



# Pauvreté, expérimentation et développement dans les travaux d'Esther Duflo

Mohamed Soual

## ► To cite this version:

Mohamed Soual. Pauvreté, expérimentation et développement dans les travaux d'Esther Duflo. Economies et finances. Université Grenoble Alpes [2020-..], 2021. Français. NNT : 2021GRALE006 . tel-03625985

**HAL Id: tel-03625985**

**<https://theses.hal.science/tel-03625985>**

Submitted on 31 Mar 2022

**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.



## THÈSE

Pour obtenir le grade de

**DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES**

Spécialité : Sciences économiques Arrêté ministériel : 25 mai 2016

Présentée par

**Mohamed SOUAL**

---

Thèse dirigée par Redouane TAOUIL, PRF.UNI.2C

Préparée au sein du Laboratoire Centre de Recherche en Économie de  
Grenoble dans l'École Doctorale Sciences Économiques

---

**Pauvreté, expérimentation et développement dans les  
travaux d'Esther Duflo.**

**Poverty and experiment: Esther Duflo work.**

---

Thèse soutenue publiquement le 13 décembre 2021

Devant le jury composé de :

**Ahmed TRITAH**, Professeur, Université de Poitiers, Président du Jury.

**Juan José JARDON URRIETA**, Professeur, Université du Mexique, rapporteur.

**Vincent SPENLEHAUER**, Professeur, École Nationale des Ponts et Chaussées  
(Ponts ParisTech), rapporteur.

**Pierre-Noël GIRAUD**, Professeur émérite, École des Mines de Paris (Mines  
ParisTech).

**Sophie SAGLIO-ROSSINI**, Maître de Conférence, Université Paris VIII.

**Redouane TAOUIL**, Professeur, Université Grenoble-Alpes, Directeur de thèse.



## Remerciements

Dans *La philosophie du non*, Bachelard écrit : « La vérité est fille de la discussion, non pas fille de la sympathie ». Aux membres du jury voudrais exprimer les remerciements les plus vifs et sincères pour le temps et l'attention qu'ils ont accordés au présent travail. Je remercie particulièrement mon directeur de thèse qui a accompagné son avènement avec confiance et bienveillance.

Il est une personne que je voudrais remercier, tout particulièrement, pour la patience dont elle a fait montre le long de six longues années de travail consacré à la thèse qui ont été autant de temps pris à la vie de famille. Il s'agit, bien entendu, de mon épouse, Amina.

Ma fille Myriam me fait régulièrement le reproche d'être, souvent, indisponible lors des rares moments de vacances pris ensemble. Qu'elle m'en excuse. Je la remercie pour sa patience et espère que cette indisponibilité forcée pourra lui inspirer le goût de l'effort et de la persévérance.

Je n'oublierai pas, non plus, tous mes amis et les membres de ma famille que je ne saurais citer un à un et qui, constamment, ont témoigné de l'intérêt pour cette aventure intellectuelle et m'ont prodigué leurs encouragements.

J'ai toujours prêté la plus grande attention aux questions liées à la recherche et à l'enseignement. Au-delà de l'intérêt intrinsèque du sujet, ce travail a permis de revisiter des questions essentielles liées à la méthodologie et aux questions épistémologiques qui représentent, en quelque sorte, la manière de dire la vérité. A ce sujet, mes nombreuses discussions avec le Professeur Pierre- Noël Giraud ont été à la fois inspirants et éclairants. Qu'il trouve, dans cette reconnaissance, le témoignage de ma très grande considération.

A toutes et à tous, j'exprime mes vifs remerciements à l'occasion de ce moment à la fois institutionnel et personnel.

# Sommaire

<b>Introduction générale</b>	<b>5</b>
<b>Chapitre I : Objet et visée du J-PAL</b>	<b>11</b>
Section 1 : De la critique de la macroéconomie du développement	14
Section 2 : Micro-fondations empiriques d'évaluations des politiques publiques	22
<b>Chapitre II : La méthode expérimentale, protocoles d'observation et terrains</b>	<b>43</b>
Section 1 : Le modèle causal de base	44
Section 2 : La causalité comme dépendance contrefactuelle	47
Section 3 : Expériences de programmes de lutte contre la pauvreté et évaluation	67
<b>Chapitre III : Limites du Gold Standard</b>	<b>78</b>
Section 1 : Les apories de l'identification causale	81
Section 2 : Les biais de randomisation entre rigueur méthodologique et remède miracle	86
Section 3 : Pléthore de l'empirie, déficit de la théorie	98
<b>Chapitre IV : Expérimentation, pauvreté et reconfiguration du consensus de Washington</b>	<b>125</b>
Section 1 : Consensus et marché des évaluations des politiques publiques	126
Section 2 : Rationalité individuelle et accès aux ressources	135
Section 3 : Pauvreté, ciblage et gouvernement du social	143
<b>Chapitre V : Expérimentation, rhétorique, performativité et fabrication des politiques publiques</b>	<b>163</b>
Section 1 : Expertise, rhétorique et performativité	166
Section 2 : Forum scientifique, politique et médiatique	177
<b>Conclusion générale</b>	<b>186</b>
<b>Bibliographie</b>	<b>189</b>

## Introduction générale

Dans une contribution, fort remarquée, lors des travaux de la session plénière de l'Académie Hassan II des sciences et des techniques portant sur l'état des sciences économiques, Jean Cartelier (2016) relève trois traits qui semblent caractériser celles-ci depuis deux décennies. D'abord, la mise en avant de la rationalité des comportements individuels et de l'équilibre comme mode de coordination des actions et comme référentiel de base. En deuxième lieu, la nette prédominance des travaux empiriques qui procèdent de la mise en cause de la figure de l'homo economicus en s'appuyant largement sur des possibilités de traitement de l'information ouvertes par le perfectionnement des techniques économétriques. Enfin, l'orientation de plus en plus affirmée de l'expertise et de l'appui à la conception de mesures de politiques publiques ou d'aide à la décision privée. Dans ce contexte, les préoccupations théoriques connaissent un déclin qui s'exprime dans l'abandon des questions nodales de l'économie néo-classique que sont la formation de l'équilibre et la coordination<sup>1</sup>.

L'approche promue par le laboratoire d'action contre la pauvreté, J-PAL (Jameel Poverty Action Lab), créé par Abhijit Banerjee et Esther Duflo, illustre avec éloquence cette configuration. L'accent est, ainsi, mis sur la réaffirmation des choix des agents comme préalable obligé à la compréhension de leurs actions et à la définition d'instruments incitatifs à la modification des comportements. En contraste avec la rationalité absolue, qui postule la maximisation d'une fonction objectif sous contrainte, sont pris en compte la limitation des capacités cognitives, le manque de temps pour la collecte d'informations et la considération des options possibles et leurs corrélats que sont les effets pervers et l'incohérence temporelle. Par le recours à l'expérimentation aléatoire de terrain, les analyses du J-PAL se rattachent au tournant empirique que Angrist et Pischke placent sous le signe de la

---

<sup>1</sup> Dans l'article intitulé "The current State of Economics: Needs Lots of Work", Bergmann, que cite Cartelier, se désole du fait que c'est l'idéologie politique qui détermine le profil des controverses. A cet effet, elle soutient que la profession d'économiste a été pervertie par deux habitudes : d'une part la théorisation basée sur de simples scénarios inventés associés aux hypothèses sur la cupidité et la rationalité humaines ; d'autre part l'absence de lien rigoureux entre la modélisation macroéconomique et la description empirique du comportement des agents.

révolution de la crédibilité. Ce tournant, incarné par les expériences contrôlées en laboratoire, les expériences naturelles et les essais contrôlés randomisés, se spécifie par la conduite d'études économétriques et l'inférence de résultats à partir de la collecte et de la construction des données. La forme d'expérimentation utilisée sous l'impulsion de Banerjee et Duflo s'assigne l'objet de mesurer l'effet causal d'un programme ou dispositif en répartissant aléatoirement, sans biais de sélection, les sujets en groupes de contrôle et de traitement : « L'un des problèmes centraux est le problème de la sélection : les participants sont en général systématiquement différents des non participants. Il est bien entendu impossible 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 » (Banerjee et Duflo, 2009, p. 693).

Comme telle, elle revendique une identification crédible des impacts des programmes mis en œuvre. Normative, cette démarche vise à établir des prescriptions destinées à inspirer les décideurs publics ou privés par la production de l'*evidence based policy*. Les résultats de terrain servent de base à la conception de dispositifs affectés à la modification des comportements au regard de l'efficacité recherchée : « Le célèbre programme mexicain intitulé Progresa- Oportunidades combine des transferts sociaux vers les femmes des familles pauvres, sous condition d'un « bon comportement » (par exemple que les femmes investissent dans l'éducation et la santé préventive) » (Banerjee et Duflo, 2009, p. 695).

Ainsi conçu, le projet du J-PAL s'est affirmé comme constitutif d'une microéconomie expérimentale au sein de l'économie du développement en opposition à la perspective macroéconomique tant au niveau d'outils économétriques qu'au plan de l'angle d'analyse. Les principaux promoteurs du J-PAL justifient cette supériorité de l'expérimentation microéconomique par la qualité des données collectées : « la méthode expérimentale encourage le développement de méthodes de collection de données originales : contrairement aux bases données traditionnelles, les échantillons sont faibles (les coûts restent

donc limités) et les questions posées sont très spécifiques [...] A l'inverse, les données non expérimentales doivent souvent compter avec des variations (changement de politique, variations induite par le marché, variations naturelles, bouleversements au niveau de l'offre, etc.) soutiennent Banerjee et Duflo dans l'article séminal destiné à répondre aux critiques adressées aux RCT (p. 700).

La réception favorable dont a fait l'objet ce projet a donné lieu à des travaux qui ont cherché à explorer la portée en considérant les expérimentations aléatoires contrôlées (RCT) comme le *Gold Standard*, ainsi que le proclament les auteurs du J-PAL. En ce sens, ils ont procédé à l'analyse de divers domaines tels que l'éducation, la santé, la finance, l'entrepreneuriat et la gouvernance. D'un autre côté, une large littérature s'est attachée à mettre en exergue les limites de la méthode en pointant les problèmes de généralisation des résultats. Ce faisant, elle tient la validité externe pour problématique mais considère que l'approche du J-PAL satisfait aux critères de la validité interne.

En centrant sa problématique sur le déploiement d'une *critique globale*, inspirée de la thèse Duhem-Quine, la présente étude remet en cause ce consensus soutenant que cette forme de validité n'est pas acquise. Sous ce rapport, l'approche du J-PAL est examinée comme un réseau interdépendant d'énoncés à travers ses analyses positives, ses composants normatifs et prescriptions et l'ingénierie sociale à laquelle elle est étroitement imbriquée.

La proposition d'analyse centrale développée ici montre que la fortune de l'approche du J-PAL et la diffusion du recours aux expérimentations ne résultent pas de la validité de son modèle d'inférence causale, mais de considérations liées aussi bien à la rhétorique de son Golden Standard, à sa symbiose avec le référentiel du consensus de Washington et ses programmes de ciblage de la pauvreté qu'à son insertion dans les réseaux de politiques publiques.

Cette proposition s'inscrit dans l'hypothèse d'admissibilité suggérée par Lévy-Leblond (2020) : « Ce qui caractérise l'admissibilité d'un énoncé dans le corpus scientifique relève moins d'une appréciation de sa validité que d'un jugement sur



sa pertinence. La contre-épreuve est d'ailleurs aisée, puisque nombre des assertions de la science telle qu'elle se fait -la majorité, sans doute- se révèlent erronées sans pour autant disqualifier les recherches qui les ont produites du moment que leur intérêt est reconnu par la collectivité savante » (p. 103). Cette affirmation, qui jette le doute sur la séparation entre l'évaluation des théories par l'examen de leur structure interne et celle qui consiste à mesurer les effets de leur contexte sur leur contenu, invite à reconsidérer la force attribuée aux critères de justification. « Ainsi, la question classique de la validation, ou de la réfutation des idées scientifiques - conclut Lévy-Leblond - le cède-t-elle en importance au profit de considérations sur leur qualification ou sur leur disqualification. De fait, bien des travaux sont abandonnés sans jamais être tombés sous le coup d'une critique explicite et rédhibitoire » (p. 103). Ce point de vue peut s'appliquer, également, à l'exemple d'approches qui restent bien admissibles alors qu'elles ne bénéficient pas de fermes titres de validité.

Mener une enquête selon cette optique requiert l'explication de la nature et la visée globale de l'approche du J-PAL en situant ses enjeux en tant que projet de facture microéconomique, déployée à l'encontre de la macroéconomie dont il récuse l'usage des régressions de croissance et le caractère foncièrement global des politiques économiques en proposant des programmes sur une petite échelle évaluée au moyen d'expérimentations aléatoires contrôlées (chapitre 1).

La reconstitution du modèle causal, qui sert de cadre à ces évaluations, par la réunion des ingrédients des méthodes de Holland, Neymar et Rubin, permet de délimiter sa structure formelle, mettre à jour le faisceau d'hypothèses et spécifier le mode de dérivation des résultats. Cette dérivation repose sur des inférences qui font appel à la dépendance contrefactuelle et la causalité statistique à travers les effets potentiels, l'effet moyen dont les affinités avec les sciences médicales visent à renforcer la crédibilité attribuée au *Gold Standard* (chapitre 2).

La critique menée à l'endroit de cette méthode est allée au-delà de la bipartition en validité interne et validité externe qui est le contexte marquant des discussions privilégiées des évaluations aléatoires. Aussi, nous nous sommes forcés à examiner

ces évaluations à l'aune de leur propre conception de la causalité. Dans la lignée des critiques de Heckman (1992, 2000) ; Manski (1995, 2003) ; Deaton (2009, 2010) ; Cartwright (2009) ; Deaton et Cartwright (2016), elle s'évertue à pointer les apories de l'identification causale en montrant comment elles affectent les conséquences déductives. Sous ce rapport, les biais sont rehaussés au rang de limites qui compromettent la validité interne de l'approche du J-PAL. Les ambiguïtés induites par les traitements de l'exogénéité et les variables instrumentales sont réinterprétées sous le thème de la contingence. De même que les effets moyens et l'hétérogénéité de l'impact des traitements, l'asymétrie des effets et des causes comme les entorses aux hypothèses sont interrogées eu égard au déroulement des expérimentations. S'agissant de la généralisation des résultats, elle implique que l'occurrence du même rapport causal se place, de ce fait, sous des procédés inductifs des énoncés observationnels. A cet égard, le J- PAL juge nécessaire de multiplier les expérimentations dans des contextes variés. Cette démarche est doublement passible d'objections : outre qu'il n' y a pas de critère pour définir le nombre d'évaluations suffisant pour valider l'élargissement du domaine de pertinence des résultats, les variations des expérimentations conduisent à renforcer l'importance des variables contextuelles et, partant, à octroyer une part prépondérante à la contingence. Ce défaut tient ses racines dans le primat accordé à l'empirie qui, faute de guide théorique, se rend aveugle et confine à l'impasse : « L'objectif de la littérature économétrique, comme celui de toute science, est de modéliser les phénomènes à un niveau plus profond, de comprendre les causes qui produisent les effets afin de pouvoir utiliser des versions empiriques des modèles pour prévoir les effets d'interventions jamais expérimentées avant, pour calculer une variété de politiques contrefactuelles et pour utiliser la théorie scientifique afin de guider les choix des estimateurs et l'interprétation des preuves. Ces activités nécessitent le développement d'une théorie plus élaborée que celle envisagée dans la littérature actuelle de l'inférence en épidémiologie et en statistique » (Heckman, 2005, p. 6). Deaton (2009) observe, d'ailleurs, que l'analyse économétrique a changé d'orientation au fil des ans, passant de l'analyse de modèles dérivés de la théorie à des spécifications beaucoup plus souples qui ne sont plus que les

représentations statistiques de l'évaluation des programmes (Deaton, 2009, p. 154). Le déficit théorique se manifeste, alors, dans la dilution de l'explication au sein de la description et de l'information du rapport causal (chapitre 3).

Au vu de ces limites de validité et d'absence de soubassement théorique, la faveur dont jouit l'approche de la pauvreté en termes d'expérimentations aléatoires tient à sa participation à l'enrichissement du consensus de Washington. Les recommandations de ciblage de la pauvreté prôné par la Banque mondiale portent l'empreinte des programmes d'évaluations aléatoires. Ce ciblage, qui recouvre l'adoption de programmes spécifiques, implique que les politiques monétaires et budgétaires centrées sur la stabilité macroéconomique comme les réformes structurelles censées accroître la flexibilité des marchés n'ont pas à prendre en charge l'action contre la pauvreté (Chapitre 4).

