



HAL
open science

Évaluation des politiques publiques : faut-il de l'interdisciplinarité ?

Etienne Wasmer, Christine Musselin

► **To cite this version:**

Etienne Wasmer, Christine Musselin. Évaluation des politiques publiques : faut-il de l'interdisciplinarité ?. 2013. hal-02140455v2

HAL Id: hal-02140455

<https://sciencespo.hal.science/hal-02140455v2>

Preprint submitted on 29 Jul 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



SciencesPo.

LIEPP

Laboratoire interdisciplinaire d'évaluation des politiques publiques

Évaluation des politiques publiques : faut-il de l'interdisciplinarité?

Methodological Discussion Paper

Septembre 2013, n°2

Etienne Wasmer

Sciences Po, LIEPP / Département d'économie
de Sciences Po

Discussion par **Christine Musselin**

Centre de sociologie des organisations
(CSO)

Sciences Po | LIEPP
27 rue Saint-Guillaume
75337 Paris Cedex 07
Tel : 01 45 49 83 61
www.sciencespo.fr/liepp

Évaluation des politiques publiques Faut-il de l'interdisciplinarité?

Etienne Wasmer

Section 1. Introduction¹

L'évaluation des politiques publiques est un grand enjeu démocratique. Les dépenses publiques dépassent la moitié du PIB et la question de leur efficacité est centrale : les dépenses qui ne sont pas « efficaces » pour la collectivité, en un sens qu'il convient de définir avec soin, sont *de facto* un simple transfert de ressources vers des groupes constitués. Comprendre l'inefficacité d'une politique permet dès lors de rendre transparent le fait qu'il s'agisse d'un transfert de ressources. Pour ne prendre qu'un exemple, l'argument principal avancé en faveur des aides à la pierre est qu'elles favorisent le cycle économique (« quand le bâtiment va, tout va »). Mais l'efficacité de ces aides est rarement évaluée, *a fortiori* jamais comparée à l'usage alternatif des fonds publics, qui pourraient servir à alléger les charges et les impôts dans tous les secteurs ou à favoriser les investissements de long terme en éducation, ou en recherche et développement.

Par ailleurs, des dépenses publiques qui ne contribuent pas à l'efficacité et dont le rendement propre est inférieur au taux d'intérêt augmentent la dette publique. Cette dette est passée aux générations futures et sa gestion est un des éléments du contrat intergénérationnel qui lie les différentes parties de la collectivité nationale : ces générations doivent donc connaître le rendement des dépenses publiques.

Enfin, l'efficacité de la dépense publique dans un contexte de crise des dettes souveraines en Europe prend une nouvelle importance : lorsque des coupes budgétaires drastiques doivent avoir lieu comme dans certains pays sud-européens, l'idéal serait d'avoir assez d'information pour sanctuariser les dépenses dont le rapport coût bénéfice est le plus avantageux.

Ces principes semblent aller de soi pour la plupart des chercheurs en sciences sociales mais se pose alors la question du choix de la méthodologie d'évaluation. Plus précisément, dans cette contribution, nous essaierons de répondre à la question : quelle est la nécessité d'éclairages multiples sur les politiques publiques par différentes disciplines au sein des sciences humaines, c'est-à-dire l'interdisciplinarité ? Faut-il aller jusqu'à un effacement des frontières disciplinaires dans le cas de la transdisciplinarité?²

1 Ce texte est issu d'une présentation en session plénière au colloque Colloque Interdisciplinaire « Aménagement - Droit - Economie - Géographie - Sociologie - Statistiques » intitulé *Territoires, Emploi et Politiques Publiques* qui s'est tenu à l'Université de Metz les 23 et 24 juin 2011. L'article qui en est issu est un témoignage de gratitude envers Pierre Morin, ancien collègue de l'Université de Metz, qui m'a dès 2000 fait comprendre tout l'intérêt qu'il y avait à réunir les enseignements de différentes sources disciplinaires avec une grande ouverture d'esprit sans sacrifier jamais à la rigueur intellectuelle, réunissant en cela les qualités des deux types d'experts décrits dans l'ouvrage de P. E. Tetlock, sans en avoir les défauts. La plupart des idées discutées dans ce texte sont le produit de rencontres et d'échanges privilégiés avec des coauteurs, des collègues du département d'économie de Sciences Po et des autres disciplines. Je tiens particulièrement à remercier Cornelia Woll qui m'a mis sur la piste féconde de P. E. Tetlock et qui est à l'origine de plusieurs des idées de ce texte, de Christine Musselin, Maurice Baslé, Etienne Lalé et Mirna Safi, pour leurs commentaires pertinents et parfois (justement) critiques, ainsi que deux rapporteurs anonymes pour leurs suggestions. Ce texte est en particulier centré sur les interactions entre économistes et non-économistes mais ne peut pas être nécessairement appliqué aux relations entre d'autres disciplines (médecine et sociologie, anthropologie et démographie etc.) même si un certain nombre de conclusions pourraient être généralisées. Je remercie enfin Guillaume Vuillemeys et Pierre Gouëdard pour leur relecture du texte. Les insuffisances restent bien entendu celles de l'auteur. Enfin, ce texte n'implique en rien les autres membres du LIEPP et n'a que la valeur d'un essai.

2 Les disciplines créent des frontières (Jacobs 2008) souvent fondées sur les méthodes ou les paradigmes. Dans le cas d'approche croisant des regards disciplinaires multiples, l'on peut proposer comme Christine Musselin, les distinctions suivantes : *pluridisciplinarité* (juxtaposition de connaissances, d'analyses, de méthodes entre plusieurs disciplines), *interdisciplinarité* (échange de connaissances, d'analyses, de méthodes entre plusieurs disciplines) et *transdisciplinarité* (effacement des frontières entre les disciplines). Jacob (2008) définit la *multidisciplinarité* comme des recherches opérant en

Cette question doit être posée car, s'il existe d'une part une littérature assez vaste sur l'évaluation et l'interdisciplinarité en dehors de la science économique, et d'autre part des méthodologies d'évaluation depuis longtemps développées au sein de la science économique, la science économique ne s'est pas réellement emparée de cette question de l'interdisciplinarité: elle semble déjà avoir suffisamment à faire avec la résolution des débats entre expériences contrôlées et estimations de modèles structurels.³

Dans cette contribution, nous ferons dans un premier temps un détour par la science politique et la psychologie (**Section 2**) pour résumer les enseignements d'un ouvrage important de Philip E. Tetlock (2005) sur la capacité prédictive des experts. Cette recherche souligne que des experts de science politique et d'économie semblent avoir de moins bons résultats prédictifs lorsque qu'ils sont plus spécialisés. Dans un second temps (**Section 3**), tout en reconnaissant que l'évaluation ne se limite pas à la prévision, nous discuterons de l'extrapolation de ce résultat à l'évaluation des politiques publiques ; puis nous passerons en revue les coûts de transaction générés par les échanges interdisciplinaires et discuterons des principes d'organisation de la science et de la tension qui existe entre le besoin de spécialisation pour la gestion des carrières et les gains de l'ouverture interdisciplinaire (**Section 4**). En conclusion (**Section 5**), nous résumons les obstacles et suggérons que le but ultime de l'évaluation des politiques publiques devrait être de pouvoir quantifier au sein de modèles de prévision ex ante, la subtilité du contexte et les enjeux de la mise en œuvre des mesures de politiques publiques, telles qu'elles sont analysées par les différentes sciences sociales. Cet article reste à ce stade prospectif : il contient de nombreuses spéculations et fait part d'expériences concrètes de la démarche interdisciplinaire. Il n'a d'autre but que de susciter une réflexion de fond sur les enjeux scientifiques liés à l'évaluation des politiques publiques.

Section 2. La prévision politique et stratégique analysée par un psychologue

Mesurer le succès (ou l'insuccès) prédictif des experts

Philip E. Tetlock, du département de psychologie de U. Penn, s'est lancé dans un projet très ambitieux consistant à suivre, au travers de questionnaires entre 1987 et 2003, plus de 20 000 prévisions de 284 « experts », puis à les analyser rétrospectivement. Dans un ouvrage magistral et original, « Expert Political Judgment » paru en 2005, il a résumé ses travaux, et cet ouvrage a été récompensé par le prix Woodrow Wilson de l'*American Political Science Association* en 2005. Selon Daniel Kahneman⁴, il s'agissait d'une avancée considérable dans la connaissance sur le jugement des experts en matière d'analyse et de prévision stratégiques.

Un expert est défini dans cet ouvrage comme « un professionnel qui vit de commentaires ou de conseils sur les tendances économiques et politiques sur des Etats, des régions du monde ou sur le système international dans son ensemble ». 76% de ces experts étaient des hommes, en moyenne de 43 ans. 52% étaient titulaires d'un diplôme doctoral. 41% étaient universitaires, 26% travaillaient pour le gouvernement, 17% pour des *think tanks* et des fondations, 8% pour des institutions internationales, et 8% dans le secteur privé. Ils ont été interrogés sur des sujets tels que la stabilité du pouvoir politique de différents pays (60 au total), leurs performances économiques et la politique de défense. Ces prévisions portaient majoritairement

parallèle mais sur des sujets similaires, comme la définition de la pluridisciplinarité ci-dessus et insiste, dans le cas de l'interdisciplinarité, sur les collaborations à partir de perspectives différentes.

3 La diffusion de ces méthodes en France s'est accélérée récemment, comme en témoignent par exemple les deux numéros spéciaux de 2011 et 2012 de la Revue Française d'Economie sur l'évaluation des politiques publiques. Il est intéressant de constater qu'aucun des articles de ces deux numéros ne fait appel à une autre discipline ni n'en évoque. Les deux introductions de ces numéros (Langot et Petit, 2011 et 2012 ou l'article de synthèse de L'Horty et Petit 2011) se consacrent surtout au débat « randomisation » ou « estimations structurelles » et sur les questions éthiques des méthodes d'assignation aléatoires, mais n'abordent pas cette question de l'interdisciplinarité, quand bien même plusieurs des auteurs de ces deux numéros la pratiquent parfois sur des sujets ponctuels.

4 « *It is a major advance in our understanding of expert judgment in the vitally important and almost impossible task of possible and strategic forecasting.* » tel que reproduit sur la 4^{ème} page de couverture de l'ouvrage de P.E. Tetlock.

sur leur zone géographique de compétence, leurs prévisions portant alors le nom de prévisions « expertes ». Certaines des questions étaient posées, à titre de comparaison, sur un ou deux pays éloignés de la zone de compétence des experts, leurs prévisions étant alors qualifiées de « dilettantes ».

La méthodologie de Tetlock a consisté tout d'abord à codifier le degré de certitude d'une prédiction d'un expert sous forme de probabilité subjective. Tetlock classe ainsi les prédictions des experts sur une échelle allant de 0 à 1. Le degré de certitude peut être la certitude absolue ($p_s=1$) ou à l'autre extrême l'impossibilité perçue de la réalisation du scénario analysé ($p_s=0$), en passant par toutes les valeurs intermédiaires : une probabilité subjective $p_s=0.1$ est associée à « très improbable », $p_s=0.33$ à « assez improbable », etc.

