
L'évaluation en France : un enjeu récent

Éléments pour une sociologie de l'évaluation des politiques publiques en France

Jean-Claude Barbier*

■ Introduction

La sociologie et la science politique en France ne se sont pas particulièrement intéressées aux activités des acteurs qui se réclament *explicitement* de l'évaluation des politiques et des programmes publics (pour simplifier la lecture, on dira dans la suite de ce texte EPP ou *évaluation* tout court¹). Le présent article, disons-le d'emblée, ne prétend pas faire une revue raisonnée des travaux sociologiques ou politistes depuis les années 1980, qui reste à conduire². Très peu de thèses traitent explicitement de l'évaluation, et elles sont éclatées entre économie, sociologie, science politique. Il n'est pas possible ici non plus, faute de données empiriques suffisantes³, de présenter une analyse détaillée des pratiques, de la « profession », des commanditaires et des enjeux financiers du « secteur ». Comme l'histoire de l'évaluation en tant que pratique est relativement peu ancienne⁴ dans notre pays (Monnier, 1992 ; Perret, 2001), il est assez commode, dans une première partie, d'en retracer les jalons qui paraissent essentiels. En second lieu, on s'efforcera de souligner, en comparaison avec d'autres pays d'Europe, ce qu'on peut considérer comme des traits spécifiques de la situation française, dont certains sont dotés de stabilité. Cela suppose de ne pas se limiter à l'examen du champ relativement étroit de l'EPP au sens strict, car l'évaluation doit

* Directeur de recherche CNRS, UMR CES Université Paris 1.

1. Le mot d'évaluation est d'une polysémie extrême ; il est largement employé en éducation et dans le langage courant du management.

2. Par exemple, dans la *Revue française de science politique*, à notre connaissance, le premier article consacré à l'évaluation est celui de J.-P. Nioche (1982). Dans les années 1990, on note l'article de P. Duran et E. Monnier (1992). Dans les années 2000, l'article de S. Jacob (2005b). Dans son ouvrage réédité en 1998, P. Muller a une section sur l'évaluation des politiques publiques (Muller, 1998, p. 117-123). La revue *Sociologie du travail* a publié avec régularité des articles sur l'évaluation depuis les années 1990 (voir ainsi, Warin, 1993 ; Spenlehauer et Warin, 2000). Dans le numéro spécial de la *Revue française de science politique* de 2005, qui comporte un bilan de l'analyse des politiques publiques, la question de l'EPP n'est pas directement abordée.

3. Les données empiriques sur lesquelles se fonde la présente analyse, ressortissent à une observation participante (ou participative) de l'auteur sur une longue période. Celui-ci a occupé, depuis 1988 et l'évaluation du RMI, à peu près constamment (et alternativement) des positions de chargé d'évaluation, de commanditaire, de membre d'instances d'évaluation. Il a également contribué à la fondation de la société française de l'évaluation (SFE) en 1999. Il a, en particulier, réalisé une série d'entretiens significatifs avec des consultants dans plusieurs cabinets offreurs de prestations d'évaluation au début des années 2000.

4. On notera en passant que l'emploi du terme « évaluation » pour l'appliquer à des programmes et des politiques, est présent dès les années 1970 en France, comme le montre l'article pionnier de J.-P. Nioche (1982) et son enquête sur les pratiques.

être replacée dans le cadre plus général de l'action de l'État (Bezes, 2009). Dans une troisième partie plus prospective, on note que plusieurs éléments intervenus récemment pourraient se combiner pour modifier la situation française marquée jusqu'à présent par une faible institutionnalisation de l'évaluation, illustrée par l'absence d'une profession d'évaluateur reconnue et par la marginalité de la « discipline » dans l'enseignement supérieur.

■ Une « brève histoire » de l'évaluation des politiques publiques en France

Les grandes étapes de l'introduction de l'évaluation jusqu'aux années 2000

Précisons d'abord des repères essentiels pour une *sociologie politique de l'évaluation* (qui reste, pour l'essentiel, à faire)¹, avant de nous centrer sur les références cognitives de l'évaluation. Par comparaison avec d'autres pays européens, et surtout avec le cas américain, l'apparition de l'évaluation en France paraît tardive : à la fin des années 1970 et, surtout, dans les années 1990, elle s'incarne pour la première fois dans des institutions formelles de l'État (central et local) ; parallèlement, les prolégomènes de la naissance d'une identité professionnelle d'« évaluateur », certes fragile, apparaissent ; un marché de l'évaluation se constitue progressivement, dont les produits (ce sont évidemment des services professionnels complexes) se distinguent difficilement d'autres, la recherche, l'expertise, l'étude, l'audit, le contrôle, le management, etc. La perspective sociologique ou de science politique dominante est de considérer ces diverses activités ensemble, comme « expertise », quelle que soit l'éventuelle revendication des praticiens engagés dans l'évaluation qui font valoir, de manière diverse, des références professionnelles distinctes. L'influence de la problématique introduite dans le débat public par la référence de l'EPP varie, en outre, selon les secteurs de politiques publiques : elle est plus récente dans certains que dans d'autres. Il est cependant logique que ce soit dans le domaine des politiques sociales, entendues au sens large², que cette influence soit la plus évidente. Au cours de la décennie 1990 et dans les premières années 2000, les formes prises par l'évaluation ont évidemment évolué. Un premier type d'opérations a été commandité par les financeurs publics (centraux, régionaux, locaux) visant

1. Pour des éléments historiques, on se reportera à Monnier (1992), ainsi qu'à Spenlehauer (1995, 1998).

2. Dans le vocabulaire de l'anglais américain, on parle, dans la littérature spécialisée de « *social intervention programs* ». L'un des manuels couramment utilisés par les évaluateurs professionnels, définit ainsi l'évaluation : « L'évaluation, ou la recherche évaluative (*evaluation research*), c'est l'application systématique des méthodes des sciences sociales (*social research procedures*) pour juger la conception, la construction, la mise en œuvre et l'utilité des programmes sociaux (*social intervention programs*) » (Rossi, Freeman, 1993, p. 5).

à objectiver les « impacts » (ou « effets nets ») des actions publiques, type associé souvent à un suivi (*monitoring*) des programmes et à l'analyse de la mise en œuvre, des systèmes d'acteurs et des enjeux de management public, en termes d'analyse des politiques publiques et de sociologie des organisations (Nioche, 1982). Plus récemment, en France et ailleurs, des évaluations ont mobilisé la microéconomie néoclassique et attiré l'attention¹.

Quoi qu'il en soit, au cours des années 1990, un nombre croissant d'acteurs français divers (hauts fonctionnaires, experts, consultants, universitaires...) ont choisi de faire référence à l'évaluation des politiques et des programmes pour qualifier certaines de leurs activités. Progressivement, bien que de manière très inégale selon les secteurs et selon les lieux, l'EPP a commencé d'être repérable parmi les activités spécialisées liées à la mise en œuvre et à la conception des politiques. Cela s'est manifesté par un processus d'*institutionnalisation*², visible, notamment, dans le fait que des administrations de l'État l'ont eue en charge. Dans le secteur social, qui concerne plus spécifiquement la *Revue française des affaires sociales*, plusieurs types d'usages de l'évaluation ont pris place dans les pratiques. Ainsi, au milieu de controverses qui n'ont pas cessé depuis, l'EPP s'est progressivement installée dans le paysage politique français dans les années 1990 (formellement, sous l'incitation de la circulaire de 1989 du Premier ministre Rocard). Bien loin d'une installation simple, cette période connaît plutôt un ensemble assez disparate d'initiatives (dont certaines étatiques) menées par une pluralité d'acteurs. Dès cette époque, commencent des controverses autour des conceptions, opposant les tenants d'une évaluation « démocratique » à une « évaluation managériale », ainsi que les partisans de l'évaluation « pluraliste » à ses adversaires.

La première controverse politique se présenta à cette époque comme l'opposition entre une conception dite « démocratique » de l'évaluation, et une conception dite « managériale ». L'exemple des réformes introduites aux États-Unis sous la présidence de R. Reagan montre que l'orientation partisane joue partout un rôle dans le recours à l'évaluation des politiques : sous sa présidence, en effet, le style d'évaluation fut réorienté vers des études centrées sur les économies budgétaires, par opposition à l'orientation de « *social engineering* » de l'administration du président Johnson (Perret, 2001, p. 62). Une tension, en partie fonctionnellement analogue, explique l'opposition thématifiée en France entre « management » et « démocratie » dans les années 1990³. En termes analytiques, il est bien sûr impossible de préciser la nature exacte d'évaluations qu'on pourrait dire « managériales » (ou aidant

1. Il s'agit de travaux qui renouent avec une approche expérimentale très ancienne, même s'ils tendent à se présenter comme entièrement nouveaux. Voir sur ce point, Nioche (1982) qui en note les limites. Pour le cas américain contemporain, voir Weiss *et al.*, (2008).

2. E. Monnier (1992, p. 63) relate « la voie française de l'institutionnalisation ». S. Jacob (2005a et b) a étudié l'institutionnalisation de l'évaluation dans quatre pays. Pour lui, c'est un « processus par lequel les dispositifs institutionnels sont créés, modifiés voire supprimés », de façon à chaque fois spécifique.