L'intérêt insigne accordée aux composants normatifs et prescriptifs explique également sa qualification de l'approche du J-PAL. Elle est à l'œuvre dans les forums de l'expertise et au sein du forum politique (Delmas, 2011 ; Jobert, 1992 ; de Maillard, 2015 ; Muller et Surel, 2018 ; ...). L'imbrication de l'entreprise J-PAL dans les réseaux d'acteurs, l'influence de la rhétorique des programmes ciblés sur les valeurs, les normes et les algorithmes des politiques publiques autant que les rapports entre ces politiques, la démocratie, la citoyenneté tendent à accréditer que le développement se résout dans des actions contre la pauvreté. Au lieu de s'attaquer aux causes structurelles des inégalités, les méthodologies promues s'attaquent, plutôt aux imperfections des marchés en pénalisant les pauvres accusés d'adopter de mauvais comportements. Dans de telles configurations, la figure du scientifique se mue en expert et, lorsque le lien avec le commanditaire de l'expertise se fait solide, la connaissance est utilisée pour limiter la prolifération des débats, afin d'orienter les politiques et rendre les contingences déterminantes (chapitre 5).

## **Chapitre I : Objet et visée du J-PAL**

L'intérêt porté aux questions relatives à la lutte contre la pauvreté coïncide avec le retournement idéologique opéré à la fin des années soixante-dix et marque l'avènement des stratégies de développement inspirées des institutions de Bretton Woods (Berr et al., 2004).

Coïncidant avec la crise des dettes et en récusant les théories du développement, les institutions financières internationales (IFI) introduiront des changements profonds dans la façon avec laquelle il conviendrait de penser les problèmes des pays en développement. L'insertion internationale basée sur les dotations en facteurs, la suprématie du marché et la remise en cause des secteurs traditionnels de l'État, notamment en matière de politiques sociales, vont être le canon à travers lequel ces IFI apporteront leur concours aux pays qui les solliciteraient. Les recommandations qui en sont issues se traduisent par la promotion de politiques macroéconomiques restrictives, l'ouverture croissante des économies à la libre concurrence par l'utilisation des leviers de la stabilisation (orthodoxie budgétaire, politique monétaire accommodante, taux de change flexibles et compétitifs) et des réformes structurelles (libéralisation, privatisation, dérèglementation et respect des droits de propriété). Sous l'impulsion et la plume de Williamson et l'appui des équipes qui s'y sont associées, ce paradigme prendra, en 1989, le nom de Consensus de Washington.

Le retournement libéral, né de la fin de l'affrontement Est-Ouest consécutivement à l'éclatement de l'Union Soviétique et du bloc de l'Est ainsi que de la crise de la dette, issue elle-même du recyclage des pétrodollars générés par les bours p  troliers des ann  es 1973 et 1979, a remis en selle les institutions financi  res internationales. La crise a, ainsi, introduit un revirement majeur en s'appuyant sur deux prescriptions    l'adresse des pays en d  veloppement : stabiliser l'  conomie en r  duisant les d  s  quilibres macro  conomiques dont en particulier les d  ficits jumel  s, d'une part, et en menant des r  formes structurelles comme la lib  ralisation des   changes, d'autre part.

Ce *profound economic mismanagement*, selon l'expression de Jeffrey Sachs, et les solutions proposées sont jugés par certains organismes internationaux, eux-mêmes, comme des préconisations « [étant] allées bien au-delà de ce qui aurait été nécessaire » (OCDE 2012).

Dans ce contexte global combinant fin de la guerre froide et promotion des politiques d'ajustements structurels, ces institutions et les pays qui y jouent un rôle dominant, États-Unis en tête, se sont libérés du poids des objectifs géostratégiques comme pays donateurs notamment par l'assèchement de l'aide publique au développement dont, en particulier, l'aide bilatérale non intermédiée par les organisations financières internationales. Cette évolution sonne la fin de « *l'ajustement des créances* » que Maynard Keynes appelait de ses vœux pour une sorte de co-développement au bénéfice des économies émergentes.

Ce profond changement paradigmatique a, par conséquent, favorisé l'introduction d'une nouvelle rationalité économique dans la production des savoirs avec des stratégies autant médiatiques que scientifiques, dépolitisant le développement au prix d'une articulation entre systèmes de connaissances et systèmes de gouvernement et visant le ciblage de pratiques de développement (Raphaëlle Parizet, 2017).

Ainsi, les grandes organisations internationales, dont la Banque mondiale et le FMI, ont-elles développé, au-delà de leur dimension normative et instrumentale, l'évaluation systématique de leurs politiques, programmes et projets opérationnels à destination des pays ayant recours à leurs interventions. Ces évaluations, qu'elles soient réalisées avant acceptation, en cours de réalisation ou bien souvent après achèvement, deviennent des instruments stratégiques et des outils impératifs nécessaires à l'élaboration des indicateurs de suivi et de comparaison. Elles ont, ainsi, « aidé à créer un paradigme technico-scientifique » garant de scientificité et d'universalité (Degeorges, Monnier et Spenlehauer, 1990).

Dans le sillage de l'évaluation et de la recherche de l'efficacité, devenues notions centrales du lexique de l'industrie de l'aide, l'on assiste à l'orée des années 2000 à

une conjonction d'un choc de « demande » émanant à la fois d'organismes internationaux de pays donateurs comme de fondations/bailleurs de fonds d'une part et d'un choc « d'offre » marqué du côté des chercheurs, d'autre part, entraînant une remarquable progression du nombre d'« évaluations d'impact » dans le secteur du développement (Stéphanie Pamies-Sumner et AFD, 2014 ; Yannick L'Horty et Pascal Petit, 2011).

Le succès des expérimentations aléatoires contrôlées (EAC), promues depuis 2003 par le J-PAL (Abdul Latif Jameel Poverty Action Lab) du MIT sous la férule d'Esther Duflo et d'Abhijit Banerjee comme instruments privilégiés d'un nouveau type de gouvernance alliant rigueur scientifique et efficacité pratique, s'explique, sans doute, par ce contexte. Ces EAC se distinguent des autres formes et méthodologies d'évaluations par leur mesure *ex-ante* d'impacts en mettant l'accent sur l'efficacité des politiques à mener. L'enjeu ne portant plus sur le bien-fondé des politiques mais sur la capacité à les mettre en œuvre à un niveau microéconomique. Si, selon une démarche positive, l'évaluation *ex-post* s'appuie sur les caractéristiques structurelles et les contraintes macroéconomiques, susceptibles de limiter les marges d'ajustement aux modifications des systèmes de prix et d'incitations pour vérifier rétrospectivement que les objectifs assignés à une politique ont été atteints, l'évaluation *ex-ante*, quant à elle, utilise des critères normatifs dans un but prospectif en s'intéressant au comportement économique des agents et à l'analyse de l'impact de chocs à un niveau microéconomique (Cling et al., 2008, p. 17). L'objet du chapitre est de montrer (section 1) l'ancrage des expérimentations aléatoires randomisées telles que promues par le J-PAL dans la critique de la macroéconomie du développement et plus généralement de l'économie politique. Ils renvoient dos à dos les défenseurs de l'aide au développement (Jeffrey Sachs, 2005) à ceux qui s'y opposent (Easterley 2006 et 2009, Moyo 2009). Ils se préoccupent moins de savoir s'il faut ou non aider, mais, comment le faire. Ils cherchent à connaître où peut aller l'aide et moins d'où elle vient, dénonçant au passage les trois tares qui affectent l'aide au pays pauvres et la lutte contre la pauvreté : « idéologies, ignorance et inertie entretenues par la pensée paresseuse » (Banerjee, 2006) et le sommeil dogmatique. Les promoteurs des RCT



s'opposent autant aux approches institutionnelles (Acemoglu et Robinson 2002) qu'aux approches structurelles (Deaton et Rodrik) en raison de la compréhension inadéquate de la pauvreté et des biais cognitifs dont souffrent les pauvres.

Se fondant sur une microéconomie empirique (section 2) et garante de la révolution de la crédibilité (Angrist et Pischke, 2000), la méthodologie « randomiste » s'inscrit fortement dans l'empirie qui cherche à trouver dans les données du terrain les « *hard evidence* » aspirant à identifier les effets causaux en répartissant aléatoirement les participants aux expérimentations de sorte à éliminer les biais de sélection et en dégagant les résultats contrefactuels issus de la comparaison entre groupes de traitement et groupes de contrôle. L'imitation du contrefactuel permet de s'affranchir de l'impossibilité d'observer en même temps un même groupe dans deux conditions différentes. A l'effet de conforter leur hégémonie dans le marché des évaluations, ils se prévalent de leur affiliation avec les expériences aléatoires, au long cours, pratiquées en médecine et garantes de la scientificité revendiquée.

## **Section 1 : De la critique de la macroéconomie du développement**

Duflo et Banerjee (2011, p. 19) invitent « ... à réfléchir non pas à deux, mais à trois fois : à nous détourner du sentiment que lutter contre la pauvreté est une tâche trop écrasante et à penser ce défi comme une série de problèmes concrets qui, une fois correctement identifiés et compris, peuvent être résolus un à un ». Les questions généralement posées se réduisent à questionner la cause ultime de la pauvreté ou la confiance que l'on peut accorder au marché. Les deux questionnements : « La démocratie est-elle une bonne chose pour les pauvres ? » ou bien « L'aide internationale a-t-elle un rôle à jouer ? », ne permettent pas selon les deux auteurs d'appréhender les causes du sous-développement ou d'évaluer efficacement les politiques menées.

Pour ces tenants des méthodes expérimentales aléatoires, les causes du sous-développement appréciées comme des déficits de croissance ne sont pas dues à des politiques macroéconomiques ou macrosociales déficientes ou encore à de mauvaises institutions mais bien à l'incapacité d'opérer des changements à la

marge. Pour mesurer la qualité de la gouvernance ou l'efficacité des institutions, lutter contre le paludisme ou mesurer l'impact effectif de l'aide au développement, il faudrait des preuves dures ou « *hard evidence* » (Banerjee, 2007) parce 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 pourrait bien s'expliquer par une troisième variable corrélée aux deux autres, constate-t-on, et par conséquent même « avec deux millions de régressions » il est difficile de 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. 23-24).

Se construit ainsi une méthodologie des petits pas et d'améliorations à la marge qui s'oppose aux approches institutionnelles ou d'ordre macroéconomique. D'essence empirique, cette approche se fonde sur le pragmatisme, qui s'oppose au sommeil dogmatique des trois i évoqués en introduction. Que ce dogmatisme soit le fait d'experts, des fonctionnaires de l'aide internationale ou de responsables politiques, il explique bien souvent l'échec des politiques et pourquoi l'aide ne peut produire l'effet attendu, considère-t-on. Commentant la désorganisation de l'aide apportée aux sinistrés du tremblement de terre ayant frappé le cachemire pakistanais en octobre 2005, Banerjee (2006) impute celle-ci à la pensée paresseuse qui domine dans l'appréhension de l'aide et la manière de la mobiliser. Selon lui et à bien des égards, l'épisode en question illustre, précisément, l'un des problèmes fondamentaux de l'aide : la paresse institutionnelle transformant l'*Aid thinking* en *Lazy thinking*.

C'est contre cette paresse et fort d'un *changement radical de perspective à l'opposé des réponses universelles* par des études menées avec *un haut degré de rigueur scientifique*, dotées de la *capacité à accepter le verdict des faits* et attachées, enfin, aux questions concrètes intéressant la vie des pauvres que les promoteurs du J-PAL s'opposent aux approches institutionnelles telles celles portées par Acemoglu et Robinson (2001, 2010, 2012) ainsi qu'aux modèles macroéconomiques notamment pour ce qui relève de l'aide au développement. Ils s'opposent également aux approches structurelles de l'évaluation des politiques publiques (Rodrik, 2008 ;



Deaton, 2009). En opérant un glissement vers une indifférenciation entre les enjeux d'ordre individuel et les enjeux d'ordre collectif, la critique adressée par ces micro-économistes expérimentateurs à la macroéconomie en général et à l'économie politique en particulier tranche avec une histoire de l'économie du développement et de l'évaluation des politiques publiques qui s'attachait aux transformations structurelles qu'exige le développement. Ils vont même jusqu'à considérer qu'il est possible de modifier aussi bien la gouvernance que les politiques sociales sans avoir à changer les structures existantes, y compris au sein des environnements les moins porteurs. D'après ces auteurs, « il y a une marge de manœuvre importante pour les améliorer dans de « bons » environnements institutionnels et une petite même dans les plus mauvais » (Banerjee et Duflo, 2011, p. 404).

Dès lors, lorsqu'il s'agit d'aide au développement, Esther Duflo et ses collègues proposent de répondre à deux questions : non pas combien d'argent faudrait-il dépenser mais comment le dépenser et non d'où vient-il mais où pourrait-il aller. Ces questionnements visent particulièrement Jeffrey Sachs (2005) qui soutient que l'aide internationale est la clé qui, par sa capacité à enclencher un cercle vertueux à même d'améliorer la productivité et à engendrer l'investissement, crée, ce faisant, une spirale bénéfique. Sachs considère, par ailleurs, que l'aide a été, très longtemps, trop parcimonieuse pour pouvoir faire sortir les pays sous-développés, les pays africains en particulier, de la « trappe à pauvreté » et plaide pour un « *big push* » (nous y reviendrons ci-après).

La critique n'épargne pas non plus William Easterley (2006 et 2009) ni Dambisa Moyo (2009), deux auteurs qui considèrent que l'aide internationale est plus néfaste que productive de développement ou de bien-être pour les populations empêchant les pays et les populations concernées de chercher leurs propres solutions en raison de la corruption et la perversion des institutions locales. Elle crée, par ailleurs, des lobbys d'organisations d'aide au développement auto-entretenus. S'agissant de l'Afrique, Moyo rapporte que ce continent a reçu d'importants montants d'aide qui ne l'ont pas empêché de s'enfoncer dans la pauvreté, et qu'en conséquence, l'aide, ne trouvant pas grâce à ses yeux, il n'y aurait rien à attendre. Pourtant, cette auteure qui, sans se soucier de l'origine des fonds qui alimentent la microfinance, plaide pour son extension et loue ses effets positifs et sa capacité à transformer les pauvres

en entrepreneurs potentiels. Ces deux auteurs ne se contentent pas de souligner l'impact négatif de l'aide sur la croissance ou sur la gouvernance, ils vont plus loin en affirmant que l'aide est tout simplement nocive pour les pays en développement.

Ces controverses ne sont pas nouvelles. Depuis le discours d'investiture de Harry Truman prononcé quelques temps après la fin de la deuxième guerre mondiale et le déclenchement de la guerre froide (« Il nous faut lancer un nouveau programme qui soit audacieux et qui mette les avantages de notre avance scientifique et de notre progrès industriel au service de l'amélioration et de la croissance des régions sous-développées. Plus de la moitié des gens dans le monde vit dans des conditions voisines de la misère. Ils n'ont pas assez à manger. Ils sont victimes de maladies. Leur pauvreté constitue un handicap et une menace, tant pour eux que pour les régions les plus prospères. »)<sup>2</sup>, la question de l'aide au développement a toujours été un sujet majeur des débats portant sur l'efficacité de l'aide comme facteur de croissance pour les pays pauvres. Le renvoi dos à dos de Sachs d'un côté et d'Easterley et Moyo de l'autre, par les tenants des ECR est à situer dans ce contexte. L'analyse des justificateurs de l'aide, dont en particulier Jeffrey Sachs, qui, avec Sen ont été les inspirateurs des objectifs du millénaire pour le développement (OMD), s'inscrit dans le sillage de l'économie du don de François Perroux. Ses justifications reposent sur une argumentation qui repère l'insuffisance de l'économie marchande à provoquer le démarrage réussi d'une croissance soutenue et durable du fait, même, que les marchés sont souvent imparfaits pour entraîner les pays en développement vers le progrès à cause des difficultés à produire les effets cumulatifs chers à Myrdal. Le plaidoyer en faveur de l'aide au développement se situe, également, dans le cadre général de la thèse de Leibenstein, pour qui, le retard de développement s'explique par la croissance retardée. Ce retard correspond à une situation de « quasi-équilibre » d'où on ne peut sortir que par un effort minimum critique (*critical minimum effort*). Cette thèse ne s'éloigne guère de la thèse du « *big push* », accroissant le rôle potentiel des externalités pécuniaires du développement, développée par Rosenstein-Rodan.

---

<sup>2</sup> Discours du 20 Janvier 1949.

Comme précisé précédemment, ces notions évoquent, en creux, la notion de trappe de pauvreté que développe Jeffrey Sachs. Pour cet auteur, la pauvreté, qui, en soi, est la cause de la stagnation économique, ajoute au problème essentiel du sous-développement le fait que la pauvreté, elle-même, peut être un piège.

Lorsque la pauvreté est extrême, les pauvres n'ont pas la capacité - par eux-mêmes - de se sortir du pétrin. Ainsi, lorsque l'on tient compte de la pauvreté causée par le manque de capital par personne, on s'aperçoit que les villages ruraux démunis vont manquer de camions, de routes pavées, de générateurs d'électricité, de canaux d'irrigation, etc. Le capital humain est très faible, avec des villageois affamés, malades et illettrés luttant pour leur survie. Le capital naturel est épuisé ; les arbres abattus et les nutriments du sol épuisés. Dans ces conditions, cette figure de proue de l'aide au développement, considère qu'il est nécessaire de pouvoir disposer de plus de capital-physique, humain, naturel et que cela exige davantage d'économies. Lorsque les populations sont totalement démunies, elles ont besoin de la totalité de leurs revenus, voire davantage, pour survivre et, partant, il n'y a pas de marge de revenu supérieure à la survie qui puisse être investie pour l'avenir (Sachs, 2005).

