérieur des articles les plus cités dans la discipline, et le nombre d'articles du quartile supérieur dans cette discipline.

Ce score peut bien sûr être calculé par type de recherche, discipline ou sous-discipline. Il permet d'examiner la position des laboratoires par rapport aux auteurs les plus actifs dans une discipline (ou dans tout autre niveau d'agrégation choisi).

Les résultats de cette recherche - que nous ne donnons pas ici - permettent de déterminer la place des laboratoires de la marine dans l'ensemble d'une discipline (ou sous-discipline). Il faut rappeler que ce travail comparatif a été rendu possible par l'existence de plusieurs analyses dont les données sont comparables : il s'agit des données bibliométriques des Sciences Indicators (1).

Les mêmes commentaires que ceux à propos des études bibliométriques des instituts du NIH peuvent être faits de nouveau ici.

(1) Rappelons que ces données peuvent être achetées sous forme de bandes magnétiques pour ordinateur. Voir DATA USER'S GUIDE, to the NSF Science Literature Data Base, NSF, Science Indicators Unit and Computer Horizons Inc., 17 november 1980.

2. L'étude de Irvine et Martin -

La deuxième étude d'évaluation que nous allons examiner ici a été présentée par ses auteurs dans "La Recherche" de décembre 1981 (1). Ils appellent leur méthode d'évaluation quantitative la "méthode des indicateurs partiels convergents".

Les auteurs utilisent des indicateurs bibliométriques en même temps que des évaluations subjectives par les pairs auprès des laboratoires d'astronomie qu'ils ont étudiés, suivant en cela les méthodes préconisées par l'étude de l'UNESCO. Cependant, leur unité d'analyse et de comparaison est un centre d'astronomie, non une "unité de recherche" ou autre groupe de chercheurs dans une institution. Irvine et Martin prétendent que leur méthode est applicable à un groupe de recherches. Les travaux de Pelz, Andrews et l'étude de l'UNESCO le confirment.

Quelques commentaires plutôt qu'une présentation de leur recherche s'imposent.

Comme une lettre de lecteur l'a montré, le choix de l'unité d'analyse est crucial. Irvine et Martin se gardent bien de regarder ce qui se passe dans les centres de recherche. Les lecteurs qui répondirent à leur article ont montré que si on ne retenait qu'une partie de l'un des centres étudiés (Nancy), celui-ci aurait reçu une notation beaucoup plus

 (1) J. IRVINE et B.R. MARTIN (1981)
 L'évaluation de la recherche fondamentale est-elle possible ? La Recherche, vol.12, n°128, décembre 1981, pp.1406-1416.

(2) Op.cit.p.1416.

élevée. De ce point de vue, l'analyse de Irvine et Martin est beaucoup moins fiable que celle de Narin (CHI) qui utilisent plusieurs comparaisons à plusieurs niveaux différents et que celle employée par les chercheurs de l'UNESCO, (qui définit des groupes de chercheurs en fonction des indicateurs recueillis). Elle a par contre l'avantage de la simplicité. En effet les auteurs ne retiennent que 3 indicateurs d'output objectif et un classement général des centres d'astronomie. Les 3 indicateurs objectifs sont :

- le compte des publications
- la fréquence de citations
- le compte des articles les plus cités
(12, 15 ou 20 fois en un an)

Beaucoup d'explications sont fournies pour expliquer les limites de ces indicateurs. On peut remarquer ainsi que les auteurs précisent bien que les citations ne sont pas des indicateurs de qualité, mais des indicateurs de l'impact d'un article.

Ils distinguent d'ailleurs l'impact effectif d'un article de son influence potentielle. "Le nombre de fois qu'un article est cité ne constitue pas un relief direct, une mesure, ni de sa qualité ni de son importance, mais un indicateur partiel de son impact" (Irvine et Martin 1981, p.1410). L'influence potentielle, ou importance de l'article est soumise à d'autres facteurs que ceux auxquels l'impact est soumis. Les réactions, parfois violentes que peuvent provoquer les études d'évaluation de ce genre, viennent en partie de l'assymétrie dans le traitement de ces concepts. En effet, seul l'impact retient l'attention dans les études d'évaluation, alors que la qualité - jugée cruciale - et l'importance - qui est de la qualité potentielle - ne font

l'objet que d'une mise en garde préliminaire puis sont rapidement oubliées.

A ce sujet notons qu'à la NSF les indicateurs partiels de citations sont appelés indices de qualité et aucune tentative de distinguer la qualité n'est faite. Narin (évaluations bibliométriques des NIH et NRL) ne considère que les indicateurs bibliométriques sont valables quand un nombre suffisant de données peut être accumulé sur une période suffisamment longue et quand on ne retient comme niveau d'analyse que des unités suffisamment agrégées pour que ces données soient significatives statistiquement.

