

Les expérimentations aléatoires et la nouvelle économie du développement

MOHAMED SOUAL
Université Grenoble Alpes.

Résumé

L'article analyse les expérimentations aléatoires contrôlées (Random Controlled Trials ou RCT)* dont ses promoteurs veulent être une microéconomie du développement basée sur la production de preuves (Evidence Based Approach) par le recours à des évaluations *ex-ante* de programmes de développement. L'essor de ces ECR coïncide avec la reconfiguration du champ de l'économie du développement et la place que l'analyse empirique y prend. Assise sur une base médiatique large, la méthode occupe une place prépondérante dans le marché de l'expertise. En comparant des groupes de auxquels les traitements sont appliqués à des groupes témoins, les ECR cherchent à inférer, sans soubassements théoriques, des causalités « incontestables » posant des problèmes épistémologiques et non dépourvus de biais statistiques. En plus des problèmes de validité externe que la méthode soulève, ses apories statistiques et ses failles épistémologiques entachant la méthodologie participent à déconstruire à être le *gold standard* que ses promoteurs revendiquent.
Mots Clés: expérimentations, microéconomie du développement, causalité, contrefactualité, effets
JEL: O12, O17, C93, C21

Random controlled experiments and the new development economics

Abstract

The article analyzes Random Controlled Trial (RTC) experiments that their promoters wish to be seen as a development microeconomy based on the production of evidence (Evidence based approach) through the use of evaluations *ex ante* of development programmes. The rise of these RTC coincides with the reconfiguration of the field of development economics and the role played by empirical analysis. Supported by a wide mediatic base, the method occupies a major place in the expert market. By comparing groups which have received these treatments to control groups, the RTC seek to infer without theoretical basis, -undeniable- causalities which pose epistemological problems which at the same time show statistical features. In addition to problems of external validity occasioned by the method, its statistical deficiency and epistemological weaknesses cloud the methodology and together deconstruct the *gold standard* which its promoters demand.

Keywords: Experiments controlled, development microeconomics, causality, confractuality, effects
JEL: O12, O17, C93, C21

Introduction

Dans le cadre de la reconfiguration de son champ et de ses contenus, initiée à partir des années 80 sous l'impulsion du FMI et de la Banque mondiale, l'économie du développement réserve de plus en plus de place à l'analyse empirique, mettant ainsi à profit les apports de l'économétrie tant dans l'évaluation des politiques économiques et la comptabilité de la croissance qu'en micro-économie de l'éducation, de la santé ou travail.

L'approche des évaluations aléatoires des politiques de développement, promue par Abhijit Banerjee et Esther Duflo, participe largement de ce tournant que d'aucuns associent, en s'appuyant sur la révolution de la crédibilité (Angrist et Pischke 2010), à l'émergence d'une nouvelle économie du développement. Prenant le contrepied de la technique des régressions à l'œuvre en macro-économie, cette approche recourt aux expérimentations contrôlées qui consistent à tester plusieurs programmes d'action sur des populations bénéficiaires à comparer à des populations témoins en vue d'identifier des effets causaux. A cet égard, elle plaide pour une refondation micro-économique des politiques publiques qui vise à appréhender le développement dans le sillage d'une ingénierie de la lutte contre la pauvreté.

Ainsi définie, elle affirme son originalité à partir d'un modelé qui réunit des ingrédients des méthodes de Holland, Neymar et Rubin et étudie les questions de causalité et de contrefactualité (partie 1). A examiner de près, on s'aperçoit que ce modèle placé en *gold standard* contient des limites qui tiennent au défaut de teneur théorique, aux biais statistiques censés être éliminés précisément par les méthodes appliquées autant qu'aux effets ambigus des procédures contrefactuelles. Examinée à l'aune de son projet, l'inférence causale prônée par la méthode expérimentale reflète les apories du primat indu octroyé à l'empirie (Partie 2).

Ainsi, cette recherche propose non seulement de décrire la méthodologie employée par les auteurs, elle vise également à souligner certains apports de la vision micro individuelle en vue d'enrichir la micro et méso économie.

Ainsi, cette recherche propose non seulement de décrire la méthodologie employée par les auteurs, elle vise également à souligner certains apports de la vision micro individuelle en vue d'enrichir la micro et méso économie. Aussi, l'article s'est-il attaché à montrer les limites, les apories et les failles épistémiques, passées souvent sous silence, et a contribué à déconstruire la prétention à la force de la validité interne conférée par l'élimination des biais de sélection montrant subséquemment que les problèmes de généralisation que connaissent les expérimentations aléatoires contrôlées (RCT) sont, précisément, dus à la faiblesse de la validité interne d'une part et à l'absence de théorie, d'autre part.

Principes de la causalité statistique et dépendance contrefactuelle

« La bonne nouvelle [...] est qu'il est possible d'améliorer la gouvernance et les politiques sans changer la structure sociale et politique existante » (Banerjee et Duflo, 2012, p. 271). En considérant les dimensions politiques et institutionnelles, macrosociales comme données, le projet du Jameel Poverty Action Laboratory (J-PAL) du MIT se place délibérément dans une optique qui privilégie la rationalité individuelle. A ce titre, il repose sur l'extension au champ de l'économie du développement de la méthode des essais cliniques, en usage en médecine, et vise à identifier les effets d'un programme face à des chaînes de causalité.

Cette méthode cherche à poser des questions de base dans les champs de l'éducation, de la sante, de microcrédit ou de l'activité de production agricole.

Parce que les participants à un programme ne sont pas comparables aux non-participants, et l'on ne dispose pas de l'information qui permet de savoir ce qui leur serait advenu s'ils n'avaient pas participé au programme, il est malaisé de séparer *l'effet causal* du programme de celui d'autres facteurs (Banerjee et Duflo 2009, p.693). Le but dès lors est de chercher à comparer la situation des bénéficiaires d'un programme à la situation qu'ils auraient vécue s'ils n'y avaient pas été exposés. La difficulté consiste par conséquent à constituer un groupe de contrôle adéquat des lors que l'on ne peut à la fois appliquer un programme et ne pas l'appliquer à un même individu et en même temps.

Pour les promoteurs de cette méthode, les évaluations conduites à partir de données d'observation non expérimentales sont entachées de biais statistiques. Aussi, seule l'évaluation par expérimentation randomisée permet-elle de s'affranchir de ces biais de sélection en proposant d'affecter le programme de manière aléatoire par tirage au sort au groupe qui bénéficiera du programme devant être évalué ou bien à un groupe témoin. L'un constitue le groupe de traitement et l'autre groupe de contrôle, dit également contrefactuel. « Il est bien entendu impossible, affirment Banerjee et Duflo, d'identifier les effets du traitement pour chaque personne individuellement. Mais une expérience randomisée, en faisant varier un seul facteur à la fois (le programme), nous permet d'obtenir une estimation valide de l'effet moyen d'un traitement pour une population donnée » (2009, p.693).

