

## **Apprendre des expérimentations aléatoires**

### **Promesse de scientificité, complications pratiques, expériences historiques**

Agnès LABROUSSE

**Les expérimentations aléatoires sont porteuses d'une forte promesse de scientificité et de renouvellement de l'action sociale. Agnès Labrousse signale ici quelques limites pratiques des expérimentations et les replace dans l'histoire plus longue de l'expérimentation sociale et du gouvernement par la preuve.**

Depuis le début des années 2000, les expérimentations par assignation aléatoire constituent assurément une méthode en vogue, portée notamment par les travaux du J-PAL (Jameel-Poverty Action Lab), et de ses fondateurs Esther Duflo et Abhijit Banerjee. Distingués à de multiples reprises, ces travaux ont rencontré un très large écho dans les revues d'économie les plus prestigieuses, dans les colonnes de la presse généraliste et sur le présent site (Mayneris, 2010 ; L'Horty et Petit, 2011 ; Bérard et Valdenaire, 2013). Ces expérimentations sont promues aujourd'hui par des organisations internationales comme la Banque mondiale, ou par des fondations comme la *Bill and Melinda Gates foundation* Elles ont connu un essor important pour la première fois en France avec le Fonds d'Expérimentation pour la Jeunesse (FEJ), initié par Martin Hirsch (Bureau *et al.*, 2013). Prenant pour point de départ les travaux de l'équipe d'Esther Duflo – qui propose une pratique parmi d'autres de ces expérimentations –, nous élargirons ensuite la perspective en évoquant l'histoire plus longue des expérimentations sociales.

Souvent présentée comme révolutionnaire, cette méthodologie est porteuse d'une forte promesse de scientificité, de transparence et d'efficacité de l'action sociale (1). Néanmoins, des travaux récents pointent d'importantes limites qui se manifestent lors de la mise en œuvre et de l'exploitation des expérimentations (2). Phénomène frappant, ces mêmes limites avaient déjà été mises en lumière dès les années 1970 aux États-Unis, lors de la précédente vague d'expérimentations aléatoires. Car nous avons affaire à un projet récurrent de gouvernement par la preuve qui donne lieu depuis les premières expérimentations sociales randomisées contrôlées dans les années 1920 à des cycles enthousiasme-déception parmi les acteurs scientifiques et politiques (3).

## **1. Une promesse de scientificité, de lisibilité et de renouvellement de l'action publique**

Les expérimentations par assignation aléatoire portent une triple promesse de scientificité, de transparence et de renouvellement de l'action sociale en général et de l'action publique en particulier.

### **Une promesse de scientificité**

Le transfert dans le champ de la lutte contre la pauvreté et de l'évaluation des politiques publiques d'une méthodologie éprouvée dans le champ médical, celle des essais cliniques randomisés contrôlés, apparaît comme une source de rigueur : « les politiques de lutte contre la pauvreté sont évaluées avec la rigueur des essais cliniques » (Duflo, 2009). Le site européen du J-PAL, plus gros laboratoire mondial de lutte contre la pauvreté, emploie le même registre : « J-PAL a fondé sa réputation sur l'utilisation exclusive d'expériences contrôlées permettant de produire des résultats d'une rigueur et d'une qualité scientifiques exceptionnelles ».

### **Une promesse de simplicité et de lisibilité**

Prenons l'exemple d'une expérimentation conduite au Maroc de 2006 à 2009, en partenariat avec l'organisme de microcrédit Al Amana. Elle visait à mesurer l'impact du microcrédit sur les revenus et la consommation des ménages en zone rurale. Elle portait sur des paires de villages aux caractéristiques similaires : un village de chaque paire a été aléatoirement choisi pour recevoir des services de microcrédit juste après l'ouverture de l'agence Al Amana, tandis que l'autre village servait de témoin (sans ces services) pendant deux ans. Au total, 81 paires de villages ont été sélectionnées sur l'ensemble du territoire marocain et 6 000 ménages ont été inclus dans l'enquête (avant l'établissement de l'agence, 1 an et 2 ans après). « En procédant à une affectation aléatoire, il est possible de constituer un groupe de bénéficiaires et un groupe témoin initialement parfaitement comparables, non seulement sur des critères observables statistiquement, mais aussi sur des dimensions [...] inobservables. Ainsi toute différence future constatée entre les deux groupes s'interprète sans ambiguïté comme l'effet de la mesure testée » (Bérard et Valdenaire, 2013)<sup>1</sup>. Cette simplicité (écart des moyennes des deux groupes) confère une lisibilité rare aux résultats. Ainsi, souligne Esther Duflo, « grâce à la randomisation, il est tout à fait transparent et aisé d'apprécier l'impact ».