La seconde étape consiste à mettre ces prédictions en correspondance avec le résultat des prédictions. Dans un environnement complexe comme celui de la science politique, au contraire des sciences exactes, il est difficile d'obtenir une certitude absolue sur le succès ou l'échec prédictif d'un expert. Par exemple, savoir si le déploiement d'un programme stratégique de parapluie nucléaire aux États-Unis (la guerre des étoiles) a *in fine* précipité la chute de l'Union Soviétique est une question complexe. Cette proposition semble plausible avec le recul du temps, mais elle n'est pas certaine. En analysant la littérature et différentes sources de données économiques et politiques, Tetlock choisit donc d'associer également une probabilité objective aux prédictions, avec le recul du temps et *a posteriori*. L'ensemble de la procédure est décrite dans l'annexe de l'ouvrage, pages 239 à 311.

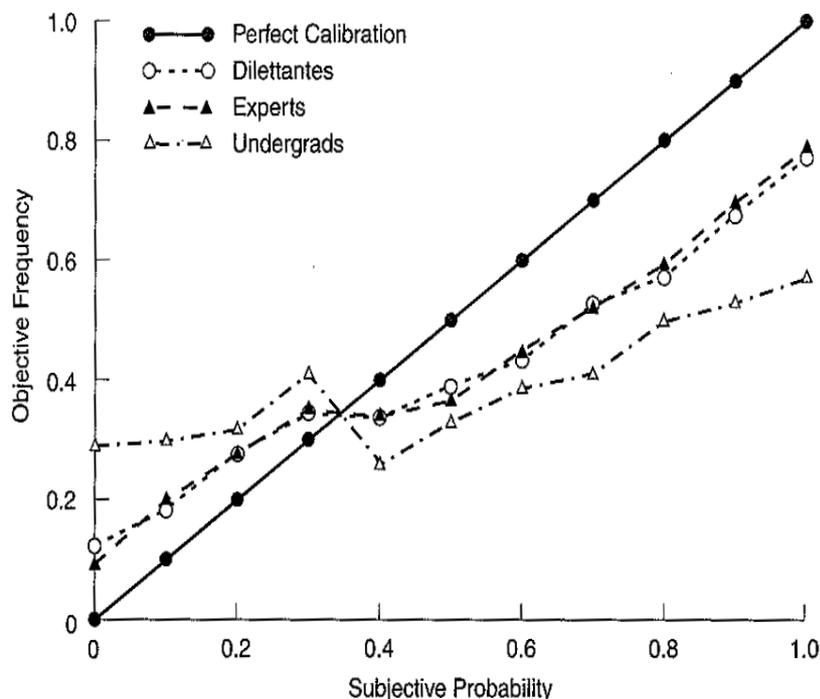
Le résultat de la mise en correspondance des probabilités objectives et subjectives est représenté sur le Graphique 1, pour différents groupes. L'axe horizontal représente les probabilités subjectives, l'axe vertical les probabilités objectives. Un expert « incertain » ou « prudent » dans ses analyses serait positionné pour chacune de ses prédictions le long de l'axe vertical à $x=0,5$. Un expert « tranché » dans ses conclusions serait au contraire positionné autour des axes verticaux $x=0$ ou $x=1$. La mise en correspondance des prédictions et de leur caractère tranché ou non, avec la réalité telle qu'elle est mesurée par les probabilités objectives, permet de tracer un nuage de points par experts, qui sont ensuite regroupés en prenant la moyenne par groupe et type de questions, et représenté sur le Graphique 1, pour différents groupes de personnes interrogées.

Le premier groupe, qui pourrait servir de groupe témoin, est celui d'étudiants « *undergraduates* », qui ont aussi participé à l'étude de Tetlock. Le second groupe est celui des experts lorsqu'ils sont en dehors de leur zone traditionnelle d'analyse : ce sont les analystes qualifiés de « dilettantes ». Le troisième groupe est celui constitué des réponses des experts interrogés sur leur domaine exact de compétences. La droite à 45 degrés est celle, idéale, de l'adéquation parfaite entre les prévisions subjectives et les probabilités objectives.

Quatre faits intéressants émergent de ce Graphique 1. D'une part, les pentes sont toutes globalement positives : les prévisions sont en moyenne corrélées à leur réalisation. Deuxièmement, les étudiants se trompent plus systématiquement que les experts, que ceux-ci soient sur leur domaine ou quand ils agissent en « dilettantes ». Le troisième, plus troublant, est que les experts dans leur domaine ne font pas nettement mieux que ces mêmes experts lorsqu'on les interroge sur un terrain éloigné de leur domaine de compétence. Cela suggérerait qu'un expert peut parler aussi bien (ou parfois aussi mal) de thèmes qu'il connaît que de thèmes qu'il ne connaît pas aussi bien. Enfin, on observe un biais classique concernant les événements certains : en moyenne, les réponses dans les trois groupes sont au-dessus de la droite à 45 degrés à gauche (pour des probabilités subjectives faibles, inférieures à 0,33) et en-dessous pour des probabilités subjectives au dessus de 0,4. Cela signifie notamment que les personnes interrogées jugent impossibles des événements à faible probabilité s'étant finalement réalisés ou au contraire, qu'ils jugent certains des événements ne s'étant finalement pas réalisés. Une interprétation est celle de la mauvaise représentation des petites probabilités, dans le prolongement de Kahneman et

Tversky (1979) et de leur *prospect theory*. On observe souvent en effet que les individus ont tendance à systématiquement sous-estimer les fortes probabilités et surestimer les faibles probabilités.

Graphique 1. Correspondance entre les probabilités subjectives et les probabilités objectives des prédictions de différents groupes d'experts. Source : Tetlock 2005.



Ces résultats pourraient dès lors nous apporter une première indication : la spécialisation d'un expert n'apporte finalement pas autant qu'il eût été permis d'espérer en termes de succès prédictif. *Pour autant, ils suggèrent que cette spécialisation a un autre rôle, celui de développer une capacité à analyser des faits.* Cette capacité améliore en effet le succès prédictif dans plusieurs domaines d'analyse, que ce soit dans le domaine de spécialisation ou dans des domaines connexes, et permettrait donc de prévoir mieux que des personnes informées, mais n'ayant pas encore pleinement développé cette capacité d'analyse (comme dans le groupe des étudiants).

Les déterminants individuels du succès prédictif

Dans une seconde étape, Tetlock analyse les déterminants individuels du succès prédictif, au regard d'une analyse de corrélation (première colonne du Tableau 1) et d'une régression multivariée (seconde colonne du Tableau 1) portant sur un sous-échantillon de 177 experts pour lesquels les informations individuelles sont disponibles au complet. Si l'on en croit ces résultats, beaucoup des caractéristiques individuelles n'ont pas d'impact significatif : le fait de détenir un doctorat, le nombre d'années d'expérience professionnelle, l'accès à des données privilégiées, ou encore l'auto-évaluation de sa compétence par rapport à la question posée n'apparaissent pas comme des variables déterminantes du succès prédictif. On approche en revanche du seuil de significativité pour trois variables : le fait de rédiger des travaux académiques, l'absence de contacts avec les médias et le fait d'être une femme améliorent la qualité prédictive.

Les orientations idéologiques ou philosophiques de l'expert, mesurées grâce à une série de questions spécifiques dans une partie préliminaire des questionnaires de Tetlock, n'ont également pas d'impact déterminant sur le succès prédictif *ex post*. C'est notamment le cas de l'orientation gauche-droite ou du

caractère plutôt idéaliste ou au contraire réaliste de la pensée des experts (on renverra à l'annexe pour la description de cette échelle de valeurs basée sur 13 questions sur les croyances des personnes interrogées).

En revanche, le fait d'être d'inspiration « croissantiste » ou « décroissantiste » a un effet fort sur les capacités prédictives. Cette dimension est définie comme le fait d'être optimiste vis-à-vis du futur de l'humanité pour ce qui concerne les ressources naturelles ou pour ce qui concerne la capacité du progrès technologique à gérer la croissance, par opposition à une vision plus malthusienne du futur. De fait, ce sont les décroissantistes qui ont tendance à avoir eu raison (non pas sur leur malthusianisme, mais en matière de succès prédictif aux questions sur le contexte politique, économique et stratégique des zones qu'ils analysent).

Tableau 1. Déterminants individuels du succès prédictif: caractéristiques démographiques et professionnelles, orientations idéologiques, traits cognitifs

Prédicteurs des différences individuelles	Précision des prédictions	
	Corrélations	Betas standardisés (et écarts-types)
I: Contrôles		
(a) Niveau d'étude (PhD ou non)	+0.02	+0.001 (0.03)
(b) Années d'expérience (1-36)	+0.00	+0.02 (0.03)
(c) Emploi universitaire ou non	-0.03	+0.05 (0.04)
(d) Accès à des informations privilégiées	+0.02	+0.01 (0.05)
(e) Exposition aux media (1-7, de jamais à chaque semaine)	-0.12	-0.09 (0.08)
(f) Genre (f=1)	+0.05	+0.08 (0.08)
(f) Auto-évaluation de sa propre expertise	+0.09	0.03 (0.07)
II : Orientation idéologique		
(a) Gauche-Droite	+0.07	+0.01 (0.05)
(b) Idéaliste-Réaliste	+0.06	-0.03 (0.06)
(c) Décroissantiste-Croissantiste	+0.20*	-0.12 (0.04)*
III : Style cognitif		
(a) <i>Hedgehog-Fox</i> (hérisson-renard)	+0.35**	+0.29 (0.04)**
(b) <i>Integratively complex thought protocols</i>	+0.31	+0.25 (0.05)**
(c) Positions modérées ou extrêmes	+0.30	+0.09 (0.06)

* : significatif à 5%

** : significatif à 1%

R² ajusté : 0.29 (N=177)

Mais c'est dans la troisième partie de ce Tableau 1 que se trouve le résultat majeur de l'analyse de Tetlock. En effet, cette partie indique que ce sont en fait les modes de raisonnement et plus particulièrement les traits cognitifs des experts qui rendent le mieux compte de leur capacité à prédire correctement l'avenir proche des zones qu'ils analysent. La ligne (a) de cette troisième partie du Tableau 1 indique notamment qu'une dimension des traits cognitifs ressort significativement. Notamment, un sous-groupe d'experts, appelé « *foxes* » (renards), a de meilleurs résultats (avec une différence significative au seuil de 1%) que le second groupe des « *hedgehogs* » (qui sont des hérissons à piquants⁵).