Le classement général des laboratoires d'astronomie dans l'étude de Irvine et Martin est une moyenne du rang attribué à chaque centre par un groupe d'experts. Les experts devaient établir un classement par valeur décroissante de neuf centres d'astronomie. On peut ici reprocher aux auteurs, comme le font Gougenheim et Kazès (1982), op.cit.p.417,(1) de ne pas avoir demandé des explications - ou de ne pas les présenter sur ce classement. Ils auraient pu, par exemple, demander aux experts d'effectuer plusieurs classements successifs sur des critères différents.

Finalement l'étude de Irvine et Martin est valable pour ce qu'elle a été conçu : comparer 2 observatoires britanniques, celui de Cambridge et celui de Jodrell Bank et donner des indications comparatives sur les 2 autres observatoires. Mais alors que beaucoup de précisions sont données

(1) L. Gougenheim, I. Kazès, Correspondance, La Recherche, 13 (n°131) mars 1982, pp.416-417.

sur le sens des indicateurs - et sur ce qu'il ne faut pas leur faire dire -, en revanche peu d'explications sont fournies sur leur signification statistique. Enfin les auteurs, comme dans toutes ces études d'évaluation quantitatives, ne tentent de donner aucune explication sur les facteurs qui peuvent influencer ces indicateurs.

3. Les études rétrospectives

Les études rétrospectives portent sur des innovations, pour mettre en évidence les travaux de recherche fondamentale et appliquée antérieurs. Deux études rétrospectives sont particulièrement connues : celle menée au département de la défense sous le nom "Programme Hindsight" (regard rétrospectif) et celle menée à la NSF sous le nom de projet TRACES (Technology in Retrospect And Critical Events in Science).

L'étude Hindsight porte sur quinze systèmes d'armement. Treize équipes d'ingénieurs et de scientifiques qui travaillaient au Département de la Défense ont identifié 713 événements critiques qui ont permis d'améliorer la rentabilité des systèmes d'armement. La différence entre le coût du matériel antérieur à cet événement et le coût du matériel après l'événement était relativement facile à effectuer car il s'agit d'innovations très bien identifiées. Ces événements ont été classés par les experts en événements scientifiques et événements technologiques. 8 % de ces événements sont scientifiques et 92 % sont technologiques. Le délai moyen séparant un produit ancien et son remplaçant est de 13 ans. Mais les auteurs du projet Hindsight ont éliminé tous les événements trop lointains (trop fondamentaux ?), ce qui peut-être explique le faible résultat des événements scientifiques dans l'étude. En effet, le délai de recherche des événements scientifiques est de 9 ans en moyenne ; il n'est que de 5 ans en ce qui concerne les 8 % d'événements scientifiques, seuls 0,3 %

ont de la recherche "pure", c'est-à-dire non-orientée et ne répondant à aucun besoin ni militaire, ni commercial.

Les auteurs du rapport concluent en disant que même si le nombre d'événements scientifiques est très faible, la base des connaissances qui agit sur les innovations est très importante. "La science ancienne organique, accumulée, parfaitement comprise et soigneusement enseignée" constitue le fondement même de la recherche, pure et appliquée⁽¹⁾.

La deuxième étude rétrospective (Traces) porte sur 5 innovations civiles (le magnétoscope, la pillule contraceptive, les ferrites magnétiques, le microscope électronique et l'isolement par matrices). 341 événements scientifiques ont été repérés par une équipe d'experts pour chaque innovation. Parmi ces événements, 70 % sont des scientifiques, attribuables à la recherche fondamentale, 20 % à la recherche orientée, 10 % au développement. Pour 90 % de ces événements, les recherches fondamentales avaient cessé depuis 10 ans avant l'apparition de l'innovation à laquelle elles avaient contribué. Mais pour 10 % des événements identifiés les recherches fondamentales continuèrent après l'apparition de l'innovation. Le rapport suggère un cycle de mûrissement des événements liés à la recherche fondamentale d'environ 30 ans. Enfin, Traces suggère que la communication entre chercheurs est un facteur essentiel dans le succès d'une innovation⁽²⁾.

(1) Le Programme Hindshigts, Le Progrès scientifique n° 120 juin 1968 pp. 43-55

(2) Traces, National Science Foundation, Illinois Institute of Technology Research Institute, SFF-C535, Washington D.C., 1968

Salomon, Jean-Jacques, Science et Politique, 1970

Kochen, Models of Scientific Output, in Elkana et al. 1977.

