

Atelier 1 - Evaluations et politiques du chiffre

Annie FOUQUET

J'ouvre cet atelier n° 1, intitulé « évaluations et politiques du chiffre », qui s'appelait dans les réunions préparatoires « évaluations, débats, décisions, des chiffonnades de statistiques ». J'aimais bien cet aspect « chiffonné ». Nous allons vous raconter comment nous sommes chiffonnés le rôle de la statistique dans l'évaluation, le débat public et la prise de décision.

Ce matin, Mireille Elbaum a employé plusieurs fois une formule que j'ai bien connue du temps où je démarrais à l'Insee, (j'ose à peine vous dire quand, disons dans les années 70, même si j'ai commencé dans les années 60... !): la statistique sert à éclairer le débat public. C'est une des raisons qui m'avait fait rentrer à l'Insee, j'avais envie de porter ma lanterne pour éclairer ce débat public; éclairer des réalités mouvantes, disait Mireille tout à l'heure.

Il y a mésusage du chiffre quand, selon l'image très connue du réverbère, l'indicateur ne va éclairer que sous le réverbère et empêche de regarder ailleurs. Je voudrais plaider pour une évaluation des politiques publiques où la statistique continue d'éclairer le débat public à l'usage des citoyens, dans un usage démocratique, et pas simplement dans une culture du résultat, où l'on ne regarde que sous le réverbère, avec une confusion de sens entre ce qu'en évaluation de politiques publiques on différencie bien entre les indicateurs de réalisations et les indicateurs de résultats. La réalisation, c'est l'activité. Bien des évaluations dites par le résultat sont des évaluations où l'on ne regarde pas autre chose que l'activité. Emmanuel Didier va nous en parler avec le mot tiré de l'argot des policiers, le chansticage. Il va nous faire rêver dans l'argot policier.

Un mot sur ce que j'entends par évaluation, pas seulement moi d'ailleurs, mais dans le cadre de mes fonctions bénévoles de militante associative, de présidente de la société française de l'évaluation. Il s'agit d'une évaluation de politiques publiques au service de la démocratie, et pas du tout une évaluation des personnes, ce qui relève d'un autre registre. Le lien entre les deux n'est pas facile à faire et d'ailleurs, il ne devrait pas l'être.

Autre élément : éviter la confusion, que j'ai faite en mon temps, entre le constat que peut donner une statistique et l'interprétation que l'on peut en faire pour éclairer la décision, avec plusieurs étapes qui me paraissent très importantes. Dans mes fonctions

actuelles d'inspection générale des affaires sociales, comme avec mes collègues de la Cour des comptes, nous continuons à faire une confusion entre le constat, la conclusion et la façon dont cela peut aider à la décision, qui sont des étapes différentes. Passer du constat à une interprétation nécessite de prendre en compte une diversité de points de vue.

Je vous donne un exemple dans le domaine social : j'ai mené une mission sur les établissements d'hébergement pour personnes âgées dépendantes. On constatait qu'il y avait en moyenne sur l'année 5 % de lits vacants. Est-ce bien, est-ce mal ? Est-ce beaucoup ? Est-ce peu ? Tout dépend du point de vue.

- Pour le directeur de l'établissement, soucieux de son équilibre financier, ce n'est pas bien du tout parce que cela constitue 5 % de chiffre d'affaires en moins. Il a donc du mal à équilibrer son budget.
- En revanche, pour le personnel, 5 %, c'est très bien car il peut passer plus de temps avec chaque personne hébergée.
- Si l'on regarde du côté du décideur, de l'élu local dont la fonction est de mettre en place une politique de maintien à domicile des personnes âgées, ce maintien à domicile suppose qu'en cas de crise on puisse héberger de façon urgente et temporaire des personnes qui, généralement, sont chez elles. Il est donc nécessaire d'avoir des places vacantes. 5 % est le minimum nécessaire pour continuer à appliquer la politique de maintien à domicile des personnes âgées.

Vous voyez sur cet exemple que le constat statistique, 5 % ne dit rien en soi. Tout dépend du point de vue duquel on le regarde. Nous, statisticiens, nous n'avons pas de point de vue. Et si nous en avons un, il est implicite. Bien des études statistiques sont fondées sur des hypothèses implicites qui mériteraient d'être discutées avant de conclure sur l'appréciation à donner à la politique ou à l'action publique observée.

Je vous donne un autre exemple. J'ai fait de l'économétrie. J'estime les paramètres de mes modèles. Estimer la valeur d'un paramètre, c'est étymologiquement « évaluer » = « estimer la valeur » de quelque chose. Quand un économètre fait un modèle pour estimer les valeurs d'un paramètre, par exemple sur les politiques d'emploi, le modèle prétend décrire la création d'emplois liée au dispositif analysé. Le statisticien dit : « J'ai estimé

la valeur du paramètre, j'ai évalué mon modèle, donc j'ai évalué la politique des 35 heures ».

Ce glissement de langage a trop souvent lieu. Quand on a fait une étude statistique qui donne un constat, ou quand on a estimé la valeur d'un paramètre, dans la mesure où un modèle économétrique part toujours d'hypothèses, il est très important que celles-ci aient été débattues.

Si je prends l'exemple des 35 heures, c'est parce que j'en ai des souvenirs cuisants : il n'y a pas eu de lieu de débat où discuter des différentes études faites, avec les hypothèses qu'elles prenaient en compte. Certaines études montraient que les 35 heures créaient des emplois ; d'autres, réalisées par l'Insee, montraient que cela en détruisait. Aucune de ces études n'était fautive en soi. Simplement, elles avaient des hypothèses et un domaine de validité différents. Il n'y a pas eu de lieu de débat serein, apaisé pour débattre des hypothèses faites dans chacun des cas.

Pour la Société française de l'évaluation, la prise en compte des différents points de vue et la création d'instances où l'on puisse délibérer des différentes études et questions qui se posent par rapport à une politique sont très importantes. La pluralité de points de vue est un des principes que la société française de l'évaluation promeut dans sa Charte. Ce sont des principes dans lesquels les statisticiens peuvent se retrouver. La pluralité des points de vue est un principe que les statisticiens ignorent. Mais les autres principes comme l'indépendance, la transparence, sont unanimement reconnus par notre communauté ici.

Pour montrer la façon dont la statistique peut être utilisée de façon chiffonnante, nous avons trois orateurs.

Emmanuel Didier, sociologue au CNRS et à l'École des hautes études en sciences sociales, va vous dire comment la politique du chiffre à New York et à Paris peut être mésusée dans la police, où l'on confond indicateurs de réalisation et indicateurs de résultat. On s'en sert pour évaluer jusqu'aux personnes, donc tout ce que, dans mes activités de militante bénévole, je dis être ce qu'il ne faut pas faire.

Daniel Le Scornet, vice-président de l'Appel des appels et ancien président de la mutuelle de France, va vous parler de la façon dont il voit les choses en matière d'usage ou de mésusage de statistiques qui nous chiffonnent.

Agnès Labrousse, maître de conférence en économie à l'Université de Picardie-Jules Verne, fera un détour par un effet de mode très important, l'expérimentation aléatoire contrôlée, qui va nous en montrer les domaines de validité et les limites.

Je donne la parole à M. Didier.

Benchmarking : mise en responsabilité et initiative dans la police nationale

Emmanuel DIDIER

Je vous remercie de m'avoir invité.

On utilise souvent la métaphore du réverbère. Ici, c'est plutôt celle du ver luisant qui convient. Comment les institutions s'éclairent-elles elles-mêmes de l'intérieur au moyen des statistiques ?