3. La sociologie politique des partisans de l'évaluation en France reste, de ce point de vue, à faire.

à la décision) par rapport à d'autres qui seraient « démocratiques » (voire, viseraient à favoriser la démocratie participative). Chacun de leur côté, leurs partisans considéraient qu'il s'agissait pourtant de la meilleure forme d'évaluation possible. Cette opposition prend un sens particulier dans la culture politique française qui tend à penser le « management » ou la gestion, comme par principe, étrangers à une mission supérieure de la politique qui ne saurait s'embarrasser de questions prosaïques, parce qu'elle tire sa valeur d'une orientation à laquelle des hommes et femmes politiques se réfèrent, qualifiée dans le vocabulaire français, de « volontariste ». Des citations de tous bords pourraient être convoquées pour illustrer ce point ; une, typique, suffira, celle de R. Forni, alors Président de l'Assemblée nationale, écrivant dans *Le Monde* (6 juin 2001) : « Faire de la politique, même dans l'opposition, ne saurait se résumer à mettre en doute des chiffres et des circuits de financement, à se transformer en une petite chambre des comptes ». Dans les termes de la polémique politique, l'évaluation « managériale » fut de droite, alors qu'elle s'appelait « aide à la décision » à l'autre bord, et l'évaluation « démocratique » fut inévitablement de gauche pour ses partisans, alors que ses adversaires tenaient pour la véritable évaluation démocratique celle qui informerait l'électeur sur l'efficacité et les coûts des politiques (*tax payer's money*). À cette période, l'approche de l'évaluation dite « pluraliste¹ » fut également l'objet de controverses entre praticiens de l'évaluation, controverses qui n'ont pas cessé. À sa définition élaborée progressivement, plusieurs auteurs et institutions ont participé ; E. Monnier lui consacre de longs développements dans son ouvrage pionnier (1987/1992) : sa définition tisse ensemble le *constat analytique* selon lequel le déploiement d'une politique publique est toujours le résultat d'interactions multiples entre des acteurs sociaux, avec l'*appréciation normative* selon laquelle la meilleure évaluation serait pluraliste, car, négociée et associant les partenaires pertinents d'une politique publique, celle-ci permettrait une meilleure efficacité et légitimité des opérations d'évaluation (*idem*, p. 123-124).

Comme elle bénéficia d'une reconnaissance auprès du Premier ministre (Conseil scientifique de l'évaluation, notamment) l'évaluation étendit son audience. Elle connut une visibilité importante dans l'évaluation de la loi expérimentale du RMI (1988-1992) et dans la discussion sur le rapport commandé à P. Viveret (1989). Cette « tradition » d'évaluation était mélangée, autant du point de vue des disciplines, des méthodes d'évaluation, des commanditaires, que des pratiques professionnelles et des acteurs. À la fin des années 1980 et dans les années immédiates qui ont suivi, il a, ainsi, pu sembler aux observateurs et aux acteurs qu'un pas décisif avait été franchi dans l'institutionnalisation de l'évaluation en France, avec la création, en 1990, du Conseil scientifique de l'évaluation (CSE). Dans son activité de surveillance et de suivi des évaluations (baptisées « interministérielles ») commanditées par les grands ministères, le CSE contribua à produire les principes d'une

1. Présentée à des lecteurs internationaux, elle a été qualifiée d'évaluation « à la française » par Duran *et al.*, (1995) ; voir aussi Toulemonde (2000, p. 356).

doctrine (Barbier, 2004). Bien qu'il n'ait jamais réuni les ressources adéquates pour asseoir une influence prépondérante, il n'en constitua pas moins un référent majeur. Le CSE privilégia souvent le choix de l'installation d'une instance pluraliste, représentant les divers points de vue qu'il qualifiait de « légitimes » à propos d'une politique, pour lui, la meilleure *stratégie cognitive*¹. Une telle instance est porteuse de légitimation pour l'évaluation, ont remarqué des sociologues, d'un point de vue à la fois politique et scientifique (Lascoumes, Setbon, 1996, p. 10-11). Cette idée est reprise par B. Perret (2001, p. 13-15 et 99-100)². Plus tard, la Société française de l'évaluation a également, après de longs débats en son sein, repris cette préférence pour l'évaluation pluraliste (*cf.* plus loin)³.

Aussi paradoxal que cela puisse paraître d'un point de vue comparatif, la France, qui n'avait pas d'organisation professionnelle de l'évaluation (ni enseignement supérieur, ni organismes professionnels) disposa, *via* cette doctrine, d'une série de « standards de qualité » qui furent effectivement utilisés pour juger les produits des évaluations coordonnées sous son égide. Ensuite, le CSE est tombé en désuétude, puis a été remplacé, en 1998, par une instance différente (le Conseil national de l'évaluation, CNE). La visibilité institutionnelle de l'organisme central de référence en matière d'évaluation en est ressortie amoindrie. Pendant que, de son côté et sans coordination de fait, le marché se développait, en particulier à travers les commandes croissantes des collectivités territoriales, le CNE tomba finalement lui aussi en sommeil et ne fut jamais renouvelé. À partir de 2002, une autre dynamique de la « modernisation » de l'État lui succéda, nous allons y revenir pour montrer qu'il faut insérer, pour vraiment la comprendre, l'apparition de l'évaluation des politiques publiques dans une dynamique beaucoup plus longue de la réforme de l'État (Bezes, 2009). Mais auparavant, il faut entrer

1. « Pour toutes les évaluations qui concernent des politiques complexes ou sensibles, et qui comportent l'utilisation d'informations et de résultats d'études de nature hétérogène, le Conseil scientifique préconise la mise en place d'une instance d'évaluation. Son rôle est de préciser le questionnement, de piloter les études, d'intégrer leurs résultats et de formuler des conclusions et des recommandations éventuelles qui peuvent être tirées de ces résultats. Sa composition doit tenir compte de la pluralité des points de vue légitimes sur les questions posées, pour des raisons qui tiennent plus à un souci d'efficacité que d'une exigence formelle de représentativité [...]. Les formes de ce pluralisme sont à inventer dans chaque situation, compte tenu de la nécessité de faire de l'instance une véritable structure de travail » (CSE, 1992, p. 14).

2. Perret (CSE, 1993, p. 67) a aussi noté : « Pour que ce travail aboutisse à des conclusions crédibles aux yeux du commanditaire politique de l'évaluation, il doit être mené selon des règles qui lui confèrent une double légitimité, scientifique et politique [...]. Dans le contexte de l'évaluation de programme, l'évaluateur professionnel assume souvent seul le poids de cette légitimation, à charge pour lui de mobiliser une information et de mettre en œuvre des méthodes crédibles aux yeux des utilisateurs de l'évaluation. Dans le cas des évaluations de politiques publiques telles qu'elles sont pratiquées en France, cette double légitimité est assurée par une instance d'évaluation qui précise le questionnement évaluatif, pilote les études, synthétise et interprète leurs résultats. »

3. Le premier principe de sa charte s'énonce ainsi : « L'évaluation s'inscrit dans la triple logique du management public, de la démocratie et du débat scientifique. Elle prend en compte de façon raisonnée les différents intérêts en présence et recueille la diversité des points de vue pertinents sur l'action évaluée, qu'ils émanent d'acteurs, d'experts, ou de toute autre personne concernée. Cette prise en compte de la pluralité des points de vue se traduit – chaque fois que possible – par l'association des différentes parties prenantes concernées par l'action publique ou par tout autre moyen approprié. »

plus précisément dans les origines de la démarche évaluative et au développement historique de ses conceptions dans notre pays.

De l'inspiration intellectuelle américaine à l'évaluation dans les collectivités territoriales

Pour schématiser les choses, l'activité baptisée EPP en français dans les années 1990 puise son inspiration dans la pratique américaine de *program evaluation*¹, particulièrement florissante au moment des programmes sociaux sous les présidences Kennedy et, surtout, Johnson. Elle plonge ses racines intellectuelles bien plus anciennes, aux USA et ailleurs, dans les pratiques de « *social engineering* » des débuts du XX^e siècle, en particulier dans les expérimentations dans le domaine de l'éducation. En France, une longue tradition d'ingénieurs économistes-statisticiens peut également être raccrochée à la généalogie de l'évaluation telle qu'elle se présente, au début des années 1970, sous un avatar dit « rationalisation des choix budgétaires » (RCB). Cette initiative pilotée par le ministère des Finances se rattache à l'EPP en ce qu'elle se fixe comme objectif de « rationaliser » les politiques par le moyen de leur fixer des objectifs quantifiables, dans l'attente de résultats escomptés, pour fonder des choix reliés à la procédure budgétaire. Des correspondants RCB furent installés dans tous les ministères et des procédures et documents nouveaux créés. Pourtant l'expérience échoua, et la RCB fut abandonnée finalement en 1984. Si l'on a été jusqu'à prétendre que « le principe de l'évaluation de l'action publique, au-delà de son nom, est pour ainsi dire aussi vieux que l'État » (Spenehauer, 1995, p. 24), on ne peut ignorer la filiation américaine des idées en matière de management public et le caractère différemment construit, différemment argumenté sur le plan technique, donc innovant par rapport à ce qui existait auparavant, des activités qui se présentèrent en France sous l'étiquette de la RCB.