Quelle est cette typologie et quelles sont les caractéristiques des *hedgehogs* et des *foxes* ? Selon Tetlock, les *hedgehogs* sont des penseurs qui fondent leur analyse sur une grande « idée » (un dogme, un principe) et étendent ensuite ce principe à de nombreux autres domaines ; ils sont peu enclins à composer avec

5 Il existe des hérissons sans piquants en Asie (gymnures).

ceux qui ne perçoivent pas ce raisonnement. Ils sont enfin assez sûrs de leur capacité prédictive.⁶ Les renards sont au contraire des penseurs qui accumulent des informations et des raisonnements variés, qui ne croient pas aux grands schémas explicatifs, qui sont plus inductifs et parfois *ad hoc* dans leurs modes de raisonnement et sont enfin plus sceptiques quant à leurs propres capacités prédictives.⁷

Cette classification des schémas mentaux est inhabituelle et peut sembler subjective. On laissera du reste le lecteur faire au préalable le test de s'y placer soi-même puis d'y positionner les experts qu'il côtoie régulièrement, que ce soit professionnellement ou dans les *media*. Pour autant, cette grille d'analyse des traits cognitifs est basée sur des éléments plus quantifiés qu'il n'y paraît, et une analyse de composantes décrite dans le Tableau 2, colonne 1 détaille la liste des facteurs affectant la dimension *fox/hedgehogs*. Le renard est d'abord celui qui s'identifie comme tel (première ligne), qui n'aime pas les explications simples (ligne 3) mais qui apprécie au contraire les questions qui permettent plusieurs réponses (ligne 8) ou qui accepte l'idée que, dans les conflits, les deux parties peuvent avoir raison (ligne 10). Il est enfin celui qui apprécie relativement plus que son collègue hérisson le fait de pouvoir interagir avec des gens qui pensent différemment de lui (ligne 12). Le hérisson pense plus fréquemment que la politique peut se décrire comme un mécanisme d'horlogerie (ligne 4, *clocklike*, ce que nous pouvons interpréter comme relevant d'une science analysant les mécanismes déterministes) plutôt qu'une science floue (*cloudlike*).

Tableau 2. Analyse des déterminants individuels des dimensions cognitives analysées.

Traits	Déterminants des facteurs	
	Hérisson-Renard (I)	Capacités à décider (Decisiveness factor) (II)
1. Auto-identification comme renard ou hérisson (définition de Berlin)	+0.42	-0.04
2. Je commets parfois des erreurs de jugement en exagérant la complexité du monde	-0.20	+0.14
3. Je préfère des modèles simples pour expliquer les processus politiques	-0.29	+0.05
4. La politique est un phénomène qu'on peut décrire comme incertain et diffus (<i>cloudlike</i>) plutôt que comme un mécanisme d'horlogerie ou déterministe (<i>clocklike</i>)	+0.26	-0.02
5. Je commets parfois des erreurs en abandonnant trop vite une bonne idée	-0.31	+0.22
6. Avoir des règles précises et une bonne organisation au travail est la clé du succès	-0.09	+0.31
7. Même après m'être décidé, je suis toujours disposé à entendre une autre opinion	+0.28	-0.07
8. Je n'aime pas les questions qui conduisent à des réponses multiples et variées	-0.35	+0.05
9. Je prends des décisions rapidement et en confiance	-0.23	+0.26
10. Dans la plupart des discussions, je perçois la façon dont les deux côtés peuvent avoir raison simultanément	+0.31	+0.01
11. Il est irritant d'écouter les gens qui ne peuvent se décider	-0.18	+0.14
12. Je préfère interagir avec les gens dont les opinions diffèrent des miennes	+0.23	-0.10
13. En essayant de résoudre un problème, je suis perturbé parce que je perçois différentes solutions contradictoires	+0.08	-0.27

Un coefficient positif reflète un profil cognitif proche du type Renard (colonne I) et des capacités à décider plus marquées (colonne II).

6 « Thinkers who know one big thing, aggressively extend the explanatory reach of that one big thing into new domains, display bristly impatience with those « who do not get it » and express considerable confidence that they are already pretty proficient forecasters, at least in the long run. ».

7 « Thinkers who know many small things, are sceptical of grand schemes, see explanation and prediction not as deductive exercises but rather as exercises in flexible « ad hocery » that require stitching together diverse source of information, and are rather diffident about their own forecasting prowess and rather dubious that the cloudlike subject of politics can be the object of a clocklike science. »

La conclusion de Tetlock est sans ambiguïté : les prédictions les plus précises proviennent des experts qui combinent des informations de sources d'origines différentes. En anglais, « *Most accurate predictions came from policy experts who combined insights from a variety of sources.* »

Analyser plus en détail la méthodologie dépasse le cadre de cet article et le lecteur pourra se reporter directement à l'ouvrage.⁸ Il serait à ce stade téméraire de vouloir généraliser les enseignements de cet ouvrage au domaine plus étroit que nous couvrons dans cet article, les évaluations de politiques publiques par les experts de différentes sciences sociales (voire notamment les limites de cette généralisation *infra*). Mais cet ouvrage pose selon nous les bonnes questions sur un point clé : les experts, qui donnent leur avis sur des questions sociales de grande importance, que la question soit limitée à une mesure d'efficacité d'une politique publique ou qu'elle soit plus globale, comme sa pertinence dans un contexte donné, ne doivent pas seulement être rigoureux scientifiquement : il faut en plus savoir si leur avis est en moyenne vérifié par l'expérience et avec le recul, ou si au contraire l'expert se trompe régulièrement, quelles que soient ses qualités scientifiques et la rigueur de ses méthodes.

Ainsi, pour ce qui ne concernerait que les experts économistes européens, des recherches passionnantes pourraient être menées sur des bases similaires à celle de Tetlock. On pourrait réaliser notamment d'intéressants travaux sur ce que les experts disaient à différents moments clés des dernières décennies ; d'autre part sur le mode de raisonnement, la formation initiale et les méthodes employées par ces experts lorsqu'ils ont conseillé les décideurs ou sont intervenus publiquement dans les débats. Par exemple, on pourrait documenter les diagnostics consistant, après le premier choc pétrolier, à proposer une relance de l'activité économique, celui consistant, en France et en 1981, à vouloir réduire l'âge de la retraite, celui consistant, au début de 2007, à nier l'existence de tensions dans le monde de la finance et à croire en la nécessité de ne pas trop la réguler, enfin celui consistant à recommander l'utilisation systématique du déficit public comme arme de lutte contre la récession tout en minimisant les tensions sur les finances publiques et les capacités d'emprunt des États, ou, *a contrario* de mener des politiques d'austérité automatiques sans prise en compte des effets récessifs comme en Grèce. Ce recensement pourrait faire l'objet de discussions passionnantes, sachant que dans au moins deux cas sur les cinq cités ci-dessus, il est difficile de dire qu'un consensus se soit totalement établi.

Section 3. L'application de cette grille d'analyse à l'évaluation des politiques publiques

L'objet du reste de l'article est de tenter de tirer des leçons de l'analyse de Tetlock en ce qui concerne l'évaluation des politiques publiques. Cette extrapolation potentielle appelle plusieurs remarques. D'une part, elle implique de définir l'évaluation des politiques publiques. On peut distinguer l'évaluation ex ante (prévision à proprement parler), l'évaluation ex post (étude d'impact) et l'évaluation liée à la mise en œuvre, l'analyse de la contextualisation ainsi que la méta-réflexion sur la non-neutralité de l'évaluation scientifique. La définition que l'on trouve le plus couramment est plus générale, plus vague aussi : il s'agit « *d'analyser si une politique publique donnée a atteint les objectifs qui lui étaient assignés initialement* », sans d'ailleurs que la notion de causalité et de contrefactuels ne soient nécessairement mises en avant.⁹

8 Un économiste en faisant la lecture détaillée ferait certainement un intéressant « rapport d'étonnement » : ainsi, le texte est souvent très imagé (comme en témoignent les classifications énoncées ci-dessus) ce qui peut déconcerter, très simplificateur (le message principal en est d'autant mieux mis en valeur) et il est presque impossible de comprendre ce qui est précisément fait sans lire en détail les annexes méthodologiques qui représentent 73 pages sur 238 du texte principal, soit 30% du texte.

9 L'article 1^{er} du décret 90-82 du 22 janvier 1990 relatif à l'évaluation des politiques publiques indique ainsi que « (...) L'évaluation d'une politique publique au sens du présent décret a pour objet de rechercher si les moyens juridiques, administratifs ou financiers mis en œuvre permettent de produire les effets attendus de cette politique et d'atteindre les objectifs qui lui sont assignés. »

Ce qui semble le plus directement relié à l'analyse de Tetlock serait l'analyse ex ante (prévision) combinée à l'analyse ex post et la mise en relation entre les deux ; ainsi, les deux axes des probabilités subjectives et objectives du Graphique 1 pourraient être remplacés par les prédictions ex ante et les résultats ex post.

Néanmoins, la question est plus complexe qu'il n'y paraît, pour deux raisons. D'une part, cela suggérerait à tort que l'analyse ex post est objective et contiendrait moins d'erreurs que l'analyse ex ante, puisqu'elle validerait ou invaliderait les prévisions ex ante. Tout praticien de l'analyse ex post sait pourtant toutes les difficultés liées à l'identification des mécanismes de causalité, que ce soit de façon empirique quantifiée ou plus qualitative, et les nombreuses hypothèses parfois arbitraires – et qui devraient être toujours transparentes – qu'il faut faire pour parvenir au résultat.

D'autre part, les aspects liés à la mise en œuvre, la contextualisation et les diverses méta-réflexion permettent justement d'affiner les diagnostics ex post mais aussi ex ante et donc ne sont pas dissociables de ces deux dimensions de l'évaluation des politiques publiques.

C'est précisément pour cette seconde raison qu'il nous semble fructueux de postuler la transposabilité de l'enseignement de Tetlock à notre question, fut-ce temporairement et avec toutes les réserves de rigueur : l'analyse des politiques publiques requiert les éclairages de différentes disciplines pour être de meilleure qualité. Ce point de vue n'est pas unanime : certains estiment que seule l'économétrie (ou *a fortiori*, seule une expérience aléatoire) permet de donner un avis informé sur les mécanismes causaux d'une politique publique, ignorant du reste qu'une partie de la sociologie cherche justement à mettre en évidence ces facteurs de causalité. D'autres domaines disciplinaires estiment à l'inverse qu'une analyse causale en sciences sociales est non seulement impossible, mais constitue un danger car les chiffres exercent un attrait qui peut faire oublier les biais ou la fragilité des données ou de leurs interprétations.

Ces précautions prises, il est possible de mener une réflexion sur le mode d'organisation de la recherche lorsque celle-ci associe différents points de vue.

Les besoins contradictoires de l'évaluation : scientificité et demande sociale

Dans les débats sur le besoin d'approches disciplinaires croisées, des perspectives différentes s'opposent régulièrement. Pour certains, l'interdisciplinarité devrait réunir plusieurs mono-disciplines qui s'associent de façon ponctuelle pour traiter d'un objet. Ce croisement peut être fécond lorsque l'objet est précis, mais il est plutôt l'exception. Jacobs (2008) déplore ainsi que la pratique de l'évaluation ne soit pas à la hauteur des attentes ou des ambitions affichées en la matière.