Pour Sachs, l'unique moyen de se sortir de ce piège est de consentir un puissant coup de main qui permettrait aux pays pauvres de s'affranchir de l'ornière du sous-développement et d'être mis sur le chemin de la prospérité déclenchant, ce faisant, le processus de développement cumulatif. Une trajectoire qui ne pourrait passer, nécessairement, que par l'utilisation d'une fraction des richesses du monde développé, à même de guérir une planète divisée, et qui mettrait un terme aux souffrances de ceux qui sont toujours piégés par la pauvreté. Ainsi, le plaidoyer devient-il messianique et sa finalité est de créer le lien d'humanité, de sécurité et d'objectifs communs entre les cultures et les régions.

Dans ce défi posé à « [sa] génération », Sachs porte l'ambition au niveau qui permettrait à une génération de sortir de la zone de pauvreté pour que les suivantes prospèrent. Partant, il propose de porter l'aide au développement de 65 milliards de dollars, son niveau de 2002 à 195 milliards pour 2015, période et moyens nécessaires pour permettre aux pays en développement d'emprunter les chemins

vertueux de la croissance. Ces montants représenteraient, d'après ses estimations, l'équivalent de 0,54 % du PIB total des nations les plus riches et se situeraient à un niveau moindre que les 0,7 % avancés dans le cadre de l'aide promise de façon récurrente, depuis que cette proportion fut suggérée par Jan Tinbergen dès 1972 et réitérée régulièrement depuis le sommet du G8 de 2005.

Les contempteurs de l'aide au développement se situent plutôt dans la lignée de ce que F. Perroux appelle l'économie marchande inscrite dans la pure tradition libérale depuis, au moins, Peter Bauer (1971) qui considérait que l'aide est l'excellent moyen de transférer l'argent des pauvres des pays riches vers les riches des pays pauvres, la performance étant, selon lui, essentiellement fonction des aptitudes et des comportements de la population et de la qualité des institutions sociales et politiques. Lesquels déterminants ou facteurs expliquent, dans une large mesure, les différences observées au niveau des résultats économiques et des taux d'amélioration des conditions matérielles.

William Easterley se positionne contre le « mythe » de l'aide au développement, porté par Jeffrey Sachs et qu'il critique sévèrement : « Les pays les plus pauvres sont prisonniers d'un piège à pauvreté (ils sont pauvres uniquement parce qu'ils l'ont toujours été) dont ils ne peuvent sortir sans un plan massif financé par l'aide internationale, plan constitué d'investissements et de mesures à même d'effacer tout ce qui empêchait jusque-là le développement ; après quoi, ces pays connaîtront un décollage menant à une croissance autonome, et l'aide internationale deviendra, ensuite, inutile » (2009, p. 53). Tel est le mythe « ... qui a donné naissance à l'aide au développement dans les années 1950 et qui continue d'être invoqué aujourd'hui pour prôner une augmentation massive de l'aide internationale ». Pour cet ancien économiste de la Banque mondiale, les plans n'atteindront jamais leurs objectifs parce que les « planistes » qu'il oppose aux « essayeurs » multiplient « indéfiniment les ressources pour atteindre les objectifs qu'ils se sont fixés... et devant l'échec d'une intervention, ont tendance à mettre la barre encore plus haut pour l'intervention suivante » (2009, p. 22) alors que les essayeurs posent la bonne question : que peut faire l'aide internationale pour les pauvres, ces derniers

s'intéressant plutôt à ce qui fonctionne réellement dans le système de l'aide étrangère ? Dambisa Moyo se situe sur la même ligne de crête pour considérer que « l'aide n'est pas la solution, au contraire, elle est une partie du problème, « en fait l'aide est le problème » (2009, p. 47). C'est un « désastre total » qui contribue à rendre les pauvres encore plus pauvres, se lamente-t-elle.

Il apparaît donc clairement que lorsque Esther Duflo et Banerjee posent la question du *comment dépenser l'argent* et non *combien faut-il en dépenser*, ils se situent résolument dans le sillage d'Easterley qui oppose les planistes, au rang desquels il place Sachs et tous les « bureaucrates » des organismes internationaux d'aide, aux essayeurs qui cherchent à trouver des solutions concrètes, et dans lesquels il situe, en creux, les expérimentateurs. Cette évocation n'étonne guère compte tenu des exemples que cite le premier à l'appui de son argumentaire (moustiquaires, scolarisation, accès à l'eau potable, etc.). De même, en écho aux analyses très critiques de W. Easterley sur les effets de l'aide, un groupe de travail, composé de plusieurs économistes, dont en particulier E. Duflo et F. Bourguignon, a produit, avec l'appui des fondations Bill et Melinda Gates et William et Flora Hewlett, le rapport « *When will we ever learn? Improving lives through impact evaluation* »<sup>3</sup>, rapport qui postule qu'après avoir dépensé des centaines de milliards de dollars dans le domaine du développement, notamment dans les programmes sociaux, les bailleurs, les ONG et les gouvernements des pays en développement n'auraient qu'une connaissance très limitée des impacts de la plupart de leurs programmes. Le dépassement de ces limites de connaissance exige de connaître précisément ce qui marche selon les bons mots de Bill Gates co-président de la fondation éponyme « *Success depends on knowing what works* ». Ainsi, le rapport en question conclut-il : « pour améliorer nos connaissances sur ce qui fonctionne dans les programmes de développement social, des évaluations d'impact rigoureuses sont nécessaires. Les méthodes appropriées pour de telles études doivent refléter la question de

---

<sup>3</sup> A signaler que Bourguignon et Gerler ont émis des réserves sur ce rapport en recommandant notamment « Que toute proposition de création d'un nouvel organisme aux fins de l'allocation de fonds garantisse que l'organisme augmente effectivement le montant total des fonds disponibles pour l'évaluation et n'augmente pas les coûts de transaction dans l'allocation de ces fonds. En particulier, nous restons préoccupés par la création d'une nouvelle bureaucratie ayant ses propres intérêts dans la gestion des fonds qui devraient être dirigés vers des biens publics mondiaux » (p. 44 du rapport).

politique particulière posée et le contexte dans lequel le programme est mis en œuvre. Les méthodes d'assignation aléatoire se sont révélées être une approche réalisable et rigoureuse de l'évaluation de l'impact dans de nombreuses situations et doivent donc être encouragées et promues le cas échéant ». En écho, Duflo et Banerjee (2011, p. 353) renchérissent : « les politiques les mieux intentionnées et les mieux conçues peuvent n'avoir aucun effet si elles ne sont pas mises en œuvre correctement ». Cette diatribe, selon la qualification de Judith Favereau (2016, p. 121), contre l'économie politique caractérisée par les travaux d'Acemoglu et Robinson révèle le peu d'importance accordée par les partisans des ECR à plus de soixante ans de théories économiques du développement en centrant leurs investigations à l'échelle locale en négligeant les dimensions institutionnelles et macrosociales de la pauvreté. Acemoglu et Robinson (2012) évoquent les soubassements théoriques qui font le lien entre les institutions économiques, les politiques inclusives et la prospérité. Ils considèrent que les institutions économiques inclusives qui font respecter les droits de propriété créent des conditions équitables et encouragent les investissements dans les nouvelles technologies. Ce faisant, les compétences sont plus propices à la croissance économique et en tout cas mieux que ne le sont les institutions économiques extractives davantage structurées pour extraire des ressources, bénéfiques pour le plus grand nombre, par un petit nombre. Ces institutions ne protègent pas les droits et n'encouragent pas non plus l'activité économique. Les institutions politiques inclusives distribuent, quant à elles et largement, le pouvoir politique de manière pluraliste qui permet d'atteindre un certain degré de centralisation politique qui a la capacité d'établir la loi et l'ordre, de promouvoir les bases de droits de propriété sûrs et l'économie de marché inclusive. Ils ajoutent, par ailleurs, que les institutions économiques extractives sont liées de façon synergique aux institutions politiques extractives, qui concentrent le pouvoir dans les mains de quelques personnes, qui seront alors incitées à maintenir et à développer les institutions économiques extractives à leur seul profit en utilisant les ressources qu'elles obtiennent pour consolider leur « *hold on political power* ». Pour les expérimentateurs randomisateurs, il n'est nul besoin d'attendre la chute des gouvernements ni



qu'arrivent de profondes transformations pour mettre en œuvre de meilleures pratiques. [La] réflexion attentive et [les] évaluations rigoureuses peuvent aider à concevoir des systèmes qui permettent de juguler aussi bien la corruption que de faire face à l'inefficacité (Duflo et Banerjee, 2011), l'accumulation de petits changements peut provoquer une révolution douce et crédible.

Que sont donc les micro-fondations de cette révolution de la crédibilité ?

## **Section 2 : Micro-fondations empiriques d'évaluations des politiques publiques**

Les micro-fondations empiriques d'évaluations des politiques publiques reposent sur deux thèmes qui concernent, d'une part, des données de terrain fortement localisées et, d'autre part, une réflexion approfondie sur les incitations et les comportements. Le premier concerne le type de preuves utilisées dans l'élaboration des politiques et le second questionne le procédé par lequel les preuves sont utilisées pour traiter. Sur le premier, la méthodologie met un poids important sur les preuves issues d'essais contrôlés randomisés (ECR). Dans le second cas, elle met l'accent sur les petites réformes politiques marginales dans les environnements existants, une « révolution silencieuse » d'améliorations minimales mais sûres. Cette approche intellectuelle de la recherche de la vérité, à partir d'observations concrètes, fait de la pratique l'unique critère de cette vérité. Ils s'inscrivent, par ailleurs, au sein d'une discipline marquée récemment et plus fortement par la progression de l'empirie qui cherche à puiser dans le terrain les résultats prouvés valides, aspirant à façonner la **manière** dont doivent désormais être menées les sciences économiques, et ainsi établir une *hégémonie gênante*, selon l'expression de Jean Michel Servet (2018).

L'objectif des expériences randomisées est d'aider à identifier les impacts causaux des interventions par la répartition purement aléatoire des participants de façon à éliminer le problème du « biais de sélection », selon lequel ceux qui sont choisis pour participer au programme, et que l'on appelle groupe de traitement, ont des valeurs de résultats contrefactuels pertinentes par rapport à ceux qui ne bénéficient pas du traitement et forment, ainsi, ce que l'on appelle le groupe témoin. La difficulté d'isoler cette inférence causale tient au fait que l'on ne peut, en même

temps, appliquer et ne pas appliquer le programme d'intervention aux mêmes personnes précisément, du fait de l'impossibilité d'observer le contrefactuel. L'estimation de cet impact causal s'obtient par la comparaison des résultats (outcomes) obtenus par les personnes ayant participé au programme à ceux qu'elles auraient obtenu si elles n'y avaient pas participé. Toute différence statistique significative entre les deux groupes peut être attribuée au programme en toute confiance (Duflo, 2005). Si la recherche évaluative *ex-post* a pour visée d'évaluer si une politique ou un programme particulier a produit ou non les résultats escomptés, celle menée *ex-ante* vise, plutôt, à exposer de façon critique les fondements conceptuels et empiriques de l'élaboration des politiques et dire pourquoi certains efforts réussissent, d'autres non. C'est la raison pour laquelle les promoteurs des expériences plaident pour les évaluations *ex-ante* à l'effet de dire ce qui marche et ce qui ne marche pas. Ils admettent, cependant, que d'autres techniques d'évaluations permettent d'imiter le contrefactuel en s'attaquant à la question fondamentale de l'impossibilité d'observer un même groupe dans deux conditions différentes et en même temps. Aussi, le travail consiste-t-il à comparer différents groupes de personnes : le groupe qui reçoit le programme (dit groupe de traitement) et celui qui ne reçoit pas le programme (dit groupe de comparaison ou groupe témoin). Parce qu'il n'y a pas de réplique exacte de traitement spécifique pour un individu dans le groupe de comparaison, on vise globalement à avoir des caractéristiques moyennes similaires dans chaque groupe. Le groupe de comparaison servant à dire ce qui se passerait en l'absence de « traitement ».

Rachel Glennerster, Kudzai Takavarasha (2013) et Esther Duflo (2005) décrivent, par le menu, ces méthodes qui tentent de trouver un groupe de contrôle qui soit aussi comparable que possible au groupe de traitement, tout au moins pour ses dimensions observables. Ces auteurs en pointent les vertus et les limites pour souligner, par contraste, les mérites des ECR, comparés à toute autre méthode, par la capacité intrinsèque à éliminer les biais de sélection. Nous examinerons ci-dessous les méthodologies mobilisées en matière d'évaluation qui constituent un coffre à outils détaillant les prescriptions portant sur leur pertinence, leurs modes d'organisation ainsi que leurs visées.



### *i- Les évaluations qualitatives d'impact*

Décrites comme probablement les plus évidentes pour comprendre comment certaines politiques publiques affectent les gens en leur parlant directement, les évaluations qualitatives d'impact disposent d'un large éventail d'outils pour obtenir des participants l'impact d'une intervention publique sur leur vie. Ainsi, ces entretiens ouverts ont-ils cet avantage de permettre aux participants de soulever des problèmes auxquels les évaluateurs n'auraient peut-être jamais pensé et qu'ils n'auraient pas soulevés dans un sondage fermé. Ces approches présentent également l'avantage de recueillir une information riche que les approches quantitatives traduisent en quelques chiffres dans l'ordre d'appliquer des techniques statistiques en vue d'appréhender les différences dans les deux groupes. Cependant cette traduction est souvent à l'origine de perte de quelques données capitales. Au demeurant, certaines évaluations quantitatives utilisent également et de plus en plus des techniques qualitatives en vue de développer des mesures quantitatives non issues de sondages pour apprécier des résultats.

Pour ces évaluations d'ordre qualitatif, il est assumé que le groupe de comparaison est le groupe lui-même, le vécu d'avant la participation au programme va servir comme contrefactuel alors même, qu'en général, le propre des évaluations qualitatives est qu'elles ne cherchent pas à tirer des conclusions à propos des impacts des programmes ou de politiques publiques, mais ont pour principal objectif la description des situations investiguées.

### *ii- Les comparaisons d'avant et d'après*

Cette technique indique comment un programme d'action publique change ou s'améliore au fil du temps et comment il affecte ceux qui y sont exposés. Méthodologiquement, il convient donc d'enregistrer, avant l'application de l'intervention, les paramètres caractérisant les participants et les utiliser comme « *Baseline Outcomes* » qu'il convient de comparer à leurs résultats après l'intervention « *Endline Outcomes* ». Il en ressort, de la même façon, que pour les évaluations d'ordre qualitatif, le contrefactuel est constitué des participants d'avant l'application du programme. Cependant, lorsque les résultats du préprogramme sont

utilisés comme hypothèse contrefactuelle, il est supposé que les résultats du groupe de traitement seraient restés les mêmes sans le programme et que par conséquent celui-ci est le seul facteur influant sur les changements dans les résultats mesurés au fil du temps. La difficulté provient du fait qu'il est souvent difficile de démêler tous les nombreux changements qui se produisent dans la société par rapport à ceux induits par le programme. Ainsi, plusieurs facteurs peuvent agir comme les changements simultanés dans l'environnement externe, parfois les changements chez les individus eux-mêmes, et enfin les effets de rebond similaires à ce que l'on retrouve dans les communautés touchées par la guerre, la sécheresse ou le chômage. La comparaison avant-après risque d'attribuer tout rebond au programme.

### *iii- Comparaison de différence dans la différence*

Cette approche combine la comparaison avant-après avec celle qui implique participants et non participants. Les changements intervenus s'obtiennent au fil du temps en comparant les résultats chez les participants à ceux qui surviennent pour les non participants. Le contrefactuel pour cette méthode est composé des personnes qui n'ont pas participé au programme mais sur lesquelles des données ont été collectées. Elle traite du problème des différences non observables entre le traitement et les groupes de traitement et de comparaison, dès lors que ces différences ont le même effet sur les résultats au fil du temps. Pour ce faire, il faudra disposer de séries temporelles longues pour les années antérieures à la mise en place du programme pour pouvoir comparer les tendances sur de longues périodes.

### *iv- Méthodes d'appariement (statistical matching)*

Le concept de la méthode est simple, : pour chaque agent du groupe de traitement, on apparie un élément (ou plusieurs) du groupe de contrôle. Le choix du matching est dicté par les caractéristiques observables. Les estimations par matching imputent le résultat potentiel manquant en utilisant seulement les individus les plus proches du groupe de contrôle. L'usage est de combiner matching et régression. La principale faiblesse de la méthode provient de l'obligation de construire des groupes équilibrés dans leurs variables observables nécessitant une quantité importante d'informations. Par ailleurs, il n'existe pas de variables omises qui pourraient être

la source de biais, les applications utilisant les méthodes d'appariement ne s'intéressent en général qu'à l'effet du traitement sur les traités permettant, ainsi, de peser sur la condition de sélection et partant sur les observables.