Le recours à cette procédure permet, selon le J-PAL, de sortir du sommeil dogmatique des « idéologies » pour entrer dans l'ère du pragmatisme et de la scientificité. Aussi, s'inscrit-il contre l'économie politique de l'aide au développement en renvoyant dos à dos les deux grandes théories de la pauvreté globale. Celle défendue par Jeffrey Sachs émet l'idée d'une « trappe de la pauvreté » dans laquelle les pays pauvres seraient enfermés et de laquelle ils ne pourraient sortir qu'à l'aide d'envois de fonds massifs. La seconde, portée par William Easterley, récuse l'idée de toute forme d'aide, responsable selon lui de corruption généralisée, d'infantilisation des pauvres et d'affaiblissement de la gouvernance locale.

Selon les promoteurs des expériences aléatoires contrôlées, au lieu de passer l'aide internationale au crible, programme par programme, afin de préciser quels sont les programmes qui marchent et ceux qui ne marchent pas, les experts s'affrontent autour de « grandes questions » comme : Quelle est la cause ultime de la pauvreté ? Quelle confiance peut-on accorder au marché ? La démocratie est-elle une bonne chose pour les pauvres ? L'aide internationale attelle un rôle à jouer ?

Jeffrey Sachs (2005) estime que ni le marché ni la démocratie ne pourront sortir les pays pauvres de leur trappe, l'aide internationale est la clé en ce qu'elle est en mesure d'enclencher un cercle vertueux à travers l'investissement qui accroît la productivité et les revenus et créent les conditions d'une croissance soutenue.

De l'autre cote, William Easterley considère que l'aide au pays du Sud ne peut résoudre le problème du sous-développement. Dans deux ouvrages consacrés à la question (Easterley, 2006, 2009), l'auteur soutient que lorsque le marché génère des incitations adaptées et stimule l'esprit d'initiative tandis que l'aide propage la corruption et les comportements rentiers, les agents parviennent à trouver des solutions pour résoudre leurs problèmes.

À la différence de Craig Burnside et David Dollar (1997) qui considèrent qu'il existe un lien fort entre croissance et aide dans les pays lorsque les institutions sont solides, il souligne que les pays qui reçoivent plus d'aide n'enregistrent pas une progression plus rapide de leur PIB.

A l'encontre de cette démarche macro-économique, Duflo et son équipe soutiennent que l'évaluation aléatoire du terrain s'impose comme une alternative pour évaluer l'impact des actions publiques. Pour appuyer son propos, Duflo n'hésite pas à affirmer qu'« Il n'est tout simplement pas possible d'isoler les mécanismes profonds de la croissance économique avec comme seul guide les expériences de croissance d'une centaine de pays. Toute variable peut être cause ou effet, ou bien pourrait être expliquée par une troisième variable corrélée aux deux autres. [...] Il faut bien se rendre à l'évidence. Même avec deux millions de régressions, nous n'arriverons pas à percer le secret de la croissance à partir d'une base de données sur l'expérience passée d'une centaine de pays (Duflo 2009, pp. 23 et 24).

Dans ce contexte, les liens entre les niveaux macro et micro soulèvent des difficultés qui tiennent à la fréquence réduite des données et leur fiabilité discutable dans les pays pauvres. Avec des échantillons faibles et des coûts réduits (à voir) et des questions posées très spécifiques, la méthode expérimentale offre une collection de données originales, et c'est ce qui permet de dépenser plus d'argent par variable et par observation contrairement aux grandes enquêtes polyvalentes auprès des ménages ou des entreprises qui doivent compter avec les variations dues aux changements de politique ou induites par le marché (Banerjee et Duflo, 2009, p.700). La micro-économie étant d'après eux le seul champ approprié au traitement de la pauvreté et à la promotion d'améliorations à la marge au moyen d'évaluations rigoureuses et ce jusqu'à produire des effets cumulatifs, les pauvres souffrant de biais de rationalité et d'incohérence temporelle. L'accent devant alors être porté sur le comportement des pauvres et sur la manière avec laquelle ils prennent leurs décisions (Favereau, 2016).

Cette nouvelle configuration de l'économie du développement a remis au goût du jour les questions de méthode et interroge sur les rapports entre théorie et expérience. Ainsi, défendre des preuves « irréfutables », mettre en avant les inférences causales « incontestables » par un travail empirique en récoltant, stockant et faisant parler des données constitue ce que d'aucuns n'hésitent pas à appeler *le tournant empirique* de l'économie. Alors qu'il était inenvisageable de publier un article utilisant les méthodes économétriques empiriques qui ne soit adossé à un modèle théorique, la nouvelle économie du développement s'affranchit de cette exigence méthodologique. Tout au plus prétend-elle aux possibilités de tester les théories par l'expérimentation et au contrôle de conditions d'interaction de phénomènes dans le but de dégager des causes ou des déterminants d'effets.

Cette transformation fondamentale du rapport entre la théorie et l'empirie pose la question du contrôle du contenu des théories par l'empirie et le délicat passage de la description statistique à la généralisation inductive.

Les processus d'élaboration des savoirs et les procédures de leur validation sont une préoccupation épistémologique permanente. Les véritables grands problèmes de l'économie du développement, qui sont aussi les grands problèmes de l'économie dans son ensemble, découlent de l'adhésion à l'individualisme méthodologique dans un cadre de choix « rationnel » d'analyse économique induisant l'élaboration des politiques qui souvent ne sont ni uniquement théoriques ni exclusivement empiriques.

Il reste simplement à constater que l'économie du développement, davantage intégrée aujourd'hui à l'économie conventionnelle au cours des trente dernières années, est entrée dans une phase fortement empirique, ceci ayant coïncidé avec une disponibilité plus grande et plus accessible des données et une puissance de calcul accrue.

Dans ce contexte, la nouvelle économie du développement a donné lieu à un vif débat où ont été convoquées des questions portant sur le statut des propositions observationnelles au regard de la théorie. Le symposium réuni par *l'Economic and*

Political Weekly (2005) autour de quatre des meilleurs économistes indiens du développement : Dilip Mookherjee, Kaushik Basu, Abhijit Banerjee et Pranab Bardhan, apporte à cet important débat méthodologique les éléments suivants :

Mookherjee caractérise quatre étapes de la recherche :

L'étape une correspond à la description empirique, l'étape deux à la formulation de la théorie, l'étape trois aux essais et enfin l'étape quatre à la prévision et à l'évaluation des politiques. Au cours des quinze dernières années, l'accent a été mis sur la troisième étape. L'économie du développement est aujourd'hui principalement axée sur la mise à l'essai, en particulier sur la difficulté de déduire la causalité de la corrélation. C'est le sujet des débats classiques, comme celui entre Keynes et Tinbergen,¹ mis en lumière par Mookherjee mais également Bernard Walliser.