Ces deux études portent des conclusions contradictoires quant au rôle de la recherche fondamentale. D'ailleurs Traces fut conçu pour répondre au rapport Hindsight. Hindsight met l'accent sur la science ancienne, la base des connaissances acquises et offre à l'enseignement un rôle moteur ; Traces met l'accent sur les recherches nouvelles et offre à la communication entre chercheurs un rôle moteur. Cependant, ces deux études ne se contredisent pas. Elles montrent que "du point de vue de la méthodologie sociologique, les résultats d'une enquête tendent à se conformer à l'attente et donc aux intérêts de l'institution qui l'a patronnée" (Salomon, 1970, p. 149). Elles montrent toujours, du point de vue sociologique, que ceux qui évaluent, qui sont interrogés, et la raison pour laquelle ils évaluent est tout aussi importante que la matière évaluée. Kochen (1978) disait à ce propos que "si des mesures ou évaluations doivent être utilisées pour la planification, connaître ceux qui évaluent, qui contrôlent les indicateurs et ceux qui les utilisent, est plus importante que la façon d'évaluer ou de mesurer" (p. 103).

Dans le cas des études rétrospectives, on aurait eu beaucoup plus d'informations si on avait connu de manière précise l'environnement de recherche, la composition sociale du groupe d'experts qui ont procédé à la catégorisation des événements critiques. Les travaux sociologiques ont pu montrer que les scientifiques universitaires ont un comportement très différent des ingénieurs et chercheurs travaillant dans un centre de recherche industriel ou un centre de recherche orientée.

Les études rétrospectives mentionnées ci-dessus montrent que les délais séparant la recherche de l'application peuvent servir de mesure de la science. Cependant "les délais d'utilisation (de la science) ne sont pas tant fonction de l'attention

qu'on lui prête en tant que source de savoir nouveau que des besoins qu'elle peut satisfaire en tant que technique réalisée" (Salomon, 1970, op. cit. p. 151). Mais là encore le besoin est notion maléable à souhait et difficile à appréhender.

Cependant, les études rétrospectives ont permis d'engager une réflexion sur les indicateurs de la science, notamment sur les produits de la science. Elles ont largement stimulé les travaux de bibliométrie et les travaux sur la filiation de la recherche avec la technologie. C'est ainsi que Kochen (1978, op. cit. p. 130) proposait d'établir des indicateurs permettant de suivre la filière intellectuelle d'un produit, retracer les antécédents intellectuels d'une technologie. C'est à ce souci que répondent les indicateurs d'output élaborés à partir des décomptes de publications, de citations des brevets, des techniques de co-citations ou de co-occurrences de mots.

Plus récemment, une autre étude rétrospective fut menée pendant six ans, par Julius Comroe et Robert Dripps (1), financée par le National Heart, Lung, and Blood Institute et du Burroughs-Wellcome Fund. Elle porte sur le domaine des maladies cardiovasculaires et pulmonaires (causes de 50 % de la mortalité aux U.S.A.). Les auteurs ont cherché à identifier les avancées cliniques dues à la recherche.

Une liste de découvertes importantes quant au diagnostic, à la prévention et à la guérison des maladies fut établie

- (1) Une présentation complète de l'étude est l'article de Julius H. Comroe et Robert D. Dripps, Scientific Basis for the Support of Biomedical Science, in T.R. Roberts, R.I. Levy, S.N. Finkelstein et E.J. Sonolick (eds), Biomedical Innovation, MIT Press, 1981, pp. 101-122

par 40 médecins et une centaine de spécialistes des maladies cardiovasculaires et pulmonaires. L'étude rétrospective porte sur les 10 "avancées" les plus importantes plébiscitées par ces experts. A l'aide de 140 experts (médecins, scientifiques, ingénieurs, administrateurs de la science et même auteurs de vulgarisations scientifiques), le "corps des connaissances" précédant ces 10 avancées fut identifié (137 ensembles de connaissance). La littérature scientifique concernant ces 137 ensembles de connaissance (listes chronologiques) fut examinée. Plus de 500 articles clés ou rapports-clés furent sélectionnés parmi 4 000 articles et rapports importants. Ces articles clés sont sélectionnés selon 5 critères assez stricts qui ne correspondent pas nécessairement à ceux, implicites, de la communauté scientifique.

Les auteurs ont vérifié la validité de leur sélection en demandant à des spécialistes d'établir une sélection d'études et articles selon les mêmes critères à partir de 42 tables chronologiques d'événements ou connaissances importantes. Les listes des auteurs et des spécialistes correspondent à peu près. Par contre les 2 classifications de ces articles en recherche clinique orientée et en recherche non-orientée, correspondent presque parfaitement.

L'ensemble des articles-clés sélectionnés ont par la suite été classés dans l'une des 6 catégories suivantes :

1. Recherche fondamentale non liée à la solution d'un problème clinique,
2. Recherche fondamentale liée à la solution d'un problème clinique,
3. Etudes non concernées par les mécanismes biologiques, chimiques ou physiques (études descriptives, observation d'un phénomène sans effectuer de recherche, tests cliniques, amélioration d'un matériel chirurgical),

4. Revues de la littérature, analyses critiques et synthèses aboutissant à de nouveaux concepts,
5. Développement de techniques ou d'appareils pour créer ou perfectionner des instruments de recherche,
6. Développement de techniques ou appareils nécessaires au diagnostic ou aux soins de malades.