On étudie comment, au sein des administrations centrales de l'Etat, de plus en plus de techniques d'évaluation des agents sont utilisées de façon parfaitement quantitative. Assez souvent, cela consiste à mettre en place des tableaux de bord et à distinguer différentes équipes dont les activités sont quantifiées. Grâce au fait qu'elles sont mises en tableau, cela produit des effets de compétition à l'intérieur des équipes pour atteindre le plus vite et le mieux des objectifs, eux aussi quantifiés.

Ce sont des techniques provenant directement du management privé, pouvant même être identifiées dans un petit nombre de livres. L'un d'eux a été extrêmement influent, le *Reengineering*, écrit par Hammer et Champy dans les années 80 aux Etats-Unis.

Le fait d'importer toutes ces techniques dans l'administration publique transforme la notion de service public car l'administration est alors placée sous le leitmotiv de l'efficacité. Comment mettre ensemble service public et efficacité ? Cela ne peut-il pas devenir contradictoire ?

Je travaille dans le cadre d'un projet global dans lequel on étudie spécifiquement le monde de la recherche, des hôpitaux. J'ai beaucoup étudié le monde de la police.

En 2001, la gauche, à l'époque M. Chevènement, a importé de New-York un outil appelé Comstat, qu'il a appliqué pour la préfecture de police de Paris. Les commissaires centraux d'arrondissement doivent produire une description quantifiée de leur activité. Tous les deux mois, ils sont confrontés les uns aux autres pour savoir quel est celui qui a le mieux pris des initiatives et fait baisser la criminalité dans son arrondissement.

L'idée est donc d'utiliser le plus de données possible, qui permettent d'identifier d'une part ce que font les policiers -les arrestations, les mises en garde à vue, les déplacements sur le terrain, etc -, et d'autre part, les ressources dont ils disposent - combien d'agents disponibles dans un commissariat, combien de voitures, etc. Le commissaire central doit ainsi montrer, grâce aux tableaux, qu'il est capable d'utiliser de façon optimale les ressources dont il dispose pour agir le plus possible.

C'est très intéressant. Cela va de pair avec une démonstration de l'activité personnelle des agents complètement ritualisée. Sur l'image qui est derrière moi, le monsieur que vous voyez habillé en noir est le préfet de police de Paris qui a convoqué trois commissaires centraux d'arrondissement avec leurs équipes rapprochées, de chaque côté de la table. Un opérateur, au milieu, projette des statistiques. Chaque commissaire central montre qu'il a bien fait baisser la criminalité. Ils sont invités par trois, ce qui permet qu'un effet de comparaison se mette en place et que chacun puisse montrer que, grâce à ses initiatives, il a été extrêmement efficace.

Un point est très important. En 2001, quand cette technique a été importée, cela a demandé peu d'innovations statistiques. Les données étaient déjà là. Mais il a été nécessaire de les regrouper dans les mains d'un seul individu, au sein de la préfecture de police de Paris, et de mettre en rapport les tableaux dont je vous ai parlé. Ces tableaux et ces données étaient donc déjà disponibles, l'innovation consistant à réaliser des rapprochements.

Un des avantages de cette technologie est qu'elle peut se décliner à tous les niveaux hiérarchiques. Une fois que le commissaire central d'arrondissement est revenu dans son commissariat, il peut mettre en place de nouveaux tableaux pour les équipes de son propre commissariat. Il met en place le même rituel grâce auquel les différentes brigades de quartier entrent en compétition les unes avec les autres. Au sein des brigades de quartier, on peut comparer les résultats chiffrés de chacun des agents de police, patrouilleurs, etc.

Cela a eu un succès assez grand. Dans la police, cela a assez bien fonctionné. Les commissaires centraux étaient au départ très contents d'utiliser ces techniques. Un des points importants est que cela leur permettait de répondre à la police judiciaire, qui est la police prestigieuse parce qu'elle traite ce que l'on appelle les belles affaires, fait des enquêtes, arrête les grands brigands, etc. Les commissaires chargés de la sécurité publique s'occupent des

bagarres dans la rue, ce n'est pas prestigieux. Ils ont donc pu dire : « *D'accord, on ne fait pas du prestige, mais on en fait beaucoup* ». Ils avaient des quantités pour montrer qu'ils prenaient des initiatives. Cela leur a tout à fait convenu. C'est pourquoi ce système s'est largement diffusé dans l'institution.

Assez rapidement, des problèmes ont été dénoncés. En France, le principal problème dont on entend énormément parler, c'est la « chanstique ». Que veut dire chanstiquer ? Cela signifie : faire en sorte que ce soit bien. Un fait peut être décrit de différentes façons, y compris statistiquement. Les policiers le disent eux-mêmes : comme il y a des façons de présenter les choses à leur honneur et d'autres qui le sont moins, ils choisissent toujours la solution la plus intéressante pour eux.

Quelques exemples. La date de clôture d'un dossier est très importante. Si vous clôturez le dossier sans avoir trouvé le brigand, avant le moment où vous allez passer au débriefing, cela vous fait un bâton négatif de plus. Il vaut mieux laisser traîner le dossier, ne pas le clôturer. Vous voyez ce que je veux dire ?...

Il y a une différence entre les faits élucidés et les faits constatés. Qu'est-ce qu'un fait élucidé ? Vous avez un cambriolage, vous pouvez mettre à côté une personne dont vous suspectez qu'elle y est liée. C'est une suspicion. C'est bien d'avoir des faits élucidés et ce n'est pas bien d'avoir des faits constatés. Fait constaté : dans le quartier, il y a beaucoup de crimes. Fait élucidé : il y a du crime, certes, mais vous luttez contre et vous trouvez le brigand.

Il existe une bonne technique pour faire en sorte que les faits constatés deviennent élucidés. Si vous voulez les détails, je vous raconterai. Ce n'est pas de la vraie triche, c'est de la chanstique. Il est rare de trouver des cas de truande totale.

Cela a été dénoncé par les sociologues. Dans le livre de Lorraine Data¹, il existe un chapitre sur la chanstique dans la police. De même, un livre a été écrit par un gendarme et un sociologue, intitulé *Police des chiffres et des doutes*² : ils décrivent les mécanismes de chanstique dont je vous parle.

Le problème est que le management de la police repose sur des données pour le moins biaisées. On ne peut pas dire tricher, mais cela représente la

¹ Data, L. (2009). *Les grand truquage. Comment le gouvernement manipule les statistiques*. Paris, La Découverte.

² Matelly, J.-H. et Mouhanna C. (2007). *Police : des chiffres et des doutes*. Paris, Michalon.

réalité de façon très étrange. Donc, évidemment, le management pose problème.

Cette critique est étonnante par ce qu'elle présuppose, si on la radicalise, que les policiers trichent pour ne rien faire pour se goberger : ils mettraient des chiffres dans les cases et ils feraient croire avec des statistiques qu'ils agissent beaucoup, mais en fait ils ne feraient rien.

Mais une autre critique, que j'ai observée aux Etats-Unis, va à l'inverse. Elle consiste à dire : « *On demande aux policiers d'en faire le plus possible, de remplir des objectifs. Qu'est-ce qui les empêche de se livrer à des abus de pouvoir et de mettre tout le monde en garde à vue dès qu'ils le peuvent ?* »

Qu'est-ce qu'un petit abus de pouvoir ? C'est lorsque vous ne devriez pas tout à fait aller en garde à vue mais que vous vous y retrouvez. Il est rare d'aller faire un procès à l'institution policière pour une nuit passée au poste. Or, il n'y a pas d'autre solution que le procès pour arrêter ce genre de chose. Aux Etats-Unis, dans certains quartiers de Brooklyn, la quantité de contrôles d'identité a explosé. Pour remplir les cases, faire monter les statistiques, les policiers font effectivement plus de contrôles, mais comme ils n'ont pas toujours de personnes en infraction sous la main, ils se mettent à harceler des innocents. Les policiers disent qu'ils en ont fait, donc qu'ils ont agi. Cela constitue de petits abus de pouvoir car ce sont des contrôles d'identité, la plupart du temps, non justifiés. Dans ce cas, c'est pire que la chanstique, ce n'est pas de tricher avec les statistiques pour ne rien faire, c'est la technique de management qui engendre des actions bien réelles, mais qui constituent des abus de pouvoir. Le management pervertit le droit.