### **Une promesse de renouvellement par la preuve des politiques sociales**

Cette rigueur et cette lisibilité des expérimentations aléatoires permettraient de « révolutionner les politiques sociales du XXI<sup>e</sup> s. comme les essais cliniques ont révolutionné la médecine du XX<sup>e</sup> s. » (Duflo *et al.*, 2004). Nous pourrions ainsi entrer dans un nouvel âge de l'expertise et des politiques publiques, un âge de l'objectivité scientifique, de l'innovation et de l'efficacité, loin des « trois I – idéologie, ignorance, inertie – » auxquels les expérimentations apporteraient enfin un remède (Banerjee & Duflo, 2011). Pour Duflo, « Il faut sortir des grandes alternatives stériles. Les évaluations sont rigoureuses. Elles ne laissent aucune place à l'interprétation. Si ça ne

---

<sup>1</sup> Pour une présentation de la méthode, on se reportera, outre les articles cités de la *Vie des idées*, à Jatteau (2013a).

marche pas, ça ne marche pas. Il reste seulement alors à tenter autre chose.»<sup>2</sup> Les expérimentations aléatoires permettraient de trancher par une analyse scientifique rationnelle, dépassionnée, des débats politiques à forte composante idéologique. C'est par la mesure objective, chiffrée, sans appel, des effets des politiques publiques, par la « pédagogie de la preuve » (Jatteau, 2013b), que les blocages de l'action publique pourraient enfin être déverrouillés.

## **2. Complications pratiques : aléas du protocole, résultats délicats à interpréter**

Néanmoins, la conduite et l'interprétation des expérimentations s'avèrent souvent délicate, au point que certains avantages théoriques de la méthode s'effacent en pratique<sup>3</sup>.

### **Une ingénierie contraignante en butte aux aléas et à la diversité du social**

La mise en œuvre des expérimentations aléatoires demande une ingénierie complexe, impliquant des acteurs diversifiés. De ce fait, elle est soumise à des aléas sociaux qui peuvent compromettre la validité interne du protocole.

Si les expérimentations sociales correspondent à un dispositif *in vivo* qui a l'avantage du réalisme par rapport aux expérimentations de laboratoire, le protocole produit des contraintes artificielles et des réactions sociales variées (résistances, désintérêt des publics concernés, détournements du dispositif, effets placebo ou nocebo, etc.).

Par exemple, la teneur du projet évalué peut susciter le rejet. Ce fut le cas du dispositif de la « cagnotte scolaire » dans l'académie de Créteil. Visant à modifier le système d'incitations des élèves en introduisant des récompenses financières pour réduire l'absentéisme scolaire, il se heurta à des résistances sociales conduisant à son abandon après la phase pilote (Bureau *et al.*, 2013). On assiste également à des réactions de désintérêt ou de défiance de la part de publics fragiles.

Les résistances peuvent également porter sur le tirage aléatoire lui-même : ainsi deux articles systématiquement mis en avant pour illustrer les vertus de la randomisation – notamment celui de Kremer et Miguel (2004) sur l'impact positif des vermifuges sur la scolarisation des enfants au Kenya – relèvent de fait de quasi-expérimentations (Deaton, 2010) : les partenaires locaux n'ont pas autorisé la randomisation dans l'affectation des écoles et lui ont préféré une assignation par ordre alphabétique.

Ces aléas sociaux se traduisent par des difficultés statistiques. Certaines expérimentations doivent faire face à un faible recours à la prestation ou à une déperdition du nombre de sujets en cours d'expérimentation (attrition), rendant délicate la constitution et le maintien d'échantillons de taille suffisante. Apparaissent aussi des problèmes de perméabilité des groupes : les individus du groupe témoin peuvent se

---

<sup>2</sup> <http://www.lejdd.fr/Economie/Actualite/Intellectuelle-de-terrain-166936>.

<sup>3</sup> D'autres limites importantes existent, notamment celles liées à la question cruciale de la validité externe, au-delà du site de l'expérimentation (Deaton, 2010 ; Jatteau, 2013a).

débrouiller pour avoir accès au traitement alors qu'il arrive à l'inverse que les individus du groupe test n'en bénéficient pas (Devaux-Spatarakis, 2014).