D'autres plaident au contraire pour l'émergence d'une nouvelle discipline, organisée comme telle. C'est notamment le crédo de plusieurs membres de la Société Française d'Évaluation. Certains points de vue sont intermédiaires entre ces deux visions opposées, mais penchent pour la constitution d'une méta-discipline ou d'une trans-discipline de l'évaluation. Maurice Baslé (2008, pp.6 et 7) écrit ainsi que « *L'évaluation a les caractéristiques d'une « cross-sectional science ». Elle est une transdiscipline comme la Statistique. (...). L'émancipation des évaluateurs suppose aussi une reconnaissance des « évaluateurs-chercheurs », une réorganisation de la recherche correspondante (...).* » Cette méta-évaluation devrait suivre un certain nombre de critères définis par exemple dans Scriven (2010) : la validité, la clarté, la crédibilité, l'éthique, le caractère général, encore que ces critères ne soient bien entendu pas propres à une transdiscipline mais s'appliquent *de facto* à toute évaluation disciplinaire. Baslé (2008) poursuit en associant implicitement la spécialisation disciplinaire avec l'absence *de facto* des chercheurs et académiques dans le champ de l'évaluation. « *Ceci est loin d'être le cas en France particulièrement où les académiques se sont jusqu'à présent contentés d'être faiblement présents dans les évaluations (...) où les évaluateurs ont accepté d'être dépassés par des hauts fonctionnaires, des inspecteurs généraux ou des juges.* »

Pour alimenter ce débat, nous pourrions relever que l'évaluation répond en fait à deux impératifs distincts et parfois contradictoires. D'une part, l'évaluation doit être scientifiquement fondée, ne serait-

ce que pour la transparence qu'elle permet : les hypothèses doivent être explicites, les marges d'erreurs indiquées s'il y a lieu, le jugement doit être fondé sur ce qui est démontré, l'analyse doit être répliquable, la falsifiabilité est la clé de la rigueur.

D'autre part, l'évaluation doit être pertinente et répondre à la demande du décideur. Or ce n'est pas toujours simple de concilier ces deux impératifs. Si le décideur demande la couleur d'un objet, on peut difficilement lui répondre que seule la longueur de l'objet peut à court terme être mesurée en chiffres et s'en tenir là.

Pour ce qui est de l'impératif de scientificité, l'organisation disciplinaire semble être la meilleure des solutions (ou du moins la moins mauvaise). Les techniques d'analyse sont de plus en plus sophistiquées, l'état de l'art en la matière est difficile à connaître dans une seule discipline, et l'on voit mal comment une méta-discipline associant trois ou quatre sciences sociales pourrait organiser l'évaluation de la scientificité d'un texte réunissant autant de composantes disciplinaires : aucun éditeur de revue, par exemple, ne pourrait être à même de juger de cela.¹⁰

La multidisciplinarité comporte donc un risque, qui est celui de conduire à terme à un affaiblissement des exigences méthodologiques, dans un processus de convergence vers le bas. Si ce n'est pas nécessairement le cas lorsque des scientifiques « *up to date* » s'associent pour former un groupe pluridisciplinaire, le temps consacré à la gestion de cette multidisciplinarité risque par contre de laisser moins de temps pour faire progresser sa discipline ou simplement se tenir au courant des évolutions méthodologiques.

A contrario, la spécialisation au sein d'une discipline, voire, dans une discipline, au sein d'un sous-groupe de chercheurs maniant et raffinant les mêmes techniques d'analyse, peut conduire à un affaiblissement de la pensée. Ainsi, certains chercheurs tenants des méthodologies d'évaluation à partir d'échantillons aléatoires estiment que les résultats de macro-économétrie, notamment en matière monétaire, n'ont pas de fondements puisque l'identification n'est pas basée sur un dispositif expérimental tel que celui des méthodes d'expériences contrôlées qui sont appliquées avec succès aux politiques de développement. On imagine pourtant mal une randomisation portant sur les politiques monétaires avec la participation volontaire des banques centrales. Le second impératif, répondre à une demande sociale (ici, comment mener une politique monétaire et quelles conséquences en attendre), ne peut pas être écarté par l'argument que seule la méthode expérimentale est valide. Les méthodes de variables instrumentales permettent de poser les hypothèses, réfutables donc, de l'identification. Inversement, la validité externe des expériences contrôlées échappe à toute réfutabilité, ce qui est en soi problématique.

Symétriquement, les chercheurs pratiquant les modèles structurels en économie expriment parfois un désintérêt frappant pour l'identification, quand bien même ils testent des scénarii de politique économique. La phrase d'Edward Prescott, « *Progress, don't regress* » en est une illustration.

On peut souligner d'autres contextes dans lesquels les deux objectifs de l'évaluation (scientificité et capacité à répondre à la demande) peuvent s'opposer. Si un corps de contrôle d'une administration souhaite mener une évaluation d'un dispositif de politiques publiques, par exemple le dispositif Scellier de défiscalisation des investissements immobiliers ou le prêt à taux zéro, il va devoir réaliser cette évaluation dans une relative urgence, typiquement trois mois, dans un contexte où les données nécessaires à la réalisation de l'évaluation requièrent comme préalable la réunion du Conseil National de l'Information Statistique (CNIS), lui-même se réunissant épisodiquement. Par ailleurs, le dispositif évalué ne se prête généralement pas à l'évaluation, pour diverses raisons liées à sa mise en œuvre (voir Ferracci et Wasmer, 2012 sur ce point). Faut-il dès lors laisser l'évaluation à des non-scientifiques ?

10 On pourrait cependant imaginer qu'il s'entoure de co-éditeurs à la frontière des différentes disciplines. Mais comme ceux-ci prennent en général leur décision sur la base de deux voire trois avis de rapporteurs différents dans leur discipline, les coûts de transaction risquent d'être prohibitifs, sauf à relâcher les exigences.

Certains le pensent, avec de bonnes raisons. D'autres pensent qu'il serait intéressant de consacrer une plus grande partie de son temps de recherche à expertiser les politiques publiques dans les domaines dans lesquels on mène des recherches, dans la mesure où cela peut contribuer à mieux comprendre les enjeux et le détail des dispositifs mis en place, et parfois de mieux connaître les données disponibles et d'en disposer.¹¹

Contradictions et complémentarités du quantitatif et du qualitatif

Il peut aussi arriver qu'une partie des enjeux d'un dispositif de politiques publiques porte sur des aspects pour lesquels il n'existe pas de données, ou pour lesquels les enjeux sont non-quantifiables, immatériels ou tout simplement symboliques. Par exemple, une étude économétrique sur le redoublement peut arriver à la conclusion que celui-ci n'a pas d'effet causal sur le succès individuel après le redoublement (Gary-Bobo et Robin, 2011). Pour en conclure que le redoublement doit être supprimé ou réduit en France, il faudrait répondre à une autre question : quel serait l'effet de la suppression du redoublement sur les élèves, mais aussi sur la logique des rapports enseignants-élèves ou enseignants-parents ? Certains sociologues estiment notamment que la peur du redoublement est un moteur du comportement dans les classes. Il s'agit donc d'un mécanisme de première importance sur l'apprentissage et les résultats scolaires, et ce mécanisme ne serait pas invariant à un changement de politique de redoublement. L'on peut en théorie imaginer un modèle économétrique structurel plus complexe et comportant notamment les effets d'équilibre général tels que celui décrit ci-dessus, ou une expérimentation permettant de tester ces différents effets. Dans tous les cas, le recours au dialogue interdisciplinaire peut se révéler efficace pour nuancer des conclusions et leur portée.

Cette dialectique du quantitatif (sans doute rigoureux car bien délimité mais parfois réducteur) et du qualitatif peut trouver de multiples applications. Par exemple, on pourrait imaginer un mode de raisonnement de type *ex ante* qui suggérerait que la loi interdisant le port du voile intégral en public est inutile, car elle ne concerne qu'un tout petit nombre de cas pour un coût important (direct ou en terme d'effort des services de police et de justice). La demande du décideur serait différente : ce serait de savoir quelle est la dynamique en termes de normes sociales qu'une telle loi va engendrer. Cette loi constituera-t-elle un signal visant à améliorer le sort des femmes dans des milieux islamistes, ou au contraire mènera-t-elle des femmes à rester confinées dans des lieux privés ? On peut imaginer différents tests de ces hypothèses, et l'alliance de méthodologies qualitatives ou quantitatives peut dès lors être utile.

On peut enfin se poser la question de la contradiction possible entre les résultats de différentes méthodes. Dans une étude sur les discriminations sur le marché du logement, Bonnet et alii. (2012) ont mené des entretiens téléphoniques auprès d'agences de location permettant de réaliser une étude économétrique d'impact sur l'influence du nom et du lieu de résidence sur les chances d'obtenir une visite d'un appartement en région parisienne. Ceci permet de révéler un résultat assez marquant : la consonance ethnique (française ou maghrébine) d'un nom n'a pas un impact très fort sur les chances de visite d'un logement disponible lorsqu'on contrôle pour l'effet de l'origine géographique. En revanche, le fait de résider dans un quartier connoté négativement (La Courneuve par opposition à Antony par exemple) a un impact très fort sur les chances d'obtenir une visite pour le postulant au logement. En voulant approfondir ce point grâce à une technique différente, qualitative, basée sur des enquêtes de terrain (30 entretiens d'une heure minimum auprès d'agents immobiliers), il en est ressorti de façon inattendue la conclusion inverse : les agents indiquent unanimement ou presque l'existence de fortes discriminations ethniques, mais ne pensent pas que le lieu de résidence des postulants joue un rôle dans

11 L'argument selon lequel il est plus difficile d'évaluer la valorisation que la science elle-même est parfois avancé. A la réflexion, il ne peut que paraître surprenant. Il est bien plus difficile d'évaluer un article soumis à *Econometrica* que de juger de la qualité d'un rapport scientifique. C'est une question d'incitations et on peut du reste imaginer d'organiser une évaluation de l'expertise et de la valorisation dans les carrières.

la sélection par les agences. Si l'article de Bonnet et alii. (2012) tente de proposer diverses explications, il n'en reste pas moins que l'alliance de différentes stratégies de recherche est difficile, quand bien même, dans le cas précis de cette étude, elle permet à nouveau d'éviter une erreur d'analyse ou du moins de nuancer une conclusion trop rapide, ce qui a bien entendu des implications pour les éventuelles recommandations de politiques publiques qui auraient pu être faites trop rapidement.

Conclusions partielles

En guise de conclusion partielle, et en forçant le trait, l'expert « renard » de Tetlock est sans doute, dans la représentation que l'on s'en fait, un producteur de savoirs plus difficilement falsifiables. Il est moins biaisé (donc plus juste en moyenne) mais en même temps plus imprécis, du moins quant à ses hypothèses de travail, et peut-être plus inductif.

En revanche, le chercheur « hérisson » produit du savoir sans doute plus facilement falsifiable, puisqu'inspiré de protocoles disciplinaires mieux établis, mais dont la pertinence peut être moins grande. Il se peut qu'il soit plus précis mais aussi plus biaisé, et qu'il ne cherche qu'au pied du lampadaire, pour reprendre l'allégorie bien connue de l'économiste qui ne recherche ses clés perdues que dans la portion éclairée de la rue, quand bien même celles-ci ont pu être perdues n'importe où.

Le croisement disciplinaire, s'il est bien mené, par des équipes dont les membres conservent une exigence disciplinaire forte, pourrait idéalement conserver la précision des analyses et la falsification, tout en permettant différents éclairages. Comment mieux évaluer ce respect de l'exigence disciplinaire que par la publication dans des revues à comité de lecture disciplinaires ? La question est ouverte.