Les résultats de cette classification sont les suivants : 350 articles parmi les 529 articles se trouvent dans la catégorie recherche (1 ou 2 de la classification), soit 61,7 % des articles classés (un article peut appartenir à plus d'une catégorie). Les études non concernées par les mécanismes de la maladie, souvent des études "inévitables", concernent 21 % de la littérature en question. Les travaux de développement (catégories 5 et 6) représentent 15,3 % et les articles de synthèse moins de 2 %.

L'étude de Comroe et Dripps montre donc l'importance de la recherche fondamentale dans les avancées médicales (et non scientifiques). D'une certaine façon elle vient conforter le projet Traces de la NSF tout en n'ayant pas les défauts soulignés ci-dessus. D'une part, l'institut qui a financé cette recherche est un organisme effectuant des recherches orientées (le National Cancer Institute). Du point de vue des hommes politiques, du législatif et de l'administration de cet institut, le résultat inverse eut été bienvenu. D'autre part, seuls 25 % environ des experts et spécialistes interrogés étaient des scientifiques de recherche fondamentale. Il est donc remarquable que les jugements de 75 % d'experts non engagés dans la recherche fondamentale aient tout de même abouti à ce résultat. Enfin, les auteurs ont mis beaucoup d'attention à ne tenir compte que des déclarations de chercheurs dans les articles analysés au moment où fut publié l'article.

Ainsi, les réorientations possibles par une reconstruction historique ("mes travaux de développement étaient en réalité de la recherche fondamentale") ont été évitées.

Des trois études rétrospectives, celle de Comroe et Dripps est probablement la plus significative. Cependant, elle se heurte à la limite inhérente à ce type de travaux : son caractère rétrospectif.

CONCLUSION

Peut-être faut-il finir ce chapitre sur les procédures d'évaluation au sein des agences et organismes de recherche en rappelant sous une forme condensée les différentes procédures, modalités d'évaluations et objets de l'évaluation.

Tout d'abord rappelons que nous ne nous sommes pas occupés de l'évaluation des chercheurs individuels. Nous avons remarqué que les agences fédérales américaines n'évaluent jamais les chercheurs ; ce sont les programmes, les laboratoires, les centres de recherche ou encore les unités institutionnelles définies administrativement qui font l'objet de l'évaluation à priori et à postériori.

L'évaluation des chercheurs individuels se fait par l'embaucheur, directeur de laboratoire, université, professeur. Elle n'est donc pas effectuée par les agences qui soutiennent la recherche, mais par ceux qui emploient directement les services de chercheurs engagés. Là réside la principale différence avec l'évaluation de programmes scolaires par exemple. Dans ce cas en effet, l'évaluation concerne autant l'élève que le professeur. Dans le cas de l'évaluation de la recherche, ce sont les unités de recherche définies qui sont évaluées, par les chercheurs qui y sont rattachés.