De ce fait, le droit devient un instrument du management aux exigences duquel il est soumis. C'est le management qui est la valeur suprême et non le droit. À mon avis, c'est problématique.

Annie FOUQUET

Merci, Emmanuel. C'est un bel exemple où l'on confond *la réalisation* (produit de l'activité) – combien ai-je fait de contrôles d'identité ou noté de cambriolages- avec *le résultat* -la baisse de la délinquance ou du nombre de cambriolages. Les indicateurs devraient aller dans l'autre sens.

Je donne la parole à M. Le Scornet.

Evaluation, quantification : les nouveaux outils idéologiques de gouvernance

Daniel LE SCORNET

Je parle au nom de l'Appel des appels, qui n'est pas l'agrégat, même pondéré, des différents appels qui se sont développés dans toutes les professions du bien et du lien public, qui se sont vu colonisées pour mettre en cause l'éthique même de leur métier, comme on vient de le voir dans la police. Elles se sont tous retrouvées subitement devoir adapter à leur propre culture, leur profession, quelque chose qui est venu d'ailleurs, du secteur privé : le *benchmarking* qui est partout. L'Appel des appels, c'est le chiffre de l'ensemble de ces appels, le chiffre transversal, en quelque sorte. Je salue mes amis de l'Appel « *Sauvons la statistique publique* ». J'aurais d'ailleurs beaucoup d'intérêt à ce que des statisticiens rejoignent l'Appel des appels plus globalement. Statistiquement, cela doit être possible sur le nombre des professionnels concernés.

Madame la présidente, vous avez voulu évaluer un peu au doigt mouillé, au téléphone, qu'en ayant à votre gauche un membre de l'Appel des appels vous n'alliez pas avoir quelqu'un qui allait charger de façon univoque, donc de peu de valeur de votre point de vue certainement, l'évaluation elle-même, en disant : il ne faut pas jeter le bébé avec l'eau du bain.

Annie FOUQUET

Ni jeter la statistique dans le débat que nous avons aujourd'hui.

Daniel LE SCORNET

J'ai esquivé la réponse. Car, oui, tout de même, nous considérons que l'évaluation des politiques publiques, l'évaluation qui devient un sport -on évalue, on s'évalue, on s'auto-évalue, on a des évaluations externes, on n'est plus évalué par ses pairs- est une idéologie qui est certainement la façon aujourd'hui de donner des ordres sous le paravent de l'objectivité que donne le chiffre, en quelque sorte.

C'est une nouvelle façon de gouverner les sociétés, une nouvelle gouvernance des sociétés, la police de la société, que l'ensemble des professions du bien et du lien public seraient amenées à accepter, voire à en être eux-mêmes les agents dans une certaine soumission volontaire à ces pratiques.

Vis-à-vis de l'évaluation elle-même, nous avons une charge conceptuelle théorique qu'il nous faut mettre en débat.

Je prends deux exemples dans mon expérience, et dans la vôtre, certainement. Je suis mutualiste et, pour un mutualiste, le syntagme mutualisation est chargé très positivement. Il signifie solidarité. Tout comme, quand on dit évaluation, c'est chargé positivement, c'est « donner de la valeur à ». J'ai demandé à un comité scientifique de faire une étude sur le syntagme mutualisation. On s'aperçoit que ce syntagme a basculé d'une position très positive à une position très négative. Si on dit que l'on va faire de la mutualisation à l'Insee, vous vous demandez ce qui va vous tomber dessus.

Annie FOUQUET

Qu'est-ce que l'on va supprimer !

Daniel LE SCORNET

La mutualisation va être réduite à une diminution des moyens, au nom de l'efficacité à trouver dans le système. Il y a une injonction à mutualiser. Normalement, la mutualisation est volontaire. Là, on vous enjoint de mutualiser des services pour réduire, nous dit-on, les redondances. Alors que tout le monde sait que, dans des systèmes complexes, avoir des redondances est une absolue nécessité. Un système complexe qui abandonne la redondance va dans le mur. C'est donc ce que l'on fait et on organise la société pour aller dans le mur.

Je pourrais donner plus de détails sur la mutualisation. Le syntagme évaluation mériterait une recherche en lui-même : comment a-t-il basculé dans le temps à ce qu'il est devenu ? C'est ce que l'on fait actuellement avec l'évaluation. C'est autre chose.

Deuxième exemple : le Japon. J'interviens dans différentes organisations sur le Japon. Au niveau de la statistique, on a des choses étonnantes. On nous dit : « *Oui mais, au Japon, il y a eu le tremblement de terre et le tsunami. Donc, chez nous, dormez tranquilles, braves gens ! S'agissant de la sismicité, on a les statistiques depuis Charles Martel. On a des niveaux au maximum à 5, on va redresser un peu à 7, si vous le voulez bien. On prend toutes les garanties. La hauteur de la vague est de 4 mètres. On va la mettre à 6* ». Vous vous rappelez qu'à la centrale du Blayais la vague faisait 20 cm et les générateurs sont tombés en panne. On a frisé exactement le même accident qu'à Fukushima. La nature a bon dos. En quoi la statistique, dans ces domaines, nous donne-t-elle une capacité de prévoir, voire de prédire, puisqu'on veut de plus en plus aller vers la prédiction ?

C'est ainsi qu'a démarré l'Appel des appels. Le premier appel était « *Pas de 0 de conduite pour les bébés de 3 ans* ». On voulait, à partir de la statistique,

décider quels étaient les enfants qui allaient certainement devenir délinquants. Formidable !

On est donc dans une situation tout à fait préoccupante sur les outils conceptuels, et pas seulement techniques, que l'on utilise au niveau de la gouvernance de nos sociétés.

S'agissant de la santé, c'est la mère de la statistique, puisque les premières études statistiques avaient pour but de mettre de côté non seulement les classes dangereuses, mais les classes contaminantes, les groupes à risque que l'on arrive à classer et à sélectionner, y compris avec les méthodes actuarielles. On utilise, en matière de santé, de mutualité, les groupes à risque pour faire une sélection en termes de tarifs, etc.

Que puis-je dire de positif ? C'est une certaine résistance à ces concepts d'évaluation et d'utilisation de la statistique. Faire notre métier, ce n'est pas faire ce que l'on nous demande, c'est faire notre métier. Ce sont deux choses différentes.