#### v- *Modèle de discontinuité des régressions*

Définie par Campbell (1969), cette approche est utilisée pour les programmes qui présentent des effets de seuil. On compare les participants qui sont juste au-dessus aux non participants qui se trouvent juste en dessous. Pour illustrer les domaines d'application du modèle de discontinuité, Campbell évoque l'exemple d'un programme portant sur l'attribution de bourses d'études au mérite et capacités où les effets de seuil introduisent ce type de discontinuité. Duflo mobilise l'étude d'Angrist et Lavy (1999) pour évaluer l'impact de la taille des classes en Israël, où un second enseignant est affecté chaque fois que l'effectif d'une classe dépasse quarante. La règle de quarante élèves est utilisée ici pour construire des estimations de variables instrumentales des effets de la taille des classes sur les résultats aux tests. La stratégie d'identification qui en résulte peut être considérée comme une application du plan de régression-discontinuité de Campbell pour la question de la taille de la classe. Pour cette approche, le contrefactuel est constitué des non participants dont le score est juste en dessous du seuil. L'avantage du modèle de discontinuité des régressions est que les règles d'allocation des programmes étant bien précisées, on se contente de vérifier qu'elles ont été bien respectées et l'on craint moins qu'un facteur de l'allocation n'ait été comptabilisé.

#### vi- *Les variables instrumentales*

L'utilisation de cette approche intervient lorsque la participation à un programme peut être prédite par un facteur accessoire ou variable instrumentale. Cette variable doit être fortement corrélée au traitement (endogène) tout en étant exogène dans le sens où l'instrument doit influencer uniquement le traitement, ce que l'on appelle restriction exclusive (exclusive restriction). Souvent, la randomisation constitue elle-même cet instrument pour identifier l'effet causal. La variable instrumentale peut également être d'ordre géographique, un facteur de distance, d'effets fixes matérialisant des chocs économiques ou tout simplement lié à des facteurs

démographiques. Le contrefactuel est constitué des personnes moins susceptibles de participer ou de bénéficier du programme en raison d'un facteur presque aléatoire. Deux exemples phare illustrent l'utilisation de cette approche : pour établir le lien entre les vers intestinaux et l'absentéisme scolaire dans l'ouest du Kenya, Miguel et Kremer (2004) utilisent des médicaments vermifuges comme traitement, l'instrument variable est supposé être la randomisation (voir plus loin au chapitre III). La colonisation comme instrument variable est donnée comme exemple par Acemoglu et al. (2001, pp. 1369-1401) dans l'établissement du lien entre la qualité des institutions (traitement) et le développement économique (outcome) par l'utilisation du taux de mortalité causée par la prévalence de la malaria.

L'avantage de l'approche des variables instrumentales est qu'elle permet d'estimer l'impact causal d'un traitement entaché de biais de sélection en utilisant un instrument prédictif qui permet de se libérer de ce biais.

Toutes ces évaluations ont pour objectif de répondre aux questions contrefactuelles essentielles suivantes : *comment l'individu qui n'a pas bénéficié d'une politique publique, aussi localisée soit-elle, aurait-il évolué en l'absence du programme y afférent ?* Et, inversement : *comment celui qui n'en a pas bénéficié aurait évolué s'il y avait été exposé ?* (Duflo, 2005 p. 188). Pour les randomistas comme les appelle A. Deaton, il arrive souvent que les règles ne soient pas appliquées de manière stricte dans les pays en développement comme pour le cas des discontinuités où des tolérances par rapport aux seuils sont permises. En conséquence, pour ces promoteurs de cette micro-économie du développement, ces techniques sont sujettes à de « larges biais » et conduisent soit à des surestimations soit à des sous-estimations des effets des programmes.

De même, est-il considéré que les estimations économétriques diffèrent significativement des résultats expérimentaux (Duflo, 2005 p. 192) puisque, contrairement à une évaluation expérimentale, les estimations non expérimentales de l'effet de formation dépendent essentiellement de la façon dont les équations de gains et de participation sont spécifiées. Si le modèle économétrique est énoncé

correctement, les estimations non expérimentales doivent être identiques (avec une erreur d'échantillonnage) à l'effet d'entraînement généré par les données expérimentales, mais s'il existe une différence entre les estimations non expérimentales et expérimentales, le modèle économétrique est par conséquent mal spécifié (Lalonde). Les autres approches n'échappent pas à la critique des chercheurs du J-PAL et leurs associés. Il convient, cependant, de relever que la littérature du J-PAL destinée à usage interne, est moins acerbe que les propos d'Esther Duflo dont la production intellectuelle intègre souvent des préoccupations médiatiques et pour qui les méthodes d'évaluations susmentionnées sont jugées non exemptes de défauts dans l'estimation des impacts pour être considérées, tout à fait, aptes à produire de l'inférence causale. Il n'en demeure pas moins que cette même littérature, à usage interne, va dans le même sens pour mettre en exergue la supériorité des RCT (Glennerster et al., 2013). Le jugement de cette école de pensée et d'action est unanime pour considérer que :

- 1) Les évaluations d'impact qualitatives ont l'inconvénient de travailler avec l'hypothèse que les priorités des chercheurs n'influencent pas la manière dont les informations sont synthétisées et que ce sont les bénéficiaires et/ou les chercheurs qui peuvent dire quel aurait été le scénario contrefactuel.
- 2) Les comparaisons avant-après supposent que tout changement au fil du temps n'est dû qu'au programme, ce qui n'est pas garanti.
- 3) Les comparaisons de différence dans les différences corrigent les différences de niveaux de départ entre participants et non-participants et comparent les changements dans le temps entre les deux groupes. Cependant, ils supposent, qu'en l'absence du programme, il n'y aurait eu aucune différence dans les changements dans le temps entre les participants et les non-participants.
- 4) Les méthodes par appariement présentent la faiblesse en ce qu'elles nécessitent des groupes équilibrés dans leurs variables observables et par conséquent nécessitent une quantité importante d'informations et de s'assurer qu'il n'y a pas de variables omises qui pourraient être source de biais. (Heckman et al., 1997) ont suggéré une association entre la méthode de différences dans les différences

et les méthodes d'appariement afin de garantir que seuls des individus comparables dans le groupe de contrôle soient pris en compte dans l'estimation de l'effet du traitement.

- 5) La discontinuité de la régression ne peut être utilisée que lorsqu'un programme a des règles d'éligibilité claires qui créent une ligne de démarcation nette entre les participants et les non-participants. Elle compare ceux situés juste en dessous et ceux situés juste au-dessus du seuil et suppose que le taux de participation est la seule chose qui compte pour franchir brusquement le seuil.
- 6) Les variables instrumentales exigent un facteur qui ne soit pas sujet à biais de sélection. Elles prédisent la participation mais n'affectent les résultats qu'en raison qu'elles prédisent précisément cette participation (endogénéité). La variable instrumentale ne peut affecter aucune autre chose d'importance par aucun autre canal (exogénéité).

Pour ces chercheurs/experts, seule la randomisation a la capacité de produire un contrefactuel valide qui, par ses caractéristiques, est jugé, en moyenne, identique au groupe qui subit l'intervention. Le fait que le groupe de traitement et le groupe de contrôle soient sélectionnés aléatoirement est bien ce qui assure l'absence, entre eux, de différences systématiques. Cette assignation aléatoire n'est pas à confondre avec l'échantillonnage aléatoire. Celui-ci n'est utile que dans la mesure où il permet, par sa fonction, de sélectionner les individus formant un groupe représentatif de l'ensemble de la population. Souvent d'ailleurs, pour sélectionner le groupe participant et celui de comparaison l'assignation aléatoire est opérée au sein de ce groupe représentatif. L'échantillonnage aléatoire peut aussi être utilisé pour décider où effectuer l'évaluation en premier lieu, de sorte que les résultats soient représentatifs d'une population plus large, à l'instar de l'utilisation de l'échantillonnage maître, bien connu des instituts de statistiques ou de sondages, pour les enquêtes nationales auprès des ménages.

Le *Journal of the Royal Society of Medicine* (2003) rapporte dans son éditorial que la première expérience contrôlée fut introduite il y a environ 250 ans par James Lind, un chirurgien de la marine écossaise, qui publiait ensuite son « *Traité du*



*scorbut* » (Tröhler, 2006). Il s'était engagé à fonder ses travaux sur des « faits observables » plutôt que sur les théories de la prise de décisions médicales de son époque, en menant son expérience par « empirisme contrôlé ». Son essai a réussi car l'un des six remèdes qu'il a administrés contenait de la vitamine C, et ce en procédant à une expérimentation comparative bien contrôlée en termes de temps et d'environnement. Pour ce faire, il avait sélectionné pour l'expérience douze marins qui se trouvaient tous à un stade clinique similaire de la maladie et il s'est arrangé pour qu'ils reçoivent le même régime alimentaire de base et soient logés dans la même partie du navire. L'examen systématique et l'expérience de Lind suggèrent que les oranges et les citrons sont plus efficaces que d'autres traitements du scorbut. Ce fait daté apparaît comme la première expérience introduisant le concept de contrôle en médecine (Glennerster et al., 2013).

La nécessité de la randomisation dans la conception expérimentale a été énoncée pour la première fois par R. A. Fisher, statisticien et généticien, en 1925, dans l'ouvrage « *Statistical Methods for Research Workers* ». Les conceptions antérieures étaient systématiques et impliquaient le jugement de l'expérimentateur pouvant conduire à la production de biais possible et/ou à une interprétation inexacte des données. Dans ces essais agronomiques à la *Rothamsted Experimental Station*, Fisher utilise la randomisation pour éliminer les biais de sélection et obtenir un test de signification valable. Avec l'aide de William Sealy Gosset, Fischer a développé ses concepts de randomisation en tenant compte, des outils des *mathematics of small samples*. Il a, depuis lors, publié de nombreux articles et ses principes de conception expérimentale ont été diffusés dans le monde entier par les nombreux « travailleurs bénévoles » issus d'autres institutions, venus à la station d'agriculture de Rothamsted pour apprendre les méthodes de Fisher (Nancy Hall, 2007).

C'est en médecine en particulier que l'essai contrôlé randomisé ou *randomized controlled trial* (RCT) trouvera son plus grand écho et y jouera un rôle prépondérant en contribuant à sa popularisation, à travers les dizaines de milliers d'essais cliniques randomisés lancés qui, souvent, combinés à des méta-analyses, visent à

mettre également en commun les résultats de nombreuses petites études en appliquant une ou plusieurs techniques statistiques (Olkin, 1995). Ioannidis (2007) met en garde contre les différences significatives sur le plan statistique pour des effets insignifiants et recommande de combiner les résultats statistiques aux méta-analyses à faible biais.

La première expérience aléatoire contrôlée en médecine, publiée en 1948, généralement attribuée à Bradford Hill, a concerné le *Traitement de la tuberculose pulmonaire au moyen de la Streptomycine*. Ce traitement a, depuis, ouvert la voie à ce que l'on a appelé la médecine fondée sur les preuves (EBM). Ces expériences considérées comme faisant partie des meilleurs standard ou *gold standard* ont le pouvoir d'évaluer les effets bénéfiques ou néfastes d'approches thérapeutiques. Duflo et Kremer (2004, p. 32) revendiquent la filiation avec cette médecine et affirment que les essais randomisés ont le potentiel de révolutionner les politiques sociales au vingt-unième siècle à l'instar de ce qui fut fait par ces RCT pour les produits pharmaceutiques au vingtième.

Dans l'article susmentionné, les auteurs observent que les évaluations d'impact crédibles sont des biens publics mondiaux et, qu'en ce sens, ils peuvent fournir des indications fiables au profit des organisations internationales, des gouvernements, des donateurs et des ONG bien au-delà des frontières nationales. Considérant qu'un déséquilibre existe au détriment des évaluations aléatoires au profit d'autres méthodes d'évaluations jugées « des modes » (*fads*) ces auteurs plaident en faveur de leur méthode, seule à produire des preuves et constituer, ainsi, des biens publics mondiaux !

Il est à noter qu'à un moment donné, observer qu'un individu est exposé ou non au programme et comparer la même personne au fil du temps, ne donnera pas, dans la plupart des cas, une estimation fiable de l'impact du programme sur cet individu. Dès lors, la question essentielle du contrefactuel à poser est : comment se seraient comportées les personnes ayant participé au programme en l'absence de celui-ci et comment se seraient-elles comportées s'elles n'y étaient pas exposées ? L'objectif essentiel de l'évaluation d'impact, par conséquent, est d'établir un groupe de



comparaison crédible qui, en l'absence du programme, aurait eu des résultats similaires à ceux des personnes exposées au programme. Ce groupe devrait donner l'idée de ce qu'il serait advenu aux membres du groupe programme s'ils n'avaient pas été exposés, permettant ainsi d'obtenir l'estimation de l'impact moyen pour le groupe en question. Aussi, le processus de randomisation doit-il observer scrupuleusement un certain nombre d'étapes pour être valide. Ainsi, lorsque les programmes relèvent de l'action de la lutte contre la pauvreté, il ne doit concerner que les ménages situés au-dessous d'un certain seuil de revenus, qu'il s'agisse d'individus, d'écoles ou de dispensaires et c'est cette population qu'il faudrait identifier avant d'opérer la randomisation. Les guides méthodologiques afférents recommandent parfois de conduire un recensement auprès de la population considérée ou par exemple auprès des écoles susceptibles d'être concernées car, argue-t-on souvent, les données administratives peuvent ne pas exister ou lorsqu'elles existent, peuvent être de mauvaise qualité.

La question qui importe, dès lors, est de pouvoir opérer l'assignation aléatoire à partir d'un nombre assez large d'unités concernées, de sorte que les groupes de traitement et témoin puissent être statistiquement identiques. L'avantage serait qu'en plus des caractéristiques observables comme les résultats par exemple, les caractéristiques inobservables comme la motivation, que l'on ne peut mesurer, soient équilibrées entre les deux groupes.

Les guides du J-PAL distinguent trois aspects lorsque l'on conduit la randomisation : l'accès, le timing et l'encouragement à la participation au programme d'intervention. L'accès est alors réservé aux personnes choisies pour accéder au programme car l'on ne peut toujours « traiter » une plus large population. Le timing d'accès détermine le moment de permettre cet accès à l'intervention puisque l'on peut varier dans le temps l'exposition à l'intervention, la différence est alors créée par ce délai d'accès au programme. Enfin, l'encouragement concerne la population qui bénéficiera de l'accès et c'est cette motivation qui peut constituer le programme où une incapacité est constatée dans l'application de la randomisation à la population elle-même. Ces trois aspects sont contrôlés et c'est pourquoi doivent-ils être alloués aléatoirement (Glennester et Takavarasha, 2013).

Les opportunités de randomisation s’offrent lorsque généralement les développeurs (ONG, bailleurs de fonds) souhaitent concevoir une nouvelle approche pour soumettre une action publique à une évaluation *ex-ante*. Il arrive, cependant, comme c’est généralement le cas dans les pays en développement, que le nombre de personnes susceptibles de bénéficier de l’action soit plus grand que ce que ne permettent les ressources allouées. Dans un tel cas, la randomisation apparaîtrait alors comme le moyen le plus équitable pour choisir les participants. On parle dans ce cas de phénomène de sursouscription. A l’inverse, lorsque le nombre de participants potentiels est inférieur à ce que pourraient permettre les ressources allouées au programme, il y a sous-souscription et l’on use souvent d’incitations pour encourager à participer. Cependant, pour les programmes non populaires telle que la conscription, l’usage de la loterie est recommandé comme le moyen le plus approprié.

Lors de la conception de l’évaluation randomisée, la décision de randomiser des individus dans un programme ou hors programme ou de randomiser des groupes entiers (ménages, communautés, écoles, cliniques ou autres unités) dans le programme ou hors de celui-ci est déterminée, en principe, par l’unité pour laquelle le programme est conçu. Il arrive cependant que les interventions puissent être opérées à différents niveaux où, par exemple, que certains organismes de microcrédit accordent des prêts à travers des groupes de prêts. La randomisation s’effectue dans ces cas au niveau individuel, des groupes de prêts de la zone géographique concernée ou par agent de crédit.

Un certain nombre de considérations doivent être prises en compte quant au niveau où s’opère l’assignation aléatoire et concernent :

- i- L’unité au niveau de laquelle le résultat sera mesuré ;
- ii- Les retombées dans la mesure où les effets des programmes peuvent être directs ou indirects et pouvant utiliser divers canaux de transmission : physiques, comportementaux ou par effets d’équilibre général ;
- iii- Les effets d’attrition où des individus sélectionnés dans les groupes de traitement s’y soustraient ;

- iv- Les problèmes de conformité (compliance) qui proviennent d'applications non conformes aux protocoles convenus ;
- v- Le pouvoir statistique et le niveau de randomisation où, plus le nombre d'unités est grand, plus les résultats statistiques peuvent être significatifs mais il arrive souvent que lorsque le niveau de la randomisation s'effectue au niveaux d'écoles ou de villages plutôt qu'aux niveaux individuels, la trop petite taille des échantillons conduit trop souvent à rejeter l'hypothèse d'un effet positif (ou négatif) du programme expérimental, parce que l'écart-type est trop large pour aboutir à une conclusion suffisamment fiable (Fougère, 2012, p. 44) ;
- vi- Enfin, les questions d'ordre éthique, qui sont traitées plus loin dans ce travail, peuvent être sérieusement problématiques.