En second lieu, la publication, en 1986, d'un rapport du Commissariat général du Plan, (Deleau *et al.*, 1986), inscrit l'évaluation dans la réflexion internationale de l'époque en termes de « crise de l'État-providence », ou de « limites du *welfare state* » (Duran *et al.*, 1995, p. 47) ; son orientation centrale défendait la nécessité d'évaluer les politiques françaises du point de vue de l'efficacité ou de l'efficience, vis-à-vis des coûts (coût-bénéfice, coût-avantages). Cela lui fut reproché par d'autres partisans de l'évaluation en France. L'impulsion donnée par le gouvernement de gauche, sous le Premier ministre Rocard, en 1989² fut, au moins en partie, différente, sensible qu'il était à l'argumentation du rapport Viveret, commandité par le Premier ministre. L'évaluation des politiques publiques y était vue comme l'un des éléments de la « modernisation de l'État », à côté de la transformation de la gestion du personnel et

1. En anglais : *programme evaluation*. *Evaluation research* est aussi le nom parfois utilisé pour la réflexion sur l'activité (Rossi, Freeman, 1993).

2. Circulaire du 23 février 1989.

la mise en place de centres de responsabilité, incitant à s'inspirer des pratiques de management privé (CSE, 1993). C. Fontaine et E. Monnier (2002) ont montré que cette reconnaissance « légale » de l'évaluation prit place à l'occasion de l'ouverture d'une fenêtre d'opportunité exceptionnelle permise par l'arrivée au pouvoir de M. Rocard, conseillé par des partisans de l'évaluation. Alors que la pratique et la légitimité de l'évaluation portant explicitement ce nom restaient marginales en France, le CSE fut installé, auprès du Premier ministre et de son Commissariat général du Plan. En théorie, le CSE fonctionnait comme auxiliaire scientifique du Comité interministériel de l'évaluation (CIME), créé par le même décret de 1990, à la suite du rapport Viveret¹. Le CIME n'était, bien sûr, pas en charge de l'ensemble des évaluations de politiques conduites : certaines politiques furent considérées par cet organe comme particulièrement dignes d'intérêt pour être évaluées selon une procédure particulière, soumise au contrôle méthodologique du CSE. Pour autant, des activités qualifiées ici ou là d'évaluation n'en continuèrent pas moins à se développer, à la frontière d'autres activités plus traditionnellement établies (inspection, étude) et sous l'égide d'un grand nombre de commanditaires non affectés directement par l'installation d'une « autorité centrale » de l'évaluation qui, faute de ressources et de statut, ne pouvait pas en devenir une, à la différence d'une agence comme le *Government Accountability Office* (GAO) américain². S'il faut en croire sa composition initiale, le CSE fut, dans un premier temps, considéré comme un enjeu relativement important par la haute fonction publique : de hauts fonctionnaires des grands corps d'inspection et de l'INSEE, de la Cour des comptes côtoyaient, en nombre à peu près équivalent, des universitaires et experts, et un seul représentant du secteur privé au départ. Le Conseil fut présidé par Jean Leca, professeur internationalement renommé de science politique. Cette composition peut être interprétée comme le signe que la haute fonction publique chargée du contrôle, des études et de l'inspection s'intéressait à l'évaluation en tant que source potentielle d'activités concurrentes. Du point de vue des conceptions présentes dans le débat français, cela donne un poids important à cette « tradition intellectuelle », qui, assez vite, perdit de son allant.

En fait, après des débuts actifs, la procédure de sélection de projets par le CIME s'embourba et il n'y eut plus, à partir de 1995, de nouveaux projets sélectionnés ; dans le même temps, les membres du CSE qui le quittaient, ne furent pas remplacés par les autorités compétentes. Au total, pendant la période 1990-1996, le dispositif CSE-CIME n'aura choisi et suivi qu'une vingtaine d'évaluations, dont la plupart étaient, certes, de grande envergure et pleines de leçons (Barbier, 2004), mais ne représentaient qu'une petite partie

1. Les préconisations du rapport Viveret, en 1989, incitaient à s'inspirer de l'exemple américain en créant une autorité indépendante, mais l'arbitrage gouvernemental ne les suivit pas.

2. Nouveau nom du *General Accounting Office*, depuis 2004.

de l'activité existant sous l'étiquette «évaluation» en France à l'époque¹. En effet, dans le même temps, le rôle des conseils régionaux comme commanditaires d'évaluations se renforçait, sous l'impulsion du cofinancement, par la Commission européenne, des actions relevant des fonds structurels (notamment le FSE et le FEDER). Une condition obligatoire du financement communautaire étant l'existence d'évaluation, les régions y trouvèrent une puissante incitation à s'y convertir. Dans un premier temps (et selon les régions par la suite aussi) une part de ces évaluations qu'il est difficile d'estimer faute de données empiriques précises, correspondait à des évaluations à caractère rituel ou formel². Une autre source d'opérations évaluatives fut fournie par la création, en 1993, de l'évaluation obligatoire des «contrats de plan État-régions³». En dépit de leur importance à de nombreux égards, on a pu considérer que, du point de vue d'ensemble de la position de l'EPP en France, ces opérations restaient modestes (Spenehauer, Warin, 2000).

Mais, du point de vue de l'influence, il faut observer que l'évaluation n'a pas atteint, à la fin des années 1990, l'audience qu'elle semblait espérer. Cette perte d'influence est très visible depuis 2002, le gouvernement n'ayant pas renouvelé le Conseil, l'État ne disposant plus d'instance «centrale». Tout se passe presque comme si l'évaluation était un parmi plusieurs «réseaux d'expertise» (Bezes, 2009, p. 267) qui se disputent l'influence dans le domaine de la réforme de l'État. Or, les milieux dirigeants ont favorisé des démarches qu'on peut, d'un point de vue technique, considérer comme «voisines» de l'EPP mais qui l'ont, en quelque sorte, «remplacée» ou assignée à des tâches spécifiques. Tout d'abord, les gouvernements ont conduit, dans les différents départements ministériels, des «stratégies ministérielles de réforme», à partir de 2003⁴. Ensuite, la procédure budgétaire a été entièrement refondue sous l'impulsion de l'adoption (en août 2001) puis de la mise en œuvre systématique de la loi organique relative aux lois de finances (LOLF) : cette réforme a conduit à transformer la vision des politiques publiques en une série de missions (composées de programmes et d'actions) axées autour d'objectifs et devant produire des résultats mesurés à l'aide d'un petit nombre d'indicateurs. Ses premiers initiateurs laissaient entendre que la LOLF allait satisfaire le besoin d'une EPP, ce qui n'a pas été le cas, comme le prouve en particulier la persistance structurelle du financement d'évaluations sectorielles

1. Au total, environ une vingtaine d'opérations. Ont été publiées, par exemple, l'évaluation du programme «nouveaux services – emplois jeunes» dans le champ jeunesse et sports (novembre 2001) et l'évaluation des mesures d'aide aux emplois du secteur non marchand (janvier 2002), l'évaluation de l'externalisation des services par l'ANPE (rapport «Balmory» en 2004). Le CNE, à partir de 1998, fit conduire, sous son autorité, une nouvelle série d'évaluations, relativement nombreuses, dont les résultats ne furent pas tous publiés.

2. Sans que cette information soit représentative, nos entretiens à la Commission européenne nous ont convaincu que les spécialistes de l'évaluation localisés dans la direction du Budget portent un regard très critique sur la qualité des évaluations conduites par les États membres dans le cadre des financements des fonds structurels, par comparaison avec les standards qu'ils appliquent aux politiques «internes».

3. Circulaire du 9 décembre 1993 «relative à la mise en œuvre de la démarche d'évaluation dans les procédures contractuelles (contrats de plan – contrats de ville)».

4. Circulaire du 25 juin 2003 du Premier ministre relative aux stratégies ministérielles de réforme.

dans de nombreux départements ministériels. Pourtant, considérée avec le recul de la longue durée, la dynamique générale ainsi enclenchée est beaucoup plus puissante que celle qui s'attache à la pratique spécifique de l'évaluation (Bezes, 2009). De manière plus conjoncturelle, le gouvernement nommé par le nouveau Président de la République, en 2007, comprenait un secrétaire d'État dont l'attribution comportait explicitement « l'évaluation des politiques publiques ». Le Président français a aussi lancé une opération ponctuelle, qu'il dit inspirée de la pratique canadienne, et qu'il a nommée « révision générale des politiques publiques – RGPP ». Cette dernière, à l'origine, a largement puisé dans le vocabulaire de l'évaluation. Alors que la première phase de l'exercice a touché à son terme (en 2009), il est clair désormais qu'il n'a pas effectivement pénétré, pas plus que la LOLF, sur les terres de l'évaluation des politiques publiques, dans le sens accepté dans le milieu international des évaluateurs.