Et c'est particulièrement vrai pour deux domaines de la littérature empirique : celui qui se concentre sur la recherche de variables instrumentales et celui qui utilise un modèle expérimental d'évaluation randomisée. Ils amènent à la frontière de la méthodologie empirique en économie où Banerjee explique, en partie, pourquoi l'économie du développement est respectée dans la discipline de la science économique : « Je pense que le fait qu'il n'y ait pas de plaidoyer particulier pour un travail empirique utilisant des données de pays en développement est la raison pour laquelle les plus brillants et les meilleurs parmi les étudiants diplômés en économie arrivent maintenant sur le terrain » (*Economic and Political Weekly*, 2005, pages 4340-4344). Selon Mookherjee et Bardhan, l'attention portée aux données empiriques pourrait attirer l'attention sur toute une série de phénomènes théoriques intéressants, tels que « les implications dynamiques des imperfections du marché des facteurs pour les agents individuels, l'interaction des processus économiques avec les institutions politiques et les normes sociales » (Mookherjee, op.cit.), ou alors « ... comment aller au-delà des modèles d'oligopole d'équilibre partiel existants pour générer une théorie viable des prix des facteurs et de la répartition des revenus dans une économie généralement oligopolistique; la nature des dynamiques de transition (par opposition à la statistique comparative) à mesure que nous nous éloignons de l'équilibre traditionnel de bas niveau » (Bardhan, op.cit.).

De plus, et ceci est un reproche particulièrement fondé contre les conceptions expérimentales randomisées, le fait même que l'inférence soit « propre » à savoir qu'un paramètre spécifique de contrôle, rende problématique toute généralisation à partir de cette inférence. Basu argumente dans la même direction. Dans sa défense de la méthode empirique employée, Banerjee réplique que « ...même si de nombreuses régressions de basse qualité disent la même chose, il n'y a aucune raison pour que les preuves de haute qualité deviennent inutiles » et qu'après tout, la même source de biais pourrait affecter tous les résultats de basse qualité, acceptant, toutefois, que de nombreuses expériences randomisées menées à différents endroits et à différents moments permettent de renforcer la confiance dans leurs résultats.

Aucun des protagonistes ne dira que l'utilisation d'une simple corrélation pour déduire une causalité devrait être le mode opératoire de l'économie du développement empirique. Par ailleurs, un compromis semble se dégager sur le fait que l'environnement, même suffisamment contrôlé et générant les données pour permettre une inférence nette, n'est valable que pour cet environnement et pas un autre. Le rôle de la théorie pour aider à tirer le meilleur parti de ce compromis n'est pas non plus contesté.

¹ Keynes critique Tinbergen et son ouvrage *Statistical Testing of Business-Cycle Theories* de postuler par le calcul économétrique l'état futur de l'économie peut être calculé à partir de données statistiques passées.

Critiques et controverses à propos du *Gold Standard*

Abhijit Banerjee et Esther Duflo (2012, p. 38) considèrent qu'en raison de l'ignorance, de l'idéologie et de l'inertie, ce qu'ils nomment les trois « i », beaucoup de politiques de développement auraient échoué. L'évaluation efficace des programmes de développement doit passer par l'expérimentation aléatoire contrôlée consacrée comme le *Gold Standard* méthodologique qui permet de produire des résultats fiables et solides et en y apportant les preuves incontestables par l'élimination de tout biais de sélection.

Le *Gold Standard* attribue aux RCT (*Randomize Controller Trial*) vient de l'envolée qu'ont connue les évaluations d'impact au cours deux dernières décennies, en particulier en matière d'impact des politiques de développement. Elles proviennent également de la revendication de la supériorité intrinsèque des RCT sur tout autre méthode (Duflo et al., 2007, cité par Bédécarrats et al., 2013).

Cet attribut vient également de la place accordée aux articles académiques utilisant les RCT dans les revues les plus prestigieuses et de l'attrait exercé de plus en plus prononcé sur les doctorants. L'attribution à Esther Duflo de la médaille Clark distinguant les économistes de moins de quarante ans exerçant aux Etats-Unis, la chaire « Savoirs contre pauvreté » au Collège de France et l'audition par le congrès américain de Dean Karla, président et fondateur de Innovations for Poverty Action (IPA), une ONG de collecte de fonds pour les RCT (Bédécarrats *et al.*, 2013, p.118), ont participé à forger ce prestige.

Le succès retentissant des RCT provient également de ses fonctions à la fois rhétorique et performative. Sa fonction rhétorique s'incarne dans toute l'énergie déployée au niveau de tous les forums dont principalement le forum politique, avec la conviction que « connaître, c'est communiquer » (Walliser et al., 1988). Communiquer auprès des milieux académiques par la diffusion des méthodes et outillages nécessaires (réseaux associés, revues y consacrées, formation et recherche, chaire, etc.). Diffuser auprès de prescripteurs ou financeurs (ONG de développement, organismes de développement et de financement ou grandes fondations bailleurs de fonds) et auprès des médias par la vulgarisation de ce que Walliser appelle les *Folk Economiques* (ouvrages à destination du grand public, article de presse, sites spécialisés ou forums à fort impact médiatique). Ce qui fait écrire à Vincent Guillin (2013) : « Quiconque a été le témoin de la diffusion en économie du développement dans le paysage intellectuel français et de l'enthousiasme qu'elle a suscité auprès des instances en charge des politiques publiques ne manquera pas d'avoir été marqué par la surface médiatique prise par ce mouvement et par sa capacité à imposer à un public donne un agenda de recherche très structure ». Sa fonction performative découle naturellement du constat de Michel Callon: « La science économique au sens large performe, formate, façonne plus qu'elle n'en explique le fonctionnement ». Les revues, les ONG associées et les organismes finançant les RCT s'en imprègnent avant de les diffuser à leur tour. Elle tient également à « l'approfondissement de la technocratisation de la représentation statistique économique et sociale, assumée par les animateurs du J-PAL qui revendiquent le statut de « co-expérimentateurs » (Bardet *et al.*, 2012, p.186).

Limites des Randomized Controlled Trial

Le *gold standard* attribué aux RCT repose sur des malentendus : la randomisation fournit un test équitable en égalisant tout sauf le traitement et permet donc une estimation précise du traitement seul ; elle est nécessaire pour résoudre la sélection sans biais qui isole l'inférence causale recherchée ; et enfin que cette inférence statistique ne nécessite que la comparaison de deux moyennes (différence des moyennes dégagées pour les groupes de

traitement et de contrôle). Aucune de ces déclarations n'est vraie. Or, sans « remonter les phénomènes à leurs principes et découvrir les structures qui les règlent » (Frank, 2007), l'absence de lien entre les RCT et d'autres connaissances scientifiques rend difficile leur utilisation en dehors du contexte exact dans lequel elles sont menées (Deaton, Cartwright, 2018).

Les données qui en sont issues sont loin d'être idéales pour estimer les paramètres structurels des modèles comportementaux. Cela rend difficile la généralisation des résultats d'une expérience à l'autre ou l'utilisation d'expériences pour identifier les paramètres structurels invariants d'une politique nécessaire à l'évaluation de la politique (Heckman, 2005, cité par Roux, 2015, p.15).