Tenir compte de ces considérations c'est agir de manière que le groupe de contrôle soit orienté vers le programme dans une moindre proportion que le groupe à traiter. C'est ce qui « permet de transporter le débat de la méthode vers le fond et ouvre les discussions sur l'interprétation plutôt que sur les hypothèses techniques » (Zamora, 2011, p. 178).

Parmi les controverses soulevées par ces discussions sur l'interprétation émergent, en particulier, les questions relatives aux comportements spécifiques imputables aux pauvres et les incitations prodiguées pour l'encouragement à la participation. Des questions traduites par les propos suivants : « Lorsque nous apprenons si les pauvres sont prêts à payer des moustiquaires, et s'ils les utilisent quand ils les reçoivent gratuitement, nous apprenons bien plus que la meilleure façon de distribuer des moustiquaires : nous commençons à comprendre comment les pauvres prennent des décisions. Une première question est ainsi de savoir ce qui fait obstacle à l'adoption plus générale des moustiquaires. Cela peut s'expliquer par un défaut d'information quant à leur utilité ou par le fait qu'elles sont trop chères pour les pauvres. Peut-être, les pauvres sont-ils si préoccupés par des problèmes immédiats qu'ils n'ont pas l'espace mental nécessaire pour s'intéresser au futur ou peut-être est-ce tout à fait autre chose qui est en jeu. Répondre à ces questions nous permet de comprendre ce que les pauvres ont de spécifique – si tel est effectivement le cas. Vivent-ils comme n'importe qui d'autre, à ceci près qu'ils ont moins

d'argent, ou y a-t-il quelque chose de fondamentalement différent dans le fait de vivre dans l'extrême pauvreté ? Et si cette situation a effectivement quelque chose de spécifique, est-ce susceptible d'empêcher les pauvres de sortir de la pauvreté ? » (Duflo et Banerjee, 2012, p. 29).

Il en ressort que les pauvres seraient une catégorie sociale nécessitant d'être « traitée » non comme une catégorie de citoyens ordinaires, pour laquelle les politiques publiques visent à traiter les hétérogénéités de destin et de revenus, mais comme une population à part. Les promoteurs des ECR mettent en cause les politiques anti-pauvreté et imputent leurs échecs à l'incompréhension inadéquate de la pauvreté. Pour eux, les pauvres souffrent de biais cognitifs, ne réfléchissent pas comme les riches, qu'il s'agisse de nutrition, de santé, d'éducation, du microcrédit, de l'épargne, voire de l'entrepreneuriat. Trop souvent, les controverses liées à la lutte contre la pauvreté se perdent en polémiques opposant ceux qui plaident « pour donner plus d'argent aux plus pauvres » à ceux pour qui « l'aide étrangère tue le développement ».

Pour Duflo, la prise de conscience que la pauvreté modifie l'ensemble des options qui s'offrent aux individus a fait naître l'économie du développement et, quand bien même le « décideur » est « néo-classique », la pauvreté affecte les comportements (Duflo, 2003). Avec Banerjee (2012, p. 14), ils « mettent en lumière la richesse du savoir économique que l'on peut tirer de la compréhension de la vie des pauvres » en proposant une économie issue de la compréhension de leurs conditions et aiderait à comprendre ce que les pauvres sont capables d'accomplir et où, et pour quelle raison, ils auraient besoin d'un coup de pouce. Lagneau-Ymonet (2013, p. 75) appelle cet angle d'analyse un « confinement dans l'individu de l'analyse des causes de son comportement » par la manière dont les randomistas appréhendent les aspects essentiels de la vie familiale des pauvres ; ce qu'ils font au sujet de la scolarisation de leurs enfants, de leur propre santé ou de celle de leurs enfants et même la taille de la famille qu'elles se choisissent.

Ces présupposés amènent une construction des micro-fondements des RCT sur l'intérêt à porter à la vie et aux choix des pauvres au sein desquels il faudrait chercher comment lutter contre la pauvreté dans le monde. L'économie politique ne

donnera pas, par exemple, les clés pour savoir que la microfinance est utile sans, toutefois, être le miracle que certains décrivent. Il faudra aussi comprendre pourquoi les pauvres recourent à des soins de santé qui leur font plus de mal que de bien et pourquoi, année après année, les enfants des pauvres vont à l'école sans rien apprendre.

Sur la base d'une étude expérimentale sur l'adoption d'engrais à base de maïs au Kenya, Duflo et Kremer (2004), cités par Mendola (2007), affirment que les théories économiques néoclassiques ne suffisent pas à expliquer le comportement des ménages agricoles. Les résultats préliminaires avaient bien montré qu'une technologie simple et efficace ne diffuse pas et que l'apprentissage social est lent, même lorsque le passage à la nouvelle technologie est subventionné et que l'accès au crédit n'est plus un obstacle. Seuls les dispositifs d'engagement semblent contribuer à augmenter le taux d'adoption. Cela signifie que le taux d'adoption augmente uniquement lorsque les ménages perçoivent le programme comme une obligation et non comme un droit.

Lorsque les choix impliquent la subsistance de la famille, les compromis sont faussés de manière particulière (notamment pour les individus) et, dans une certaine mesure, les pressions exercées par la famille élargie ou les voisins peuvent également exercer une influence (Duflo, 2003). Selon les économistes du comportement, les communautés rurales ne recherchent pas toujours les meilleures options car elles sont contraintes par des normes psychologiques et/ou sociales. Ainsi, l'idée d'une prise de décision individuelle basée sur une rationalité limitée et non sur un simple intérêt personnel et des coûts de décisions psychologiques, constitue alors une meilleure explication au comportement des ménages dans les économies à faibles revenus.

Aussi, et quand bien même la lutte contre la pauvreté est une tâche peu aisée, les randomistes suggèrent ceci : une aide bien ciblée ne l'est que par une information appropriée ou un coup de pouce ou *Nudge* popularisé par Thaler et Sunstein dans l'ouvrage *La méthode douce pour inspirer la bonne décision* qui constitue la référence incontournable sur le sujet. Au-delà de l'apport scientifique et théorique sur le *coup de pouce*, il comporte une visée politique pour un nouveau type de *paternalisme libertaire* que Frenkiel (2009) appelle paternalisme bienveillant

tendant à influencer les comportements en refusant les contraintes et les interdictions.

La théorie économique classique repose sur deux hypothèses : un agent économique, consommateur ou producteur, cherche à maximiser sa satisfaction ou son profit tout en cherchant à minimiser ses coûts ou ses contraintes comme le décrit Simon (1972) dans sa critique de la théorie de l'homo-œconomicus substantivement rationnel : « Rationality denotes a style of behavior that is appropriate to the achievement of given goals, within the limits imposed by given conditions and constraints » (p. 161). Avec sa théorie de la rationalité limitée, Simon, constatant le profond décalage avec l'homo-œconomicus décrit par l'économie standard, développe un modèle alternatif de prise de décision en l'imputant à la rationalité limitée (Ferrière, 2011, pp. 37-45).

Bien que les promoteurs de la micro-économie du développement basée sur l'expérimentation aléatoire ne se réfèrent que rarement à cette économie comportementale, ne serait vrai que ce qui fonctionne et pour le savoir, il faudrait l'expérimenter (Servet, 2018). Schultz (1964), cité par (Duflo, 2003, Mendola, 2007 et Sergot, 2009), évoque le cas du paysan pauvre mais efficient, qui est, cependant, bloqué par des marchés défaillants et incomplets et par des politiques défavorables face auxquels il fait du mieux qu'il peut. Selon lui, il est rationnel pour un paysan pauvre de ne pas chercher à améliorer son sort car les effets générés par d'éventuels investissements sont aléatoires et les coûts de production lui échappent en partie (coûts des intrants). Son statut le conduit à pondérer très fortement le risque de perdre le peu qu'il possède. Dès lors, il est rationnel de ne pas chercher à faire évoluer les techniques culturales ou à étendre les champs cultivés, de ne pas utiliser des outils performants et, ainsi, rester pauvre. La logique de raisonnement peut être étendue à tous les pauvres vivant dans les PED, qu'ils soient ou non agriculteurs : tenter de s'extraire de la pauvreté est aléatoire et les coûts attachés à cette tentative, comme les risques encourus, sont très élevés. Pour Schultz, l'attitude du paysan pauvre n'est, ni plus ni moins, que de la gestion efficiente des moyens disponibles : le pauvre est rationnel et ses décisions conduisent à une situation optimale (donc efficace) compte tenu des contraintes qui pèsent sur ses choix (d'Agostino et Duvert, 2008).



Dans un article intitulé « Poor but rational » Duflo (2006) s'appuie sur des études menées en Côte d'Ivoire pour apporter des explications de nature à saper l'espoir de trouver une explication au comportement des ménages dans le cadre du « *pauvre mais néoclassique* ». Pour cette leader du J-PAL, bien qu'il soit admis que les pauvres sont confrontés à des environnements fort risqués (climat incertain, mauvaises récoltes, volatilité des prix, etc.) et que cela rende l'assurance à la fois plus précieuse et plus facile à mettre en œuvre, des biais qui ne sont pas toujours rationnels empêchent les pauvres de s'assurer mutuellement. Ainsi, elle rapporte que même dans un environnement très soudé (le « village ») où la menace de couper l'assurance d'une personne qui se fait prendre à tricher est très forte, où l'information circule facilement et où les gens ont la possibilité d'exercer des sanctions les uns contre les autres s'ils se rendent coupables d'abus du système, l'institution villageoise ne constitue guère le terrain d'essai fertile pour l'hypothèse « pauvre mais efficace ». Au sein de la famille même, elle donne l'exemple de la culture de l'igname, traditionnellement cultivée uniquement par les hommes (certaines opérations spécifiques ne peuvent pas être réalisées par les femmes, et les droits de propriété sur les champs d'igname sont clairement attribués aux hommes). Une norme sociale forte limite l'utilisation légitime du produit des ignames pour nourrir les membres du ménage et s'occuper des enfants.

Le groupe cultivant l'igname fut traité comme un groupe distinct et il a été examiné comment les dépenses pour divers biens réagissaient à la variation du revenu issu de cette culture. Les bonnes années, sa récolte est en effet associée à une augmentation des dépenses en nourriture, achetée et consommée à la maison. Les dépenses en biens privés (alcool, tabac, vêtements ornementaux et bijoux) ne réagissent nullement à l'augmentation ou à la diminution des revenus issus de cette culture. Les dépenses d'éducation ne réagissent qu'à l'évolution du revenu de l'igname, et non à celle du revenu des femmes ou des hommes. Les hommes semblent traiter différemment les revenus provenant de différentes sources et n'utilisent pas la manne d'une source pour compenser tout manque à gagner provenant d'une autre source. Le ménage semble donc tenir des « comptes mentaux » séparés (Thaler, 1994, p. 188), traitant différemment les différents types

de revenus, conclut Duflo en constatant que les comptes réagissent différemment aux chocs observables du revenu et il est difficile de concilier avec une observabilité imparfaite, un aléa moral ou des limites aux régimes d'assurance qui s'imposent d'eux-mêmes. Ainsi, la complexité des arrangements de partage intra-ménage semble-t-elle résister aux explications fondées uniquement sur l'information et les incitations.

Dans le même élan de démonstration, Duflo évoque l'exemple du travail d'une ONG kenyane qui promeut auprès de paysans pauvres l'utilisation de fertilisants. Contrairement à la thèse de Schultz sur la rationalité du « pauvre efficient » elle avance, comme raison de la non-utilisation des engrais par les paysans pauvres, le manque de savoir-faire nécessaire à leur usage en dépit de la modicité de leur prix. Pour elle, ce n'est que par des mesures adaptées comme le paiement juste au moment des récoltes en numéraire ou en nature, qu'il a été démontré qu'un fort accroissement a été enregistré. L'incohérence temporelle est due à la procrastination, comportement caractérisant particulièrement les pauvres.

Mieux, comprendre les pauvres requiert, pour les randomisateurs, une approche méthodologique qui emprunte à l'économie comportementale et intègre des éléments psychologiques dans l'analyse des comportements. Sans même revendiquer une affinité particulière avec l'économie comportementale, ces promoteurs des RCT subissent l'influence de cette dernière. Les retombées des travaux de V. Smith “ *pour avoir fait de l'expérience en laboratoire un instrument d'analyse économique empirique, en particulier dans l'étude de différentes structures de marché* ” et D. Kahneman “ *pour avoir introduit en sciences économiques des acquis de la recherche en psychologie, en particulier concernant les jugements et les décisions en incertitude* ”, et de Thaler en 2017 pour avoir “ *montré comment des caractéristiques humaines affectent les décisions individuelles et les orientations des marchés* ” sur les expérimentations aléatoires en économie sont d'autant plus datés que l'un des trois fondateurs du J-PAL aux côtés de Duflo et Banerjee n'est autre que S. Mullainathan, co-auteur avec Richard Thaler de l'article séminal “ Behavioral economics ” (2000). L'approche qualifiée de *paternalisme libertarien* s'appuie sur la critique du choix individuel et l'efficience



du modèle de l'agent rationnel. Sa visée est de doter les individus d'une plus grande responsabilité en résolvant les problèmes à partir de l'analyse des biais cognitifs. Plutôt que de confier cette responsabilité à des structures contraignantes et défaillantes, la marge de choix devant être laissée aux individus. Les remèdes proposés prennent toujours pour hypothèses la réduction des défaillances consubstantielles à la prise de décisions individuelles, l'environnement social n'étant pris en compte que parce qu'il accentue l'effet des biais comportementaux. Pour ces économistes comportementalistes, selon la qualification de Servet (2018), les pauvres sont plus exposés aux biais cognitifs que les autres en raison de l'absence de politiques pro-pauvres et de structures collectives qui prennent les décisions à leur place.

Aussi, appliquer fidèlement les théories conçues pour les pays développés à l'analyse des décisions des pauvres dans les pays en développement relève-t-il de la même erreur d'appréciation que celle commise par les partisans du cadre « pauvres mais efficaces ». Ainsi, essayer de réduire le comportement d'un agriculteur kényan qui n'utilise pas d'engrais et celui d'un employé américain sur le même modèle peut être aussi infructueux que d'essayer de se convaincre que les agriculteurs guatémaltèques sont à la limite de l'efficacité. Être pauvre affecte presque certainement la façon dont les gens pensent et décident. Peut-être que lorsque les choix impliquent la subsistance de la famille, les compromis sont faussés différemment de la question de savoir combien d'argent on aura à la retraite. La pression exercée par les membres de la famille élargie ou les voisins est également plus forte lorsqu'ils risquent de mourir de faim. Ou bien, la prise de décision peut être simplement influencée par le stress (Duflo, 2003).

Pour ce courant, il conviendrait de mettre en place de façon paternaliste des services et des biens en faveur des pauvres avec des modalités de mise en place qui ne fassent pas reposer sur eux la décision et la responsabilité de se les procurer. On estime, en effet, que si les biais cognitifs affectent l'ensemble des individus, les personnes en situation de pauvreté y sont plus particulièrement exposées ; l'absence de structure adaptées ne permet guère d'annuler les effets de ces biais. Ce constat est largement partagé par des figures importantes des RCT: "We propose that the behavioral patterns of the poor may be neither perfectly calculating nor especially deviant.

Rather, the poor may exhibit basic weaknesses and biases that are similar to those of people from other walks of life, except that in poverty, there are narrow margins for error, and the same behaviors often manifest themselves in more pronounced ways can lead to worse outcomes” (Bertrand, Mullainathan et Shafir, 2006, p. 8).

La lutte contre la pauvreté s’analyse, dès lors, à travers les mécanismes qui ne partent pas du point de vue de ceux qui en sont l’objet mais bien de ceux qui les conçoivent (Élisa Chelle, 2012). Elle se construit notamment à partir de *conditionnalité comportementale* qui relève de « composé de solutions technocratiques, réformistes et libérales... loin d’être simplement techniques, répondent d’une orientation politique » (p. 15) et nécessite un travail conséquent de dépolitisation qui passe, selon Chelle, par la *neutralisation de l’arène publique*. Le modèle de la *conditionnalité comportementale* tend à remplacer la logique économique issue de la responsabilité collective par une organisation pratique des politiques sociales conditionnelles se diffusant grâce à des *savoirs pratiques* qui renforcent *l’autonomie institutionnalisée* en invitant les pauvres à se gouverner par eux-mêmes (p. 147). C’est ainsi que des zones spécifiques sont choisies pour expérimenter les dispositifs censés orienter les conduites à travers des structures réunissant chercheurs, société civile et bailleurs de fonds alliés dans un souci de coopération.

Ainsi, agir pour que l’achat d’engrais s’effectue juste après la récolte, offrir un kilo de lentilles pour inciter les mères à vacciner leurs enfants ou bien installer un distributeur de chlore à côté d’un point d’eau et conduire des expérimentations capables d’identifier les effets causaux du « traitement » relèvent, donc, de cette gouvernance de la pauvreté. Éric Durant et Charlotte Nordman (2011, p. 29) évoquent un « réductionnisme microéconomique » qui occulte les conflits d’intérêt et la domination à l’origine de la pauvreté.