Y a-t-il une pluralisation des donneurs d'ordre au niveau statistique ? Puisqu'il existe un lien très fort entre Etat et statistique, l'Etat perd-il actuellement le monopole de la statistique ? C'est beaucoup moins sûr. Les collectivités territoriales, qui sont donneuses d'ordre dans le social et le médicosocial, rejouent la même chose au niveau décentralisé. On est dans les mêmes utilisations. On a très peu de statistiques plus citoyennes, qui pourraient demander des indicateurs, pas seulement macro avec ce que l'on connaît autour du PIB, mais qui tiennent compte d'une des apories me semble-t-il de l'évaluation actuelle -j'en terminerai là, pour montrer que l'on est peut-être dans une crise, un malaise de la rationalité et de la rationalisation sur lequel la statistique doit pouvoir intervenir : c'est qu'aujourd'hui on est dans la démesure. On vous demande, par exemple, d'évaluer des politiques publiques au millimètre près, à la décimale près. Dans le même temps, on lance une campagne de vaccination, avec une allocation de moyens considérable, sans aucune évaluation ni *ex-ante* ni *ex-post*. Pour déconsidérer l'évaluation et la statistique, c'est superbe ! Je ne dis rien de 2008 et de l'allocation de ressources faite aux banques. On est dans la démesure absolue. Il n'y a eu aucune évaluation de l'efficacité de cette allocation de ressources, aucune ! On est là dans des problèmes de démesure. Le calcul des probabilités est en train de changer, certainement. C'est difficile à dire parce que la science ne change pas. Peut-être que si.

Aujourd'hui, on est sûr que l'improbable va arriver. Il y a une certitude sur l'improbable.

Vous vous rappelez du tremblement de terre de Lisbonne en 1755. Il y avait eu de grandes controverses sur les rapports culture/nature entre Voltaire et Rousseau, ce qui a ouvert la rationalité moderne. Aujourd'hui, les tremblements de terre multiples, les catastrophes -car nous ne sommes pas avant la catastrophe, nous sommes dans la catastrophe au sens fort du terme- nous permettent-ils d'ébranler justement la raison, les conditions épistémologiques de la raison ?

Je reviens au Japon. Ce qui aurait été plus intéressant, que l'on n'a pas, qui n'est pas connu, c'est d'avoir la clinique du cas singulier des différents accidents qui se sont produits dans l'histoire en matière de catastrophe atomique. Il y en a eu quelques uns. Qu'est-ce qui permet de comprendre ? Ce n'est ni la sismicité, ni la hauteur de la vague, ni même les fuites, etc., c'est toujours le système humain.

Alors que l'on a aujourd'hui ce développement non contrôlé –et j'insiste auprès de vous parce que l'on est dans une situation folle aujourd'hui- de 4 réacteurs en fusion, il est incroyable de voir que ni la communauté internationale, ni l'OMS, ni l'Agence internationale de l'énergie atomique ne bougent. Formidable ! On est dans un monde formidable ! Les statisticiens ne bougent pas non pour mettre en cause la façon dont on traite ce problème, qui à mon avis le demande.

Que fait-on de ce que l'on ne peut pas quantifier dans une société de la création, qui devient déterminant pour l'évolution des systèmes globaux ? On est à la fois dans la démesure et dans la faible dose, ou la dose non quantifiable. Une réflexion sur l'ensemble des concepts que nous manions est, me semble-t-il, nécessaire.

Annie FOUQUET

Merci d'avoir réintroduit les fondamentaux de la statistique probabiliste dans la situation où l'improbable est prévisible. Cela remet en cause pas mal de choses, y compris sur le cœur du métier.

Je passe la parole à Agnès Labrousse, pour d'autres usages de la statistique, notamment les expérimentations aléatoires.

Les évaluations aléatoires : de l'evidence-based medicine à l'evidence-based policy ?³

Agnès LABROUSSE

Je remercie les organisateurs pour cette invitation. On va passer des vers luisants au vermifuge. Cela va être, hélas, beaucoup plus scolaire que les deux précédents orateurs. Je suis une ennuyeuse *mulier economica et academica*.

Je suis partie de quelques déclarations d'Esther Duflo qui met en avant la rigueur des essais cliniques appliqués au développement, à la lutte contre la pauvreté. Elle explique que cela devrait permettre de révolutionner les politiques sociales du 21^e siècle, comme les essais randomisés contrôlés avaient révolutionné la médecine du 20^e siècle. Elle conclut : « L'évaluation aléatoire est la meilleure méthode ».

Face à ces déclarations assez frappantes, je suis partie d'un double constat.

Le premier est que l'on a là une sorte de transfert de technologie. On transfère une technique qui provient de la médecine. Cela va de pair avec un certain transfert de légitimité scientifique. Ce qui est mis en avant par Banerjee et Duflo est l'idée d'« hard evidence », de preuve dure que l'on va opposer à des preuves sans consistance que l'on peut obtenir par les régressions en coupe internationale ou, pire encore, par des études de cas.

Le deuxième élément est qu'il s'agit d'un transfert de technologie, comme souvent, partiel. Je n'ai pas trouvé trace de débats, très riches en médecine, autour de l'*evidence-based medicine*. Ce sont ces débats que je souhaite réintroduire rapidement avec une hypothèse : les travaux en économie utilisant cette méthodologie devraient connaître des limites comparables à celles identifiées et admises de longue date en médecine.

Mon objectif ici va être d'essayer de mieux délimiter les champs de validité de cette méthodologie. Je vais, en particulier, essayer de cerner quelques limites. Il est bien évident que ces expérimentations apportent énormément de choses. Ces apports sont bien mis en avant par le principal laboratoire qui les pratique, le J-PAL, associé au MIT.

³ On trouvera plus de précisions sur nombre de thèmes évoqués dans cette intervention dans l'article en ligne et en libre accès : Agnès Labrousse, « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement », *Revue de la régulation*, n°7, 1^{er} semestre 2010, <http://regulation.revues.org/index7818.html>

Je vais brièvement parler des débats en médecine sur les types de preuve et d'évaluation (I) pour, ensuite, mettre en évidence quelques zones floues, quelques points aveugles de ces expérimentations (II). Dans un troisième temps, on verra que ces essais cliniques, loin d'une vision fantasmée des sciences dures, sont aussi des construits sociaux. Cela implique d'autres limites (III).

Tout d'abord, il s'agit d'un transfert de technologie. On transfère le principe des essais cliniques randomisés avec des groupes d'intervention et de comparaison, avec une méthode qui repose sur l'assignation aléatoire pour inclure les patients dans l'essai. On a exactement la même chose en économie du développement, domaine sur lequel je m'attarde ici. Je n'entre pas dans les détails. C'est aujourd'hui bien connu, je pense.

Je voudrais mettre en évidence que des débats existent en médecine sur les différents types de preuves. On distingue des preuves d'existence, des preuves de causalité, des preuves d'efficacité et, enfin, des preuves d'innocuité. Cela permet de mieux délimiter l'apport du J-PAL qui se situe principalement au niveau de l'évaluation de l'efficacité des programmes.

Deuxième élément, on peut distinguer plusieurs types d'évaluations :

A) Tout d'abord, des évaluations comme processus d'apprentissage collectif, ce qui correspond à l'introduction faite par Annie Fouquet, avec une dimension très participative de l'évaluation.

B) Ensuite, l'évaluation permettant de mesurer la performance d'une action, d'un programme spécifique.

C) Enfin, des évaluations visant à identifier des mécanismes causaux, qui vont développer une théorie systématique et cohérente.

On peut voir que l'apport du J-PAL se situe principalement au niveau des évaluations qui mesurent la performance d'une action donnée, même si les dimensions A) et C) sont partiellement présentes. Le J-PAL cultive une relative ambiguïté sur ces questions et se présente volontiers comme une approche complexe et autosuffisante des évaluations, alors que c'est une seule dimension certes très importante, mais une dimension parmi d'autres qui prime.

S'agit-il d'une évaluation comme processus d'apprentissage collectif ?