Les expériences aléatoires contrôlées entrent dans ce que l'on peut appeler « les stratégies d'acteurs et de construction des notions de preuve et de vérité » que nous analysons sous les grilles de lecture de Michel Callon et Bruno Latour dans *la traduction et l'acteur réseau* » ou bien chez Bruno Latour dans *la science en action* et où la formule *acteur-réseau* désigne à la fois un réseau hétérogène d'intérêts alignés entre eux et le processus qui mène éventuellement à la production d'un artefact sociotechnique. Ce cadre théorique repose sur certaines notions-clés. L'une est justement la distinction entre le concept d'« acteur » central, dont dépendent d'autres éléments, dont, il traduit la volonté dans son propre langage, et celui d'« actant », désignant à la fois les humains et les non-humains d'un même réseau. Une autre notion clé est la « controverse », qui est une condition nécessaire à la constitution du réseau et à sa traduction par l'acteur (Mahil et Tremblay, 2015, pp. 234-237). Bref, un contexte où la production scientifique s'apparente à un bricolage doublé de l'art consommé de la négociation.

Dans ce contexte de révolution néolibérale où les programmes se substituent à la loi, prospérant sur les bases d'un nouvel idéal normatif et visant la réalisation efficace d'objectifs mesurables plutôt que l'obéissance à des lois justes, les construits sociaux portés par la révolution numérique et les modèles sont le nouvel imaginaire se voulant être l'outil de gouvernement et de construction de la preuve (Alain Supiot, 2015, pp.47-48).

Outils de « preuve et de gouvernement » selon la formule d'Alain Desrosières (2008) les expérimentations aléatoires sont porteuses à la fois de promesses de scientificité et de renouvellement de l'action sociale (Labrousse, 2010) ; leurs limites pratiques replacées dans la longue histoire de l'expérimentation sociale et du gouvernement par la preuve sont largement pointées du doigt aussi bien au niveau théorique, statistique, de validité externe et de généralisation des résultats que sur le plan épistémologique relatif à l'inférence causale en économie dont celle induite par la contrefactualité sur laquelle nous reviendrons plus loin.

Les controverses portées sur les limites de ces méthodes trouvent leur origine dans ce que Deaton (2009) et Rodrik (2008) ont souligné à propos d'absence de modèle théorique sous-jacent. Pour Deaton, les RCT accumulent des faits et non de la connaissance, le problème n'est tant pas ce qui marche mais comment ça marche. Rodrik ne dit pas autre chose lorsqu'il écrit : « On expérimente, mais qu'apprend-t-on ? ». Les limites de la validité interne due à la méthodologie de la randomisation qui permet d'éliminer les biais de sélection sont largement soulignées par la littérature qui y est consacrée (Favereau 2016, Labrousse 2010, l'Horty et Petit 2011, Zamora, 2011). Elles trouvent, selon les hypothèses que nous émettons, leur origine dans l'absence de ce cadre théorique. Il n'en demeure pas moins que la randomisation elle-même n'est pas dépourvue d'apories d'ordre statistique également largement soulignées (Heckman et Smith 1995, Manski, 1997 ; Fougère 2012). Ces apories statistiques génèrent au moins trois ordres de biais que la randomisation est censée éliminer (Fougère, 2012).

Faiblesse des effectifs. Pour des raisons de coûts ou de temps, les chercheurs expérimentateurs recourent à des effectifs réduits pour mener leur expérience et les résultats dans ce cas peuvent être imprécis car l'écart-type peut être trop large pour aboutir à une conclusion suffisamment fiable.

Effet Hawthorne. Bien connu des psychologues du travail, l'effet dit Hawthorne décrit la situation dans laquelle les résultats d'une expérience ne sont pas dus aux facteurs expérimentaux mais au fait que les sujets ont conscience de participer à une expérience dans laquelle ils sont testés, traduisant ainsi une plus grande motivation, et en tout cas le comportement des sujets étudiés est modifié par le fait d'avoir été ou non retenu pour participer à l'expérience et non par le dispositif ou le programme lui-même. Cet effet est à rapprocher de l'effet Pygmalion, que l'on observe chez des élèves dont les résultats s'améliorent du simple fait que l'enseignant s'intéresse davantage à eux.

Les effets d'attrition et de substitution. Ces biais sont dus souvent, dans le cas d'expériences qui se déroulent à une échelle de temps longue, au fait que des membres du groupe de traitement ou du groupe de contrôle décident de quitter l'expérience (effet d'attrition). C'est le cas par exemple de l'exemple de la valise des parents, expérience menée par le fonds d'expérimentation de la jeunesse menée dans le cadre des missions confiées à Martin Hirsch, Haut-commissaire à la Jeunesse. L'effet de substitution concerne généralement les membres des groupes de contrôles qui se soumettent à d'autres programmes comme ceux destinés au retour à l'emploi. S'y ajoutent les problèmes de perméabilité ou, selon Devaux-Spatarakis, les individus du groupe-témoin peuvent s'arranger pour avoir accès au traitement et ou à l'inverse les individus du groupe-test y échappent (Devaux-Spatarakis, 2014).

Par ailleurs, les influences politiques et les interactions chercheurs expérimentateurs et partenaires peuvent poser problème par le recours aux « bricolages » qui affectent le crédit attribué aux RCT. C'est le cas de l'expérimentation au Cambodge relative à l'évaluation d'un programme de micro-assurance (Quentin et Guerin, 2013) ou de l'expérience tant vantée et citée sur les vermifuges, où les partenaires locaux n'ont pas autorisé la randomisation dans l'affectation des écoles et lui ont préféré une assignation par ordre alphabétique (Labrousse, 2016).

L'expérimentation a servi également comme transposition d'une méthode scientifique hors de son cadre d'origine ou la notion d'expérimentation introduite dans la révision de la Constitution française en 2003 fut utilisée alors même que le dispositif d'évaluation était en cours : une loi généralisant le Revenu de Solidarité Active (RSA) a en effet été promulguée sans attendre les résultats de cette expérimentation (Massieu, 2011). Ce fut également le cas pour le programme Progreso de l'aide ciblée au Mexique (Faulkner, 2012).

L'inférence causale en économie : contre factualité et essai contrôlé randomisé

L'interdiction formulée par Hume de déduire des jugements sur ce qui doit être à partir de propositions factuelles sur ce qui est, a souvent servi en philosophie des sciences à

caractériser le travail scientifique, tout particulièrement en ce qui concerne les approches normatives.

Le Connaitre et le Comment sont des questions régulièrement réactualisées dans l'ensemble des sous-champs de l'économie (économétrie, économie expérimentale, théorie de la décision, modélisation macroéconomique, etc.). (Courgeau 2000). Le problème posé par la nouvelle économie du développement s'appuyant sur les expérimentations aléatoires relève du manque de données à un instant t où un individu ne peut être à la fois bénéficiaire et non bénéficiaire d'une politique, rendant inobservable l'impact causal d'une politique. Ce problème est connu sous le nom de « situations inobservables issues de l'analyse contrefactuelle ».