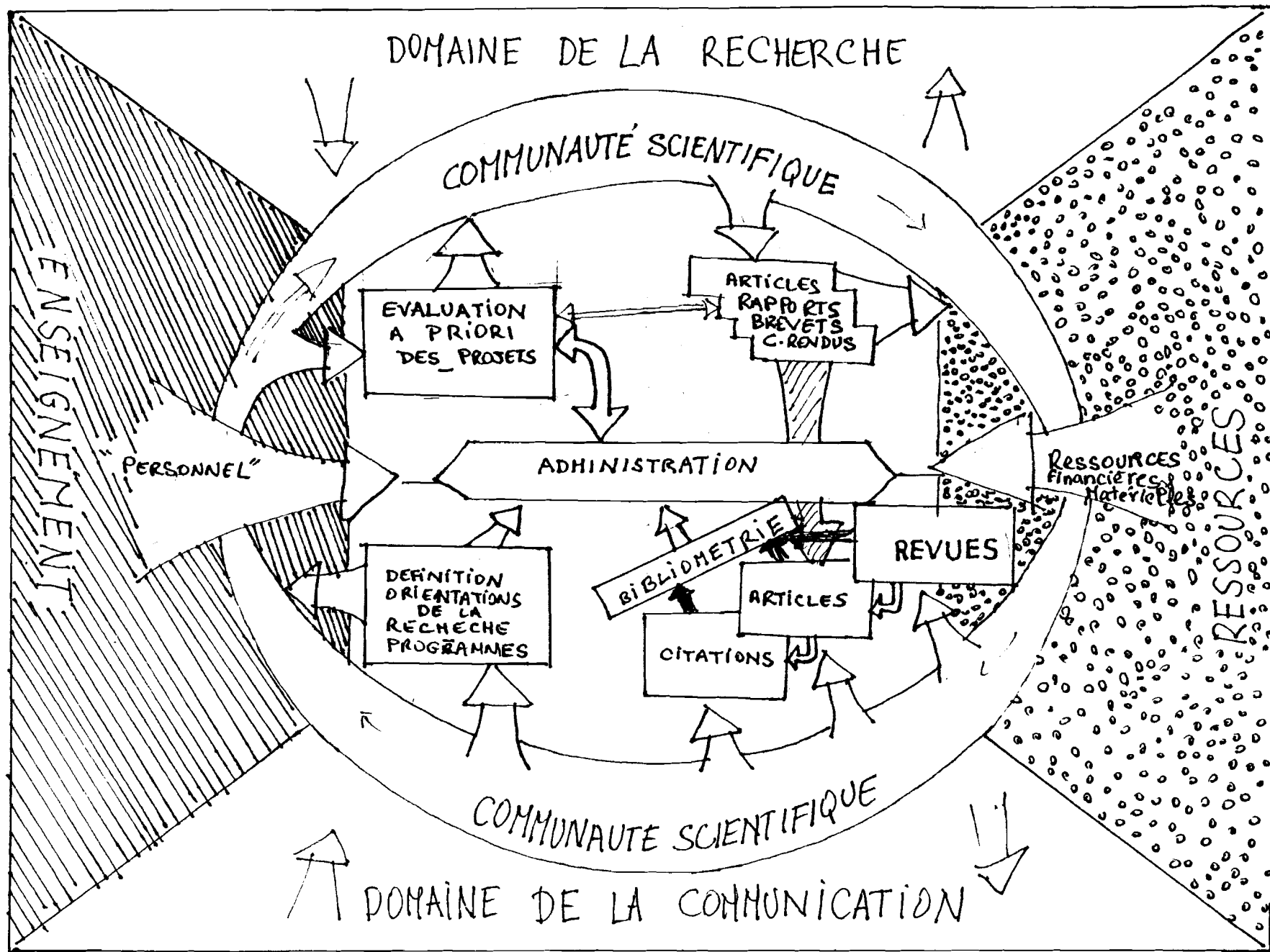
On peut distinguer deux types d'évaluations. L'évaluation par les pairs, collègues, ou supérieurs hiérarchiques que l'on peut appeler évaluation subjective, et l'évaluation quantitative "objective", ou discrète comme disent les chercheurs en organisation⁽¹⁾.

 (1) Unobstrusive Measures. Une bonne discussion de cette notion se trouve dans Eugène Welb et Karl E. Weick, Unobstrusive Measures in Organizational theory, Administrative Science Quarterly, décembre 1979, pp. 650-659, notamment 651-653

Dans les agences fédérales américaines la plupart des évaluations, qu'elles soient à priori ou a posteriori, sont des évaluations par les pairs. Ces évaluations sont numériques et le chiffrage s'effectue au moyen de notations, des projets ou programmes ou publications, ou tout autre objet évalué. L'évaluation hiérarchique, que l'on rencontre très souvent dans les centres de R et D, de recherche orientée ou dans des centres de recherche militaires et industriels ne s'effectuent pas dans les centres de recherche fondamentale civile. Nous avons essayé de montrer dans les pages qui précèdent que l'évaluation par les pairs a été massivement plébiscitée par les scientifiques, car elle permettait d'institutionner le fonctionnement implicite de la communauté scientifique : les rôles, nous avons dit, ont pu ainsi être transformés en statuts.

Nous proposons une présentation graphique du fonctionnement de l'évaluation et de l'articulation des divers éléments qui interviennent dans le fonctionnement de l'évaluation.

Ce schéma permet de représenter une idée implicite mais centrale dans l'évaluation par les pairs. La communauté scientifique est ce qui permet de faire circuler les différents éléments institutionnels, les ressources financières et humaines, l'enseignement, l'information, la documentation scientifique et les autres moyens de communication (colloques, réunions) et les recherches elles-mêmes, les centres de recherches, laboratoires et autres lieux où s'effectue la recherche. Le terme de communauté scientifique est trompeur car il évoque un ensemble homogène de scientifiques, ce qui est loin d'être la réalité, mais il permet de cerner un ensemble d'hommes, d'activités, d'institutions qui disposent des mêmes moyens. Cela ne signifie nullement que tous les membres de la communauté scientifique disposent



des mêmes quantités de ressources. C'est précisément les différences dans la distribution des ressources qui rendent nécessaires les études d'évaluation. Les évaluations à posteriori furent conçues pour rendre compte des critères de la distribution des ressources pour la recherche et de leur influence. L'évaluation à priori a pour objet la définition même des critères et la distribution des ressources.

Les études d'évaluation bibliométrique sont particulières. Elles sont contraires à l'esprit de l'évaluation par les pairs non pas à cause de la quantification, mais surtout parce qu'elles détournent les produits de la recherche. Tout le monde sait qu'une des activités principales des chercheurs est d'écrire des articles, et que chaque fois qu'un collègue lit un article, il le jugera. Les études bibliométriques exploitent ce mécanisme implicite d'évaluation par les pairs "mis en place" par les scientifiques eux-mêmes.