Ce croisement permet-il de réduire les biais ? Le *trade-off* semble bien de gagner en pouvoir prédictif, ne serait-ce qu'en évitant les erreurs de type I (false positive) sur l'effet d'un phénomène social mais génère également des coûts de transaction : d'expérience, le travail interdisciplinaire est rendu compliqué par le fait que les chercheurs de différentes disciplines ont leur propre langage, leurs propres biais et leurs propres logiques de carrière. On pourrait imaginer de revenir sur ces obstacles en créant une nouvelle société scientifique pour gérer ces carrières, mais avec le risque de l'affaiblissement progressif de l'exigence méthodologique par éloignement de la frontière du savoir méthodologique disciplinaire.

Qui plus est, le contexte de l'évaluation en France est particulier. Il est celui d'un relatif manque de confiance réciproque entre trois acteurs distincts¹² : les membres de cabinets ministériels qui font l'interface avec le décideur politique ; les administrations qui ont longtemps été dans une situation de monopole des données ; et les chercheurs qui ont parfois eu un monopole temporaire de compétences – le mot compétence est entendu au sens large.¹³ Un dialogue interdisciplinaire dans une évaluation de politiques publiques comportera donc au minimum quatre acteurs, puisqu'au moins deux universitaires interagissant de façon stratégique, ce qui augmente la complexité et la gestion de l'évaluation.

12 Un exemple de ceci est la clause d'accord préalable avant diffusion des résultats que certaines administrations tentent d'imposer (voir à ce sujet la charte dite des trois Labex co-rédigée par le LIEPP, IPP et l'IDEP en mars 2012 et disponible sur le site de ces institutions). On trouve une analyse poussée de ces interactions entre les trois mondes dans Lascoumes et Setbon (1996) : l'administration et les inspections ont toujours tenté d'être les acteurs et les évaluateurs des politiques publiques.

13 Par exemple, lorsque Thomas Piketty (Piketty, 2001) a analysé la dynamique des inégalités sur le long-terme dans les millimes supérieurs de la distribution des revenus, la compétence était non pas tant dans les techniques utilisées que dans l'intuition que ces données révéleraient des faits importants et des enseignements précieux sur l'interaction entre la fiscalité et les inégalités. Les administrateurs d'un ministère disposent en général de moins de temps pour se lancer dans des projets de cette ampleur, ce qui est *a fortiori* plus vrai encore dans les corps de contrôle.

Section 4. Les coûts de transaction des dialogues interdisciplinaires et une méthodologie pour les gérer

S'entendre sur les concepts

Partons d'une question qui a animé un dialogue interdisciplinaire fécond mais parfois complexe, et qui s'est posée lors de la co-rédaction d'un projet de laboratoire interdisciplinaire d'évaluation des politiques publiques (campagne 2011) dans le cadre des investissements d'avenir (Labex). Cette question est pourtant LA question préalable : elle tient en quelques mots: « *Pourquoi faut-il des politiques publiques?* »

Un économiste bien formé (les collègues d'autres disciplines diraient bien formaté !) répondrait spontanément qu'une politique publique est – ou devrait être – la réponse de la collectivité publique à des défaillances du marché. Par exemple, des externalités conduisent à une situation spontanée, en l'absence d'intervention publique, soit de surproduction (dans le cas de la pollution) ou de sous-investissement (dans le cas des externalités de capital humain). Un autre exemple est celui des biens publics qui sont difficilement financés par les individus en raison des comportements de passerager clandestin. Une troisième situation est celle des problèmes informationnels qui conduisent à une inefficacité relative en raison des phénomènes d'aléa moral, ou à une inefficacité dramatique lorsque l'anti-sélection conduit à la disparition d'un marché. Enfin, les rendements d'échelle croissants conduisent à la trop grande concentration de l'activité et aux pertes sèches. Dans chaque cas, il est conceptuellement très simple pour l'économiste d'imaginer une politique optimale résolvant l'inefficacité, même si sa mise en œuvre est généralement une gageure.

La réponse des non-économistes est généralement très différente. Une politique publique, pour les politistes ou les sociologues, est en général une solution *politique* à un problème donné. Ce problème donné est donc identifié par l'existence même d'un conflit entre des intérêts divergents. En un sens, la politique publique est une façon de dénouer ce conflit et constitue un arbitrage rendu. Cet arbitrage ne conduit pas nécessairement à la résolution de l'inefficacité au sens des économistes, mais il est au contraire rendu suite à un rapport de force politique, militaire ou dans le cadre plus feutré d'un jeu d'influence (lobbying).

Bien entendu, ces deux perspectives ne sont pas irréconciliables ni exclusives l'une de l'autre. Les économistes ont raison en un sens normatif, les politistes en un sens positif. Le mot normatif lui-même est sujet à de nombreuses incompréhensions entre les économistes et les non-économistes. En effet, on utilise ici le mot dans son acception habituelle en économie : il décrit ce qui *devrait être* une fois qu'on a précisé le critère normatif retenu. Ce critère normatif et son intérêt sont un autre sujet de débat sans fin entre disciplines. Les économistes pensent habituellement qu'il est possible de classer différentes politiques entre elles, et qu'il existe une meilleure politique, qu'on qualifiera de politique optimale (et par une tentation coupable de la pensée, que le monde serait meilleur si les décideurs les écoutaient enfin).

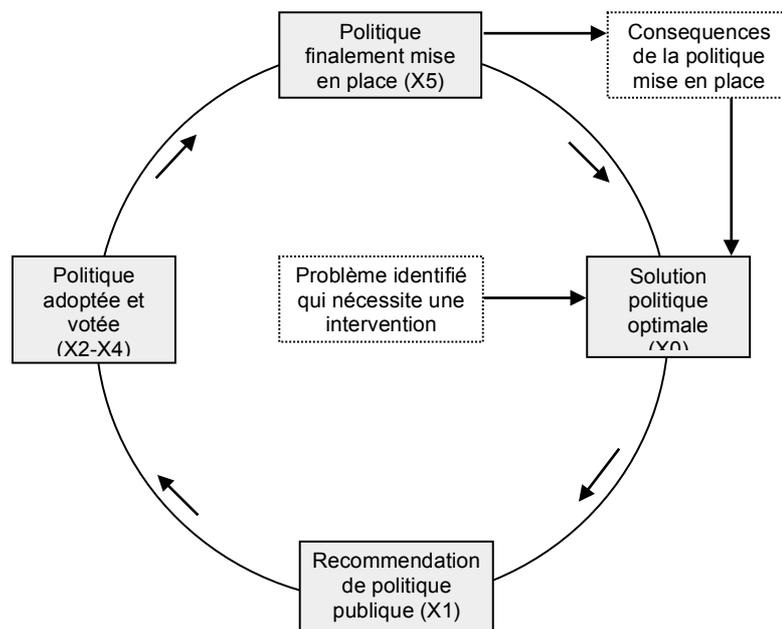
D'autres disciplines adoptent quant à elles une vision descriptive de cette question : les décisions politiques échappent sans doute grandement à toute rationalité économique. L'émergence de l'économie politique au sein du corpus néoclassique, comme au sein du paradigme post-moderniste, suggère en effet que la décision collective peut s'analyser différemment de la réponse collective optimale à une défaillance du marché.

Le Graphique 2 issu de Wasmer et Woll (2011) représente par ailleurs les différentes étapes d'un processus politique qui permet le consensus entre différentes disciplines. Au centre se trouve, non pas la défaillance de marché, mais le problème qui mérite intervention (qui inclut donc aussi la résolution d'un conflit entre intérêts divergents). Il en découle une solution optimale pour l'économiste, ou une préconisation idéale, notée X0. Il semble aussi utile de la distinguer au moins conceptuellement de ce que sera la recommandation de politique économique X1. Compte tenu de ce que le chercheur en

sciences sociales sait du lien qui peut exister entre la nature de X1 et les probabilités de succès de cette proposition X1 à travers les étapes X2-X4 du processus démocratique (respectivement le programme politique d'un parti ou d'un candidat avant l'élection, la proposition de loi déposée par le parti élu puis le processus effectivement voté), il se peut que le chercheur s'autocensure : plutôt que de prôner l'abandon d'une mesure symbolique forte mais aux effets négatifs avérés, il pourrait atténuer la proposition en imaginant une mesure améliorant la situation compte tenu de la probabilité de sa proposition d'être adoptée plutôt que rejetée d'entrée.¹⁴ En tous les cas, conceptuellement, X0 et X1 n'ont pas de raison d'être identiques.

L'étape suivante noté X5 est celle de la mise en place de la mesure effective, après la parution des dispositions d'application (décrets, circulaire) et de la mise en place des budgets.

Figure 2: Les étapes du développement d'une politique publique. (Source Wasmer et Woll, 2011).



Ce type de graphique semble *a posteriori* évident. Il est en fait une des étapes préalables de constitution d'un langage commun minimal. Il permet à tout le moins l'identification des sujets qui « hérissent » les uns et les autres. Une fois ceci réalisé, il est possible d'avancer dans un cadre d'analyse plus complet des politiques publiques.¹⁵

S'entendre sur les avantages comparatifs et la complémentarité des approches

Pour simplifier sans doute exagérément l'analyse des politiques publiques, il est utile de distinguer trois moments : l'évaluation ex ante ; l'évaluation ex post ; et enfin les processus de mise en œuvre et la contextualisation, avec la réserve que ces moments ne sont jamais totalement disjoints dans la pratique.

L'évaluation ex ante consiste en une prévision et une anticipation des effets, qui peut être chiffrée, basée sur des modèles économétriques ou sur des modèles théoriques dits structurels. Elle peut

14 Certains reconnaîtront ici l'objection de Charles Péguy « *Le kantisme a les mains pures, mais il n'a pas de mains.* » (Pensées, octobre 1910) à l'impératif catégorique de Kant.

15 L'économiste Robert Solow rapporte que lorsqu'il débutait sa carrière, les économistes de différents pays utilisaient des termes identiques pour des concepts différents et que les premières étapes de discussion étaient de s'assurer que les concepts utilisés étaient identiques, ce que la synthèse néo-classique a permis de réaliser. Une telle étape de constitution d'un langage commun entre différentes sciences sociales reste à franchir.

également être moins quantitative, basée sur une connaissance fine des processus susceptibles de jouer ou enfin basée sur une analyse comparative internationale.

L'évaluation ex post consiste à faire le bilan d'une expérience passée, en se basant sur une identification économétrique (expérience naturelle, variables instrumentales) qui permet de construire un contrefactuel, exercice toujours délicat ; ou sur des techniques plus qualitatives (audit, entretiens).

Les aspects de mise en œuvre, ou monitoring, apparaissent comme importants pour mieux comprendre ce qui pourrait faire échouer une prévision ex ante qui ne tiendrait pas compte de l'importance de mécanismes tels que l'appropriation ou non des réformes par les agents ou les usagers. Ainsi, on peut faire une étude épidémiologique de type ex ante sophistiquée sur l'impact d'une politique de vaccination sur le taux de diffusion d'une grippe saisonnière ou exceptionnelle (type H1N1). Si les médecins généralistes et les patients refusent massivement la campagne, cela change radicalement le processus de diffusion. Le modèle reste valide au sens de sa cohérence interne : le modélisateur indiquerait que si les médecins avaient vacciné comme sa paramétrisation l'avait indiqué, ses résultats auraient été validés. Pour le décideur, le fait que le modèle n'ait pas pris en compte cette dimension est évidemment un aspect majeur.