Nous reviendrons plus loin sur les plus récents changements liés à la modification de la Constitution, en été 2008, mais il faut, avant de prendre du recul par rapport à cette histoire française très schématisée, noter qu'en 1999 le « monde » français de l'évaluation s'était enrichi d'un nouvel acteur, la Société française de l'évaluation (SFE). Cette société a regroupé, au cours des dernières années, entre deux et trois cents adhérents, y compris des collectivités locales et des organismes divers. Face à la diversification et à la croissance du marché de l'évaluation, elle ne représente pas l'ensemble des acteurs sociaux concernés par l'EPP : c'est le cas notamment parce que certains secteurs de « politique sociale », qui pratiquent l'évaluation n'y sont pas – ou marginalement – représentés (ainsi, la politique d'éducation) ; mais la situation de représentativité limitée s'explique, aussi, par le fait que certains acteurs en sont absents ou y sont sous-représentés (par exemple les universitaires, les statisticiens). La SFE a, en 2003, adopté une « Charte de l'évaluation des politiques publiques et des programmes publics », actualisée en 2006, qui a pour ambition de constituer l'énoncé de lignes directrices professionnelles et éthiques pour les participants des opérations d'évaluation en France¹.

Notre trop brève présentation illustre la présence malaisément reconnue des activités dites d'évaluation en France depuis les années 1980. Dans d'autres pays, où les controverses ne sont évidemment pas absentes, la situation est souvent différente : le recours à l'évaluation apparaît plus banalisé. C'est surtout le cas dans les pays où des pratiques et références professionnelles sont largement reconnues, qui donnent crédibilité à une *existence distincte* de l'activité d'EPP, appuyée sur des communautés professionnelles relativement identifiables. Pourtant, en France, au cours des dernières années, des formations supérieures à l'évaluation sont apparues dans plusieurs masters professionnels (par exemple, un cursus spécifiquement centré sur l'EPP à l'Institut d'études politiques de Lyon et une formation « finances-évaluation »

1. La SFE est le correspondant français d'un vaste réseau de sociétés dans les différents États européens, mais aussi en Amérique et en Afrique.

à l'université de Rennes). Dans la section suivante, nous proposons de réfléchir sur ce qui explique la situation observée : on ne se met pas à la recherche ici d'une quelconque « exception française » car chaque « culture politique d'évaluation » possède ses caractéristiques particulières en Europe et aux États-Unis.

■ Traits de la situation française en comparaison internationale

Les activités portant l'étiquette de l'évaluation dans le système politico-administratif français du management public

Sans envisager ici une analyse approfondie de l'idée et des pratiques d'évaluation en France, qu'il faut replacer dans une genèse longue et des rivalités pour la représentation du monde (Jobert, 1992), on peut souligner quelques traits qui contribuent à expliquer cette situation.

Il y a, tout d'abord, des traits qui s'imposent à l'ensemble de la période de la gestion de l'État depuis l'après-guerre et d'autres qui sont encore plus enracinés dans l'histoire (cf. sur ce point Bezes, 2009, notamment p. 434-442). Au titre des caractéristiques du système politique français, il faut ainsi signaler la faiblesse du Parlement face à l'exécutif (depuis 1958), ce qui expliquait, jusqu'à la réforme constitutionnelle, son investissement marginal dans les questions de l'évaluation, malgré l'existence d'organes spécialisés¹. On peut aussi ajouter l'institution du cumul des mandats pour les hommes et femmes politiques, qui aboutit à ce que l'investissement dans les arènes autres que nationales soit toujours partiel et relatif, les hommes et femmes politiques d'importance étant tous visibles dans l'arène nationale : l'opposition, par exemple, entre élus régionaux et État central est toujours relative, ce qui diminue la demande d'évaluations contradictoires entre niveaux de gouvernement légitimes. Cet état de choses se trouve congruent avec l'organisation et l'histoire de l'administration française, qui, *structurellement*, présente une tendance à exclure, d'ailleurs comme d'autres activités d'étude ou d'expertise, l'évaluation au bénéfice du contrôle et de l'expertise de l'État *central*, lesquels aiment à se légitimer, dans une tradition jacobine, comme les porteurs exclusifs de l'intérêt général. Comme l'a montré un rapport rédigé par des sénateurs (Bourdin, 2001²) qui comparait

1. Il existait, avant 2008, auprès du Parlement français plusieurs « offices parlementaires d'évaluation » spécialisés et une mission d'évaluation et de contrôle (MEC).

2. Ceux-ci ont noté qu'il n'existait en France à l'époque que quatre petites organisations indépendantes productrices d'expertise (dont l'OFCE est la plus connue) (Bourdin, 2001, p. 18-25). En Allemagne par exemple une grande variété de fondations contribuent à la production d'une expertise pluraliste.

les situations américaine et française, on est en présence en France d'un quasi-monopole des systèmes statistiques, des études et de l'expertise : s'il ne s'agit d'un quasi-monopole de production que dans le cas des services de l'INSEE, l'expertise a beau être confiée à des consultants ou à des universitaires¹, cela ne se fait qu'exceptionnellement dans un cadre de commande indépendante (Jobert, 1994). Pour les raisons indiquées précédemment, le poids des régions et autres collectivités territoriales n'est pas en mesure de contrebalancer ce quasi-monopole.

La thématique de « l'intérêt général » permet de comparer la situation française à celle des autres pays développés. L'attitude vis-à-vis de cet « intérêt général », sa représentation, sa rhétorique distinguent les sociétés en Europe en tant qu'un des éléments de cultures politiques restées nationales malgré l'europanisation des élites (Barbier, 2008a). Par exemple au Royaume-Uni et dans le monde anglophone ou scandinave, le système politique est évalué, au moins au niveau de la rhétorique, au regard de son « *accountability* », c'est-à-dire de la façon dont il se soumet à la reddition des comptes de son action au nom de l'intérêt public. Significativement, l'organisme indépendant américain, détenant une grande puissance légale et légitime à la fois – y compris celle de contraindre le Président – s'appelle désormais *Government Accountability Office*. Qu'il soit ardu de traduire le terme d'*accountability* en français (Perret, in CSE, 1993, p. 72), et d'ailleurs aussi en allemand, s'explique par le contexte politico-administratif. Dans la tradition républicaine plurisécularisée, les agents de l'État doivent, certes, rendre des comptes aux citoyens, comme le précise l'article 15 de la Déclaration des droits de l'homme et du citoyen de 1789 : « La société a le droit de demander compte à tout agent public de son administration. » Mais le système français se caractérise par le rôle éminent joué par les « grands corps² » de l'État dans les tâches d'inspection, d'expertise et de contrôle³. Une telle organisation, fondée sur l'expertise étatique (comme celle des inspecteurs des finances ou des administrateurs de l'INSEE) est étrangère à la construction d'une connaissance partagée avec la société civile, puisque les agents du contrôle y représentent des incarnations de l'intérêt général (Jobert, 1992). Il était ainsi logique que, dans les années 1990, ces mêmes corps jouent un rôle stratégique dans l'EPP, comme on l'a vu pour la création du CSE. Pour ce qui concerne la reconnaissance de l'EPP, aux côtés des disciplines « installées » (contrôle, inspection, etc.) les grands corps détiennent inévitablement un pouvoir de *gate-keeper*, assis sur des ressources institutionnelles et sociales ancrées dans le système politico-administratif. Pour que l'évaluation se développe en tant

1. On ne souligne pas assez que les universitaires français sont aussi des « hauts » fonctionnaires.

2. Duran *et al.* (1995, p. 54) ont justement noté : « *The French state has always relied simultaneously upon the figure of the engineer and the legal scholar in order to 'produce' a society in the name of public authority. Historically, this has legitimised the state's superiority over civil society, which at the same time is assumed to be outside state's control [...]. In this manner, actual or potential policy evaluation structures find themselves closely linked to the civil servants who have traditionally been in charge of administrative control.* »

3. Ces grands corps ne sont, au demeurant, pas homogènes, par exemple les corps de l'inspection et les corps d'ingénieurs, etc. Voir sur ce point Bezes (2009).

que champ particulier d'une profession, concurrente et complémentaire à la fois du contrôle ou de l'inspection, sur la base de standards indépendants, il faudrait logiquement que ces acteurs traditionnels y trouvent un intérêt, ou que la dynamique d'un changement soit impulsée « de l'extérieur », par exemple par la concurrence du secteur privé. L'expérience française montre qu'aucune de ces conditions n'a été réunie. Le rôle du privé, du recours aux consultants, dans le cadre de démarches de management et de réforme, ne s'est jamais significativement centré sur la question de la démarche particulière de l'évaluation, un marché qui est resté toujours relativement incertain pour les grands groupes, dont l'offre épouse au contraire, de façon flexible, la demande de la haute administration.