Les expérimentateurs doivent déployer des trésors d'énergie et d'ingéniosité pour tenter de contrôler le protocole de bout en bout. Car le dispositif expérimental pour maintenir sa validité interne implique une forte rigidité et une standardisation très poussée des « traitements ». Or rigidité et standardisation peuvent entrer en conflit avec la flexibilité et la diversité des relations sociales en jeu dans un dispositif (relation de conseil pour retrouver un emploi, parrainage d'individus en difficulté, relation de crédit etc.). C'est ce que montre une étude qualitative conduite en parallèle à l'expérimentation Al Amana. Le microcrédit n'est pas qu'un dispositif technique : il s'enchaîne dans des représentations sociales et religieuses associées au crédit et à la dette, il s'inscrit dans des configurations agro-écologiques diversifiées, il fait appel à des interactions sociales avec les opérateurs de crédits et les leaders locaux qui varient considérablement d'un lieu, voire d'une personne à l'autre. Al Amana était perçu selon les régions comme un organisme assimilé à l'État central, source de crainte ici, illégitime ailleurs (« l'argent des voleurs ») ou comme un organisme de don et non de prêt (Morvant-Roux *et al.*, 2014). Cela explique la très forte hétérogénéité de l'impact selon les régions et à l'intérieur de celles-ci, le taux de recours pouvant varier de 5 à 43% au sein d'une même région. La mise en œuvre d'un programme peut différer dans l'espace, selon les sites, les valeurs et les routines des opérateurs, mais aussi dans le temps : le programme lui-même peut évoluer en cours d'expérimentation parce que les acteurs s'aperçoivent pragmatiquement de certaines difficultés (dans le cas d'Al Amana, abandon en cours de projet d'un quota de crédit pour les femmes, introduction d'une formule de crédit à des individus et non seulement à des groupes comme prévu). Difficile alors de savoir ce qui est testé et sur quelles populations cible (Bernard *et al.*, 2012). La nécessaire standardisation du protocole s'accommode donc mal des bricolages locaux et de la variabilité du social.

Ce problème se pose aussi en biologie, où les expérimentations doivent faire face à la variabilité intrinsèque du vivant. Jean-Paul Gaudillière (2006) rappelle les efforts de standardisation des animaux de laboratoire nécessaires pour assurer la reproductibilité et la comparabilité des expériences : standardisation à la fois physiologique et génétique des lignées de cobayes, instauration dans les élevages de condition de vie et de nutrition homogènes. En dépit de cette standardisation, les interactions entre les manipulateurs et les cobayes peuvent perturber les expériences : ainsi la façon dont le personnel de laboratoire prend en main la souris (avec douceur ou brusquerie) a une influence significative sur les résultats des expérimentations. C'est encore plus vrai des essais cliniques qui sont des constructions sociales soumises à des influences multiples (Labrousse, 2010).

Pour ces raisons, on constate fréquemment un décalage important entre le protocole planifié et son avatar de terrain. Maintenir le protocole peut relever de la mission impossible. Un assistant de recherche du J-PAL en Afrique évoque ainsi des pratiques expérimentales qui s'apparentent à des « ROCT : *Randomized Out of Control Trials* » (Jatteau 2013b : 20), plutôt qu'à des RCT (*Randomized Controlled Trials*). En France, un chercheur évoque des difficultés telles qu'il a dû recourir à des techniques quasi-expérimentales « il y a tellement d'adaptations, il y a tellement de rustines là-

dedans que c'est aussi contestable, ou si ce n'est pas contestable, c'est difficilement probant » (Devaux-Spatarakis, 2014 : 455). Mises à l'épreuve du terrain, les expérimentations nécessitent de multiples expédients qui les éloignent de la pureté du discours de la méthode de certains « randomistas ». Toute méthodologie – quantitative comme qualitative – met en jeu des bricolages. Ces arrangements ne font-ils pas de cette technique un outil ordinaire, demandant un surcroît de réflexivité de la part des expérimentateurs ?