L'approche contrefactuelle en matière de causalité en sciences sociales théorisée par David Lewis (1973) utilise une théorie des mondes possibles qui considère qu'ils sont aussi concrets et réels que celui dans lequel nous vivons, et c'est ce qui permet de parler d'évènements contrefactuels : bien qu'ils ne se soient pas produits, ils sont aussi concrets que ceux que nous observons.

Ainsi, pour signaler la disproportion des causes des effets, Pascal avait écrit dans les *Pensées* : « Le nez de Cléopâtre s'il eut été plus court, toute la face de la terre aurait changé ». On est dans le cas de figure où, par exemple, si A est considéré comme l'antécédent d'un énoncé et C son conséquent, l'énoncé contrefactuel prend la forme suivante :

« $A > C$ » ou « si A avait été le cas, alors C aurait été le cas » ; $>$ étant le symbole du conditionnel contrefactuel. Tout le raisonnement va reposer sur la condition de vérité lewisienne suivante : On considèrera que si un évènement C cause un autre évènement E et l'on ira se demander s'il existe un monde possible dans lequel E est le cas, bien que C n'y soit pas le cas et si ce monde est plus similaire au monde actuel que ne l'est le monde possible, le plus proche du monde actuel dans lequel ni C ni E ne sont le cas.

Application de la contre factualité aux expériences aléatoires contrôlées

Cette approche part d'une hypothèse qui dès le départ est non testable. En effet, si l'on applique à un individu u un traitement à évaluer, sa réponse au traitement étant $Y_t(u)$, ou si on lui applique un traitement classique c appelé contrôle, $Y_c(u)$ étant sa réponse au contrôle, on mesure l'effet du traitement de contrôle par la différence : $Y_t(u) - Y_c(u)$. Comme indiqué plus haut, l'on ne peut simultanément appliquer au même individu les deux traitements, et si on les applique à la suite l'un de l'autre, lorsque c'est possible, on ne peut généralement plus parler de causalité. Cependant, si l'on constitue deux groupes par tirage au hasard, tels que deux individus pris dans chacun des groupes, u_1 et u_2 , soient strictement comparables, à l'exception du traitement donné, l'effet causal du traitement serait : $Y_t(u_1) - Y_c(u_2)$.

Pour passer de l'individu au groupe, le modelé de Ruben est convoqué : Le modelé causal contrefactuel prend ses racines dans les travaux de Jerzy Neyman et Fisher cités par Holland (1986, pp. 945-960) dans le cadre d'expériences aléatoires ; il a ensuite été étendu par Rubin (Cité par Holland 1986) en proposant un cadre plus général applicable aux sciences sociales et que Holland appelle « modelé causal de Rubin ». Ce modèle causal permet de mesurer ce que l'on nomme « les résultats potentiels fonction de l'exposition de l'agent au traitement ».

Ces deux piliers sont, d'un côté, le paramètre d'intérêt, qui représente le phénomène économique que l'on cherche à analyser, de l'autre côté, l'analyse causale qui fait suite au choc que l'on impose au paramètre d'intérêt au travers de la modification

des facteurs d'influence. Si le sujet a subi le traitement, on a $T_i = 1$, et s'il n'a pas reçu le traitement, on note $T_i = 0$, si bien que pour chaque i appartenant au groupe de traitement ou au groupe de contrôle, on observe l'un des Y_{0i} et Y_{1i} mais pas les deux. Comme on ne peut connaître la répartition des effets du traitement, $Y_{1i} - Y_{0i}$, grâce à la randomisation on pourra observer la différence entre le résultat moyen dans les traitements et le résultat moyen dans les contrôles, que l'on note :

$$E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 0) \quad (1)$$

Cette différence peut être divisée en deux termes :

$$E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 0) = [E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 1)] - [E(Y_{0i} | T_i = 1) + E(Y_{0i} | T_i = 0)] \quad (2)$$

Soit :

$$E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 0) = E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 1) \quad (3)$$

De sorte que la différence entre le résultat moyen traité et le résultat moyen non traité représente la différence entre les deux résultats observables dans le groupe de traitement et dans le groupe de contrôle, le dernier terme à droite étant non observable.

Nous considérons en conséquence la moyenne de la différence plutôt que la différence de moyennes qui se trouve actuellement à droite de (3) qui est une estimation de l'effet moyen du traitement chez les personnes traitées, laquelle, étant donné que le traitement et les témoins ne diffèrent que par randomisation, est une estimation de l'effet moyen du traitement pour tous :

$$E(Y_{1i} | T_i = 1) - E(Y_{0i} | T_i = 0) = E(Y_{1i} - Y_{0i} | T_i = 1) \quad (4)$$

Ainsi, les problèmes posés par la généralisation des résultats des expérimentations aléatoires contrôlées proviennent-ils du fait que leur but n'est pas de comprendre la structure sous-jacente du système de relations générant ces résultats, mais uniquement de mesurer l'impact statistique pour prouver l'effet causal attribuable à l'intervention. En l'absence de théorie qui peut imposer des corrections, le pouvoir causal qui peut justifier l'inférence s'en trouve affaibli car si les RCT fournissent des informations sur la moyenne des effets du traitement, ils n'identifient pas d'autres caractéristiques de la distribution.

Dans l'un des textes les plus récents consacrés à la causalité, que nous empruntons à Bernard Paulré, (2007) qui se prononce surtout sur la fonction de la notion de cause, B. Saint-Sernin (2002), cité par Paulré, observe : « Plus on réfléchit à la notion de cause, plus elle paraît obscure. Pourtant, sa fonction est claire : elle sert à reconstituer des enchaînements, à inscrire des événements à première vue disjoints dans une trame temporelle ». Selon cette idée, c'est la succession temporelle des événements qui constitue, selon B. Saint-Sernin, la manifestation typique d'une explication causale : « De Platon et Aristote jusqu'à aujourd'hui, le concept de causalité occupe une place centrale dans la recherche de la vérité, car il a pour fonction de rendre intelligibles les enchaînements temporels. » Toujours selon Paulré, citant J. Piaget, c'est l'idée de production, utilisée dans un sens très général c'est-à-dire qui ne se limite pas nécessairement à l'idée physique de cause efficiente, qui est privilégiée mais en

l'associant a l'idée de nécessité : « La causalité apparaît a tous ses niveaux et sous toutes ses formes comme impliquant à la fois la production d'une nouveauté, parce que l'effet est nouveau par rapport à la cause et qu'il y a donc transformation et, d'autre part, une liaison nécessaire sans quoi on ne peut parler de causalité ».

Ainsi deux positions semblent se dégager : soit on se place au niveau des objets physiques, la causalité serait alors une caractéristique ontologique du lien qui pourrait y avoir entre deux objets (quelle est la nature d'une cause ?), soit on se positionne au niveau épistémologique des représentations du monde physique, on chercherait, donc, à expliquer et non à révéler des principes intrinsèques (comment est-il possible d'inférer l'existence d'une cause à partir de l'observation ?).