Du côté des corps d'inspection et de statisticiens, l'idée d'une spécificité de l'évaluation reste relativisée, parfois ignorée ou même combattue. E. Audebrand (1999) a, par exemple, noté que lors de l'élaboration du nouveau décret créant le CNE, en 1998, alors que certains membres de la Cour des comptes entendaient participer, à leur façon, à l'évaluation, des membres du Conseil d'État défendaient la thèse selon laquelle « on ne savait pas ce qu'était l'évaluation ». L'ignorance de l'évaluation est aussi incorporée dans l'éthos traditionnel des agents de l'INSEE. Ainsi, lorsque les syndicats de l'institut organisèrent un colloque en 2001 sur le thème « Statistique publique, évaluation et démocratie », la question de l'évaluation ne fut évoquée qu'à la dernière table ronde du colloque. Un intervenant, administrateur de la DARES, fut contesté par des représentants syndicaux quand il souligna qu'il fallait « être trois » pour l'évaluation, c'est-à-dire le pouvoir exécutif, ses services techniques, et la société civile (CGT, CFDT, 2001, p. 73). Dans un conclave qui rassemblait un échantillon probablement représentatif des chargés d'études et de statistiques du niveau central produisant des données utilisables pour l'évaluation des politiques sociales, la question de l'évaluation avait été uniquement considérée comme celle de l'indépendance des services techniques de l'État vis-à-vis de l'exécutif. Ces exemples doivent être compris comme des manifestations d'une certaine culture politique française. On peut aussi ajouter que les universitaires n'ont pas plus d'autonomie dans ce jeu stratégique : soit, comme c'est le cas des économistes, ils sont inévitablement alliés aux ingénieurs-économistes de l'État (Barbier, 2008b) ; soit pour les sociologues ou politistes, leur participation éventuelle à des opérations d'évaluation les met « sur le marché » comme les autres consultants privés.

Pourtant, le secteur privé occupe une place dominante parmi les producteurs d'évaluation. L'offre de conseil, sérieusement segmentée, n'est pas centrée sur l'EPP, un produit qui reste marginal pour elle. Les grands cabinets internationaux de conseil en management, contrôle de gestion, comptabilité et audit ont, certes, accru leur présence sur les marchés publics au niveau central et territorial. Mais ce phénomène a concerné surtout le management public, l'audit, l'organisation, à la faveur, comme en d'autres pays, de l'influence du *new public management*. De l'autre côté, les offreurs de produits d'évaluation en France se divisent en cabinets moyens et petits qui

réalisent des études à visée évaluative dans le prolongement de leurs niches construites à partir d'activités d'études sectorielles¹, de contrôle de gestion ou de conseil. En outre, contrairement à ce qui se passe par exemple aux États-Unis, les cabinets moyens ou petits exclusivement spécialisés dans la vente de services d'évaluation sont restés marginaux, et fortement dépendants des financements européens. Certains acteurs professionnels de l'évaluation en France ont pensé que l'europanisation allait entraîner une assimilation des pratiques de l'évaluation en France (Fontaine, Monnier, 2002 ; Toulemonde, 2000). Mais, jusqu'à présent, le marché de l'évaluation, que l'Union européenne impulse, il est vrai de façon décisive, a pu se développer sans que, pour autant, des coalitions d'acteurs suffisantes se forment pour introduire, au sein du système politico-administratif relativement hostile qu'on a décrit, des conceptions professionnelles équivalentes à celles qui ont largement cours dans de nombreux autres pays de l'Union². Le processus européen n'est, en tous les cas, pas celui d'une simple convergence universaliste vers l'utilisation progressive de l'évaluation, par apprentissage, qui ferait passer tous les pays d'un état « d'ignorance » ou de « refus » de l'évaluation à un état, envisagé comme idéal par certains évaluateurs, de son usage pour le débat démocratique. Dans ce dernier cas, que l'on peut repérer dans certaines pratiques en vigueur dans les pays scandinaves, les rapports de l'EPP et de la légitimité des politiques sociales se posent de façon très différente de ce qui se passe en France. On peut ici penser que l'une des raisons principales de cette « banalisation » tient dans la « symbiose » existant entre évaluation et autres techniques de management, de même que la liaison avec les processus budgétaires. Or, la première évaluation institutionnalisée qu'on a décrite dans la première partie s'est présentée comme alternative aux techniques existantes.

Controverses, coalitions

Les controverses à propos de l'évaluation ont, clairement, des traits nationaux. Cela n'est pas pour surprendre, car, activité pratique, l'évaluation ne saurait, au-delà des guides généraux rédigés en anglais international, se développer de façon indépendante d'une culture politique enchâssée et reproduite dans des institutions formelles. Depuis les années 1990, de nombreux acteurs de l'évaluation ont regretté son développement incertain (par exemple, Martin, 2001 ; Perret, 2001, p. 85-93). Du point de vue d'une sociologie de l'évaluation, cette situation de « retard » doit être vue de la façon la plus « neutre » possible. C'est le point de vue de P.-A. Sabatier (1998), quand il souligne que le cœur de la question relève de la possibilité (ou de l'impossibilité) de la construction, dans une communauté d'évaluateurs,

1. Des entretiens réalisés dans les cabinets moyens nous ont montré que les conceptions de l'évaluation étaient très diverses. En 2001, un responsable d'un des principaux cabinets moyens intervenant dans l'EPP était à la fois incapable de définir l'évaluation et de dire quelle était la proportion du chiffre d'affaires réalisé dans ce domaine.

2. Les cas espagnols et italiens sont assez proches du cas français, sous l'angle de leur réticence à instaurer des systèmes d'évaluation.

de standards professionnels, autour desquels les différents groupes se trouvent contraints de contester les conceptions de leurs adversaires, et de se mettre d'accord sur des conceptions communes de l'évaluation, scientifiques et professionnelles¹. Indirectement, cet auteur rejoignait l'analyse de Jean Leca (in CSE, 1992, p. 13). Il s'ensuit qu'aujourd'hui, *de facto*, le champ des activités et acteurs se réclamant de l'EPP est, malgré plus d'une décennie d'histoire, marqué en France par une grande variété de conceptions, de méthodes et de points de référence ; cet éclectisme fait que la question des frontières entre l'EPP et d'autres activités, présente dès les années 1990² a été, sans cesse, remise sur le métier³.

Toutefois, l'éclectisme et la variété ne s'expliquent pas uniquement en termes de champs professionnels plus ou moins différenciés. Par comparaison, là aussi, avec d'autres pays étrangers, la situation française est marquée par des divergences politiques qui, tout en évoquant des controverses comparables à l'étranger, se présentent sous des formes spécifiques. À ce propos, on peut noter l'influence durable, quoique très minoritaire, de ce qu'on peut appeler la « doctrine du CSE » dans la situation française. Le rôle de J. Leca dans sa création a été vraisemblablement déterminant, et il n'est pas inutile de rappeler le postulat principal qu'il a systématiquement argumenté. Pour lui, parce que l'information est un instrument à caractère stratégique, « l'évaluation doit être conçue comme une zone d'autonomie par rapport aux acteurs (y compris l'État) auxquelles elle apporte de l'information. Sinon, elle devient, dans son déroulement même (et non pas seulement dans ses usages ultérieurs), un élément même de la manipulation de l'information que les stratèges *ne peuvent que pratiquer* » (1993, p. 189-190) ; au sein de la zone d'autonomie, il serait possible ainsi de « neutraliser un temps l'énonciation polémique dans le processus d'évaluation et de faire de ce dernier un processus de connaissance partagée par ceux qui prennent la responsabilité de ses conclusions, afin que cette connaissance puisse être, ultérieurement ou en même temps, appropriée ou utilisée par l'ensemble des protagonistes qui peuvent parfaitement rester opposés sur telle ou telle interprétation ou telle mesure à prendre » (*ibid.*, p. 192-193)⁴.

1. "Given that most policy evaluators will be members of different coalitions, the goal should be to develop institutions that force them to confront their adversaries in forums dominated by professional/scientific norms" (Sabatier, 1998, p. 143).

2. « En l'absence de milieu professionnel les ayant intégrées, la diversité des traditions disciplinaires mobilisées par l'évaluation produit un éclectisme de fait, qui conduit à admettre que pour un même type de question, on puisse utiliser une pluralité d'approches sans tenter de les confronter, de les articuler, ou de les référer précisément à des finalités différentes de l'évaluation » (CSE, 1992, p. 13).

3. Notons qu'en décembre 2007, la société française de l'évaluation a pris l'initiative d'organiser une réunion publique sur « L'évaluation et les activités voisines ».