### **Un processus d'interprétation délicat**

De même, l'interprétation des résultats n'est pas aussi tranchée et dénuée d'ambiguïté que sur le papier. Il est tout d'abord difficile d'isoler l'impact testé. Loin d'un simple "verdict des données", repérer ce qui a été véritablement testé dans l'expérimentation n'a rien d'évident (Bernard *et al.*, 2012). Dans le cas de l'expérimentation Al Amana, les résultats étaient apparemment limpides : ce programme de microcrédit en zone rurale avait échoué (très faible taux de recours, impact insignifiant sur la pauvreté, la consommation et la diversification de l'activité). L'étude qualitative a cependant montré que le calendrier des remboursements, conçu pour des zones urbaines, ne correspondait pas aux contraintes du calendrier agricole. Ce qui est donc testé, c'est un dispositif qui – même simple en apparence – correspond à un ensemble de propositions explicites et implicites dont il est difficile de savoir lesquelles jouent dans le succès ou l'échec. Il s'agit d'un problème épistémologique connu : le problème de Duhem-Quine. Il est impossible de tester une hypothèse de manière isolée car tout test empirique de cette hypothèse (ici, le dispositif de microcrédit est-il un outil efficace de lutte contre la pauvreté ?) nécessite une ou plusieurs hypothèses auxiliaires (ici, le calendrier n'a pas d'influence causale). Les expérimentations visent à isoler un effet pur mais l'isolation s'avère souvent équivoque. L'idée d'*experimentum crucis*, à savoir d'expérimentations permettant de trancher définitivement les débats, semble utopique.

Il est également délicat de saisir le cheminement causal (comment, par quels mécanismes ?) qui a conduit aux résultats constatés (ça marche ou pas). En effet, en dehors des cas de mono-causalité simple (une cause entraîne un effet, sans rétroaction de l'effet sur la cause), les expérimentations aléatoires fournissent des preuves d'efficacité (tel effet est observé) et non de causalité (quels sont les mécanismes générateurs de cet effet ?). Des essais cliniques montrent ainsi l'efficacité de l'acupuncture pour prévenir les nausées post-opératoires mais les mécanismes générateurs de cet effet ne sont pas connus (Labrousse, 2010). Dans les cas de causalités complexes, cumulatives, multifactorielles, non linéaires, les chaînes causales relèvent d'une boîte noire pour les expérimentateurs. Mener systématiquement des enquêtes qualitatives complémentaires permettrait d'ouvrir cette boîte noire.

### **Une pertinence circonscrite à un type particulier d'action publique**

Comme les expérimentations aléatoires sont relativement rigides et adaptées à une causalité simple, elles s'avèrent pertinentes pour des projets où le lien causal entre le traitement et son effet s'exerce de manière relativement rapide et linéaire comme dans le cas des vermifuges étudié par Kremer et Miguel (2004). Ce traitement simple (un comprimé tous les 6 mois) améliore rapidement l'état de santé des enfants touchés par les vers et contribue à réduire l'absentéisme. Les expérimentations

paraissent en revanche moins adéquates pour tester des ensembles de mesures complexes et évolutives impliquant de longs apprentissages. Bernard, Delarue et Naudet (2012) ont étudié de près ces questions à l'Agence Française de Développement. Ils qualifient les projets en adéquation avec les expérimentations aléatoires de « projet-tunnel ». Ils précisent ainsi les prérequis de la randomisation : le programme testé doit porter sur « (i) une période cohérente avec la chaîne causale supposée, (ii) un nombre limité de traitements, homogènes et précis, (iii) un processus d'administration préalablement testé, (iv) une chaîne causale courte et indépendante d'événements extérieurs, (v) une appropriation rapide et stable dans le temps par les bénéficiaires du traitement, (vi) une participation large et stable dans le temps des bénéficiaires, et (vii) un ensemble d'effets mesurables à court ou moyen terme, qui couvre les principaux aspects du traitement ». Comme beaucoup d'actions sociales s'écartent de ces prérequis, le champ de pertinence des expérimentations est finalement très circonscrit. Ces conclusions rejoignent de manière troublante la littérature des années 1970-80 faisant le bilan de la vague expérimentale précédente (Monnier, 1992).

### **3. Un projet récurrent et cyclique de gouvernement par la preuve**

De fait, les problèmes pratiques rencontrés par les expérimentations ne sont pas nouveaux. Ils relèvent d'une histoire ancienne, celle de la « société de l'expérimentation », des espoirs qu'elle a soulevés et du relatif désenchantement qui a suivi.