Cet exemple suggère que le modèle ex ante est améliorable si les aspects liés à la mise en œuvre sont mieux compris. Dans la mesure où l'analyse ex ante quantifiée est basée sur des travaux d'économistes, d'économètres ou d'épidémiologistes, et où l'analyse de mise en œuvre fait appel à des éléments de la sociologie des organisations, les aspects de collaboration interdisciplinaire peuvent se révéler féconds.

De la même façon, l'analyse quantitative ex post, qui permet d'apprécier les effets d'une politique donnée et d'en faire un bilan coût-bénéfice, peut bénéficier de l'apport des éléments de contextualisation et des divers mécanismes portant sur des variables qualitatives ou difficilement mesurables. Par exemple, lorsque l'on mesure le contenu en emploi des zones franches urbaines (Mayer et alii., 2012), le bilan quantitatif est généralement positif, mais avec un rapport coût-bénéfice faible. L'emploi créé et les implantations d'entreprises nouvelles sont parfois le résultat d'un effet de substitution, des déplacements d'activité en provenance d'autres zones proches. Les acteurs de terrain ne sont pas sensibles à ces considérations quantifiées, et revendiquent quant à eux la redynamisation du quartier et les effets induits qui cassent la dynamique de ségrégation des quartiers. Les dimensions qualitatives ou quantitatives, mais relevant d'autres dimensions (criminalité, santé publique), sont dès lors susceptibles d'affiner le diagnostic à destination du décideur.

Un autre exemple où une analyse du contexte peut être très utile est celui de l'évaluation réalisée par Behaghel, Crépon, Gurgand (2009). Ceux-ci ont analysé le placement des demandeurs d'emploi à travers différents dispositifs institutionnels, issus de l'Unedic ou émanant du secteur privé. Ils ont trouvé que le programme Cap Vers l'Entreprise, qui était issu des services de l'ANPE, avait eu de meilleurs résultats que les organismes privés de placement (OPP). Ce résultat battait en brèche des idées reçues, mais ne pouvait pas conclure sur les causes profondes de ces différences d'efficacité. Il suggérait cependant qu'un effet de concurrence locale entre public et privé pouvait améliorer la motivation des acteurs publics. Si cela s'avérait être le cas, ce qu'une démarche qualitative aurait pu révéler dans une étude subséquente, cela aurait enrichi considérablement l'analyse des conséquences possibles d'une généralisation des mesures de mise en concurrence du secteur public par le secteur privé.

Un autre exemple intéressant de la combinaison de différentes approches est celui de l'expertise considérable autour du protocole du projet canadien d'autosuffisance expérimenté entre 1992 et 1995 au Canada (*Self Sufficiency Project*). Ce projet consistant à verser un complément salarial conséquent aux personnes retrouvant un emploi à temps plein, en ciblant les mères célibataires, a été testé sur des populations choisies aléatoirement dans deux provinces canadiennes, au Nouveau-Brunswick et en Colombie Britannique. L'expérience a fait l'objet de nombreuses évaluations ex post purement économétriques (Michalopoulos, 2002). D'autres travaux ont tenté de modéliser la dynamique de

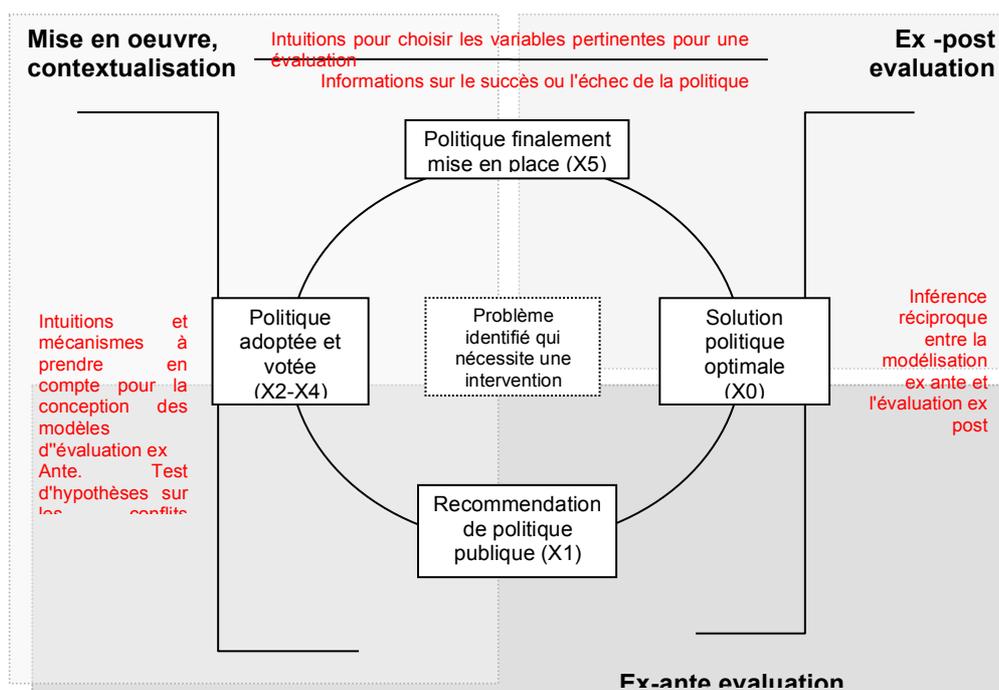
formation des salaires d'une éventuelle généralisation du dispositif (Lise, Seitz et Smith 2005). Enfin, des travaux de sociologie de la santé et de sciences de l'éducation ont tenté d'appréhender le développement cognitif des enfants dans les foyers où l'adulte isolé reprenait un travail en comparaison avec les foyers où l'adulte isolé n'avait pas été sélectionné aléatoirement pour bénéficier de l'expérimentation.

Les liens entre l'analyse ex post et l'analyse ex ante sont déjà en soi une source d'enrichissement considérable. C'est le cœur de la démarche d'Attanasio et alii. (2005) à partir d'expérimentations aléatoires et de modèles d'équilibre général permettant de capturer les effets de bouclage d'un programme de développement au Mexique (Progresa).

Dans Chemin et Wasmer (2011), les auteurs ont utilisé une autre méthode d'évaluation ex post, l'expérience naturelle que constitue l'Alsace-Moselle, pour évaluer l'effet de l'implémentation du RMI, qui pré-existait sous la forme de l'aide sociale depuis 1908 dans les trois départements du Haut-Rhin, du Bas-Rhin et de Moselle. Les effets désincitatifs se mesurent grâce au taux de retour à l'emploi et nous permettent de calibrer un modèle, notamment le coût de l'effort de recherche des populations éligibles au RMI, qui est habituellement difficile à mesurer. Une fois cette calibration effectuée, il est possible d'utiliser le modèle pour faire des expériences de politiques économiques de type ex ante, par exemple sur le RSA. Le RSA diminuant le taux marginal implicite de taxation des revenus des travailleurs à faible revenu, notre modèle prédit une diminution de moitié par le RSA des effets désincitatifs propres au RMI.

Cette triangulation des allers-retours entre les trois moments de l'analyse des politiques publiques, et le caractère interdisciplinaire assumé de cette démarche, est au cœur du projet scientifique du LIEPP (Laboratoire Interdisciplinaire d'Évaluation des Politiques Publiques). Les détails se trouvent disponibles dans Wasmer et Woll (2011). Elle reprend les éléments des différentes étapes des politiques publiques (Graphique 3).

Figure 3: Les liens entre les trois moments de l'évaluation. (Source Wasmer et Woll, 2011).



S'entendre sur une méthode de travail

Les différents moments de l'évaluation ne peuvent être croisés que si des disciplines distinctes (et organisées comme telles, cf. supra) peuvent utilement échanger entre elles. Cela implique de réfléchir au jeu d'incitations qui peut animer les différents chercheurs. À partir du moment où l'évaluation des chercheurs se fait au travers de la reconnaissance des pairs, et où les publications dans des revues à comité de lecture jouent le rôle essentiel, le fait de travailler avec des chercheurs d'autres disciplines implique une réallocation du temps disponible qui pourrait décourager les chercheurs les plus investis dans leur champ disciplinaire ; il convient donc d'analyser le système d'incitations permettant de corriger cette tendance naturelle. On peut bien entendu espérer que les gains de la collaboration soient suffisamment forts pour que les chercheurs espèrent publier mieux dans leur discipline lorsque des collaborations interdisciplinaires ont permis des avancées scientifiques. Mais comme tout investissement à haut rendement potentiel, celui-ci est sans doute plus risqué d'une part, d'autre part sujet à des économies d'échelle, en raison de la structure des coûts fixes. En particulier, les coûts de transactions liés au manque d'un langage commun entre différentes disciplines est un obstacle *a priori*.

Une option incitative serait alors de rémunérer les chercheurs pour leur participation à des travaux réunissant des équipes différentes, sous forme de primes de direction de programmes de recherche. Une perspective plus dirigiste serait de recommander aux départements disciplinaires de tenir mieux compte des revues hors de la discipline lorsque celles-ci sont de qualité. Après tout, Gary Becker a connu l'énorme succès qu'on lui connaît, d'une part après des échecs au sein de la discipline économique (et notamment la façon négative dont son travail sur le capital humain a initialement été perçu au sein de l'*American Economic Association* dans les années 1960, voir Becker – 1993), et d'autre part a publié dans les meilleures revues de démographie, d'éducation ou de criminologie, ce qui a assuré la diffusion de ses idées et de sa méthodologie. La question de savoir si la décision d'accorder la titularisation à un professeur assistant doit prendre en compte les publications dans des revues hors du champ, lorsqu'il existe un classement consensuel de ces revues et de leur qualité. C'est une question ouverte mais il est facile de trouver des arguments en faveur et en défaveur d'une telle solution. Le risque de prendre en compte les publications hors discipline est de donner une incitation à délaisser son propre champ dès lors que les premiers échecs de publication surviennent, ou à tout le moins de réduire l'exigence sur les strictes publications académiques disciplinaires. Il peut au contraire être intéressant de valoriser des publications hors du champ si le chercheur fait preuve de sa capacité, d'une part de publier dans les meilleures revues disciplinaires (disons les « top 5 » ou « top 10 ») et d'exporter son thème d'intérêt dans des revues d'excellence hors de son champ. S'il fallait prendre partie, on pourrait plaider en faveur d'une forte complémentarité entre des publications des deux disciplines (il est nécessaire de faire ses preuves dans son champ d'origine) mais aussi qu'un certain degré de substituabilité est souhaitable : au risque du dissensus, deux publications « top 10 » dans deux champs disciplinaires pourraient bien valoir une « top 5 » dans un champ dans la mesure où ces deux publications de champs différents sont à la fois une preuve d'excellence et d'ouverture d'esprit.

Dans tous les cas, le terrain fertile sur lequel un vrai travail interdisciplinaire de qualité peut se construire semble être la capacité à réunir une masse critique de chercheurs de très haut niveau et de mettre à leur disposition une structure leur permettant de travailler ensemble sans que les enjeux administratifs ou corporatistes ne prennent le dessus. L'idée est bien d'éviter une convergence vers le bas et de susciter des avancées scientifiques réelles.