Les deux caractéristiques de l'économie expérimentale résident dans deux aspects primordiaux : le « contrôle » qui renvoie à la capacité de l'expérimentateur à maîtriser ce qui se passe au cours de l'expérience pour produire des données de bonne qualité, et la « réplication » qui permet de reproduire une expérience analogue afin de tester la « robustesse » des résultats. Ces deux caractéristiques garantissent en général la production de connaissances expérimentales. Mais c'est précisément ce qui renvoie aux problèmes de validité interne et externe soulevés de façon récurrente par les critiques des RCT.

Jatteau (2013) évoque la pression à la publication qui conditionne la carrière des chercheurs pour qu'un résultat positif ait davantage de chances d'être publié ainsi que le problème de conflit d'intérêts dans la mesure ou la distinction évaluateur/expérimentateur dans les expérimentations aléatoires n'est pas prise en compte sur le plan scientifique, pour espérer mettre au jour de façon objective l'effet d'un programme. Toutes ces manipulations ajoutées aux biais statistiques, dont nous en avons rapportés quelques exemples ci-dessus, jettent le doute pour le moins sur la rigueur tant louée de ces expériences et parfois discréditent même le bien-fondé de la validité interne reconnue aux RCT.

Quant aux incertitudes entourant la validité externe tant décriée, nous pensons qu'elles sont organiques à la fois à la faible validité interne qu'à l'absence de cadre théorique les encadrant. La question de la validité interne qui précède forcément celle de la validité externe devrait garantir la « cohérence interne », en ce sens que les résultats de l'expérience sont bien imputables au « traitement » qui est opéré. La perte de contrôle d'une expérience grève la validité interne des résultats. Et ce, même si ces résultats sont largement reproductibles : replicabilité et validité interne sont des propriétés distinctes – sachant qu'un manque de replicabilité laisse toutefois présager une faible validité interne. Dans une vue d'ensemble fort éclairante sur les méthodologies en économie expérimentale, Daniel Serra décrit remarquablement cette tension validité externe/validité interne en signalant : « On pressent immédiatement l'existence d'une sorte de tension entre l'exigence de validité interne et celle de validité externe, dans la mesure ou les designs expérimentaux les mieux aptes à conférer la première – *via* un niveau élevé de contrôle – sont susceptibles de jouer en défaveur de la seconde, en accentuant le caractère *aseptisé* du cadre expérimental » (Serra, 2012, p. 36).

Après avoir considéré que le cadre théorique est bien un mauvais guide, Banerjee et Duflo (2009, p.693) disent plus loin dans le même article avoir entendu les critiques : « les expérimentations sont en train d'apparaître comme un outil puissant pour tester des théories ».

Cependant, lorsque l'on examine la question relative au test de la théorie par un test empirique, on ne peut jamais savoir si le résultat est imputable à l'une des hypothèses auxiliaires implicites ou non identifiées ou à la théorie testée tant les données rendant compte des effets joints de l'hypothèse qui est testée sont difficilement isolées de

l'ensemble d'hypothèses qui y sont associées selon le projet de critique globale inspirée du critère Duhem-Quine.²

Critique de l'analyse contrefactuelle

Par ailleurs, les réflexions de Smith (1990) sur cette approche permettent de douter de la robustesse de cette inférence causale utilisée par les promoteurs des expériences aléatoires contrôlées : « Comment le chercheur peut-il savoir que les variables convenables ont été incorporées dans l'étude expérimentale ? De même, pour la spécification des modèles en vue de l'analyse des données obtenues d'études expérimentales, la théorie est le guide ultime. » Pour lui, si l'on mesure au niveau micro, l'intervention s'effectue toujours au niveau macro, et les analyses au niveau individuel n'ont guère d'intérêt pour l'action. Dawid (2000, pp.407-448) va plus loin dans la critique de l'approche contrefactuelle en précisant qu'il y a deux types de problème d'inférence causale, déjà indiqués par Holland (1986) : d'une part, les *effets des causes* consistant à comparer les conséquences attendues de différentes interventions possibles dans un système et, d'autre part, les *causes des effets* ou l'on cherche à comprendre la relation causale entre un résultat déjà observé et une intervention antérieure. Ce qu'il illustre par l'interrogation suivante : « J'ai une migraine. Serait-il utile que je prenne de l'aspirine ? » Et d'ajouter : « Ma migraine est passée. Est-ce parce que j'ai pris de l'aspirine ? ». Tout le problème consiste alors à comprendre la relation causale entre un phénomène déjà arrivé et une intervention antérieure. En recherchant l'effet des causes par la méthode contrefactuelle, Dawid (2000) considère donc que l'analyse est alors basée sur une attitude qu'il appelle *fatalisme* pour qui les différentes réponses potentielles $Y_i(u)$, lorsque le traitement i est appliqué à l'unité u , comme des attributs prédéterminés de l'unité u , attendent seulement d'être retrouvés par l'expérience. Il note par ailleurs : « Parce que l'unité u est considérée comme individuelle et non répétable, il n'y a jamais aucune possibilité de tester empiriquement cette hypothèse fataliste, qui peut ainsi être considérée comme métaphysique ».

L'inférence causale issue des approches contrefactuelles induisant une induction par élimination n'est, et plus encore que d'autres formes d'inférence inductive, guère qu'une tentative, et la considération d'entités non observables en affaiblit les fondements empiriques (Dawid, 2000, 2007).

Nancy Cartwright parle d'un angle à la fois ontologique et pragmatique, l'exigence ontologique opérant comme garantie de l'efficacité pragmatique (Guillin 2013, p.82) : « Pour répondre à cette exigence pratique et à cette visée pragmatique, il est indispensable que les explications mobilisées dans les expériences aléatoires contrôlées (RCT) puissent servir de base à une intervention efficace dans le réel, c'est-à-dire qu'ils fournissent des preuves suffisantes de l'applicabilité du programme teste (la livraison de l'engrais juste après la récolte, l'exemple des fermiers kenyans) en dehors du contexte de

² Cités par Daniel Serra in : « Principes méthodologiques et pratiques de l'économie expérimentale : une vue d'ensemble », Vrin, *Revue de philosophie économique*, 2012/1, vol. 13, p. 21 à 78. See also Quine, W. V. O. (1951: 20-43) et Duhem, P. (1914, 2016).

mise à l'épreuve expérimentale. Tout le problème est alors de savoir comment s'assurer de cette applicabilité ». Pour N. Cartwright, la solution passe par la clarification du statut ontologique des *capacités*, des entités ou des agents considérés dans les RCT. Dans les expériences aléatoires telles que pratiquées par le J-PAL, ces capacités ne réunissent pas les conditions de robustesse dans la mesure où elles seraient plutôt des expérimentations qui servent à les mettre en évidence. Elles n'ont pas non plus la capacité d'être irréductibles dans la mesure où il n'est pas encore possible d'effectuer l'inventaire de ces dispositions pour qu'une théorie unifiée établisse une hiérarchie des dispositions et des mécanismes à l'œuvre. Enfin, elles sont dépourvues de lois de compositionnalité qui décriraient les mécanismes équivalents, sans parler de compositionnalité qui théorise les rapports entre mécanismes psychologiques et mécanismes non psychologiques constitutifs des phénomènes sociaux. Pour elle, même quand les RCT ont une validité interne élevée, la méthodologie formelle pose de sérieuses contraintes sur les hypothèses qu'une population cible doit respecter pour justifier l'exportation d'une conclusion de la population testée vers la cible.