#### **Des années 1920 à nos jours : l'*experimenting society*, projet ancien et récurrent**

Quand on élargit la perspective à l'histoire longue des expérimentations sociales, il apparaît que les psychologues y ont joué un rôle pionnier, bien avant les expérimentations de Fisher en agronomie dans les années 1930 et l'avènement des essais cliniques en médecine à partir des années 1940. Dès les années 1910, les thématiques éducatives et les élèves ont été les sujets privilégiés de ces expérimentations : « les enfants, comme les rats, sont disponibles en quantité et sans coût » et leur soumission à l'autorité des enseignants favorise le respect des contraintes du protocole expérimental (Boring, 1954, in Dehue, 2001 : 290). Les revues de psychologie américaines publient alors de nombreuses expérimentations contrôlées (avec groupes test et témoin) portant sur l'impact de la variation de la taille des classes, du sexe des enseignants ou encore des modalités d'aération des classes etc. Pour constituer des groupes comparables, ces psychologues mettent d'abord en place des méthodes d'appariement, pour recourir ensuite au tirage aléatoire. Au milieu des années 1920 a lieu à Chicago la première expérimentation sociale randomisée contrôlée d'envergure. Elle teste l'impact d'une campagne d'information – au sujet des dates et procédures de vote, rédigée dans la langue maternelle des citoyens – sur la participation au vote : 6 000 citoyens de diverses origines y participèrent. On le voit, ces expérimentations pionnières rejoignent dans leurs modalités et leurs thématiques les expérimentations contemporaines. Elles s'inscrivent dans un même projet d'une société de l'expérimentation gouvernée par la preuve et semblent favorisées par des contextes similaires.

Car on observe aux États-Unis dès les années 1920-30 une conjonction favorable à l'avènement progressif d'une société de l'expérimentation (Dehue, 2001) : tout d'abord, une montée des politiques sociales et des rationalités administratives (et donc des préoccupations sociales et du modèle de l'expertise rationnelle, standardisée et impersonnelle dans le gouvernement des populations) ; elle va se conjuguer au soupçon libéral tenace de l'inefficacité des deniers publics, soupçon qui justifie l'exigence d'une évaluation objective de leur utilisation. Une conjoncture comparable se retrouvera lors de la seconde grande vague expérimentale dans les années 1960-1970 en Amérique du Nord : programmes sociaux de la *war on poverty* engagée par l'administration Johnson et rationalisation des choix budgétaires en vue d'une utilisation efficace de l'argent public (Monnier, 1992). En France, au mitan des années 2000, c'est également dans le double contexte d'une rationalisation des politiques budgétaires et d'une volonté de renouveau des politiques sociales sous l'impulsion de Martin Hirsch que les expérimentations sont mises en place pour la première fois, dans le cadre de la troisième grande vague expérimentale.

Autre facteur important dans le développement des expérimentations : l'économie mobilisée de la seconde guerre mondiale. Elle permettra le développement des essais cliniques par la mobilisation des ressources, la mise à disposition de sujets nombreux et la coordination publique des essais. Elle favorisera aussi les expérimentations psychologiques au sein de l'armée américaine : la section expérimentale de la « Morale Division » était notamment composée de psychologues chargés d'évaluer les motivations des soldats et de tester au moyen d'expérimentations randomisées l'impact du cycle de films *Why we fight* sur ces motivations (Dehue, 2001). Et Donald Campbell, figure majeure de la seconde vague expérimentale, y fera ses premières armes en tant que psychologue de l'armée.

### **Les cycles enthousiasme-déception des évaluations expérimentales**

Introduites en économie à la fin des années 1960, les expérimentations randomisées contrôlées vont connaître un essor important dans les années 1970 suivi d'une normalisation et d'un tassement dans les années 1980, avant le regain des années 2000. Elles semblent sujettes à des cycles : une phase initiale d'enthousiasme, suivie d'une phase de relative déception.

Dans la phase ascendante, celle du « paradigme expérimental flamboyant » (Monnier, 1992), des expérimentations d'envergure, souvent coûteuses, prennent place. On pense alors qu'elles vont reconfigurer en profondeur les programmes sociaux comme les sciences sociales. Les expérimentations y sont présentées comme un outil dépolitisé de l'action publique, l'expérimentateur incarnant la figure de l'expert impartial, agnostique s'appuyant sur des faits purs. Ainsi, Campbell fait au début des années 1970 du chercheur en sciences sociales le « serviteur méthodologique de la société de l'expérimentation », « une société scientifique, non dogmatique, honnête, responsable » (Monnier, 1992).