Le travail n'a du reste pas besoin d'être réalisé en commun, à plusieurs mains : la confrontation systématique entre disciplines, idéalement au début, au milieu et à la fin d'une étude donnée, peut dans un premier temps suffire à significativement améliorer la qualité et susciter de nouvelles questions, voire des collaborations effectives.

Une question pratique qui se pose est également de savoir si les économistes doivent favoriser les collaborations avec des non-économistes qui seraient plutôt d'inspiration quantitative, ou au contraire avec ceux qui sont inscrits dans des démarches qualitatives. Les sociologues et politistes quantitatifs ont un mérite, celui de disposer des éléments communs du méta-langage de la statistique et de plus en plus, du langage de l'identification causale. Le dialogue avec les qualitatifs est plus complexe en raison de l'absence de langage commun, mais peut se révéler très fécond, pour la simple raison qu'il n'a jamais été tenté et permet précisément de documenter ce que les économistes oublient faute de données.¹⁶

Il reste enfin une problématique à aborder, qui d'ailleurs explique ce qui rend si difficile le dialogue entre les disciplines. En réalité, il existe un fossé paradigmatique entre ceux qui estiment rechercher une vérité cachée que les chiffres permettent de révéler, qui se retrouvent sous le label de scientifiques prônant l'« *evidence-based policy* » ; et ceux qui estiment, à des degrés divers, qu'une telle recherche est rarement neutre, soit parce que les chiffres eux-mêmes sont des constructions sociales, soit parce que l'évaluation elle-même est un processus politique, parfois, si ce n'est toujours, instrumentalisé. On ne tranchera pas dans ce débat. On se contentera de dire qu'une donnée, même « construite socialement », semble toujours plus objective qu'une appréciation non réfutable, et que la démarche qualitative devrait conduire à éliminer le plus grand nombre des assertions non-réfutables.

Section 5. Conclusion

Cet article, après avoir résumé le principal enseignement d'un ouvrage de P.E. Tetlock sur la démarche cognitive des experts et le coût relatif de la spécialisation en termes de succès prédictif qu'il suggère, a ensuite passé en revue les avantages et les inconvénients de la pluralité des approches. Les contradictions entre les divers objectifs des praticiens de l'évaluation (scientificité d'une part, capacité à répondre à des questions posées par le décideur d'autre part) et entre diverses méthodologies et leurs différences de langage, engendrent de nombreux coûts de transaction. En particulier, les coûts de transaction liés à des questions de langage d'une part, de gestion des carrières d'autre part, nécessitent une réflexion sur la mise en œuvre, si elle est souhaitable, de la pluridisciplinarité.

La question initialement posée dans cet essai était double : d'une part, quelle est la nécessité d'éclairages multiples dans des évaluations de politiques publiques, et d'autre part, faut-il aller jusqu'à effacer les frontières disciplinaires dans la pratique de la recherche ? La réponse apportée ici à la première question est que toute évaluation avec un regard multiple sera de bien meilleure précision si les points de vue se confortent, et évitera d'être biaisée, et donc fautive, si les points de vue s'opposent. Cela permet donc de mieux peser, d'un point de vue scientifique, les conclusions atteintes, comme il a été montré dans l'un des exemples cités dans ce texte, à propos de la discrimination sur le logement. La réponse à la seconde question semble sans ambiguïté négative, tant le sacrifice de la gestion disciplinaire de l'excellence est important au regard des éventuels avantages d'une pratique scientifique multidisciplinaire : l'évaluation des chercheurs par leurs pairs n'a de sens que s'ils sont sur des champs bien délimités.

Pour pouvoir répondre qu'une seule discipline, en l'occurrence la science économique, pourrait faire l'économie des apports d'autres disciplines, elle devrait déjà parvenir à cet objectif très ambitieux, qui serait de pouvoir arriver à incorporer dans la paramétrisation de modèles d'évaluation ex ante, à la fois les enseignements des évaluations ex post (expériences naturelles ou contrôlées) et les enseignements plus qualitatifs liés à la mise en œuvre ainsi que la dimension politique de l'appropriation des réformes sur lesquels les points de vue sociologiques et de science politique sont les plus précis. De fait, peut-être y-a-t-il pour les économistes une certaine urgence à prendre mieux en compte des dimensions non mesurables mais qui peuvent affecter profondément la justesse des recommandations.

16 De même que la psycho-histoire rêvée par l'auteur de science fiction Isaac Asimov permettait la prévision du futur en alliant mathématiques, sciences sociales et histoire, la prise en compte des facteurs politiques et sociologiques de la mise en œuvre d'une politique publique pourrait constituer le graal du modélisateur structurel dans son approche ex ante !

Références

- Attanasio, Orazio, Costas Meghir et Ana Santiago. (2005) « Education choices in Mexico: Using a structural model and a randomized experiment to evaluate Progresa », wp IFS 05/01.
- Baslé, Maurice. (2008). *Suivi et Evaluation des Politiques Publiques et des Programmes*, Economica.
- Behaghel Luc, Bruno Crepon et Marc Gurgand. 2011. *Evaluation d'impact de l'accompagnement des demandeurs d'emploi par les opérateurs privées de placement et le programme Cap vers l'entreprise*.
- Gary S. Becker, "Nobel Lecture: The Economic Way of Looking at Behavior", *The Journal of Political Economy*, Vol. 101, No. 3, (Jun., 1993), pp. 385-409
- Bonnet, François, Etienne Lalé, Mirna Safi and Etienne Wasmer (2012). What is really "discriminatory"? Disentangling ethnic and spatial stigma in the measurement of housing discrimination in France
- Chemin, Matthieu et Wasmer, Etienne (2009). "The employment effects of 35-hour workweek regulation in France: using Alsace-Moselle local laws to build a diff-in-diff", *Journal of Labor Economics* October 2009, Vol. 27, No. 4: 487-524
- Chemin, Matthieu and Etienne Wasmer (2011). "Ex-ante and ex-post evaluation of the 1989 French welfare reform using a natural experiment : the 1908 social laws in Alsace-Moselle", mimeo Mc Gill et Sciences Po.
- Ferracci Marc et Etienne Wasmer. (2010). *Etat Moderne, Etat Efficace*, Odile Jacob.
- Gary-Bobo, Robert et Jean-Marc Robin. 2011. "La querelle des redoublements: analyse économique et problèmes statistiques", *Revue Economique*.
- Jacobs, Steve (2008). Cross-Disciplinarization : A New Talisman for Evaluation ?, *American Journal of Evaluation*, pp 175-194, vol. 29.
- Kahneman Daniel et Amos Tversky, « Prospect Theory: An Analysis of Decision under Risk », *Econometrica*, vol. 47, no 2, mars 1979, p. 263-291.
- Langot, François et Pascale Petit, 2011. Introduction du numéro spécial L'Évaluation des Politiques Publiques, *Revue Française d'Économie*, Vol. XXVI, juillet.
- Langot, François et Pascale Petit, 2012. Introduction du numéro spécial L'Évaluation des Politiques Publiques, *Revue Française d'Économie*, Vol. XXVII, juillet.
- L'Horty, Yannick et Pascale Petit, 2011. Évaluation aléatoire et expérimentations sociales, *Revue Française d'Économie*, Vol. XXVI, pp13-48
- Lascombes Pierre et Michel Setbon (1996). *L'Évaluation Pluraliste des Politiques Publiques. Enjeux, Pratiques, Produits*, CNRS-Commissariat Général au Plan, 1996.
- Lise, Jeremy, Shannon Seitz and Jeffrey Smith. (2005). " Equilibrium Policy Experiments and the Evaluation of Social Programs ", working paper, U. Michigan.
- Mayer, T., Malgouyres C., Mayenris, F. et L. Py, Évaluation des effets des politiques ciblées sur les territoires : le cas des Zones Franches Urbaines Rapport remis à la DARES le 27 janvier 2012
- Michalopoulos Charles et alii. (2002). « Making Work Pay. Final Report on the Self-Sufficiency Project for Long-term Welfare Recipients », Social Research and Demonstration Corporation.
- Piketty Thomas. : *Les Hauts revenus en France au 20e siècle : inégalités et redistribution, 1901-1998*, Paris : B. Grasset, 2001, 807 p.
- Scriven, Michael. (2010). *Evaluating evaluations : a meta-evaluation checklist*, mimeo, Claremont Graduate University.
- Tetlock, Philip E. (2005). *Expert Political Judgment. How Good Is It? How Can We Know?*, Princeton Univ. Press.
- Wasmer, Etienne et Woll, Cornelia, 2011. "Scientific project, Sciences Po | LIEPP laboratoire interdisciplinaire d'évaluation des politiques publiques or in english, interdisciplinary research center for the evaluation of public policies," MPRA Paper 31434, University Library of Munich, Germany.

Discussion par : Christine Musselin

Longtemps objet privilégié de la science politique et du management public – et plus marginalement de la sociologie –, l'évaluation des politiques publiques est aujourd'hui revendiquée comme un champ de recherche de l'économie, ce dont on peut se réjouir si les travaux menés capitalisent sur ceux conduits sur le même objet par les disciplines qui ont été à son fondement. Dans le texte qui précède, Etienne Wasmer explique avec grande clarté et conviction ce qu'apporte une lecture économiste des politiques publiques et comment elle pourrait être combinée avec d'autres disciplines. Il nous permet ainsi de mieux comprendre ce qui la distingue d'autres perspectives et du coup invite à prolonger le dialogue. Je le ferai en développant trois questions : Prévoir est-ce évaluer ? L'évaluation gagne-t-elle à mobiliser plusieurs disciplines ? Conflits de discipline ou opposition entre approche objectiviste et constructiviste ?

1. Prévoir est-ce évaluer ?

Mon premier point de discussion porte sur l'assimilation que font les économistes entre prévision et évaluation et qu'Etienne Wasmer justifie en s'appuyant sur les travaux de Philip E. Tetlock. Se demander quels effets pourrait avoir une politique publique et étudier les effets qu'elle a eus relèveraient ainsi de la même catégorie d'opérations. Il n'est pourtant pas évident que les compétences requises ou les méthodes mobilisées doivent être les mêmes selon que l'on fait un bilan ou que l'on cherche à estimer les effets à venir. Les experts de la prévision étudiés par Tetlock n'étaient d'ailleurs probablement pas les évaluateurs qui lui ont permis de comparer ce qui est effectivement survenu et ce qui avait été annoncé. Deux raisons me semblent par conséquent justifier de maintenir deux verbes distincts pour désigner les deux opérations, quelles que soient les disciplines.