4. Voir aussi Leca (1997, p. 15) : « Dans la mesure où "l'évaluation de politique" n'est pas le même type de système et de jeu que "l'évaluation politique", elle peut (et doit) être conçue comme une "zone d'autonomie" par rapport aux autres systèmes (politique et politico-administratif) auxquels elle apporte de l'information. Sinon, elle devient dans son fonctionnement même (et non pas dans ses usages ultérieurs) partie de l'évaluation politique et des manipulations propres à celle-ci que ses stratèges ne peuvent que pratiquer, quelles que soient leur moralité et leur bonne volonté. »

L'essentiel semble donc reposer, dans cette conception, sur la possibilité de la construction de cette zone d'autonomie, à l'écart, nous dit-il, du système « politico-administratif ». Or, cette possibilité n'est pas donnée et, c'est le moins qu'on puisse dire, elle est difficile à construire par des acteurs qui ont bien d'autres intérêts à défendre que cette *orientation civique* de renonciation, ne fût-ce que pour un temps, à la « polémique » et à la manipulation de l'information. Il n'est pas rare que des instances pluralistes aient fonctionné, dans la réalité, différemment, comme des forums de représentation des intérêts. Plus généralement, s'il est observé constamment que l'évaluation, comme le disent Spenlehauer et Warin (2000, p. 251), « ne peut exister que si elle est perçue comme étant fondamentalement d'une nature autre que politique », cela n'entraîne pas qu'elle *parvienne effectivement* à l'autonomie vis-à-vis de la politique. À la limite, on peut voir dans la conception d'une « zone d'autonomie » un écho de la discussion wébérienne à propos de la « neutralité axiologique », même si la connaissance que l'évaluation cherche à construire et partager reste toujours relative à ce que décident « des protagonistes » situés participant à la zone d'autonomie (imposant des limites à la connaissance en raison de la situation de commandite et de l'obligation finale de porter un jugement de valeur). Quoi qu'il en soit, cette vision particulière de l'évaluation n'a pas trouvé suffisamment de continuateurs de la « doctrine du CSE ». Sans doute, est-ce au sein de la SFE que cette tendance trouve, aujourd'hui, une expression privilégiée parmi d'autres ¹.

■ L'éclairage porté par les évolutions des années 2000

Dans la deuxième moitié des années 2000, en l'absence d'enquêtes empiriques approfondies, il est difficile de juger les mouvements qui affectent les milieux et acteurs qui se réclament de l'évaluation en France. Au contraire, les mouvements de fond de la gestion de l'État laissent penser, s'il faut en croire les spécialistes, à un tournant relativement tranché en matière de « configurations de réformes » et de « répertoires » (Bezes, 2009). Dans les milieux composites de l'évaluation, en raison de l'importance du secteur privé, il faudrait situer plus précisément l'activité (du point de vue de son économie) dans le plus vaste domaine des « études et du conseil », un secteur particulièrement sensible aux variations du marché, mais aussi au phénomène de la mode managériale et à la « flexibilité » de l'offre. Dans le secteur social qui nous intéresse ici, une évolution se présente relativement clairement, qui contraste fortement l'État central et les collectivités régionales. En

1. Les principes généraux de la charte de la SFE sont inspirés des chartes et standards étrangers, que les rédacteurs ont analysés et adaptés, mais ils sont présentés d'une façon originale et comportent des particularités. Ce sont les suivants : principe de pluralité (prise en compte des différents points de vue) ; distanciation (autonomie et impartialité de l'évaluation) ; compétence (des participants) ; respect des personnes (protection de leur intégrité et sécurité) ; transparence (justification des conclusions et diffusion publique) ; opportunité (justification de l'utilité) ; responsabilité (division des rôles et responsabilité conjointe des participants).

schématisant de façon forte : la tendance de l'État central des années 2000 semble le recours systématique à des démarches d'audit, pour la gestion des divers départements ministériels et leur « modernisation » ; en même temps, pour les politiques sociales, le gouvernement s'est tourné de façon croissante vers l'évaluation économique, en particulier pour l'évaluation *ex ante*. Au niveau régional, comme aux autres niveaux territoriaux, les démarches de management public n'écartent pas le recours à des travaux d'évaluation inspirés de la sociologie des organisations, qui semble toujours une source essentielle du management public. Ces tendances sont observées de façon très générale et mériteraient une enquête beaucoup plus fine.

D'une certaine manière, depuis le début des années 2000, l'évaluation des politiques publiques s'est en quelque sorte « banalisée », dans la diversité des méthodes et des références professionnelles, mais aussi dans la diversité des institutions concernées, alors qu'il n'existait plus d'organisme à vocation spécifique auprès du gouvernement. Sur le plan du marché des collectivités territoriales, les prestations d'évaluation sont souvent référées, voire gérées en commun avec des activités de contrôle et de management public classique, mais des enquêtes empiriques systématiques manquent. Au plan de l'État central, parallèlement à l'évolution signalée plus haut, on a assisté à une diffusion rapide de l'application de l'économie dominante (*mainstream*), à un éventail croissant de politiques et de réformes (contrat de travail, prime pour l'emploi, services de l'emploi, incitations au travail, revenus minimum...) à des fins de conseil du gouvernement – un conseil économique qui s'est présenté aussi comme évaluation. La prééminence du raisonnement économique dans les opérations d'évaluation est évidemment liée, à ce niveau central, à des acteurs privilégiés (« économistes d'État » de l'INSEE et économistes des universités). Cette nouveauté paraît dans les débats à propos du « social » dans les médias (qui participent, pour employer les catégories de B. Jobert, au forum de la communication politique). Elle manifeste la tenace progression des travaux d'économistes qui se sont saisis, au fur et à mesure, d'objets qui n'étaient, il y a dix ans, que marginalement traités par cette discipline. Ainsi, le langage et le champ de la recherche appliquée à propos du social ont été profondément altérés. Le changement tient aussi dans l'importance prise, au cours de la période considérée, par la dimension comparative, européenne et internationale des débats dans les divers forums ; ces derniers sont marqués par l'influence croissante des institutions internationales, qui mélangent inextricablement discours politique et discours économique alimenté par l'évaluation de « bonnes pratiques » décontextualisées (Barbier, 2008a ; Merrien *et al.*, 2005).

Cette transformation aboutit, dans les débats à propos du social et de sa réforme, dans la majorité des cas, à faire circuler une denrée intellectuelle composite, qui joue un rôle important dans la légitimation des politiques. Cette denrée combine indistinctement connaissance scientifique, information, description journalistique, discours politique, prises de positions normatives et résultats d'évaluations. Il est vrai que les économistes de l'école

dominante se sont intéressés au « social », en particulier à la relation entre protection sociale et marché du travail. Utilisant la microéconométrie, leur approche, très différente de celle préconisée à la fin des années 1980 par le Commissariat du Plan ou le CSE, fait désormais partie du paysage actuel de l'EPP. Elle offre l'avantage de simplifier les problèmes du jugement complexe sur l'*impact*, l'*efficacité* et l'*efficience* des mesures prises. L'apport propre d'autres disciplines s'en trouve contesté. Pour l'économie majoritaire, le caractère composite de la denrée intellectuelle qui circule dans les forums, ne pose pas de problème majeur nouveau, d'autant qu'elle est spontanément normative. Pour les autres sciences sociales, celles que J.-C. Passeron (2006) qualifie de « sciences historiques », qui entendent travailler empiriquement sur le « social », le forum scientifique semble de plus en plus incapable de réunir les conditions d'un fonctionnement autonome. Dans le domaine qui nous occupe ici, cela a de nombreuses conséquences pour les débats à propos des politiques, et cela dépasse de loin les enjeux circonscrits de la démarche de l'évaluation. Par ailleurs, ces réorientations en termes de modèles disciplinaires, ne doivent pas faire oublier que l'usage de la démarche d'évaluation doit toujours être référée à la dynamique plus englobante, et autrement plus puissante, de la réforme de l'État, laquelle est désormais complètement conduite par l'analyse économique qui relève du « néolibéralisme gestionnaire » (Jobert, 1994).

Ainsi, pour ne prendre qu'un exemple, l'analyse des politiques visant à contrecarrer la pauvreté a été bouleversée. À la suite du discours international prescriptif de l'économie standard, le pauvre a eu tendance à être désormais considéré comme un *homo oeconomicus* qu'il faut inciter à travailler car « le travail paie » (*work pays*)¹. C'est ce que deux sociologues ont pointé en mettant en cause la réduction de la rationalité de l'acteur (Dubet, Vérétoit, 2001). Le point de vue dominant à l'époque de l'invention et de l'évaluation du RMI, en 1988-1992, est aujourd'hui marginalisé : le pauvre était un citoyen qu'il convenait d'aider à s'intégrer dans la société car il était la victime de circonstances défavorables (Morel, 2000). Au total, l'évaluation microéconométrique offre des réponses « réduites » et simplifiées à de nombreuses interrogations des responsables politiques et elle est désormais devenue, au moins pour un temps, une des techniques privilégiées de la légitimation des politiques sociales. Dans la presse et dans les débats politiques, les citoyens se voient, par des canaux divers, proposer des « politiques qui marchent », pendant que les organisations internationales produisent des discours de légitimation décontextualisés, promouvant des types de politiques censés réussir dans tous les pays (Barbier, 2008a). Ainsi, du côté de l'évolution de la science économique aussi, l'évaluation pratiquée a tendance à se banaliser.

1. Cette conception est profondément congruente avec des méthodes économiques d'évaluation qui réactivent des méthodes inventées aux États-Unis, il y a trente ans (Monnier, 1992; Weiss *et al.*, 2008), utilisant la comparaison entre groupes « de traitement » et « de contrôle », dans lesquels on affecte au hasard les personnes éligibles au programme considéré.

Au total, en ce début 2010, on est bien loin du ton que donnait l'ex-CSE, et il s'opère peut-être une double « banalisation » : banalisation de l'évaluation comme l'une des cordes du référentiel réformateur du *new public management*, mais certainement pas sa corde principale ; banalisation, d'autre part, de l'évaluation comme l'une des formes de l'analyse économique généralisée. En contrepoint, il restera dans l'avenir à mener à bien l'analyse sociologique de ce que fut l'EPP des années 1990, une analyse qui n'est ici qu'esquissée.