De l’examen des buts et des visées des travaux du J-PAL, le laboratoire du MIT s’est nettement distingué par une critique de la macroéconomie du développement appréciée non pas par les déficits de croissance ou des politiques macroéconomiques et macrosociales déficientes mais par son incapacité à opérer des changements à la marge. Les micro-fondations empiriques d’évaluation des

politiques publiques sur lesquelles reposent les expérimentations aléatoires contrôlées sont les seules à même de « rendre aux pauvres la lutte contre la pauvreté... en leur donnant les moyens de l'autonomie » par l'action appropriée sur les biais cognitifs qui les pénalisent et, partant, en inférant des causalités appréciées comme dépendances causales.

## Chapitre II : La méthode expérimentale, protocoles d'observation et terrains

La crainte de ne pouvoir déterminer de façon crédible les impacts causaux des interventions justifie l'utilisation des expériences aléatoires contrôlées comme réponse aux problèmes généralisés d'identification faible en économie empirique, en raison principalement des doutes quant à l'efficacité du contrôle des variables confusionnelles dans les régressions fondées sur des données d'observation (Imbens et Wooldridge, 2008). L'objectif de ces expérimentations est d'utiliser l'assignation aléatoire aux fins de s'assurer que l'hypothèse essentielle à l'identification d'un effet de politique publique se vérifie .

La force présumée de la randomisation par l'élimination de tout biais de sélection est sa capacité (*gold standard*) à identifier des effets des causes et de fournir les explications causales afférentes. Cette recherche causale bien ancrée, aussi bien, dans la tradition scientifique que dans le débat philosophique depuis, au moins, Hume où dans « *Enquête sur l'entendement humain* » (1748) considère qu'entre deux évènements, l'un est cause et l'autre est effet lorsque l'on observe une succession temporelle d'évènements contigus dans l'espace où le premier, la cause, engendre le suivant, l'effet.

Il revient, cependant, à David Lewis (1973) de considérer qu'une chaîne causale peut être considérée en termes de dépendance contrefactuelle : C cause E si et seulement si, si C n'avait pas eu lieu, alors E n'aurait pas eu lieu. Pour ce faire, Lewis propose d'interpréter les énoncés contrefactuels dans les termes des mondes possibles considérés tout aussi réels et que celui dans lequel nous vivons dans la mesure où, bien qu'ils ne se soient pas produits, ils sont tout aussi concrets que ceux que nous observons (Courgeau, 1997).

Le fondement de cette causalité repose sur des imputations de nature empirique inférant un lien causal issu de deux évènements successifs définis en termes d'expérience sensible marquée par une succession d'évènements supposés produire des effets reproductibles. L'affiliation des RCT avec la médecine est censée opérer un transfert de scientificité à la méthodologie (Labrousse, 2010).

Nous passerons en revue le modèle de base comme dépendance contrefactuelle pour inférer les causalités recherchées (section 1) et procéderons ensuite à une critique de la causalité inférée par le modèle de causalité établi, précisément, comme dépendance contrefactuelle (section 2). En section 3, nous passerons en revue quelques expérimentations aléatoires randomisées promues par leurs auteurs comme exemples de production de la preuve.

## **Section 1 : Le modèle causal de base**

Le modèle de base utilisé, issu d'un transfert du domaine biomédical à l'évaluation des politiques publiques, participe d'une démarche d'identification des conditions causales. La pierre angulaire en est la formulation des résultats potentiels (outcomes potentiels), esquissée par Neyman (1923) et mise au point par Rubin (1974) et Holland (1986), qui consiste à examiner l'efficacité d'une intervention à l'aide de procédures contrefactuelles. L'identification de l'effet causal repose sur l'estimation de l'impact d'un traitement  $D$  sur une variable de réponse  $Y$  pour une population  $U$  pour :  $i = 1, 2, \dots, N$ .

En supposant que le traitement est binaire, chaque unité de la population est associée à un couple de résultats  $y_{1i}$  et  $y_{0i}$  qui expriment respectivement la valeur de la variable  $y$  en présence de traitement ( $x=0$ ) et en son absence ( $x=1$ ). La réponse de chaque individu s'écrit sous la forme :

$$d_i y_{1i} + (1 - d_i) y_{0i}$$

Les participants à l'expérimentation peuvent être décrits par l'ensemble

$$(y_{0i}, y_{1i}, d_i, x_i, \mathcal{E}_i).$$

Les termes  $x_i$  et  $\mathcal{E}_i$  représentent respectivement les attributs spécifiques des unités et les caractéristiques non observables.

La population est segmentée en un groupement de traitement,  $B = \{(i \in U | d_i = 1)\}$  et un groupe de contrôle.  $N = \{(i \in U | d_i = 0)\}$

L'effet du traitement individuel noté  $\delta$  est mesuré par :

$$\delta = y_{1i} - y_{0i}$$

Comme tel, il ne peut être identifié que parce que les réponses sont conditionnelles à deux conditions d'expérimentation qui ne sont pas passibles d'observations simultanées. Une même unité ne peut appartenir aux deux groupes des participants. Pour échapper à cet écueil de l'inférence pointée par Holland, le recours à la dépendance contrefactuelle est nécessaire. La recherche de l'effet causal s'appuie sur la distribution de l'impact de l'intervention sur la population à partir de l'effet moyen du traitement (*Average Treatment Effect* : ATE) :

$$\delta = E(y_1 - y_0) = E$$

Cette différence représente l'espérance du résultat potentiel et celui où l'individu n'est pas soumis à l'intervention. La manifestation d'un effet causal présuppose la condition :  $E$

L'estimation du changement induit par la réponse d'un individu participant s'obtient par la différence des moyennes observées :

$$\delta = \frac{1}{n_b} \sum_{i \in B} y_i - \frac{1}{n_n} \sum_{i \in N} y_i = (\bar{y} | i \in B - (\bar{y} | i \in N))$$

$n_b$  est le nombre d'unités de l'échantillon traité et  $n_n$  celui de l'autre échantillon. Cet estimateur est affecté d'un biais de sélection dans le cas où le couple de caractéristiques  $(x_i, \mathcal{E}_i)$  entre les deux sous-populations est différent.

L'existence d'une différence à priori a pour conséquence :

$$E(y_1 - y_0) \neq E$$

La détermination de ces paramètres d'intérêt requiert la réunion d'hypothèses identificatrices.

H1 : l'indépendance statistique des résultats potentiels vis-à-vis de l'accès au traitement,  $(y_0, y_1 \perp d)$ , implique que chaque individu a la même probabilité de recevoir le traitement. Cette propriété permet d'éliminer le biais de sélection.

ATE devient :

$$\delta = E(y_1 - y_0) = E(y_1) - E(y_0) = E$$

L'effet moyen du traitement sur les traités est, quant à lui, donné par :  $\delta^B = ATT = E(y_1 - y_0/d = 1) = E$

L'identification de ce second paramètre appelle une hypothèse moins forte : il suffit que les résultats  $y_0$  soient indépendants en moyenne de  $d$ . Dans ce cas, il vient :

$$E(y_0/d = 1) = E(y_0/d = 0)$$

Les deux paramètres d'intérêt sont égaux sous les deux conditions afférentes à chacun et l'estimateur n'est autre que la différence des moyennes observées des deux groupes :

$$E(y_1 - y_0/d = 1) = E$$

H2 : Les résultats du traitement doivent être constants dans le temps :

$$Y^B = Y_i^B, Y^N = Y_i^N$$

Pour tout  $i$  et  $t$ .

H3 : Cette invariance temporelle est associée à l'homogénéité des unités que l'effet produit par une intervention sur unité  $i$  est égal à celui correspondant à l'individu  $k$  :

$$Y_i^B = Y_k^B, Y_i^N = Y_k^N$$



H 4 : Cet impact est supposé être le même sur chaque unité de la sous-population traitée. Comme tel, il implique que les effets moyens sont égaux :

$$\text{soit } Y_i^B - Y^N = Y_K^B - Y_K^N$$

Ces réquisits, qui découlent de l'isolement de l'impact d'une variable sur une autre, vont de pair avec l'hypothèse que l'effet produit sur l'échantillon témoin ne se répercute pas sur les bénéficiaires. Ce faisceau d'hypothèses écarte la variabilité intra-individuelle et entre individus en vertu de la clause *ceteris paribus*. L'identification comprend, en définitive, deux moments : la génération des données expérimentales et l'estimation de l'effet causal par une mise en variable centrée sur la densité des observations.

## **Section 2 : La causalité comme dépendance contrefactuelle**

Pour Paulré, c'est Hume qui serait à l'origine de l'approche contrefactuelle de la causalité et cite, en appui, ce passage de *l'enquête sur l'entendement humain* : « On peut définir la cause comme un objet, suivi par un autre, et où tous les objets, semblables au premier, sont suivis par tous les objets semblables au second. En d'autres mots, où, si le premier n'avait pas existé, le second n'aurait jamais existé » (2013, p. 10). Pour les historiens, cette approche désigne une forme particulière pour dire ce que serait le récit historique si la réalité avait été différente. Se demander, par exemple, si la première guerre mondiale aurait eu lieu sans l'assassinat de François-Ferdinand d'Autriche à Sarajevo ou que la seconde guerre n'aurait pas eu lieu, non plus, si Hitler avait péri durant la première guerre, c'est élever au rang d'objets de recherche de plein droit que l'histoire contrefactuelle permet d'étudier à partir de données d'archives et de preuves d'autres types (Reiss, 2012, pp. 113-127). Lewis, dans son approche des mondes possibles, soulève l'hypothèse qui consiste à considérer que le passé aurait pu être différent si un acteur avait agi autrement ou si un facteur historique serait survenu autrement. Plus précisément, ce que l'on aurait pu faire et que l'on n'a pas pu faire, quelqu'un d'autre, d'un autre monde qui nous ressemble aurait pu le faire. Ce que Lewis traduit par la métaphore du Kangourou qui fonde le conditionnel contrefactuel : si les

kangourous n'avaient pas de queue, ils se renverseraient, cependant, il existerait un monde qui ressemblerait à l'état du monde réel et où ceux, même sans queue, se renversent. Lewis développe encore davantage cette notion contrefactuelle dans *De la pluralité des mondes* (2007), quand il envisage les possibles en s'affranchissant des lois, de causalités et leurs corollaires de régularités par le rapport d'accessibilité logique aux mondes possibles.

Ce genre d'uchronie, décrite dans un genre littéraire que Renouvier (1876) présente par une écriture contrefactuelle de l'histoire, sous la forme d'une utopie des temps passés mêlant fiction et ambition philosophique pour déboucher sur une approche apocryphe du développement de la civilisation européenne non tel qu'il a été mais tel qu'il aurait pu être (David, 2015). Par la démarche contrefactuelle, sa visée consistait à démontrer le rôle que joue la liberté face au « fatalisme historique, le déterminisme, le rôle de la prédestination et la providence » (Deluermoz et Singaravélou, 2012, p. 72).

Aux côtés de l'histoire où l'usage du contrefactuel fut le plus mobilisé, d'autres sciences sociales y ont eu recours de manière explicitée et théorisée comme le droit, les sciences politiques, voire l'économie.

Ainsi, en matière de droit, le raisonnement contrefactuel a-t-il consisté à mobiliser deux formes de raisonnement contrefactuel. En premier lieu, la règle du *sine qua non* avec pour objet la détermination d'un facteur particulier pour connaître s'il pourrait être déterminant à l'avènement donné. En second lieu, la règle issue du processus de la fiction juridique qui partage avec la dépendance juridique la question du « si » et celle de « si ne pas » (Le Guennec et al., 2017) à l'instar de Cynus de Pistoia pour qui : « La fiction prend pour vrai ce qui est certainement contraire au vrai » (Thomas, 2007, p. 128).

En sciences politiques, l'exemple du déclenchement de la première guerre mondiale, met en exergue les liens avec l'histoire dans la mesure où la détermination des chaînes causales par le raisonnement contrefactuel tend à expliquer le temps présent ou spéculer sur le développement des relations internationales.

La *New Economic History*, école s'appuyant sur les méthodes économétriques et statistiques et l'application de la théorie économique, a largement fait à appel à la notion de contrefactualité à l'effet d'appréhender les interactions non linéaires de manière à dégager des liens entre des événements sans qu'il existe des contiguïtés dans le temps et l'espace. (Deluermoz et al., 2012, p. 75).

Ainsi, Fogel (1964) cité par (Reiss, 2012 ; Livet, 2012 ; Deleurmoz et al., 2012) fait usage de l'analyse contrefactuelle pour évoquer le rôle des chemins de fer en faveur de la croissance économique aux États-Unis d'Amérique. En s'attaquant à ce qu'il appelle « l'axiome de l'indispensabilité », Fogel considère que les USA auraient eu le même niveau de croissance économique sans les chemins de fer ; les transports fluviaux et maritimes ayant pu constituer un substitut produisant, à peu près, le même essor économique. Cette histoire contrefactuelle refaite telle qu'elle aurait pu être n'a pas manqué de susciter des controverses par rapport au degré de vraisemblance d'événements qui ne sont pas survenus. Contre la comparaison entre les données chiffrées factuelles relatives à la réalité impactée et les données chiffrées contrefactuelles, Paul David (1969, p. 520) rétorque que le rendement des investissements dans les transports fluviaux et par canaux étaient bien moins élevés que ne l'étaient les investissements dans les chemins de fer (Livet, 2012). La principale difficulté signalée des raisonnements contrefactuels en histoire est qu'ils ne peuvent être soumis à expérimentations (Drouet et al., 2012, pp. 81-112).

Il revient cependant aux statistiques appliquées aux sciences sociales, à la psychologie et à la médecine, toutes disciplines pouvant faire l'objet d'examens expérimentaux, de recourir à un large usage de la notion de contrefactuel à l'effet de dégager des inférences causales. Neyman (1923 et 1935), Fisher (1935), Cochran et Cox (1950), Cox (1958), sont à l'origine des travaux de conception expérimentale. C'est à Donald Rubin qu'il revient, cependant, de formaliser les données d'observation dans une série d'articles, dont particulièrement, les articles de 1974 et 1990. Dans la tradition statistique, le modèle est souvent appelé cadre des résultats potentiels en référence aux rendements potentiels des travaux de Neyman en statistiques agricoles (Gelman et Meng, 2004 ; Rubin, 2005). Le modèle contrefactuel trouve également des racines dans les travaux importants de Heckman (1974, 1978, 1979, 1989, 1992, 2000) et de Manski (1995, 2003).

Le cœur du modèle contrefactuel d'analyse de données d'observation est, somme toute, assez simple. Ainsi, lorsqu'un individu d'une population d'intérêt est exposé à deux états alternatifs d'une cause, chacun de ces états se caractérise par un ensemble distinct de conditions auxquelles l'exposition peut avoir une incidence sur le résultat d'intérêt. Dans la tradition contrefactuelle, ces états causaux alternatifs sont appelés traitements alternatifs. Lorsque seulement deux traitements sont considérés, ils sont appelés traitement et contrôle. L'hypothèse clé du cadre contrefactuel est que chaque individu dans la population d'intérêt a un résultat potentiel pour chaque état de traitement, quand bien même un individu ne peut être observé que dans un seul état de traitement à un moment donné. Toutefois, on ne peut jamais observer le résultat potentiel dans l'état de traitement pour ceux observés dans l'état de contrôle et on ne peut jamais l'observer dans le résultat potentiel sous l'état de contrôle pour ceux observés dans l'état de traitement. C'est cette impossibilité qui explique le fait que l'on ne peut jamais calculer d'effets causaux au niveau individuel.

Dans son introduction largement dédiée au modèle contrefactuel et à l'inférence causale, Holland (1986) décrit ce défi comme le problème fondamental de l'inférence causale. Balayant toutes les préventions philosophiques quant à l'usage des statistiques à l'objet d'inférer des relations causales, il critique la réaction des nombreux statisticiens, qui, face à la possibilité que leur profession puisse contribuer à une discussion sur le lien de causalité, nient cette possibilité. Critique qui vise Barnard (1982, p. 387) qu'il cite : « Cette corrélation n'est pas la causalité est peut-être la première chose à dire » mais également Smith (1980, p. 998) « si les statistiques ne peuvent pas établir une relation de cause à effet, elles peuvent certainement ajouter du poids à la réalité » (Smith, 1980, p. 998). Pour Holland, cette action évasive pourrait être une réponse à tous ces petits titres insignifiants qui apparaissent dans les endroits les plus inattendus, constate-t-il (p. 945).