Il s'agit d'une technique aujourd'hui bien documentée, bien définie, mais la logique d'investigation et de découverte pratiquée par le J-PAL reste comparativement floue. On est dans l'ordre de l'interprétation. Certains éléments témoignent d'une sorte de moment abductif où l'observation et la surprise sont fondamentales. Cela contraste avec quelque chose de classique en économie : le schéma déduction-prédiction-vérification. Cela donne un rôle important à l'observation et à la surprise. Esther Duflo insiste, à juste titre, sur le pouvoir subversif de ces expérimentations et sur les débats qu'elles suscitent. Elle insiste également sur le rôle de la co-expérimentation, sur le partenariat avec des acteurs de terrain dans la construction même des expériences.

On a donc là des éléments allant dans le sens de l'évaluation dans le processus d'apprentissage collectif.

Néanmoins, le rôle du terrain, dans la conception des enquêtes et la remontée d'hypothèses, est limité par plusieurs facteurs.

Tout d'abord, par un biais de publication : parfois, les thématiques choisies le sont dans une sorte de quête de nouveauté gadget, pour publier de l'inédit dans les revues majeures en économie. Cela peut se faire au détriment de la pertinence locale des actions envisagées.

Autre élément : la séparation de fait entre la conception des enquêtes, qui est le fait de chercheurs établis du J-PAL, et la réalisation sur le terrain des enquêtes effectuées le plus souvent par de jeunes assistants de recherche et des enquêteurs locaux à qui elles sont sous-traitées. Bien souvent, c'est un frein à l'émergence par le bas de nouveaux objets et à une réelle réflexivité de terrain.

De plus, Esther Duflo a une définition relativement étroite des acteurs de terrain, limitée -c'est ce que j'en comprends- aux responsables des ONG, aux acteurs politiques locaux, n'englobant pas les sujets des expérimentations eux-mêmes.

Enfin, une dernière limite : l'absence de recours à des techniques plus qualitatives, pourtant relativement peu coûteuses, dans l'élaboration et l'interprétation de ces expérimentations. On pourrait ainsi avoir recours à des enquêtes de type ethnographique, à de la *soft system methodology*, etc.

S'agit-il d'évaluations pour comprendre des mécanismes, débouchant sur des théories intégrées ? Il y a de la théorie dans ces expérimentations,

théorie qui va notamment intervenir dans l'infirmité de lois prétendument universelles. Par exemple, en termes d'économie de l'offre, des propositions vont être infirmées de manière éclatante, en particulier l'idée que la gratuité d'un bien, comme des moustiquaires imprégnées pour lutter contre le paludisme, n'induit pas un usage moindre de ces biens. C'est ce que montrent une série d'expérimentations répétées.

Néanmoins, en dehors de cet aspect, le rôle de la théorie est relativement limité. Il y a un nombre réduit de références théoriques, une absence relative de réflexion conceptuelle, points sur lesquels je ne vais pas m'attarder.

Il manque également une théorie de l'acteur et de la connaissance véritablement articulée, déployant un ensemble cohérent de propositions sur les formes de la rationalité et les processus cognitifs. Il existe des éléments sur les biais cognitifs, sur la procrastination, etc. Mais tout cela ne constitue pas système théorique pour autant.

Le rapport à la théorie est fréquemment sous-déterminé ou partiel dans les explications des résultats obtenus, parce que ces résultats s'écartent souvent de ce qui est prévu par les théories, le plus fréquemment *mainstream*. Les explications des mécanismes causaux restent bien souvent embryonnaires, voire anecdotiques. Des chaînons théoriques sont manquants dans l'argumentation du J-PAL, ce qui en limite la portée. Sans théorisation approfondie des mécanismes à l'œuvre, les effets de connaissance sont réduits. Il ne suffit pas de répliquer les expérimentations pour accumuler les connaissances.

Une autre question est ambiguë : celle de l'objectif scientifique de ces évaluations. Toute une série d'affirmations relativement vagues selon lesquelles le contexte peut avoir une influence, coexistent avec des postulats contradictoires selon lesquels des incitations produisent exactement les mêmes résultats, quel que soit l'endroit.

Le J-PAL vise-t-il à dégager des lois économiques universelles valables en tout temps et en tout lieu ou simplement des régularités situées dans le temps et dans l'espace ?

Ce n'est pas un hasard si un cas paradigmatique est toujours mis en avant. Ce sont les vermifuges, avec les travaux de Kremer et Miguel de 2004 notamment. Dans ces expérimentations, on est en présence d'un objectif intermédiaire et d'un remède relevant du domaine médical. Cela repose sur une

relation physiologique universelle dans les zones où ces vers sont présents. Les vermifuges permettent d'améliorer la santé des enfants, qui, ainsi, peuvent aller en bonne santé à l'école. Il existe une adéquation parfaite entre cette technique et le domaine de l'expérimentation.

C'est très différent d'autres expérimentations où les structures sociales, les processus d'apprentissage, les institutions jouent un rôle fondamental. Cela pourrait expliquer les raisons pour lesquelles les mêmes expérimentations conduites dans des pays différents aboutissent à des résultats différents, tout comme des expérimentations conduites dans un même pays mais auprès de professions différentes comme des infirmières et des enseignants.

D'autres ambiguïtés concernent les processus de généralisation. Duflo et Banerjee concèdent bien volontiers que leurs expérimentations microéconomiques ne peuvent être extrapolées au niveau macro en raison d'effets d'équilibre général. Mais il demeure une ambiguïté parce qu'ils partent d'un modèle de mécano où ils expliquent que les blocs microéconomiques sont les blocs fondamentaux. Ils partent donc de cette idée, qui ne tient pas épistémologiquement, que l'on peut passer par un agent représentatif pour aller du micro au méso et au macro. C'est donc une forme de réductionnisme micro qui laisse dans l'ombre les phénomènes d'émergence.

Or, des éléments majeurs, en matière de développement de lutte contre la pauvreté ou de politique de l'emploi, nécessitent d'être analysés à l'échelle macro ou méso. Cela m'amène à un point très important : ces essais randomisés constituent un outil d'isolation de l'effet d'une action spécifique, mais ils ne saisissent pas les interactions. C'est la même chose en pharmacologie où l'on sait très bien que les essais cliniques ne permettent pas de saisir les interactions médicamenteuses qui sont une cause très forte de mortalité : environ 7 000 personnes aux Etats-Unis chaque année.

Ces essais cliniques laissent de côté ce que l'on appelle les effets de synergie, de potentialisation et d'antagonisme entre traitements, effets qui existent également en économie. L'économie passe par des processus complexes, non linéaires. Il y a des combinaisons de facteurs, des effets de seuil, etc. Les micro-dispositifs ne suffisent pas, même quand ils sont étendus à des pays entiers. Il y a loin du vermifuge au développement d'un pays.

Autre limite : la sensibilité des résultats à la durée de l'observation. On le sait en matière d'essais

cliniques. Après les essais cliniques, après l'autorisation de mise sur le marché des médicaments, des effets indésirables apparaissent. C'est le cas pour un médicament sur 5, ce n'est pas négligeable. C'est la même chose pour la lutte contre la pauvreté qui s'inscrit également dans la durée. Toute une série d'expérimentations montrent que les résultats évoluent dans le temps, notamment en matière de microcrédit, de promotion des cultures destinées à l'exportation. Cela rejoint des expérimentations menées dans les années 60 et 70 en matière de politique de l'emploi. Cela rejoint aussi ce qui s'est passé pour le RSA en France.

Cette sensibilité à la durée de l'expérimentation est cruciale face à des acteurs qui apprennent, interagissent et évoluent, de même que le cadre institutionnel des activités économiques. Or, là, on est face à un problème fondamental. En raison des coûts conséquents des expérimentations, celles-ci sont souvent conduites sur une durée limitée, d'où la nécessité de les assortir d'un suivi, tout comme la pharmacovigilance est cruciale dans le domaine médical. Malheureusement, c'est le parent pauvre. Il ne faudrait pas que cela le soit pour les expérimentations en économie.