■ Conclusion

Le passage en revue des quelques caractéristiques sociologiques essentielles de l'EPP en France permet des conclusions analytiques provisoires, à partir d'un point de vue qui ne saurait trancher entre différentes conceptions normatives¹. Tout d'abord, trois éléments nous semblent essentiels : le champ des acteurs et pratiques se réclamant en France de l'EPP est habité de controverses, irréductibles à ce jour, qui empêchent son unification comme champ professionnel homogène ; dans ce champ « éclectique² », les économètres ont pris une importance récente, qui peut n'être que conjoncturelle ou se révéler durable ; les tenants d'une évaluation « pluraliste » ont une audience maintenue, mais très minoritaire. En second lieu, la conception privilégiée par l'ex-CSE a laissé en France une place particulière, singulière par rapport à ses homologues étrangères, dans la construction d'une doctrine « à la française », aujourd'hui vivement contestée pour des raisons hétérogènes (lourdeur et lenteur, contestation du principe même des « instances », du pluralisme, pratique effective, etc.) ou tout simplement ignorée. En troisième lieu, compte tenu de cette variété méthodologique et théorique, normative et politique, il n'est pas facile de situer l'EPP parmi d'autres activités qui concourent à la fabrication et à la mise en œuvre, ainsi qu'à la légitimation des politiques publiques (inspection, contrôle, audit, management). Le contexte est donc le suivant : plusieurs coalitions concurrentes et/ou complémentaires pratiquent l'EPP, pendant que les caractéristiques du système politico-administratif français sont relativement stables et se reproduisent³. Les développements précédents ne montrent pas de tendance forte quant à la modification de la situation actuelle : marché segmenté du côté de l'offre comme de la demande,

1. Ceci n'interdit nullement d'afficher par ailleurs, dans un autre moment que celui de l'analyse, des préférences pour telle ou telle conception de l'évaluation.

2. Voir ici l'appréciation portée en 1992 dans le rapport du CSE : « En l'absence de milieu professionnel les ayant intégrées, la diversité des traditions disciplinaires mobilisées par l'évaluation produit un éclectisme de fait, qui conduit à admettre que pour un même type de question, on puisse utiliser une pluralité d'approches sans tenter de les confronter, de les articuler, ou de les référer précisément à des finalités différentes de l'évaluation » (CSE, 1992, p. 13). G. Martin (2001) a parlé à ce propos d'*auberge espagnole*.

3. Dans son article, J.-P. Nioche (1982) notait déjà l'existence de traits qui sont restés pratiquement peu changés. Bezes (2009, p. 474) avance prudemment une hypothèse de rupture durable dans les années 1990.

diversité des disciplines de référence, obstacle structurel constitué par l'organisation administrative (elle-même en profond mouvement) et le système politique français, absence de références professionnelles partagées par un nombre suffisant d'acteurs, tous ces éléments laissent penser que, dans un proche avenir, la variété continuera de prévaloir, et, avec elle, sans doute cet aspect de « banalisation » (ou d'éclectisme relativiste) signalé plus haut. Nous n'avons pas, cependant, évoqué un facteur institutionnel, celui de la réforme de la Constitution, par la loi du 23 juillet 2008. Les deux articles modifiés portent les numéros 24 et 47-2 ; selon le premier, est introduit le principe que le Parlement « évalue les politiques » ; selon le second, le rôle de la Cour des comptes, est modifié et précisé : « la Cour des comptes assiste le Parlement dans le contrôle et l'action du Gouvernement. Elle assiste le Parlement et le Gouvernement dans le contrôle de l'exécution des lois de finances et l'application des lois de financement de la sécurité sociale *ainsi que dans l'évaluation des politiques publiques* [souligné par nous] ». Ces dispositions seront-elles à l'origine d'un changement du paysage de l'EPP en France ? Cette réforme ne saurait avoir produit encore d'effets mesurables à l'heure où ces lignes sont écrites : les institutions parlementaires se sont mises en place, une doctrine a été définie par la Cour des comptes. Il est évidemment prématuré de répondre à la question de savoir comment le Parlement pourrait prendre un pouvoir nouveau dans l'évaluation face à l'exécutif, et comment la Cour des comptes¹, qui participe déjà à l'EPP, va ou non tirer parti de cette modification de la loi fondamentale pour changer son rôle, à côté des grands départements ministériels.

Pour conclure ce texte, qu'est-ce qui caractérise finalement l'*activité évaluation dans le cas français* ? Quel est le noyau distinctif qui résiste quand on compare l'évaluation et les autres techniques de la gestion publique dans ce pays ? Laissons de côté celles qui sont contrôlées en France par des corps professionnels clairement constitués : inspection, contrôle, conseil, audit, pour l'essentiel. On peut distinguer *a minima* quatre types, parmi les activités dont l'action publique tire des « informations évaluatives ». Le premier type est la recherche scientifique appliquée (évaluée par les pairs de chaque discipline). Le second, c'est l'étude/recherche finalisée, qu'on pourrait dire « caméraliste »², à destination d'un ou plusieurs financeurs intervenant plus ou moins dans la commande et dans l'étude elle-même. Ensuite, il existe au moins deux catégories d'évaluations : des opérations, souvent multidisciplinaires, qui embrassent des politiques assez larges, sous l'égide d'instances *relativement* autonomes, plus ou moins formalisées, comme les commissions d'évaluation mises en place par l'ex-CSE ou le Commissariat du Plan ou d'autres plus récentes ; et, enfin des évaluations (économiques, sociologiques surtout) à destination d'un commanditaire unique, sur une base contractuelle avec un expert/consultant.

1. Voir sur ce point le discours de son premier président lors du colloque organisé à Marseille, par la SFE, en juin 2009. Le discours est accessible sur le site www.sfe-asso.fr

2. Pour reprendre la formule de R. Boudon (2003).

Cette typologie, simplificatrice, sépare clairement l'EPP de la recherche et des études. Elle fait en général consensus. La recherche est à part, car elle pose ses propres questions et l'étude, de son côté, n'est pas aimantée par le questionnement de la valeur et des effets de l'action publique. Ensuite, elle distingue *au moins* deux formes différentes d'évaluation parmi celles qui sont empiriquement pratiquées dans notre pays : soit les EPP ont un commanditaire unique, soit elles en ont plusieurs (dont, souvent, même implicitement, l'opinion publique au sens de « l'Öffentlichkeit » ou de « l'intérêt général »). Entrent dans la deuxième catégorie, bien sûr, celles qui parviennent à construire la « zone d'autonomie » dont parle J. Leca. La présence de plusieurs commanditaires (de participants ayant voix au chapitre au cours de l'évaluation et, partant, dans le partage effectif des connaissances produites, ce qui peut aller jusqu'à inclure des représentants des citoyens et toute autre « partie prenante »¹) autorise sans nul doute à ne pas assimiler évaluation et contrôle, évaluation et inspection, évaluation et management. D'un point de vue analytique, en revanche, une « évaluation » qui se déroule dans la relation fermée d'un commanditaire et de l'évaluateur est plus difficile à distinguer du conseil et de l'expertise classique². Il faudrait entrer ici plus dans les détails techniques de la professionnalité, qui comprend indéniablement des savoir-faire, en partie spécifiques, de *conduite des processus* d'évaluation. Cependant, l'un des points centraux – peut-être est-ce le seul – que la plupart des évaluateurs considèrent comme spécifique de leur activité est de modéliser l'action publique *en termes de causes et d'effets* et d'utiliser, pour objectiver ces causes et ces effets, les méthodes des sciences sociales. Dans la « communauté internationale » des évaluateurs, il est pratiquement exclu que le questionnement évaluatif échappe à la nécessité de rechercher des liens de causalité entre le déploiement des politiques et leurs « effets », « impacts » ou « résultats »³ : sous cet angle, l'EPP peut être considérée comme *l'instrumentation particulière et explicite des sciences sociales, dans le but de construire une représentation des effets des politiques conduites, à des fins de décision politique ultérieure*. En termes de la construction de référentiels, au sens de B. Jobert et P. Muller, l'EPP ainsi entendue, prend très au sérieux l'étude des « algorithmes » des politiques publiques, leurs « théories normatives d'action ». À côté de ce premier trait spécifique de l'évaluation en France, un second tient dans les modalités de son « pilotage », dans le « processus » qui tend à favoriser une certaine ouverture sur les parties prenantes concernées. Ni l'audit, ni l'inspection, ni le contrôle, ni la recherche appliquée, ni les études finalisées et encore moins le management public, ne répondent à ces deux critères. Le critère de la recherche des causalités

1. Voir sur ce point l'article de P. Warin (1993).

2. Il n'en reste pas moins qu'une étude demeure, pour ses responsables, à visée évaluative, se distinguant ainsi d'une étude en général, si son commanditaire la considère comme telle, sous réserve que des références professionnelles légitimes viennent apporter la preuve professionnelle contraire.