Holland va plus loin en soutenant qu'une expérience randomisée bien conçue peut constituer un outil puissant pour étudier les relations de cause à effet et remettre en question la nécessité d'une telle posture défensive de la part des statisticiens. Pour

cet auteur, les expériences randomisées ont transformé de nombreuses branches de la science, et les premiers partisans de telles études étaient les mêmes statisticiens qui ont fondé l'ère moderne de ce domaine. Les statistiques auraient beaucoup à dire, selon lui, sur certains problèmes d'inférence causale et devraient, même, jouer un rôle plus important dans les analyses philosophiques de la causalité qu'auparavant. Il cherche, en outre, à montrer pourquoi les modèles statistiques sont utilisés pour tirer des inférences d'association causale : « ... Premièrement, les modèles statistiques appropriés pour les inférences d'association et causales seront discutés et comparés. Ensuite, ils seront appliqués à diverses idées sur la causalité qui ont été exprimées par plusieurs auteurs sur le sujet. Une des difficultés qui se pose lorsque l'on parle de causalité est la variété des questions qui sont résumées dans le titre. Certains auteurs se concentrent sur le sens ultime de la notion de causalité. D'autres se préoccupent de déduire les causes d'un effet donné. D'autres encore souhaitent comprendre les détails des mécanismes de causalité. L'accent sera mis ici sur la mesure des effets des causes, car il semble que ce soit un endroit où les statistiques relatives à la mesure peuvent apporter une contribution. À mon avis, mettre l'accent sur les effets des causes est en soi une conséquence importante de l'intégration du raisonnement statistique dans l'analyse de la causalité et s'oppose directement aux analyses plus traditionnelles de la causalité » (p. 945).

Cet auteur considère que l'expérimentation constitue un outil scientifique et statistique puissant, introduisant souvent de la clarté dans les discussions sur des cas spécifiques de causalité, bien qu'il ne considère pas que l'expérience soit le seul cadre approprié pour en discuter. Il s'appuie sur le modèle de Rubin et explique : « Le but est de construire un modèle suffisamment complexe pour nous permettre de formaliser les intuitions de base concernant les causes et les effets. Le point focal part de l'analyse des effets de causalité donnée par Rubin (1974, 1977). Il suffira, cependant à nos fins, de traiter une version simplifiée du modèle de Rubin au niveau de la population » (p. 946). Ce modèle simplifié a été utilisé dans Holland et Rubin (1980) pour analyser l'inférence causale dans des études de cas-témoins rétrospectives utilisées dans la recherche médicale et dans Holland et Rubin (1983) pour analyser le paradoxe de Lord sur l'analyse de la covariance. Holland précise,

encore, que lorsqu'il évoque le « modèle Rubin », il sait que Rubin arguerait, lui-même, que les idées à la base du modèle existent depuis Fisher et qu'il n'en demeure pas moins que, pour lui, ce sont les idées de Rubin qui furent appliquées pour la première fois à l'étude de la causalité (p. 946).

La causalité découle de la comparaison d'une situation observée avec une situation contrefactuelle issue du milieu médical et basée sur l'approche du résultat potentiel (*Potential Outcome Approach : POA*), situation qui aurait été observée si le traitement n'avait pas été mis en œuvre. Cependant l'application de la contrefactualité aux expériences aléatoires contrôlées part d'une hypothèse qui dès le départ est non testable. En effet, si l'on applique à un individu un traitement à évaluer, sa réponse au traitement est  $Y_t(u)$ , alors que si on lui applique un traitement classique, c, appelé contrôle,  $Y_c(u)$  est sa réponse au contrôle. On mesure donc l'effet du traitement de contrôle par la différence :  $Y_t(u) - Y_c(u)$ .

Comme indiqué ci-dessus, 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 :

$$Y_t(u_1) - Y_c(u_2) .$$

De cette différence, on cherche à dégager le résultat potentiel entre le résultat que l'individu ( $u_1$ ) aurait pu obtenir s'il avait bénéficié du traitement administré  $Y_t(u_1)$  et celui qu'il aurait présenté s'il n'avait pas bénéficié du traitement  $Y_c(u_2)$ . Cet effet causal individuel introduit pour la première fois par Neyman est difficile à estimer pour une même unité d'observation. Dès lors que les unités observées sont assignées aléatoirement aux conditions d'expérience, l'effet causal moyen peut être estimé par la différence entre les espérances des valeurs de la variable-réponse conditionnellement aux deux conditions expérimentales  $x_i$  et  $x_j$  par une évaluation



de la différence (Paulré, 2003, p. 30). Les résultats potentiels ne sont jamais observés simultanément du fait que l'épreuve contrefactuelle du résultat observé n'est jamais réalisée. Théoriquement, lorsque le tirage est aléatoire puis effectué avant le début de la politique à évaluer (raison expliquant pourquoi les évaluations menées par le J-PAL sont toujours *ex-ante* et revendiquées en tant que tel) les groupes test et témoin ont statistiquement les mêmes caractéristiques observables et non observables. Ainsi, la comparaison des moyennes des résultats de ces groupes permet d'identifier l'effet du traitement dont l'objectif est d'évaluer, en conséquence, un dispositif de politique publique unique par son effet (ou par son absence d'effet), par une variable-résultat unidimensionnelle (p. 32).

Il s'avère, cependant, que deux situations fréquentes altèrent l'identité des caractéristiques observables et non observables statistiquement garanties par la randomisation. Premièrement, les effets directs des programmes objet de l'évaluation peuvent avoir des effets indirects et c'est typiquement le cas de l'expérimentation de Michael Kremer et Eduard Miguel sur les effets des vermifuges quant aux performances sur l'assiduité à l'école (Kremer et Miguel, 2004) du fait de la diminution de la contamination des élèves non traités par le bénéfice que d'autres élèves ont reçu le médicament contre les vers intestinaux.

Deuxièmement, un certain nombre de programmes peuvent générer des effets d'équilibre général, situation souvent observée pour les programmes de retour à l'emploi. Nous reviendrons dans le chapitre III sur ces aspects tout comme nous reviendrons sur les questions de conformité qui font que certains individus affectés au traitement n'y participent pas et que d'autres alloués au groupe témoin y participent.

Ainsi, les nombreux problèmes posés par la généralisation des résultats des expérimentations aléatoires contrôlées proviennent-ils du fait que leur objectif est de mesurer, uniquement, l'impact statistique tendant à prouver l'effet causal attribuable à l'intervention et non de chercher à comprendre la structure sous-jacente du système de relations générant ces résultats. En l'absence de théorie qui peut imposer des corrections, le pouvoir causal attribué pour justifier l'inférence s'en trouve affaibli, les expérimentations aléatoires contrôlées fournissent des



informations sur la moyenne des effets du traitement, mais, n'identifient pas d'autres caractéristiques de la distribution.

## **2-1 : critique de la causalité par l'approche contrefactuelle**

Prenant le contre-pied de la technique des régressions à l'œuvre en macroéconomie, l'approche du J-PAL recourt aux expérimentations contrôlées qui consistent donc à tester plusieurs programmes d'action sur des populations bénéficiaires puis à les comparer à des populations témoins en vue d'identifier des effets causaux. Elle plaide à cet égard pour une refondation microéconomique des politiques publiques qui vise à appréhender le développement dans le sillage d'une ingénierie de la lutte contre la pauvreté.

Laissant de côté pour l'instant le discours séculaire sur les récits de causalité en philosophie - on y revient plus bas - les discussions sur la causalité et les variétés de modèles causaux se sont intensifiées dans divers disciplines touchant aux domaines de l'économétrie, des statistiques, de l'épidémiologie et de la sociologie mais également dans l'informatique avec l'émergence des machines Learning et de l'intelligence artificielle. On dénombre en matière de causalité, dans les domaines susmentionnés trois approches qui dominent actuellement le débat sur l'inférence causale : les modèles d'équations structurelles, les modèles de résultats potentiels, et enfin les modèles issus des graphes acycliques dirigés. Nous nous concentrerons ici sur l'inférence causale issue des modèles d'identification des résultats potentiels utilisant un contrefactuel.

Comme évoqué ci-dessus, l'approche contrefactuelle de la causalité a à voir avec la sémantique des mondes possibles due à D. Lewis. Remontons à la citation de David Hume dans *Enquête sur l'entendement humain* rapportée précédemment. Elle est souvent présentée comme le fondement de l'analyse de la causalité : « On peut définir la cause comme un objet, suivi par un autre, et où tous les objets, semblables au premier, sont suivis par tous les objets semblables au second. Ou, en d'autres mots, si le premier objet n'avait pas existé, le second n'aurait jamais existé ». De cette définition, l'on peut tirer les constatations suivantes : premièrement, les imputations causales sont de nature empirique, le lien causal concerne deux événements successifs et est défini en termes d'expérience sensible et non en termes

d'affirmation analytique. Deuxièmement, deux propriétés principales caractérisent cette définition de la causalité : asymétrie que marque la succession temporelle des événements et régularité en ce qu'elle entraîne des effets reproductibles. Enfin, la définition n'explicite pas la nature des éléments empiriques qui mettent en évidence la cause et l'effet, et l'on ne sait pas comment s'articulent expérimentation et expérience (Bono et Trannoy, 2012, pp. 35-53). Il faudra, cependant, se garder d'évoquer la thèse essentielle de Lewis des mondes les plus proches sans prendre en compte les risques liés aux productions mécaniques de proximité qui ne génèrent pas nécessairement des déclarations causales et nettes (Kluve, 2004).

La thèse de Lewis (1973) apparaît, elle-même, reliée à l'analyse de Hume sur la causalité et l'affirmation que *si le premier objet n'avait jamais existé, le second n'aurait jamais existé* et prend la forme d'approche contrefactuelle enchâssée dans la métaphysique humienne liée à l'intuition de régularité observée marquée, alors, par une priorité temporelle, une contiguïté spatio-temporelle et une conjonction constante de composants constitutifs d'une relation de cause à effet.

Pour Dawid (2000, pp. 407-448) l'approche contrefactuelle est métaphysique et les inférences qui en résultent ne peuvent être tirées de données empiriques. Pour tirer des conclusions sur les effets probables des causes appliquées, cet auteur soutient que les arguments contrefactuels sont inutiles et même potentiellement trompeurs, dans la mesure où les inférences qui en sont issues peuvent être sujettes à un degré d'ambiguïté irréductible, l'association n'étant pas un lien de causalité. En cherchant à mettre en évidence les « causes » des phénomènes et à reconstituer les relations qui lient ces causes aux effets qu'elles entraînent, l'approche contrefactuelle se base sur une attitude que Dawid appelle un *fatalisme* dans la mesure où l'attitude en question considère les différentes réponses potentielles  $Y_i(u)$ , lorsque le traitement  $i$  est appliqué à l'unité  $u$ , comme des attributs prédéterminés de l'unité  $u$ , attendant seulement d'être retrouvés par l'expérience. Reprenant les questionnements posés par Holland (1986), Dawid attire par ailleurs l'attention sur les confusions commises par différents problèmes d'inférence causale et appelle à distinguer deux types de requête causale. Ainsi lorsque l'on s'interroge : « J'ai mal à la tête. Cela m'aiderait-il si je prends de l'aspirine ? », cette requête nécessite une inférence sur *les effets*

*des causes* ; c'est-à-dire des comparaisons entre les conséquences attendues des différentes interventions possibles dans un système. Ces questions ont longtemps été au cœur de la théorie statistique standard du plan d'expérimentation qui, soutient-il, n'a généralement guère suscité l'intérêt des analyses contrefactuelles.

Dawid donne l'exemple de la prise d'aspirine en cas de mal de tête : mon mal de tête a disparu. Est-ce parce que j'ai pris de l'aspirine ?

La question concerne, en revanche, *les causes des effets* et le lien de causalité entre le résultat déjà observé et l'intervention antérieure et c'est, précisément, ce que l'on cherche à comprendre. Les deux requêtes portent sur des circonstances singulières, les preuves directement pertinentes pour la première requête sont souvent utilisées de manière inappropriée pour adresser la seconde omettant de porter une attention particulière à la différence entre les requêtes. Pour distinguer entre ces deux types de questions, Dawid avance l'idée d'inférence prédictive considérant que l'objectif de l'expérience est d'aider à la prise de décision sur les traitements à affecter (Lecoutre, 2004, pp. 193-245). En procédant ainsi, il appelle à réduire l'ambiguïté inhérente à l'inférence sur les causes des effets en approfondissant les mécanismes cachés des unités par l'observation de variables supplémentaires concomitantes appropriées. En matière statistique, les effets des causes relèvent du passé et de l'analyse rétrospective et que la science statistique « *fait plus de bien au monde lorsqu'elle se concentre sur la mesure minutieuse des effets des causes que lorsqu'elle tente d'expliquer les causes des effets* » (Holland, 1988, p. 186). Par ailleurs, trouver des explications présuppose que les effets des causes sont les mêmes pour toutes les unités de la population (Lecoutre, 2004, p. 225). Dès lors, on se demande si la réponse observée pourrait être attribuée, après coup, à un traitement spécifique (Dawid, 2000 ; Lecoutre, 2004) et, qu'au fur et à mesure que la connaissance s'approfondit, on élabore de plus en plus de variables qui s'enchaînent pour les présenter, alors, comme les vraies causes (Holland, 1986, p. 959). Déduire une distribution commune à partir de deux distributions marginales est impossible du fait que celles-ci limitent l'utilité de la distribution conjointe. Cette limite constitue, à son tour, un handicap supplémentaire pour les RCT (Heckman et Smith, 1995).

De son côté, Deaton (2010, p. 439) considère que la différence de moyennes entre les traitements et les témoins étant une estimation de l'effet moyen du traitement parmi les personnes traitées, le traitement et les témoins ne diffèrent que par la randomisation et, de ce fait, elle ne renseigne que sur l'estimation de l'effet moyen du traitement pour tous. Une des conséquences de cette dérivation pointée par les critiques des ECR, qui n'est informative que sur la moyenne des effets du traitement  $Y_{i1} - Y_{i0}$ , est qu'elle n'identifie pas d'autres caractéristiques de la distribution. Pour lui encore, la médiane de la différence n'est pas la différence des médianes susceptible d'intéresser les décideurs politiques, en particulier par l'identification de la fraction de la population pour laquelle l'effet est positif, ce que, bien entendu, ne peut fournir l'effet moyen du traitement. En d'autres termes, l'essai pourrait révéler un effet positif moyen bien que la quasi-totalité de la population soit touchée et qu'un petit nombre d'entre elle reçoive des bénéfices très importants, situation qui ne peut être révélée par l'ECR, et pourrait être désastreuse si elle était mise en œuvre. Ravi Kanbur (2001), cité par Deaton (p. 339), fait valoir qu'une grande partie du désaccord sur la politique de développement est due à des différences de cette nature. Étant donné les hypothèses minimales qui entrent dans un ECR, il n'est pas surprenant qu'il ne puisse pas nous dire tout ce que nous aimerions savoir.

Matalon (1988) met en garde contre la confusion qui, souvent, découle de la constitution d'un échantillon aléatoire qui aspire à la généralisation pour population parente et la randomisation qui n'a pour visée que d'assurer la comparabilité. Il rappelle, à cet effet, qu'au sens statistique, un échantillon représentatif n'est pas défini par ses caractéristiques mais par la procédure utilisée pour sa construction. On verra plus loin, comment certains arrangements et bricolages rapportés altèrent les procédures utilisées pour certaines expérimentations randomisées que mènent parfois les chercheurs affiliés au J-PAL.

## **2-2 : Affinités avec les expérimentations en médecine**

Lorsque les randomistas, ainsi nommés par Deaton, revendiquent l'affinité avec la médecine comme gage de scientificité, souvent, ils occultent les controverses que la randomisation déclenche à propos de son usage dans la discipline. Il revient à Claude Bernard (1865) d'avoir jeté les bases de la médecine expérimentale dans

son essai éponyme. Cependant, Lind (1947) y avait recouru avec son essai clinique pour traiter et ensuite prévenir le scorbut qui frappait les marins. Il a été évoqué au premier chapitre l'expérience aléatoire contrôlée menée par Hill en 1948 pour le traitement de la tuberculose pulmonaire par la streptomycine, ouvrant, ainsi, la voie à ce que l'on a appelé, plus tard, la médecine fondée sur les preuves. Il convient, néanmoins, de signaler que ce n'est qu'à l'orée des années quatre-vingt que la médecine basée sur les preuves est née au sein de l'université McMaster dans l'Ontario visant, d'abord, le développement d'un enseignement fondé sur la recherche de preuve et utilisé, à la fois, comme méthode de raisonnement et source d'information faisant partie du cursus de formation des étudiants. Dès le début des années quatre-vingt-dix, l'EBM fut étendue comme méthodologie en usage par les praticiens. Un tutoriel élaboré par l'universités Descartes (2014) définit la médecine basée sur les preuves comme technique consistant à associer de manière judicieuse l'expertise clinique individuelle aux pratiques prouvées cliniquement et issues de la recherche systématique. Ainsi, expertise individuelle et pratique clinique prouvée se conjuguent pour prodiguer les soins les mieux appropriés. Cette technique se décline en quatre étapes : formulation claire et précise du problème médical à résoudre, recherche dans la littérature académique pertinente en rapport avec le problème posé, évaluation et applicabilité des résultats obtenus pour enfin intégrer les conclusions retenues aux questions posées. Aux étapes susmentionnées, il faudrait ajouter les conditions visant à formuler les hypothèses avant la collecte des résultats de l'expérimentation, à intégrer le raisonnement statistique nécessaire à la prise en compte de la variabilité de la maladie pour qu'enfin la preuve causale puisse se fonder sur la sélection aléatoire des groupes de traitement et témoin au regard de la typologie des patients en tenant compte des effets du traitement (Jaillon 2007).