L'enthousiasme initial s'estompe ensuite. « Les commanditaires [politiques], déçus par l'incapacité des universitaires à formuler des conclusions en terme d'implications politiques plus larges, ont renoncé à la mise en place de dispositifs lourds étroitement focalisés sur un seul objectif » (Monnier, 1992 : 45). C'est la « période

des petits pas », les expérimentations se font moins intrusives et plus modestes : elles visent à mieux se fondre dans le fonctionnement administratif et social ordinaire afin de limiter les réactions de résistance, elles sont conduites sur des durées plus brèves et engagent moins de fonds. Elles testent des changements incrémentaux au sein de dispositifs existants plutôt que d'ambitieux nouveaux programmes (Greenberg et al., 1999). Quant aux promoteurs scientifiques des expérimentations, ils se montrent plus mesurés et précautionneux. Ainsi le manuel de Cook & Campbell (1979) fait-il passer le nombre de « menaces potentielles à la validité » des expérimentations de 12 à 33 par rapport au Campbell & Stanley (1963). Si l'on manque du recul nécessaire dans le cas français, certains indices donnent à penser que nous entrons dans cette phase. Le nombre d'expérimentations s'est tari depuis le départ de Martin Hirsch et on note une faible utilisation par les administrations des résultats produits par cette méthode (Devaux-Spatarakis, 2014). Initialement présentées comme l'étalon-or autosuffisant de l'évaluation, les expérimentations aléatoires deviennent « un outil de plus dans la panoplie des évaluateurs » (L'Horty et Petit, 2011). Et sont privilégiées de manière croissante des expérimentations moins « intrusives » dans lesquelles les sujets ne sont pas nécessairement conscients de l'évaluation.

Néanmoins, les expérimentations aléatoires poursuivent aujourd'hui leur ascension dans les pays en voie de développement, auprès des organismes internationaux, des fondations et dans la sphère académique en économie. Cette dernière vague est portée par des déterminants propres à la discipline économique : tournant empirique d'une partie du *mainstream*, liens croissants avec la psychologie expérimentale, valorisation dans les revues économiques les plus cotées, ce qui n'était pas le cas de la précédente vague. En témoigne l'essor concomitant à celui du J-PAL des travaux expérimentaux de John List à l'université de Chicago, plus imprégnés de comportementisme et de référents néoclassiques que ceux de Duflo.

### **Les expériences répétées des expérimentations : quels apprentissages ?**

L'histoire des expérimentations semble parfois bégayer. Les expérimentations aléatoires avaient fait une entrée fracassante dans le domaine économique en 1968 avec le lancement du *New Jersey Income Maintenance Experiment* sous l'impulsion d'une doctorante au MIT, Heather Ross (Greenberg et al, 1999). C'est une autre doctorante en économie au MIT, Esther Duflo, qui jouera un rôle considérable dans la formation d'une seconde vague expérimentale en économie une trentaine d'années plus tard.

Pourtant, il est peu fait référence à l'expérience des vagues précédentes dans les écrits de la nouvelle vague expérimentale, tout particulièrement ceux qui portent sur les pays en développement, comme si leur application à de nouveaux territoires ouvrait une page blanche. Développée au début du XX<sup>e</sup> s. dans la volonté d'améliorer les programmes sociaux dans les pays développés, la méthode a acquis, semble-t-il, une nouvelle virginité dans les pays du Sud, légitimant, par effet de boomerang, son retour en force au Nord : « L'un des messages essentiels de la professeure au MIT est que l'évaluation expérimentale a fait ses preuves pour analyser les causes de la pauvreté dans les pays pauvres, et qu'il convient désormais de l'utiliser pour le même objectif dans les pays riches et notamment en France. » (L'Horty et Petit, 2011). Au sein de la société d'évaluation américaine, des protagonistes de la période précédente s'étaient émus de ce « retour vers le futur ». Ainsi Nick L. Smith s'étonne-t-il : « Lorsque les



discussions sur le rôle du design expérimental dans les évaluations sont revenues sur le devant de la scène il y a quelques années, j'ai pensé : mais n'avions-nous pas réglé la question ? Il y a près de 25 ans, en 1981, j'avais organisé un débat lors du Congrès [de ce qui deviendra] l'*American Evaluation Association* "Le gouvernement fédéral doit-il rendre obligatoire l'usage de méthodes expérimentales dans l'évaluation ?" » (Devaux-Spatarakis, 2014 : 109). Cette association mettra en garde contre l'idée d'un étalon-or de l'évaluation et contre les limites des expérimentations aléatoires, sans rencontrer d'écho notable parmi les économistes.