Autre limite fondamentale : la question des objectifs intermédiaires pour les essais cliniques que l'on va utiliser plutôt que l'expérimentation des résultats par rapport à l'objectif final, et ce pour accélérer le processus. C'est le cas, par exemple, dans le domaine des accidents cardio-vasculaires. On va prendre comme objectif intermédiaire la baisse de l'hypercholestérolémie, parce qu'elle est associée à une réduction du risque cardio-vasculaire, pour une classe de molécules appelées les statines. Cela pose problème quand on va tester des molécules qui ne font pas partie de cette classe ; par exemple, l'Ezetimibe, approuvée en 2002 par la FDA. Il s'avère que ce médicament n'offre pas les mêmes bénéfices que les statines. Il y a des mécanismes causaux sous-jacents différents, et des études cliniques récentes indiquent même un lien potentiel avec un risque fortement accru de cancer.

C'est à mon avis la même chose en économie du développement. Les expérimentations à sélection aléatoire se concentrent sur les objectifs intermédiaires réputés servir le développement, par exemple l'utilisation accrue d'engrais chimiques par les agriculteurs. On va supposer que cette utilisation accrue va être un facteur de développement. C'est l'objectif final pour les agriculteurs, *via* l'augmentation des rendements occasionnée qui peut se traduire par une augmentation des revenus.

Or, il n'existe pas de garantie que cette chaîne causale soit toujours observée. Le problème est que cette focalisation sur des objectifs intermédiaires se traduit bien souvent par une limitation des débats. Il n'existe pas de lien mécanique entre la réalisation de l'objectif intermédiaire et celle de l'objectif final. La pertinence des objectifs intermédiaires est insuffisamment mise en débat. Ceux-ci sont présentés comme données et les alternatives ne sont pas envisagées. En l'occurrence, pourquoi ne pas envisager notamment l'introduction de filières biologiques qui utiliseraient moins d'engrais ?

En conclusion, s'agit-il d'un nouveau *gold standard* en économie ? Il s'agit d'un complément très utile. Ce n'est en rien un substitut à une mésoéconomie ni à une macroéconomie. Même si le niveau de preuves fournies par ces expérimentations par sélection aléatoire est élevé, les expérimentations ne s'appliquent qu'à un nombre limité de situations. Il ne faudrait pas sacrifier des preuves pertinentes mais de niveaux moins élevés à des preuves de niveau élevé. Se pose la question de la concurrence entre preuves. Il existe un risque d'effet d'éviction à l'égard d'autres approches. Cela devient un nouveau *one best way* qui truste énormément de financements en matière de lutte contre la pauvreté et de développement.

D'autres pratiques de ces expérimentations par assignation aléatoire sont envisageables. Il serait par exemple souhaitable de compléter la démarche du J-PAL par des évaluations de type processus d'apprentissage collectif, par des évaluations visant à identifier plus rigoureusement les mécanismes causaux et construire une théorie d'ensemble. Pour ce faire, d'autres usages de ces expérimentations qui seraient moins *top-down*, moins *mainstream*, pourraient être promus dans le cadre d'une collaboration étroite avec les acteurs de terrain au sens large et les sciences sociales autres que l'économie. Il faudrait enfin clarifier le processus de construction de l'objet et les limites des expérimentations, et ne pas se réfugier derrière une certaine forme de scientisme. Quand Esther Duflo déclare : « *Les évaluations sont rigoureuses, elles ne laissent aucune place à l'interprétation* », c'est une position difficilement tenable compte tenu des débats en sciences dures, et en médecine particulièrement.

Débat avec la salle

Annie FOUQUET

Ces trois exposés sont trois exemples de l'usage de la statistique pour éclairer le débat public qui nous

chiffonnent un peu. L'exemple du pilotage interne avec la chaussette, la question de l'expérimentation qui devrait être un standard et qui ne l'est pas pour les raisons montrées par Agnès, et une remise en cause de la statistique probabiliste, dans un monde complexe, en perpétuelle soumission à des aléas incontrôlés.

J'ouvre le débat. Qui veut intervenir sur tel ou tel de ces points, exposer des expériences personnelles ou réintervenir ?

Elisabeth REIGNIER

Je me pose une question par rapport à ce qui se passe dans la police. Nous avons ici des propos d'experts. Comment les syndicats de police s'approprient-ils ces propos d'experts ? Comment font-ils le lien avec les éventuelles revendications des policiers soumis à ces évaluations ? Que font-ils de cette synthèse pour lutter contre cette normalisation que l'on peut juger par moment excessive des organisations sociales ?

Michel EURIAT

J'étais président du Comité du label du Cnis récemment. Je voudrais dire tout l'intérêt que j'ai trouvé dans le troisième exposé. Plusieurs enquêtes ont été présentées au Comité du label, mais uniquement la partie enquête auprès des différents échantillons, dans le cadre d'expérimentations contrôlées. Je suis intervenu au Cnis pour cela. J'ai considéré que c'était une sorte d'instrumentalisation car l'enquête en elle-même était très bien conçue par les professionnels, mais elle ne constituait qu'un aspect tout à fait partiel de tout le processus. C'est tout le processus qu'il aurait fallu voir pour porter un jugement, d'autant plus que se posaient, un peu comme en médecine, les questions d'exclusion. À partir du moment où quelqu'un était dans un échantillon ou un autre, il ne bénéficiait pas de la mesure que l'on était sensé expérimenter.

Je plaide encore une fois, même si je ne suis plus au Cnis, pour que l'ensemble du dispositif soit présenté, qu'il puisse y avoir le débat nécessaire autour de chaque opération de ce genre, quelle que soit son utilité par ailleurs.

Mireille ELBAUM

Je voudrais faire deux remarques. Je ne vais pas développer la première, elle va dans le même sens que ce qu'a dit Annie. Il est dangereux d'assimiler indicateurs de performance et évaluation. Les démarches d'évaluation lancées dans les années 90 se distinguaient très profondément de ces

indicateurs de performance qui en sont une dérive pathologique : on voit bien qu'ils sont partiels, manipulables et qui, s'ils ne le sont pas au départ, le deviennent du fait de la focalisation des acteurs et de l'interaction entre eux. L'évaluation est fort heureusement bien autre chose, on va très vite le voir. Je l'avais analysé de mon côté sur les indicateurs de la Lolf en matière sociale. Je regrette vraiment cette assimilation.

Le deuxième point porte sur les expérimentations contrôlées, qui deviennent effectivement très importantes dans le domaine social. C'est une technique comme une autre qui, statistiquement, a des avantages, mais qui ne saurait en rien être l'alpha et l'oméga de l'évaluation. Tous les moyens ne doivent pas être concentrés là-dessus.

Regarder et évaluer ces expérimentations, pourquoi pas ? En déduire à pieds joints des choses sur la généralisation constitue vraiment un abus et encore plus en matière sociale : d'une part parce qu'en matière économique et sociale on a les effets de cannibalisme, de concurrence, les effets d'aubaine extrêmement importants qui jouent sur les résultats ; d'autre part, parce que l'on a ce que l'on appelle l'effet conforme : le fait que les gens qui participent à l'évaluation, acteurs comme expérimentés, adoptent des comportements différents du fait même d'être dans l'expérimentation. Évaluer la chose et la regarder, pourquoi pas. Mais sauter à pieds joints sur la généralisation pose un autre type de problème.