Si les études bibliométriques n'ont de sens qu'à un niveau statistique agrégé⁽¹⁾, c'est bien parce que tous les chercheurs n'ont pas le même rôle, ni la même attitude face à ce mécanisme implicite.

Contrairement à ce que l'on dit habituellement, les statistiques portant sur les articles et les citations sont beaucoup plus simples à récolter que les statistiques d'"input". Mais il est vrai que les indicateurs bibliométriques sont loin de représenter numériquement l'ensemble de l'output de la recherche scientifique.

 (1) Les études bibliométriques portant sur les citations par exemple ne sont statistiquement fiables que si elles portent sur au moins 20 à 25 articles ou sur 2 ou 3 ans.

Mais la même remarque est valable également pour les statistiques d'input. Qui peut prétendre que les ressources en homme et en argent sont les seules nécessaires à la recherche ?⁽¹⁾.

Nous n'avons mentionné dans ces pages que les évaluations soit du contenu des recherches, soit des résultats des recherches. Comme nous l'avons déjà dit, l'évaluation est principalement préoccupée par les résultats (évaluation a posteriori) actuels ou futurs (évaluation à priori).

Les audits et les évaluations du processus de recherche (visite des laboratoires) n'ont pas été étudiés ici⁽²⁾. Cependant, aux yeux des scientifiques, ces audits, etc... sont secondaires quant au contenu de la recherche. Sous n'importe quel arrangement institutionnel, entend-t-on dire, un bon chercheur sera bon, une bonne recherche sera bonne et trouvera certainement une façon d'aboutir. Il faut toutefois remarquer que certaines études d'évaluation, comme celle des laboratoires MRL commandée par la NSF à la société MITRE, ont démarré à partir de considérations administratives : il s'agissait bien de faire un audit du système de financement des MRL. Mais comme on a pu le voir bien d'autres éléments ont dû être évalués avant de pouvoir parler de l'efficacité du système de financement.

- (1) Même les économistes s'accordent pour dire que le capital monétaire et humain n'est pas la seule ressource, par exemple dans l'évaluation. Ils contournent cependant la question de l'imperfection de l'information sur les inputs en y ajoutant l'incertitude. Par exemple, récemment W.E. Becker Jr, The Educational Process and Student Achievement Under Uncertainty in Measurement, American Economic Review, 72 (1), mars 1982, pp. 229-236.
- (2) Mais voir R. Arvanitis, L'évaluation des chercheurs aux U.S.A., C.P.E., avril 1983

Enfin il faut remarquer que dans le cas des recherches financées par la NSF et le NIH, au moyen de bourses individuelles il n'existe pas de sanction externe à celle de la communauté scientifique au moment de l'évaluation à priori d'un renouvellement de bourse ou de continuation de la bourse, ou au moment du dépôt d'une nouvelle demande. Il n'y a donc pas d'évaluation qui se fait par l'aval, c'est-à-dire d'éventuels utilisateurs des résultats.

Ce qui nous a semblé important, et que le schéma présenté ci-dessus veut mettre en évidence, c'est qu'au moment des discussions, désaccords, controverses, négociations, évaluations, présentations, la partie de la communauté scientifique intéressée par ce qui se dit fait fonctionner un flux d'informations, d'hommes, de produits, d'articles, etc...

Ceci a une conséquence mal perçue, mais extrêmement importante. C'est que toute étude d'évaluation utilisant le jugement par les pairs rentrera nécessairement dans le flux qu'organise la communauté scientifique. Il est donc illusoire de vouloir imposer à cette communauté des directives contraires à celles qui sont les siennes. Seule une discussion avec l'administration peut éventuellement faire varier ce flux d'informations qui est organisée par la communauté scientifique.

Les études strictements bibliométriques ne sont par contre rattachées au jugement de la communauté que par l'intermédiaire d'objets, qui sont médiats⁽¹⁾, et qui

 (1) Un objet est une chose existant en dehors de nous même, "médiateur du corps social" (A. Moles, Théorie des objets, éd. Universitaires, Paris 1972, p. 21), donc médiat contrairement aux jugements par les pairs qui sont immédiats et donc immanents aux juges, à la communauté scientifique.

peuvent donc être détournés de ce flux (les articles, les revues). D'où la méfiance des scientifiques et leurs reproches : "Vous ne tenez pas compte du contenu", "vous ne pouvez pas mesurer la qualité", etc... Leurs reproches sont certes justifiés : mais ils "oublient" que les études bibliométriques ne font que transcrire en des termes quantitatifs, la structure de la communication écrite au sein même de la communauté scientifique.