Cartwright, citée par Guillin (op.cit.) est en fait bien plus sévère parce qu'elle conclut son jugement par : « En l'absence d'une théorie des capacités des causes que nous étudions, la valeur probatoire d'une RCT n'est pas seulement affaiblie, mais bien...réduite à rien ».

Enfin, James Heckman décrit ses méthodes d'empirisme aveugle qui, non guidées par un cadre théorique d'interprétation des faits, ne conduisent nulle part. Pour lui, le cadre Neyman-Rubin se concentre exclusivement sur les *effets causaux moyens* et les effets du traitement au niveau de la population pour les politiques réellement appliquées mais ne fournit aucun cadre pour l'extrapolation de la découverte de nouveaux environnements ou pour les prédictions pour de nouvelles politiques. Il ajoute que puisque la randomisation est utilisée pour définir les paramètres d'intérêt, elle conduit parfois à la confusion que celle-ci est la seule façon ou au moins la meilleure façon d'identifier les paramètres causaux à partir de données réelles. En vérité, ce n'est pas toujours le cas, assène-t-il.

Il décrit cet usage de l'inférence causale contrefactuelle comme une approche basée sur le voile de l'ignorance et largement utilisée dans le travail appliqué pour évaluer différentes distributions de revenus, par exemple. Les résultats peuvent également dépendre de ce que les personnes qui bénéficient d'une politique peuvent également bénéficier d'une autre. En conclusion, les hypothèses dépourvues de cadre théorique impliquant l'existence d'entités imaginaires dont on postule l'existence ne sont pas fondées épistémologiquement (Thom, 1986), et dans ces cas les sauts inductifs dans une situation réelle semblent un mauvais pari (Nancy Cartwright, 2009).

En ne s'attaquant pas aux causes structurelles générant les inégalités, ces méthodes figurent comme des marqueurs agissant dans la lutte contre la pauvreté des politiques néolibérales dont le consensus de Washington constitue l'avatar le plus emblématique mais aussi l'échec le plus patent. Leur visée se positionnant contre l'approche macroéconomique a pour effet la primauté accordée aux comportements et à la rationalité limitée au détriment des questions globales comme les inégalités débouchant sur la dissolution de la question du développement. Prétendre ainsi faire l'inventaire de tous les phénomènes observables pour contribuer à édifier les bases d'une nouvelle économie du développement en faisant bon marché de l'économie politique ne fait que déboucher sur des propositions qui, puisqu'elles s'intéressent uniquement aux capacités individuelles et non aux capacités collectives, ne s'attaquent ni aux processus d'exclusion ni aux causes de la marginalisation et maintiennent en l'état les causes structurelles de la pauvreté. La prétention des RCT à participer à un tournant de l'économie du

développement ne s'avère pas fondée au regard des limites que nous avons montrées ci-dessus. Les pauvres sont culpabilisés, et leur sort est davantage du a des « mentalités » inadéquates héritées induisant leurs décisions erronées. L'aversion au global et général explique largement les bricolages tendant à régénérer une économie du développement (Servet, 2018).

Conclusion

Le recours systématisé à l'évaluation des politiques publiques englobant aussi bien les dimensions normatives que positives encouragé, particulièrement, par les organisations internationales (FMI et Banque Mondiale), les donateurs et autres bailleurs de fonds a favorisé le développement d'un marché de production de savoirs autant scientifiques que médiatiques articulant systèmes de connaissances et pratiques de gouvernement en vue de cibler des pratiques de développement tout en dépolitisant le développement par des experts cherchant à remplacer « le gouvernement des hommes » par « la gouvernance des choses » selon l'expression de Dupuy (2012, p.5).

Cet article s'est attaché à montrer les limites, les apories et les failles épistémiques passées souvent sous silence et a contribué à déconstruire la prétention à la force de la validité interne conférée par l'élimination des biais de sélection montrant subséquemment que les problèmes de généralisation que connaissent les RCT sont, précisément, dus à la faiblesse de la validité interne d'une part et à l'absence de théorie, d'autre part.

References

- ANGRIST J. & PISCHKE J. (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con Out of Econometrics", *Journal of Economic Perspectives*, pages 3 à 30.
- BANERJEE A. & DUFLO E. (2009), « L'approche expérimentale en économie du développement », *Revue d'économie politique*, 2009/56, vol. 119, pages 691 à 726.
- BANERJEE A. & DUFLO E. (2012), *Repenser la pauvreté*, le Seuil. 432 pages.
- BANERJEE, A., BARDHAN P., BASU K. & MOOKHERJEE D. (2005), *New Directions in Development Economics: Theory or Empirics? A Symposium in Economic and Political Weekly*, in *Economic and Political Weekly*, 2005, pages 4340-4344.
- BARDET F. & CUSSO R. (2012), « Les essais randomisés contrôlés, révolution des politiques du développement ? Une évaluation par la Banque mondiale de l'empowerment au Bangladesh », *Revue française de socio-économie*, 2012/2, n° 10, pages 175 à 198.
- BEDECARRATS, F., GUERIN, I. & ROUBAUD, F. (2013), « L'étalon-or des évaluations randomisées : du discours de la méthode à l'économie politique », *Sociologies pratiques*, N° 27, pp. 107 à 122.
- BURNSIDE, C. & DOLLAR, D. (1997), "Aid, Policies and Growth", The World Bank Research Department, pp. 1 à 64.
- CARTWRIGHT, N. (2009), "If No Capacities, then No Credible Worlds, But Can Models Reveal Capacities", *Erkenntnis*, 70, pp. 45-58.
- COURGEAU, D. (2000), Réflexions sur la causalité en sciences sociales, in *Recherches et Prévisions*, N°60, 2000, Chômage et famille, pp. 49-60.