D'une part, la prévision est par nature balistique alors que l'évaluation ne devrait pas l'être. La prévision a en effet pour mission de juger si le problème visé par une politique donnée sera ou non réglé par celle-ci, ou, pour reprendre le cadre d'analyse proposé par Etienne Wasmer de dire aux décideurs s'il est utile et efficace d'engager les dépenses publiques envisagées. L'évaluation (ex post) ne peut se contenter d'une

opération aussi simple (dans ses attendus) et doit en revanche tenir compte de l'évolution des objectifs au cours de la mise en œuvre, des effets induits que cette dernière a provoqués, mais aussi du processus lui-même et pas seulement de ses résultats rapportés aux objectifs initiaux. Autrement dit, comme nous y invitent les travaux issus de la *public policy analysis*, l'évaluation (ex post) ne peut se limiter à mesurer la distance entre l'objectif visé et l'objectif atteint (Kessler et al. 1998). Aussi consiste-t-elle en un processus complexe qui comprend plusieurs phases engageant différemment l'intervention d'experts et de décideurs. Tout d'abord, un état des lieux, qui vise à produire à un moment « t » une photographie d'une situation donnée et mobilise surtout de l'expertise méthodologique. La seconde phase peut alors être engagée afin de produire un diagnostic, ou si l'on préfère de porter une appréciation critique sur la photographie produite : correspond-elle à ce qui était attendu ? D'autres objectifs ont-ils été atteints ? Y a-t-il eu des effets inattendus bénéfiques (ou pervers) ? C'est aussi au cours du diagnostic que seront recherchées et identifiées les causes des éventuelles dérives, qu'elles soient liées à la plus ou moins grande robustesse de la théorie implicite qui a justifié la mesure ou aux dynamiques propres à la mise en œuvre de celle-ci (Majone et Wildavski 1984). Enfin le diagnostic permet de confronter les résultats,

non pas aux objectifs d'hier mais à ceux d'aujourd'hui et il suppose de se demander si la photographie obtenue est satisfaisante par rapport aux objectifs que l'on se fixe désormais. Ce n'est qu'en fonction de la réponse à cette dernière question qu'une troisième phase pourra être engagée, qui pourra conduire à de nouvelles mesures... dont on pourra chercher à prévoir les effets.

Une seconde raison pour distinguer prévision et évaluation tient à ce que la prévision intègre mal les dimensions temporelles. Non seulement les résultats attendus de la mesure envisagée n'intéressent les politiques que s'ils surviennent dans un délai compatible avec les échéances électorales, mais surtout la précision et la fiabilité de la prévision s'érodent quand elle s'inscrit dans le long terme tant les évolutions du contexte institutionnel au sein duquel une mesure va se développer sont imprévisibles et incertaines. Au contraire, l'évaluation (ex post) intègre la temporalité : elle doit tenir compte des délais de mise en œuvre pour étudier les effets d'une mesure donnée mais aussi prendre en considération l'éventuelle transformation de la définition initiale du problème, des reconfigurations des équilibres politiques survenus entre temps etc. La dimension temporelle est donc partie intégrante de l'évaluation (ex post) et celle-ci doit autant porter sur ce qui a été accompli que sur la dynamique de ce qui s'est produit. Toute évaluation ne vaut que située dans le temps. Ses conclusions sont de ce fait rarement définitives et peuvent varier selon le moment où elle se produit.

2. L'évaluation gagne-t-elle à mobiliser plusieurs disciplines ?

Le texte d'Etienne Wasmer ne soulève pas seulement la question de la distinction entre évaluation et prévision. Il s'interroge aussi, au vu des travaux de Tetlock sur les mérites comparés des renards et les hérissons, sur les bienfaits et les méfaits de la pluralité des disciplines pour les évaluations. Pour Etienne Wasmer, cette pluralité pose deux problèmes. Le premier tient à ce que la reconnaissance scientifique est attribuée sur des bases monodisciplinaires et que les chercheurs ont donc intérêt à s'enfermer

dans leur discipline plutôt qu'à pratiquer la pluralité et le second à ce que différentes approches peuvent aboutir à des résultats différents voire opposés. Je laisserai de côté le premier problème qui renvoie à des questions épistémiques sur ce qui « fait science » dans différentes disciplines.

En ce qui concerne le second problème, il me semble que les exemples que donne Etienne Wasmer montrent en fait les limites de la pluridisciplinarité, c'est-à-dire de la juxtaposition d'approches disciplinaires différentes. Dans le cas des redoublements comme dans celui des agences immobilières, l'analyse repose sur la conduite parallèle d'enquêtes différentes. Chacun travaille de son côté avec ses méthodes et ses concepts et apporte une explication complémentaire qui peut être éclairante (comme dans le cas des redoublements) mais aussi contradictoire, ce qui ne permet pas de conclusions communes (comme dans le cas des agents immobiliers). Ce principe de juxtaposition se traduit aussi par un partage (plus ou moins équitable et négocié) des tâches qui confie à la sociologie, la science politique ou l'histoire ce qui est « résiduel », c'est-à-dire ce qui reste « inexplicé » une fois établies des régularités statistiquement significatives.

Ces limites, mais aussi des expériences de recherche récentes, me conduisent par conséquent à plaider pour l'interdisciplinarité, à laquelle le « I » du LIEPP nous invite, c'est-à-dire pour des démarches plus intégrées. D'une part, l'interdisciplinarité repose sur un traitement symétrique des disciplines concernées et non sur une répartition des rôles, ce qui facilite les partenariats. D'autre part, la construction d'une recherche (ou d'une évaluation) interdisciplinaire articule les disciplines sans les dissoudre. Il s'agit de collecter et de traiter des données qui fassent sens pour les unes et les autres et surtout de construire le travail empirique et méthodologique conjointement et non séparément : l'élaboration du design de recherche est plus longue ; elle suppose de part et d'autres de l'écoute et des efforts, mais les résultats que permettent des éclairages croisés plutôt que complémentaires ou contradictoires méritent largement cet investissement.

3. Conflits de disciplines ou opposition entre approche objectiviste et constructiviste ?

Il me semble pour finir qu'une tension plus difficile à surmonter que les conflits de disciplines est celle qui traverse certaines disciplines (notamment la sociologie ou la science politique) et oppose approche objectiviste et constructiviste. Comme l'observe Bernard Lahire (1999) dans le premier chapitre de son ouvrage sur l'illettrisme, certains chercheurs considèrent qu'un phénomène social est « donné » et que le travail du scientifique est d'en mesurer l'ampleur, d'en déterminer les causes, d'en apprécier les effets. Il s'agit alors de s'inscrire dans une démarche objective, qui est le propre des économistes aujourd'hui, mais aussi de nombreux sociologues, politistes ou historiens. Ils considéreront ainsi que la consommation de drogues est un phénomène social, et qu'il s'agit de caractériser statistiquement ceux qui s'y adonnent (âge, sexe, classe sociale, ethnie...), d'identifier les facteurs aggravants ou au contraire prévenant la prise de stupéfiants, d'évaluer l'efficacité de mesures publiques visant à contrer ce phénomène...

Pour d'autres chercheurs, tout phénomène doit d'abord être appréhendé comme un construit.. Il ne s'agit pas, sauf à tomber dans la relativité la plus extrême, de nier l'existence même des faits, mais de s'intéresser aux processus par lesquels ces derniers (chômage, toxicomanie, inflation...) ont été identifiés, définis d'une certaine manière, qualifiés et nommés. Les travaux portent alors sur les controverses et les conflits, par lesquels un groupe d'acteurs (*claimmakers* pour Malcolm Spector and John I. Kitsuse, 1977 entrepreneurs de cause pour Howard Becker, 1966 et 1985) a rendu ce phénomène visible, lui a donné un nom, en a imposé une définition. Cette construction sociale des problèmes qui a notamment été défendue par l'école de sociologie de Chicago, a été largement reprise par les analystes des politiques publiques sous l'angle de la « construction des problèmes publics » et de la « mise sur agenda ». Or cette

perspective n'est pas sans importance pour l'évaluation des politiques publiques au moins pour trois raisons. D'une part une telle approche conduit l'analyste à s'interroger sur l'évidence des catégories qu'il mobilise et ce faisant sur « l'objectivité », la « naturalité » des données dont il dispose : pour prendre un exemple trivial, le nombre de chômeurs renvoie à la définition que l'on retient du chômage. D'autre part, cette approche attire l'attention sur le caractère politique, c'est-à-dire inéluctablement négocié, produit de rapports de pouvoir et soumis à redéfinition, de toute mesure publique. La manière dont un problème est construit et défini a une incidence sur le registre de solutions qui sera mobilisé (ou au contraire exclu), sur les groupes d'acteurs ou les facteurs auxquels on imputera la cause du problème, et sur les services ou secteurs qui seront chargés de trouver et de mettre en œuvre des solutions : définir l'obésité comme un problème de comportement individuel suscitera des mesures de contrôle des pratiques alimentaires par le ministère de la santé, alors que la définir comme un problème lié à la composition des produits de consommation entraînera des mesures visant l'industrie alimentaire et mobilisera le ministère concerné. Les mesures à prendre ne sont pas seulement (Hall 1989) viables en théorie mais aussi politiquement (au sens où elles satisfont les intérêts en présence) et administrativement (c'est-à-dire compatibles avec les capacités d'action dont dispose l'administration) (Hall 1989). Enfin, une telle posture suggère que le bon fonctionnement d'une démocratie passe par la qualité et la nature des processus de mise sur l'agenda (qui a accès à l'agenda, qui définit les problèmes, qui est chargé de les traiter), et par celle des processus de mise en œuvre, et ne peut être réduit à la question de l'efficacité des actions publiques.

Aussi, le problème à surmonter est-il peut-être moins celui de la combinaison entre les disciplines que celui de la capacité, dans l'évaluation des politiques publiques à articuler, au-delà des disciplines, approche objective et subjective.

Références

- Spector M., Kitsuse J.I. (eds.), *Constructing Social Problems*, Cummings, Menlo Park, 1977
Becker H.S. (ed.), *Social Problems : a modern approach*, John Wiley, New York, 1966
Becker H.S., *Outsiders : études de sociologie de la déviance*, Paris, Métaillié, 1985
Lahire B., *L'invention de « l'illettrisme » ? Rhétorique publique, éthique et stigmat*, Paris, La Découverte, 1999
Majone, G., Wildavsky A., "Implementation as evolution" dans Pressman J. et Wildavsky A, *Implementation*, Los Angeles, University of California Press, 1984 (3ème ed.)
Hall P., « Conclusion : The politics of keynesian ideas », dans Hall, P. (1989) : *The political power of economic ideas. Keynesianism across Nation*, Princeton, Princeton university press
Kessler M.-C., Lascoumes (P.), Setbon (M.) et Thoenig (J.-C.) *Évaluation des politiques publiques*, Paris : L'Harmattan, 1998.
-



***Le LIEPP (Laboratoire interdisciplinaire d'évaluation des politiques publiques) est un laboratoire d'excellence (Labex).
Ce projet est distingué par le jury scientifique international désigné par l'Agence nationale de la recherche (ANR).
Il est financé dans le cadre des investissements d'avenir.***

(ANR-11-LABX-0091, ANR-11-IDEX-0005-02).

www.sciencespo.fr/liepp

Si vous voulez recevoir nos publications et rester informés de nos activités, merci d'envoyer un courriel à :
liepp@sciences-po.fr

Directeurs de publication:

Etienne Wasmer & Cornelia Woll

Maquette:

Alexandre Biotteau & Iana Markevitch

Sciences Po - LIEPP
27 rue Saint Guillaume
75007 Paris - France
+33(0)1.45.49.83.61