Pourtant, tirer les leçons des expériences passées aurait peut-être permis d'accélérer les effets d'apprentissage des acteurs politiques comme scientifiques. Outil utile, l'expérimentation aléatoire n'est ni un *gold standard*, ni une méthode révolutionnaire. Elle gagnerait à être plus insérée dans des approches multi-méthodes (Morvant-Roux *et al.* ; Labrousse, 2016), pour en développer un usage plus contextualisé (pour qui ? dans quel contexte ?). La question du choix des « traitements » et de leur pertinence pour les populations est aussi fondamentale que celle de leur efficacité. Plutôt qu'un outil dépolitisé de l'action sociale, l'enjeu est d'en faire un instrument au service des débats démocratiques, passant d'une « politique fondée sur la preuve » à une « politique informée par les preuves ».

### Références bibliographiques

Banerjee Abhijit & Esther Duflo (2011), *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, New York: Public Affairs.

Bérard Jean & Mathieu Valdenaire (2013), « L'expérimentation pour renouveler les politiques publiques ? », *La Vie des idées*, 25 juin 2013, <http://www.laviedesidees.fr/L-exp experimentation-pour-renouveler.html>

Bernard Tanguy, Jocelyne Delarue & Jean-David Naudet (2012), « Impact evaluations: a tool for accountability? Lessons from experience at Agence Française de Développement », *The Journal of Development Effectiveness*, vol. 4, n° 2 : 314-327.

Bureau Marie-Christine, Sarfati François, Simha Jules & Tuchsirer Carole (2013), « L'expérimentation dans l'action publique », *Travail et Emploi*, 135, 41-55.

Deaton Angus (2010), « Instruments, randomization, and learning about development », *Journal of Economic Literature*, vol. 48, n° 2 : 424-455.

Dehue Trudie (2001), « Establishing the experimenting society: The historical origin of social experimentation according to the randomized controlled design », *American Journal of Psychology*, 114-2, p. 283-302.

Devaux-Spatarakis Agathe (2014), *La méthode expérimentale par assignation aléatoire : un instrument de recomposition de l'interaction entre sciences sociales et action publique en France ?*, Thèse pour le Doctorat en Science politique, Université de Bordeaux, octobre 2014.

Duflo Esther, Rachel Glennerster & Michael Kremer (2004), *Randomized Evaluations of Interventions in Social Service Delivery*, MIT.

Duflo Esther (2009), *Expérience, science et lutte contre la pauvreté*, Paris, Fayard.

Greenberg David, Mark Shroder & Matthew Onstott (1999). "The Social Experiment Market", *Journal of Economic Perspectives*, 13 (3) : 157-172.

Jatteau Arthur (2013a), *Les expérimentations aléatoires en économie*, Paris, La Découverte.

Jatteau Arthur (2013b), « Expérimenter le développement ? Des économistes et leurs terrains », *Genèses*, 4 (93) : 8-28.

L'Horty Yannick et Pascale Petit (2011), « L'évaluation aléatoire : un succès qui ne doit rien au hasard », *La vie des idées*.

Labrousse Agnès (2016), « Not by Technique Alone. A Methodological Comparison of Development Analysis with Esther Duflo and Elinor Ostrom », *Journal of Institutional Economics*, Issue 12(2), June 2016, [en ligne](#) (octobre 2015).

– (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. <http://regulation.revues.org/7818>

Mayneris Florent (2009), « L'économie du développement à l'épreuve du terrain – Entretien avec Esther Duflo », *La vie des idées*.

Monnier Eric (1992), *Évaluations de l'action des pouvoirs publics*, 2<sup>e</sup> éd., Paris, Economica.

Morvant-Roux Solène, Guérin Isabelle, Roesch Marc & Moisseron Jean-Yves. (2014), “Adding Value to Randomization with Qualitative Analysis: The Case of Microcredit in Rural Morocco”, *World Development*, 56, 302–312.

Publié dans [laviedesidees.fr](http://laviedesidees.fr), le 4 mars 2016

© [laviedesidees.fr](http://laviedesidees.fr)