- COURGEAU, D. (2012), « Peut-on parler de causalité en sciences sociales ? Les approches contrefactuelle et mécaniste », in *Les rencontres de statistique appliquée : la causalité en sciences humaines et sociales : « expliquer, modéliser, mesurer »*, Colloque conjoint entre l'INED et la Société française de statistique.
- DAWID, P. (2000), Causal inference without counterfactual, *Journal of American Statistical Association*, 95, pp. 407-448.
- DAWID, P. (2007), *Fundamentals of Statistical Causality*, RSS/EPSRC Graduate Training Program University of Sheffield.
- DEATON, A. (2010), “Instruments, Randomization, and Learning about Development”, *Journal of Economic Literature*, 48, pages 424-455.
- DEATON, A. & CARTWRIGHT, N. [2018], “Understanding and Misunderstanding Randomized Controlled Trials”, *Social Science & Medicine*, Elsevier. Volume 210, August 2018, pp. 2-21.
- DESROSIERES, A. (2008), *Pour une sociologie historique de la quantification. L'argument statistique*, tome I, Presses des Mines, 329 pages.
- DUFLO, E., GLENNERSTER, R. & KREMER, M. (2007), “using Randomization in Development Economics Research: A toolkit”, in *Handbook of Development Economics*, Elsevier, p.3895-3962.
- DUFLO, E. (2009), *Expérience, science et lutte contre la pauvreté*, Fayard, Paris, 60 pages.
- DUHEM, P. (1914), *La théorie physique, son objet, sa structure*, B/2S ENS éditions (2016).
- DEVAUX SPATARAKIS, A. (2014), « La méthode expérimentale par assignation aléatoire : un instrument de reconstitution de l'interaction entre sciences sociales et action publique en France ? ». Thèse de Doctorat. Université de Bordeaux.
- DUPUY J.P (2012), *l'avenir de l'économie*, Flammarion, Coll. « Essais », 291 pages.
- ECONOMIC AND POLITICAL WEEKLY, 2005, pages 4340-4344.
- EASTERLY W. (2009), *le Fardeau de l'homme blanc : l'échec des politiques occidentales d'aide aux pays pauvres*, ed. Markus Haller.
- EASTERLY W. (2006), *les Pays pauvres sont-ils condamnés à le rester ?* Editions d'organisation, Groupe Eyrolles.
- FAULKNER, W. (2012), “Norm, Mistake, or Exemplar? A complexity Approach to IFPRI-PROGRESA in Mexico”, (http://wnfaulkner.files.wordpress.com/2012/08/faulknerw_negotiatedperspective_s.pdf).
- FAVEREAU J. (2016), in *Dictionnaire des conventions : autour des travaux d'Olivier Favereau*, édité par P. Batifoulier, F. Bessis, A. Ghirardello, G. de Larquier, D. Remillon, Presses universitaires du Septentrion.
- FOUGERE, D. (2010) “ Les méthodes économiques d'évaluation ”, *Revue française des affaires sociales*, 2010/1, pp. 105-128.
- FOUGERE, D. (2012) “ Les méthodes d'expérimentation en question ”, *Education et formation*, n° 81, pp. 41-47.
- FRANK R. (2007), « Peut-on accroître le pouvoir explicatif des modèles en économie? », in A. Leroux, P. Livet, *Leçons de philosophie économique*, t. III, Paris, Economica.
- FAVEREAU J. (2016), in *Dictionnaire des conventions : autour des travaux d'Olivier Favereau*, édité par P. Batifoulier, F. Bessis, A. Ghirardello, G. de Larquier, D. Remillon, Presses universitaires du Septentrion.
- GUILLIN V. (2013), « De quoi les essais contrôlés randomisés sont-ils capables ? Evaluation, mécanismes et capacités en sciences sociales », *Cahiers philosophiques*, 2013/2, n° 133, pages 79 à 102.
- HECKMAN, J. & SMITH J. (1995), “Assessing the Case for Social Experiments”, *The Journal of Economic Perspectives*, vol. 9, n° 2, American Economic Association, pp. 85-110

- HECKMAN J. (2005), "The Scientific Model of Causality", *Sociological Methodology*, 35(1), pages 1 à 97. <https://doi.org/10.1111/j.0081-1750.2006.00164.x>.
- HOLLAND, P.W.(1986), «statistics and causal inference», *Journal of The American Statistical Association*, Vol 81, N 396, pp. 945-960.
- JATTEAU, A. (2013), « Expérimenter le développement ? Des économistes et leurs terrains », *Belin/Genèses*, 2013/4, n° 93, pp. 8 à 28.
- LABROUSSE, A. (2010), « 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.
- LABROUSSE, A. (2016), « Apprendre des expérimentations aléatoires : promesse de scientificité, complications pratiques, expériences historiques », *la vie des idées* (4 mars 2016).
- LEWIS, D. (1973), *Counterfactuals*, Blackwell Publishers, Harvard University Press, 156 pages.
- L'HORTY, Y. & PETIT, P. (2011), « Evaluation aléatoire et expérimentations sociales », *Revue française d'économie*, 2011/1, Volume XXVI, pp. 13 à 48.
- MAHIL, A. & TREMBLAY, D.G. (2015), « Théorie de l'acteur-réseau » in *Sciences, technologies et sociétés de A à Z*, Presses de l'Université de Montréal, pp. 234-237.
- MANSKI, C. (1997), "Monotone Treatment Response", *Econometrica*, Vol 65, N° 6, pp. 1311-1334.
- MASSIEU, V. (2011), « L'expérimentation comme exemple de transposition d'une méthode scientifique hors de son cadre d'origine », intervention – 8^e congrès de droit constitutionnel, 2011, Nancy, (<http://www.droitconstitutionnel.org/congresNancy/com N3/MassieurT> ».pdf).
- PAULRE, P. (2007), « La causalité en économie : mises en perspective et approches opérationnelles », in *Leçons de philosophie économique. Sciences économiques et philosophie des sciences-Tome 3*, *Economica*, pp. 211-282.
- QUINE, W. V. O. (1951), «Two Dogmas of Empiricism», *The Philosophical Review* 60 (1951), pp. 20-43.
- QUENTIN, A. & GUERIN, I. (2013), « La randomisation à l'épreuve du terrain, l'exemple du projet de micro-assurance au Cambodge », *Revue Tiers Monde*, N° 213, pp. 179-200.
- ROUX, S. (2015), « Approches structurelles et non structurelles en micro-économétrie de l'évaluation des politiques publiques », *Revue française d'économie*, 2015/1 Vol. (XXX), pp. 13 à 65.
- RODRIG D. (2008), *The New Development Economics: We Shall Experiment, but How Shall we Learn ?* John F. Kennedy School of Government, Harvard University, RWP08-055, October.
- SACHS J. (2005), *The End of Poverty*, New York, The Penguin Press.
- SAINT-SERNIN, B., (2002), *La causalité*, in Andler, D., Fagot-Largeault, Saint-Sernin, B., *Philosophie des Sciences*, Tome II, Gallimard.
- SERRA, D. (2012), « Principes méthodologiques et pratiques de l'économie expérimentale : une vue d'ensemble », *Revue de Philosophie Economique*, 13, pp. 21-78.
- SERVET, J. M. [2018], *L'économie comportementale en question*, Paris, Editions Charles-Léopold-Mayer, 208 pages.
- SUPIOT, A. (2015), *La gouvernance par les nombres*, Fayard, 512 pages.
- THOM, R. (1986), « La méthode expérimentale : un mythe des épistémologues (et des savants ?) », in *La philosophie des sciences aujourd'hui*, sous la direction de Jean Hamburger. Paris. Gauthier-Villars.

- WALLISER, B. & PROU, C. (1988), *La science économique*, Editions du Seuil, 395 pages
- ZAMORA, P.(2011), « La méthode d'évaluation aléatoire : apports et limites », Tracés.
Revue de Sciences humaines, Hors-série 2011, À quoi servent les sciences humaines
(III), pp. 175-186.