J'en viens à un point important dont il me semble que vous n'avez pas parlé. Ce n'est pas tant l'utilisation scientifique en matière de politique sociale de ces expérimentations qui me gêne que leur utilisation politique. Ce que l'on a vu, notamment avec le RSA, est vraiment l'exemple pur d'une utilisation politique déviée du résultat, de la notion d'expérimentation, de ce que l'on pouvait apprendre de cette notion d'expérimentation. C'est absolument splendide, parce que l'on a d'abord pris des décisions de généralisation alors que les expérimentations n'étaient pas terminées, *a fortiori* pas évaluées.

Ce qui est extraordinaire, c'est que l'on a, comme d'habitude, focalisé l'observation sur les comportements d'incitation à l'activité des bénéficiaires de minima sociaux. On a complètement laissé de côté la novation majeure du RSA qui consistait à s'adresser à d'autres types de populations que l'on ne connaissait absolument pas,

à savoir les travailleurs pauvres qui n'étaient pas dans le champ des minima sociaux.

On s'aperçoit maintenant que tous les comportements de non-recours sont concentrés sur cette population et liés au fait, en grande partie, qu'on a cru qu'ils allaient accepter spontanément de présenter des dossiers d'aide sociale dans lesquels ils ont besoin de justifier de leur situation familiale, de leur situation de ressources dans les détails. On n'a même pas eu l'idée que, concernant cette population que l'on ne connaissait pas, c'était ce qu'il fallait regarder avant de mettre en place ce type de politique, c'était cela qui allait poser problème. On s'est focalisé sur des comportements vus, revus, sur-examinés, qui polarisent toute l'attention des pouvoirs publics, à savoir les comportements d'activité des bénéficiaires de minima sociaux.

C'est plein d'enseignements, pas tant sur les expérimentations et leur évaluation, qui a été ce qu'elle est dans les délais impartis, mais sur la façon dont les pouvoirs publics se sont ou non mal emparés de ces résultats d'expérimentations.

Emmanuel DIDIER

En réponse à la première question, la réaction des syndicats a été assez surprenante. J'ai pris rendez-vous avec eux. Les syndicats de commissaires ont des intérêts différents des syndicats d'officiers. Ils sont intéressés par cette technologie. Le déclencheur a été le fait d'être payé à la prime, car les policiers sont de plus en plus payés à la prime. Ils préféraient donc largement que ces primes soient distribuées en fonction de résultats quantifiés plutôt que de résultats non quantifiés.

Par ailleurs, je trouve délicat, Madame Elbaum, de considérer que l'on peut faire une distinction radicale entre indicateurs de performance et évaluation. Là, en l'occurrence, il est difficile de faire une distinction entre les deux.

Mireille ELBAUM

Tout dépend de ce que l'on évalue. Je parle de l'évaluation des politiques publiques, pas d'un commissariat en tant que tel.

Emmanuel DIDIER

C'est très délicat. Cela paraît difficile.

Nicolas FLEURY

J'avais une question en lien avec les syndicats de la société civile. Les différentes politiques du chiffre et de l'évaluation présentées par les intervenants incitent-elle les syndicats ou la société civile à se

doter de compétences en termes d'évaluation de politiques publiques ? Ces organisations syndicales ou la société civile ont-elles une demande spécifique en termes d'évaluation ?

Florence AUDIER

Je voulais insister sur un aspect qui n'est pas suffisamment évoqué : l'influence de ces indicateurs, des procédures d'évaluation sur le contenu même du service public. Celle-ci est pourtant directe. À mesure que des indicateurs apparaissent et à partir du moment où ils prévalent, on le voit dans la police et dans d'autres domaines, le contenu même du service rendu au public, et donc la relation avec l'usager, en est complètement modifié.

Si, par exemple, le chiffre retenu est le nombre de gardes à vue, on voit que la fonction de la police change. C'est la même chose dans les domaines de la santé, de la justice, etc. Tous ces effets pervers - plusieurs livres sont sortis sur ce sujet - montrent bien que c'est un point fondamental. C'est pourquoi certains indicateurs sont retenus et pas d'autres.

Cela me permet de passer aisément à l'approche expérimentale et aux travaux de Mme Duflo. Elle insiste beaucoup, notamment dans ses deux petits livres, sur le fait que, finalement, la politique est exclue de tout cela. Il n'y a plus de choix politique, de débat politique, il y a des expériences. On tire de ces expériences des politiques publiques. C'est autre chose qu'un débat politique. C'est appliqué sur le développement mais aussi sur bien d'autres domaines, y compris chez nous. Il y a les expérimentations de Martin Hirsch. Les portées sont plus ou moins lourdes, graves, irréversibles. Et le constat qui me semble être incontournable est que la politique est éjectée de tout cela.

Fabrice BARDET (Université de Lyon)

Dans l'intervention de Mme Elbaum, je n'ai pas saisi la distinction entre expérimentations scientifiques et expérimentations politiques. Je me posais la question de la responsabilité des scientifiques. Je rapportais le débat à notre assemblée et je rebondissais sur l'intervention d'Agnès Labrousse, qui a souligné l'absence de cadrage théorique de la réflexion, de l'expérimentation telles qu'elles se développent au J-PAL. Les conséquences de cette absence de cadrage théorique sont colossales. Cela renvoie à l'invite de l'intervenant précédent, qui estimait qu'il était temps de repenser les conditions épistémologiques de la raison, formule qui m'a bien plu. Il semble qu'un courant nous emmène vers des endroits où l'on n'a pas envie d'aller.

Je voulais renvoyer cette problématique de l'absence de cadrage théorique au fait que la structure du J-PAL est une structure dont les financements doivent nous poser un certain nombre de questions, avec une entreprise s'opposant à d'autres entreprises qui défendent un autre bien public, celui de la statistique. Tu l'as souligné, Agnès, dans la fin de ton propos, en disant que cela pose les problèmes de financement de la recherche. Tu évoquais les régressions en coupe internationale qui nécessitent un certain nombre de financements, même si elles ont des problèmes. Ce sont des financements trustés par ces entreprises privées que l'on voit émerger, qui se présentent de manière hégémoniques comme défendant un *gold standard*. Dans les ouvrages de Duflo, les précautions prises par rapport à ce risque d'hégémonie sont très relatives, très problématiques.

Par rapport au titre de notre réunion, cette question mérite un mot.

Vincent GUILLON

J'étais content d'entendre l'ensemble des interventions. La troisième m'a particulièrement intéressé. L'année dernière, on a assisté aux expérimentations de Martin Hirsch, inspirées par Esther Duflo. Cela m'a scandalisé à plusieurs niveaux. On connaît depuis 1927 ou 1929 les expériences d'Elton Mayo. Dès 1936, on avait en psychothérapie des gens qui disaient : tous les traitements sont équivalents. C'est l'effet dodo : tous ont concouru, tous peuvent avoir des prix.

De plus, venant de la part de scientifiques, il existe des méta-analyses montrant qu'une hirondelle ne fait pas le printemps. On ne s'arrête pas sur une expérimentation, avec tous les effets d'attente que l'on a, pour généraliser, alors que l'on n'a encore que des objectifs intermédiaires. C'est un non-sens scientifique absolu ! Se pose la question des ordres de grandeur : en psychothérapie ou en médecine, où le placebo est aussi important, quand il peut jouer, que le traitement, on peut obtenir au mieux deux points d'élévation du niveau de la moyenne d'une classe, de 10 à 12, quand quelque chose marche bien. C'est la moyenne.