C'est d'ailleurs là que se trouve la principale faiblesse des études bibliométriques. Elles ne peuvent que reproduire une structure déjà existante. Là encore, les études bibliométriques quoique indépendantes des scientifiques ne peuvent pas faire varier le flux.

Mais les études bibliométriques et les autres indicateurs de science et de technologie peuvent éventuellement engager un débat entre scientifiques et administrateurs. L'exemple de la NSF et du NIH le prouve abondamment.

ANNEXE 5.1.CRITERES POUR L'EVALUATION DES PROJETS DE RECHERCHE
PAR LES PAIRS DE LA NSF

La NSF demande aux évaluateurs de donner une note générale, ou globale, pour le projet qu'ils ont à évaluer. Mais elle demande aussi des commentaires détaillés sur le projet. Ces commentaires doivent tenir compte des critères suivants :

- A. La capacité de recherche,
- B. La valeur scientifique du projet,
- C. L'utilité ou applicabilité éventuelle des résultats de la recherche,
- D. L'influence du projet sur la science américaine.

Ces quatre grandes catégories sont décomposées en 11 critères :

A. Capacité de recherche

1. Diplôme du chercheur en chef (Principal investigator), ses études, son expérience de la recherche, les résultats éventuellement des recherches précédentes, une estimation d'ensemble des dispositions à la recherche.
2. Les connaissances du PI sur le domaine de recherche dans lequel se trouve son projet, a-t-il suffisamment une connaissance suffisante des recherches dans lesquelles on peut trouver des éléments pour sa propre recherche.
3. Adéquation du matériel qui est proposé et du support technique.

B. Valeur scientifique

4. Probabilité pour que la recherche proposée apporte des résultats importants dans son propre domaine, et dans le meilleur des cas, dans d'autres domaines (par résultats on entend des découvertes théoriques, des conceptualisations nouvelles).
5. Probabilité pour que la recherche proposée débouche sur des améliorations notables d'une méthode, ainsi que la possibilité d'appliquer cette méthode à d'autres domaines de recherche.

C. Utilité de la recherche

6. Probabilité pour que la recherche soit utile à la recherche appliquée (nouvelle technologie).
7. Probabilité pour que la recherche en question permette une meilleure évaluation d'une technologie existante (estimer effets directs et indirects de la technologie existante).
8. Cette recherche permet-elle d'identifier une aire de recherche nouvelle en recherche orientée.

D. Influence à long-terme et impact du projet

9. Influence probable de la recherche en question sur la carrière universitaire et professionnelle des jeunes chercheurs qui collaborent dans ce projet.
10. Diffusion probable de résultats techniques, des normes et protocoles de recherche, du savoir-faire et de l'image du domaine ou de la discipline dans lesquels se trouve le projet.
11. Effets anticipés de la recherche sur la science américaine.

La NSF demande aux évaluateurs de tenir compte de ces différents critères, mais de donner le plus de poids aux critères n° 4, 5, 6 et 7. Les critères les moins importants étant les critères 9, 10, 11. Les commentaires des évaluateurs suivent ces critères de façon différente selon la discipline. D'autre part, la note globale qui est attribuée au projet doit tenir compte de ces critères, qui sont spécifiés en face de chaque note attribuée (Excellent, Bon, etc...).

(NSF, Criteria for the selection of research projects by the National Science foundation, NSF 74-300, Février 1975).

ANNEXE 5.2.CRITERES POUR L'EVALUATION DES PROJETS DE RECHERCHE
PAR LES PAIRS DE LA NIH

Contrairement à la NSF, les critères ne sont pas spécifiés par écrit, puisque la procédure d'évaluation du NIH est orale. Les projets sont classés selon un ordre de mérite qui doit tenir compte des critères suivants :

1. Probabilité pour que le projet apporte des connaissances nouvelles, en cas de succès de la recherche.
2. Justesse des notions utilisées, de la méthodologie proposée.
3. Capacité de recherche des chercheurs, appréciée sur la base de ses travaux précédents et de sa productivité antérieure.
4. Choix de la période de recherche, en fonction du plan de recherche et de son budget (ne pas oublier que chaque projet est comparé aux autres projets en même temps),
5. Moyens et matériels, facilités offertes dans l'institution de recherche.

(Voir NSB, Basic Research in Mission oriented Agencies, NSB 10th report, 1979, p. 279. Ces critères ont été donnés au NSB par Ann Kaufman).