3. Pour autant, en raison des controverses sur les méthodes, les disciplines, et les cadres institutionnels, plusieurs conceptions s'affrontent à propos de ce qu'on peut considérer comme « effets », « impacts », etc.

et imputations semble très distinctif. Celui concernant l'ouverture sur les «parties prenantes», la dimension démocratique, l'*accountability* devant l'intérêt général public (l'*Öffentlichkeit* allemande serait ici l'expression idoine), est certainement plus controversé.

Évidemment, une grande partie de l'interrogation sur un futur de l'EPP, dans un pays comme la France, chez qui on a noté le monopole de l'expertise, la puissance des grands corps et la faiblesse du Parlement, réside dans l'incertitude des évolutions en cours, sociales et institutionnelles. La question est ouverte, comme celle de la création des futures institutions politiques. L'empreinte des origines et la persistance des traits des cultures politiques françaises n'a aucune chance cependant de disparaître, ce qui n'équivaut pas à prédire un déterminisme de la «dépendance du sentier». Mais, en tous les cas, contrairement aux espoirs (ou craintes) d'une universalisation des pratiques du management public et de la politique, ce ne sont pas la présence des «standards» communautaires, ni celle des références internationales qui décideront de *l'évaluation à la française*. Ce sont les acteurs divers qu'on a évoqués dans cet article, et l'efficacité de leurs coalitions et concurrences, au sein des contraintes héritées du passé, au moyen d'essais et d'erreurs. Les deux tendances observées de banalisation (raisonnement économique, intégration dans le *new public management*) sont à l'œuvre. La qualité du gouvernement, à n'en pas douter, en sera affectée, mais les décisions ne peuvent venir que de l'intérieur de la société française.

Bibliographie

- AUDEBRAND E. (1999), *L'évaluation en réforme : la relance du dispositif national d'évaluation des politiques publiques*, mémoire de DEA, J.-C. Thoening (dir.), ENS Cachan, 128 p.
- BARBIER J.-C. (2004), « Devising and Using Evaluation Standards: The French Paradox », in Schwartz R., Mayne J. (eds), *Quality Matters, Seeking Confidence in Evaluating, Auditing and Performance Reporting*, Rutgers, New Jersey, Transaction Publishers, p. 21-41.
- BARBIER J.-C. (2007), « Evaluation and Political Communication in the French Public Employment Service », in R. Boyle, J. Breul, P. Dahler-Larsen (eds), *Open to the Public: Evaluation in the Public Arena*, Rutgers, New Jersey, Transaction Publishers, p. 21-41.
- BARBIER J.-C. (2008a), *La longue marche vers l'Europe sociale*, Paris, PUF, coll. « Le lien social ».
- BARBIER J.-C. (2008b), « La transformation des forums de politique sociale en France depuis les années 1990 », in O. Giraud, P. Warin (dir.), *Politiques publiques et démocratie*, Paris, La Découverte, p. 217-239.
- BEZES P. (2009), *Réinventer l'État, les réformes de l'administration française (1962-2008)*, Paris, PUF.
- BOURDIN J. (2001), *Rapport d'information fait au nom de la délégation du Sénat pour la planification sur l'information économique aux États-Unis*, Sénat, Session 2000-2001, n° 326, Paris, 382 p.
- BOUDON R. (2003), *Existe-t-il encore une sociologie ?*, Paris, Odile Jacob.
- CGT, CFDT (2001), *Statistique publique, évaluation et démocratie*, Actes du colloque du 21 mars 2001, miméo, Paris (accessible sur le site www.cgt.insee.fr).
- CSE (Conseil scientifique de l'évaluation) (1992), *L'évaluation, de l'expertise à la responsabilité*, Paris, La Documentation française.
- CSE (Conseil scientifique de l'évaluation) (1993), *L'évaluation en développement*, Paris, La Documentation française.
- DELEAU M., NIOCHE J.-P., PENZ P., POINSARD R. (1986), *Évaluer les politiques publiques : méthodes, déontologie, organisation*, Paris, La Documentation française.
- DUBET F., VÉRÉTOUT A. (2001), « Une "réduction" de la rationalité de l'acteur. Pourquoi sortir du RMI ? », *Revue française de sociologie*, 42-3, p. 407-436.
- DURAN P., MONNIER E. (1992), « Le développement de l'évaluation en France, nécessités techniques et exigences politiques », *Revue française de science politique*, vol. 42, n° 2, p. 235-262.
- DURAN P., MONNIER E., SMITH A. (1995), "Evaluation à la française. Towards a New Relationship between Social Science and Public Action", *Evaluation*, vol. 1 (1), p. 45-63.
- FONTAINE C., MONNIER E. (2002), "Evaluation in France", in Furubo J.-E., Rist R., Sandahl R. (eds), *International Atlas of Evaluation*, Transaction Publishers, New Brunswick and London, p. 63-75.
- JACOB S. (2005a), *Institutionnaliser l'évaluation des politiques publiques, étude comparée des dispositifs en Belgique, en France, en Suisse et aux Pays-Bas*, Bruxelles, PIE Peter Lang.
- JACOB S. (2005b), « La volonté des acteurs et le poids des structures dans l'institutionnalisation de l'évaluation des politiques publiques (France, Belgique, Suisse et Pays-Bas) », *Revue française de science politique*, vol. 55, n° 5-6, 2005, p. 835-864.

- JOBERT B. (dir.) (1992), « Représentations sociales, controverses et débats dans la conduite des politiques publiques », *Revue française de science politique*, n° 2, vol. 42, p. 219-234.
- JOBERT B. (éd.) (1994), *Le tournant néolibéral en Europe*, Paris, L'Harmattan.
- LASCOUMES P., SETBON M. (1996), « L'évaluation pluraliste des politiques publiques : enjeux, pratiques, produits », Note pour le Commissariat général du Plan, miméo, Paris, 17 p.
- LECA J. (1993), « Sur le rôle de la connaissance dans la modernisation de l'État et le statut de l'évaluation », *Revue française d'administration publique*, n° 66, avril-juin, p. 185-196.
- LECA J. (1997), « L'évaluation comme intervention, sciences sociales et pratiques administratives et politiques », in *L'évaluation des politiques publiques*, Actes du colloque GAPP-CNRS-ENS Cachan des 5 et 6 février, miméo, 19 p.
- MARTIN G. (2001), « Évaluation, organisations sociales et politiques sociales, clarifier les termes, repositionner les fonctions et les rôles de chacun et s'interroger sur les limites de l'exercice », in *Actes du colloque de la SFE*, juin.
- MERRIEN F.-X., PARCHET R., KERNEN A. (2005), *L'État social, une perspective internationale*, Paris, Armand Colin, collection U.
- MONNIER E. (1992), *Évaluations de l'action des pouvoirs publics*, Economica, 2^e éd., Paris [1^{re} éd. 1987].
- MOREL S. (2000), *Les logiques de la réciprocité, les transformations de la relation d'assistance aux États-Unis et en France*, Paris, PUF.
- MULLER P. (1998), *Les politiques publiques*, Paris, PUF, coll. « Que sais-je ? » (3^e éd.).
- NIOCHE J.-P. (1982), « De l'évaluation à l'analyse des politiques publiques », *Revue française de science politique*, vol. 32, n° 1, p. 32-61.
- PASSERON J.-C. (2006), *Le raisonnement sociologique, un espace non poppérien de l'argumentation*, Paris, Albin Michel (2^e éd. revue et augmentée).
- PERRET B. (2001), *L'évaluation des politiques publiques*, Paris, La Découverte, coll. « Repères ».
- ROSSI P. H., FREEMAN H. E. (1993), *Evaluation, a systematic approach*, Newbury Park, Sage.
- SABATIER P.-A. (1998), "The Political Context of Evaluation Research, an Advocacy Coalition Perspective", in Kessler M. C., Lascoumes P., Setbon M., Thoenig J. C. (dir.), *Évaluation des politiques publiques*, L'Harmattan, Paris, p. 129-146.
- SPENLEHAUER V. (1995), *L'évaluation de politique, usages sociaux. Trois études de cas d'évaluations*, L'Harmattan, Paris.
- SPENLEHAUER V. (1998), *L'évaluation des politiques publiques, avatar de la planification*, Thèse de doctorat de science politique, Grenoble, 586 p.
- SPENLEHAUER V., WARIN P. (2000), « L'évaluation au service des conseils régionaux », *Sociologie du travail*, n° 2, avril-juin, p. 245-262.
- TOULEMONDE J. (2000), « Évaluation culture(s) in Europe : Differences and Convergence between National Policies », *Vierteljahrshefte Zur Wissenschaftsforschung*, 3, p. 350-357.
- VIVERET P. (1989), *L'évaluation des politiques et des actions publiques*, Rapport au Premier ministre, Paris, La Documentation française.
- WARIN P. (1993), « Vers une évaluation des services publics par les usagers ? », *Sociologie du travail*, n° 3, p. 309-331.
- WEISS C. H., MURPHY-GRAHAM E., PETROSINO A., GANDHI A. G. (2008), "The Fairy Godmother and Her Warts: Making the Dream of Evidence-Based Policy come true", *American Journal of Evaluation*, vol. 29-1, p. 29-47.