Et puis, -et là c'est plutôt un plaidoyer pour continuer dans les statistiques- quand on va jusqu'au bout des réflexions, et il en existe beaucoup en psychologie clinique, en psychothérapie, autant qu'en médecine, on redécouvre au bout du compte l'importance du facteur humain, l'alliance de travail, le thérapeute, le conseiller. Les chiffres le montrent par les méta-analyses. Le pourcentage de variance expliquée est toujours 8 à 10 fois plus important, ce

que peut expliquer un psychothérapeute. On est dans la nature humaine. Mais il ne faut pas renoncer aux chiffres. Il faut bien sûr dénoncer les usages politiques, mais ce serait une erreur de... je n'ai pas envie d'utiliser l'expression « *jeter le bébé avec l'eau du bain* » parce qu'elle est trop souvent employée par les tenants des modèles médicaux !

Ana MARQUES

Ma réflexion concerne les implications entre le recueil de données et les restitutions aux personnes qui font ces recueils dans le cadre particulier d'un établissement de santé. Quand cela part dans les ministères, que l'on ne sait pas ce qui en est fait, que l'on n'a pas de retour, les gens du terrain ne veulent pas recueillir. On a des variations extrêmement importantes, de l'ordre du triple, d'actes cotés de façon différente en fonction des endroits. Que pouvez-vous en dire, par exemple dans le cas de la police ? Que pouvez-vous dire des usages que font les policiers eux-mêmes de ces données, en dehors des usages par des politiques ou d'autres experts ?

Annie FOUQUET

Nous allons essayer de répondre à ce lot de questions.

Sur les compétences syndicales en matière d'évaluation, je me permets une publicité. La Société française de l'évaluation a tenu des réunions avec chacune des organisations syndicales et patronales représentatives pour leur demander, compte tenu de l'inscription de l'évaluation dans les accords nationaux interprofessionnels récents et dans la loi de novembre 2009, si cette association qui est un lieu d'échange pouvait les aider. Nous organisons la semaine prochaine, le 6 avril, au Conseil économique, social et environnemental, une journée sur le thème « *cheminer dans l'évaluation* », avec deux exemples d'évaluation de la formation professionnelle pour discuter sur ces sujets d'évaluation avec les partenaires sociaux. Si cela vous intéresse, vous pouvez aller sur le site www.sfe-asso.fr pour vous inscrire. Car, sans inscription, on ne peut pas rentrer. J'ai rencontré les syndicats : ils sont à des niveaux de réflexion différents, mais avec une appétence réelle.

Je vous laisse répondre.

Agnès LABROUSSE

Merci pour ces compléments tout à fait importants sur les expérimentations, l'effet Hawthorne et surtout leur dimension politique. J'abonderai dans votre sens. C'est un enjeu fondamental que de

replacer les expérimentations dans une vision d'ensemble du processus d'acquisition de connaissances, en associant ces expérimentations à d'autres méthodes, avoir un pluralisme de ces méthodes précisément pour éviter ces effets de focalisation, de débats tronqués que l'on a eus de manière flagrante dans le cas du RSA. En exposant le processus de construction de l'objet, en restituant une dimension interprétative, en ne la confisquant pas en se faisant passer pour un expert neutre et objectif, cela permet de redonner une place au débat politique et au débat démocratique. Cela me semble fondamental. Il faut sortir de cette exclusion du débat politique, de cette vision en termes d'expert neutre, extérieur au choix, pour replacer l'expert dans le débat démocratique, de manière très large. Il était question, à propos des hôpitaux, de restitution. La restitution est aussi une méthode utilisée en sciences sociales. Elle pourrait être mise en place dans le cadre de ces expérimentations.

La remarque de Fabrice Bardet sur la question du financement de la recherche et la dimension multinationale de la recherche du J-PAL est juste. C'est devenu une multinationale de la recherche avec des filiales, des organismes affiliés, dont le Crest. Ce qui peut être problématique, c'est que le J-PAL a un quasi monopole scientifique de ces expérimentations par assignation aléatoire. Il faudrait introduire plus de pluralisme dans ce domaine.

J'ai découvert l'effet dodo. Je connaissais l'effet Hawthorne, l'effet placebo, mais pas l'effet dodo. Donc je vous remercie.

Merci également d'avoir parlé de méta-analyses. C'est très présent dans l'*evidence-based medicine*, et cela demanderait à être introduit dans les expérimentations en matière de lutte contre la pauvreté ou de politique de l'emploi.

Merci d'avoir introduit la dimension clinique, la question du colloque singulier entre le patient et le médecin. L'*evidence-based medicine* n'a jamais voulu abolir d'autres formes de preuves médicales, dont celles issues de ce colloque singulier.

Je vous remercie encore pour vos compléments.

Emmanuel DIDIER

Un point sur les usages locaux des données : c'est une question très intéressante. Souvent, les données formatées pour ce genre de mécanismes d'évaluation internes sont difficilement utilisables par les acteurs eux-mêmes pour autre chose que cette évaluation. S'ils ont besoin de données pour

faire autre chose, très souvent de nouvelles bases de données apparaissent. Si vous avez envie de vous opposer, d'avoir des revendications syndicales, c'est une autre base de données qui est mise en place.

Daniel LE SCORNET

Je vais être trivial pour terminer parce que j'ai fait quelque chose d'un peu austère. Sur l'expérimentation, des choses m'ont choqué profondément, politiquement. Comment se fait-il que cette loi sur l'expérimentation, qui était justement en direction des acteurs de terrain, sociaux, etc., soit déjà monopolisée par l'Etat ? C'est une loi sur l'expérimentation où seules les expérimentations d'Etat sont prises en compte. Aucun acteur de la société civile n'est redevable de ces expérimentations.

J'ai été un des initiateurs de la couverture maladie universelle : autant on voyait pourquoi on pouvait la généraliser, autant il existe un manque au niveau statistique qui me laisse pantois concernant cette histoire d'expérimentation sur le RSA jeune : il nous manque des statistiques de politiques publiques que l'on devrait avoir, comme le nombre de personnes qui devraient bénéficier de la mesure. La complémentaire de la CMU plafonne à 4,2 millions de bénéficiaires. Or, normalement, plus de 6 millions de personnes devraient en bénéficier. On devrait avoir une statistique qui parte des politiques publiques et non pas des personnes qui les utilisent. Il y a un manque extraordinaire de ce point de vue.

Sur le RSA jeune, il aurait été facile de savoir dès le départ, avec les conditions que l'on mettait, le nombre de jeunes pouvant y accéder. On aurait vu tout de suite le côté absurde d'une telle mesure. Quand, dans les Hautes-Alpes, on me dit qu'il y a un seul titulaire du RSA jeune, on ne sait même pas combien il devrait y en avoir. Il serait intéressant de le savoir.

Il y a des politiques explicites. Ce sont des décisions politiques fermes, on utilise ensuite les techniques.

Le dernier point concerne le facteur humain. Évidemment, c'est la grande question. S'agissant de la causalité au Japon, d'après les informations que j'ai - je suis ces problèmes de près - ce n'est pas du tout cela. Lorsque le problème des réacteurs a commencé, on risquait d'être hors contrôle. Le facteur humain, dans des situations comme celle-là, hyper critiques, ne doit pas jouer parce qu'il va avec l'émotivité et le lucre - essayer de sauver les réacteurs. Il était facile de noyer les réacteurs dans les premières heures. Or, il y a eu décision de ne pas

le faire, en contradiction avec toutes les règles de la sûreté nucléaire.

Le facteur humain, guidé par le lucre, est décisif.

Je terminerai par une citation d'Adorno qui dit :
« *L'objectivité dans les relations entre les hommes qui fait place nette de toute enjolivure idéologique est déjà devenue elle-même une idéologie qui nous invite à traiter les hommes comme des choses.* »

Annie FOUQUET

C'est une belle conclusion, nous allons en rester là.