TABLE DES MATIERES

Avertissement au lecteur pressé	i
Sommaire	ii
Remerciements	iii
CHAPITE 1 L'évaluation de la recherche : pratiques et études évaluatives aux Etats-Unis	
1. Objectifs et plan de cette étude	1
2. Evaluation des projets et des programmes	3
3.1. L'évaluation des projets de recherche	3
3.2. Enseignements issus de la pratique américaine de l'évaluation des projets	6
4.1. L'évaluation des programmes de recherche	9
4.2. Origine des évaluations de programmes et des travaux quantitatifs	11
4.3. Les composantes du "mouvement quantitatif"	13
a) Sociologie et histoire des sciences	13
b) Les études retrospectives	16
c) les évaluations a-posteriori	18
4.4. Quelques enseignements des études d'évaluation américaines	19
4.5. Les perspectives d'avenir en matière d'études évaluatives	21
a) Topographie de la recherche	21
b) Qualité et critères de sélection	22
c) lier la sociologie, l'évaluation de la recherche et l'économie de la R&D	23
Conclusion	24
CHAPITRE 2 L'évaluation et la sélection de projets dans les les entreprises	
RESUME	26
1. Distinction entre la recherche fondamentale et la recherche-développement	28

2. La recherche-développement	31
3. Critères d'évaluation de la recherche-développement	33
4. Critères d'évaluation de la recherche fondamentale dans les entreprises	35
5. Méthodes d'évaluation	
5.1. Méthodes non-formelles	39
5.2. Méthodes formelles et quantitatives	48
5.3. L'utilisation des méthodes formelles d'évaluation hiérarchique	55
Conclusion	69

CHAPITRE 3 La productivité des chercheurs : les résultats des travaux sociologiques

Résumé	72
1. Les problèmes de méthode	74
1.1 La "loi" de Lotka	74
1.2 Reconnaissance et productivité scientifique	76
1.3 Remarques sur la sociologie des sciences quantitative: les prémisses d'un modèle	81
1.4 La méthode de Pelz et Andrews	83
1.5 La méthode de l'enquête de l'UNESCO	85
1.6. Rapports entre mesures objectives et mesures subjectives	89
2. Les résultats de ces travaux	91
2.1. Les performances individuelles et la liberté des chercheurs	91
2.2. Les performances des équipes et la structure du pouvoir	94
2.3. Les performances individuelles et les caractéri- stiques des chercheurs	97
2.4. La nature des recherches et les performances	99
2.5. Les performances sont liées à d'autres activités que la seule recherche	100

2.6. Les performances en fonction de l'âge	101
2.7. Les performances des groupes de recherche	104
3. Les études de cas	108
3.1. Productivité et perception dans un laboratoire de recherche militaire	108
3.2. Etudes de cas non-quantitatives	111
Conclusion	117
CHAPITRE 4 L'évaluation de la science et la politique scientifique	
Résumé	119
1. Evaluer et coordonner	121
2. L'expertise de la NSF	123
3. Les documents d'analyse issus de la NSF	124
4. Le rapport annuel au Congrès	127
5. Le Five Year Outlook : rapport participatif	130
6. Les "Science Indicators"	131
a) Naissance des S.I.	132
b) Le rôle des S.I.	133
c) les qualités de l'unité S.I. à la NSF	134
d) Les S.I. et l'évaluation	135
e) Les mesures de la science contenues dans S.I.	137
f) Les indicateurs bibliométriques de S.I.	139
g) les indices composites de technologie	147
Conclusion	150
CHAPITRE 5 L'évaluation de la recherche à la NSF et au NIH	
Résumé	152
Introduction	154
I. L'évaluation de la recherche à la NSF	157
1. L'évaluation a priori des projets	158
1.1. Le "Mail Review"	160

1.2. Le débat sur l'évaluation par les pairs	164
2. Les "Program Reviews"	168
3. Les "Status of Science Reviews"	172
4. L'évaluation a posteriori à la NSF	175
4.1. Evaluation du programme MRL de la NSF	179
a) Capacité de recherche des laboratoires MRL	184
b) Evaluation des types de recherche	190
c) Cohérence des recherches	191
d) Jugements d'experts à partir des publications	191
e) Analyses bibliométriques	195
f) Production des laboratoires	196
4.2. Evaluation a posteriori du programme d'océanographie	197
4.3. Le projet d'évaluation a posteriori du programme de chimie	202
4.4. Etudes sur l'équité des décisions de financement à la NSF	204
4.5. Les évaluations a posteriori : un outil fort contesté	205
II. L'évaluation de la recherche au NIH	210
1. L'évaluation apriori au NIH	210
2. Comparaison du mécanisme NSF et NIH	213
3. L'évaluation des programmes NIH	216
a) Les évaluations des programmes	216
b) Evaluation de l'efficacité des programmes	217
c) études sur "l'état de la question"	218
d) Evaluation des programmes éducatifs	218
e) Travaux méthodologiques; f) données statistiques	218
4. Indicateurs bibliométriques	219
III. Autres exemples d'évaluations a posteriori	229
1. Analyse bibliométrique des laboratoires de la marine	229
2. Les travaux de Irvine et Martin	233
3. Les études retrospectives	236

258.

Conclusion

243

Annexes:

5.1 Critères pour l'évaluation de la recherche à la NSF 250

5.2 Critères pour l'évaluation des projets de recherche
par les pairs au NIH 253

Table des matières

254

°

°

°