La différence fondamentale entre RCT appliquées en médecine et celles appliquées en sciences sociales tient au fait que les premières reposent sur deux traitements : effectifs et placebos souvent en aveugle, voire en double aveugle. La comparaison peut également comparer les effets de deux médicaments et non seulement une médication et un effet placebo. Ainsi, chez le même individu un seul effet est observé, l'autre n'ayant qu'un statut hypothétique, le sens à donner aux résultats d'un traitement qui aurait pu être appliqué est le but de l'investigation. En sciences

sociales, le recours au raisonnement contrefactuel pour inférer les effets d'une politique publique repose sur les observations qui associent à chaque individu (traitement ou contrôle) son résultat, là où seul l'un des deux est observé, l'autre n'ayant qu'un statut hypothétique et débouchant sur des données inobservables (Lecoutre, 2004). Pour cet auteur, à l'instar de Dawid, le risque est de glisser de l'évènement potentiel à des propensions que pourrait avoir le sujet d'expérience et que celle-ci ne soit pas une donnée d'observation. Rubin, une autorité reconnue en matière d'identification et d'estimation des effets causaux des traitements qu'ils relèvent de la psychologie de l'éducation, voire de l'effet de politiques publiques, met en garde contre la tentation d'ajuster pour n'importe quelle variable présentée, autrement dit n'importe quelle estimation peut être obtenue (1974, p. 700). Il considère, par ailleurs, que si la randomisation peut être utilisée chaque fois que possible sans jamais omettre que l'utilisation de données non randomisées soigneusement contrôlées pour estimer les effets causaux est une procédure raisonnable et nécessaire (*reasonable et necessary*) dans de nombreux cas (p. 688).

La réclamation d'affiliation des RCT appliquées en microéconomie du développement avec la médecine est illustrée par la fixation de l'objectif assigné à cette méthodologie telle qu'appliquée par le J-PAL de révolutionner aux vingt-unième siècle les sciences sociales comme l'ont réalisé les essais cliniques en médecine durant le vingtième siècle. Duflo (2010, p. 17) avance l'exemple de l'industrie pharmaceutique qui procède aux tests par choix d'échantillons aléatoires pour isoler l'effet comparaison relatif à l'efficacité d'un médicament pour fonder la parenté avec une « science dure ». Ce faisant, elle déplore que les politiques relatives à la santé ou à l'éducation ne soient pas évaluées rigoureusement avant d'être généralisées et recommande de procéder à des essais pilotes pour conduire l'évaluation des programmes en s'inspirant des essais cliniques, qui font usage d'un outil riche et polyvalent apportant transparence et flexibilité nées d'une position à l'intersection du monde politique et de la recherche. La revendication de la parenté avec ce qu'il est convenu d'appeler *l'evidence based medicine* se heurte, pourtant, aux principes des pères fondateurs de cette dernière. Sackett, autorité reconnue, distingue bien ce que l'EBM est et ce qu'elle n'est pas. Pour cet auteur et ses co-auteurs (Sackett et al., 1996) l'EBM n'est pas un livre de recettes mais une approche



ascendante qui intègre les meilleures preuves externes et l'expertise clinique individuelle alliée au choix du patient. Cette approche ne se fonde aucunement sur ce qu'ils appellent une approche servile puisant dans un livre de recettes et qui indique les soins à prodiguer. Les preuves cliniques externes peuvent informer mais ne remplacent jamais l'expertise individuelle à même de décider si ces preuves externes s'appliquent ou pas pour traiter le patient. Tout au plus, indiquent-elles comment elles doivent être intégrées à la décision clinique en fonction de l'état du patient. L'EBM ne se limite pas aux essais randomisés et aux méta-analyses. Elle recherche les meilleures preuves qui déterminent l'exactitude d'un test de diagnostic et, pour ce faire, il faudra chercher dans les études transversales appropriées sur les patients cliniques soupçonnés d'héberger le trouble concerné et non par un essai randomisé. Ces auteurs poursuivent en indiquant que les preuves dont les cliniciens ont besoin sont à chercher dans les sciences fondamentales telles que la génétique ou l'immunologie. L'altération de la recherche de l'efficacité des médicaments est le fait de gestionnaires et d'acheteurs visant constamment la réduction des coûts des soins alors que le rôle du médecin est la recherche de l'intervention la plus efficace et la maximisation de la qualité et la durée de vie du patient. Cet état de fait soulève des objections de la part de cliniciens qui ont le sentiment d'être dépossédés et pour beaucoup d'entre eux administrer un placebo à un malade pose un certain nombre de soucis d'ordre déontologique (Matalon, 1988, p. 246). Nous développerons plus loin les questions d'ordre éthique soulevées par les essais cliniques. Black (1996) fait observer que l'opinion largement partagée, selon laquelle les méthodes expérimentales (essais contrôlés randomisés) sont le *gold standard*, conduit à dénigrer les méthodes non expérimentales dans la mesure où les organismes de financement de la recherche et les éditeurs de revues les rejettent automatiquement. De nombreux problèmes liés à la réalisation d'essais randomisés pourraient, en théorie, être résolus mais les implications pour les organismes de financement d'un côté et les chercheurs de l'autre font que cela n'est, souvent, pas possible.

Lorsque l'on croise les deux points de vue, on s'aperçoit que d'un côté il y a la revendication non nuancée de la supériorité des RCT comme outil riche et polyvalent par les chercheurs du J-PAL ; de l'autre on se confronte à moult



précautions chez les cliniciens qui lient la validité des résultats cliniques à de meilleures preuves pouvant répondre aux questions cliniques en puisant dans les études transversales appropriées et adaptées aux patients en fonction du trouble concerné. Lorsque Duflo évoque les RCT appliquées pour la mise sur le marché de nouveaux médicaments, elle semble faire l'impasse sur la réalité de la mise sur le marché des médicaments qui nécessite un phasage obligatoire en quatre étapes : une phase préclinique précédant le traitement de l'animal à l'homme qui définit l'effet thérapeutique du produit ; la phase qui suit concerne la première administration pour évaluer les conditions de tolérance humaine et isole les effets secondaires des effets souhaités. Cette phase confronte généralement la pratique médicale et l'évaluation du médicament à l'effet d'apprécier l'efficacité du traitement et durant cette phase sont étudiées les conditions de mesure de l'efficacité du médicament par le recours aux essais sur un petit nombre d'individus volontaires en vue de déterminer les modalités thérapeutiques optimales en termes de prescriptions et de dosage. Cette étape est décrite par Bouvenot (2006), que cite Favereau (2014), comme la phase de l'étude de l'efficacité du traitement à proprement parler par la caractérisation de l'essai clinique. En ultime étape, intervient le suivi des effets du traitement sur le long terme. Seule la troisième étape apparaît comme phase d'ordre quantitatif. Toutes les autres sont de nature qualitative. L'on s'aperçoit, alors, que le J-PAL ne transfère que l'aspect quantitatif de l'essai clinique sans jamais qu'il fût considéré dans sa fonction thérapeutique évoquée par Sackett et ses co-auteurs cités plus haut. La pharmacocinétique susmentionnée qui peut durer sept ans, voire davantage, se distingue nettement de la manière dont cette technologie est promue en microéconomie du développement.

Nous proposons d'analyser la filiation entre la médecine basée sur les preuves et les expérimentations sociales telles que pratiquées par les promoteurs de la randomisation à travers l'analogie assumée par ces derniers. Judith Favereau se référant à Robert Nadeau (1999), propose une grille d'analyse de cette analogie qu'elle transfère de la médecine à l'économie du développement : au cœur des similarités se trouve d'abord la randomisation qui permet d'assurer la comparaison entre les résultats potentiels dégagés entre groupe traité et groupe de contrôle. L'utilisation du tirage au sort en vue d'attribuer le traitement ou pas (placebo pour

la médecine) fait rapprocher l'économie de la scientificité attribuable à la médecine. Ainsi, la microéconomie du développement promue par les « randomistes » cherche-t-elle à s'ancrer dans les sciences dures (l'Horty et Petit, 2011 ; Jatteau, 2019) et opère, du coup, un transfert de scientificité (Agnès Labrousse, 2010). Vient ensuite le constat que les évaluations menées dans le cadre des deux domaines visant à mesurer les effets strictement attribuables à une intervention s'effectuent *ex-ante*. L'évaluation se situe alors au cœur même du déroulement de l'action et relève, en conséquence, d'une recherche-action « chemin faisant » pour un programme se déroulant et s'adaptant au fil de l'avancée du projet expérimental. Ainsi, la mise en circulation d'un médicament, par exemple, ou la recommandation d'une politique publique requièrent-elles de mesurer leurs effets avant leur implémentation et le cas échéant leur généralisation. Et c'est, alors, au niveau de la généralisation ou la validité externe que se posent tous les problèmes liés aux considérations de contextes, lesquels contextes recouvrent, une variété de situations où l'argument d'absence de biais de sélection n'opère pas.

Abondantes sont les études qui indiquent qu'il existe de nombreux cas d'interventions de soins de santé, en usage, auxquels s'ajoutent des interventions qui, pour la plupart, comportent de nombreux éléments. Dans une opération chirurgicale simple plusieurs actes médicaux basiques se combinent : tests préopératoires, anesthésie, approche chirurgicale, gestion des plaies, soins infirmiers postopératoires et la pratique de sortie. Il sera difficile de soumettre, même pour un nombre limité d'éléments, des évaluations expérimentales simultanément pour inférer un effet causal (Black, 1996). Tirer parti d'autres méthodes, pour tenter de combler les énormes lacunes qui existent dans les résultats expérimentaux publiés, devient alors nécessaire. Apparaissent, ensuite, les difficultés liées à la généralisation des résultats faisant que la validité externe des essais randomisés en microéconomie de développement est assez largement admise comme faible de l'aveu même de ses propres promoteurs. En vue d'y remédier, ils proposent de procéder à la réplique des expériences comme suggéré pour le cas de l'expérience menée par Cohen et Dupas (2010) au sujet de la distribution gratuite des moustiquaires (nous y reviendrons au Chapitre III). En matière de RCT, l'autre aspect de similarité entre les deux disciplines réside dans l'effet de contamination

valable aussi bien entre médecins (adoption ou rejet) ou entre patients que pour les effets d'équilibre partiel ou général que l'on observe consécutivement au déploiement d'une politique sociale. L'expérience menée au Kenya par Kremer et Miguel (2004) portant sur la distribution des vermifuges et visant à accroître la présence des enfants en classe en est une illustration. Un des aspects ayant conduit à la réduction de l'absentéisme scolaire est dû à la réduction de la transmission ou l'infection des élèves affectés par les vers intestinaux à leurs camarades sains ou aux membres immédiats de leur communauté. Cette réduction est un acquis de la médecine déployée pour soutenir des politiques publiques liées aussi bien à la santé qu'à l'éducation.

Enfin dans les deux cas (médecine et politiques économique-sociales), les expérimentations peuvent dépendre assez largement des financements liés à des donateurs qui pourraient ne pas être désintéressés. Le dynamisme de la recherche et l'engagement des acteurs concernés se mesurent à l'intensité et à l'efficacité des financements nécessaires. Les changements paradigmatiques dans ce domaine, comme dans d'autres, obéissent pour une large part à une logique d'offre aussi bien de la part des chercheurs que des acteurs impliqués.

Souvent, l'essor des évaluations qui engagent le contrefactuel comme méthode dépend largement de leur financement rendant nécessaires des structures ad-hoc de financements. Au push des chercheurs, répond le pull des bailleurs qui mettent en place des structures dédiées auxquelles se lient des réseaux de praticiens qui se spécialisent dans les évaluations d'impact.

Pamies-Sumner (2014) en dresse la liste quasi exhaustive de ces « knowledge brokers » tels le 3IE créé en 2008, spécialement dédié aux financements des programmes d'évaluations. Cet organisme bénéficie de ressources issues aussi bien du DFID britannique que de fondations caritatives comme celles de Bill et Melinda Gates ou de William et Flora Hewlett. Le travail mené par cette auteure au sein de l'agence française de développement signale une multitude de fonds dédiés aux évaluations d'impacts où organismes internationaux de financement, recherche académique, ONG et grandes fondations d'entreprises joignent leurs efforts pour

faire prospérer les RCT. Parfois, les frontières entre évaluateurs et fondations sont ténues et, ainsi, trouve-t-on Duflo siéger dans certains conseils d'administration ou au sein de réseaux d'experts engagés par certaines de ces fondations comme c'est le cas pour Danone ou Véolia. Comme nous le signalerons plus loin, le J-PAL bénéficie largement de ces concours et va jusqu'à créer une structure dédiée : l'Innovation for Poverty Action (IPA) à la fois comme puissant outil de communication et d'instrument de *fund rising*. Annie Duflo, sœur d'Esther, y assume la fonction de directrice exécutive, Karlan le président-fondateur et les principaux dirigeants du J-PAL en sont les administrateurs.

S'agissant des problématiques posées par le financement de la recherche médicale, et plus généralement, les questions posées par l'influence systémique de l'industrie pharmaceutique, Paul Scheffer détaille par le menu, (Monde Diplomatique, octobre 2015), le rôle que joue cette industrie dans l'orientation des publications qui lui sont favorables pour mobiliser ce qu'il appelle la *complicité active* des Key Opinions Leaders (KOL) engagés pour conseiller l'industrie et « éduquer » les confrères. La carrière de certains parmi ces KOL s'en trouve accélérée par l'industrie pharmaceutique qui n'hésite pas à soumettre à leurs signatures des articles qu'elle a elle-même rédigés. Cette pratique porte le nom de *ghostwriting* et qui appelle à revoir l'ensemble de la chaîne du médicament, de la recherche à la prescription en passant par la régulation sans oublier la formation des médecins. Flanagin et al. (1998) montrent, à l'issue d'une enquête menée pour mesurer la prévalence d'articles avec des auteurs prestigieux ou fantômes dans trois journaux médicaux américains à grands tirages ou forts impacts, qu'un article sur quatre ne présentait pas les bons critères d'attribution, voire une attribution inappropriée de la paternité. Ioannidis (2016) soutient que la recherche clinique en capacité de générer des preuves a disparu des radars de la plupart des bailleurs de fonds publics pour lui faire substituer le soutien de l'industrie dont les services marketing sont plus puissants que leurs services Recherche-Développement et ce n'est point pour procéder à des découvertes majeures, même à hauts risques, mais plutôt pour faire plus d'argent !

Le même auteur (2007) va encore plus loin en se demandant s'il était inévitable que les résultats de la recherche soient faux, ajoutant qu'il est impossible de connaître avec certitude la vérité dans toute question de recherche et qu'à cet égard le « *gold* » standard pur est inatteignable. Cette autorité reconnue de l'*evidence based medicine* propose des approches nouvelles pour améliorer la probabilité de l'après-étude à l'aide d'études de grande envergure ou méta- analyses à faible biais qui peuvent être utiles et se rapprocher, ainsi, de cet étalon-or. Ioannidis pense, au demeurant, qu'il est impossible d'obtenir des preuves à grande échelle pour les millions de questions posées par la recherche actuelle, les preuves à grande échelle devraient être ciblées, plutôt, sur les questions de recherche pour lesquelles la probabilité pré-étude est déjà considérablement élevée, de sorte qu'un résultat de recherche significatif puisse conduire à une probabilité post-test qui, à ce moment, pourrait être considérée comme définitive (p. 700). Les preuves à grande échelle sont particulièrement indiquées lorsqu'elles permettent de tester des concepts majeurs plutôt que des questions spécifiques et étroites. Un résultat négatif peut, alors, réfuter, non seulement une allégation spécifique proposée mais un domaine entier ou une partie considérable de celui-ci. La sélection des performances des études à grande échelle sur la base de critères étroits, comme la promotion commerciale d'un médicament spécifique, est une recherche largement inutile. Il faudrait, en plus, demeurer prudent car des études importantes peuvent révéler une différence significative sur le plan statistique mais pour un effet insignifiant. Il est intéressant, par ailleurs, de noter que la plupart des questions de recherche sont abordées par de nombreuses équipes et il peut s'avérer erroné de mettre l'accent sur les résultats statistiquement significatifs d'une seule équipe.

Banerjee, intervenant au symposium organisé par la revue indienne *Economic and Political Weekly* (2005), mentionnée plus haut, suggère de procéder à l'enregistrement préalable des études et mettre ensuite en réseau la collecte de données, pour les confronter à celles que pourrait apporter chaque expérience considérée isolément. Il s'engage à soutenir l'initiative visant à créer un site web où chaque évaluation aléatoire est enregistrée avant d'être lancée. Cette transparence garantirait, selon lui, que toutes les études (y compris celles qui échouent ou

présentent des résultats « problématiques ») soient enregistrées, les auteurs ne pouvant plus décider *ex-post* de mettre l'accent sur des résultats intéressants auxquels ils n'avaient jamais pensé au début de leur travail.

Le rôle que jouent les journaux médicaux dans la collusion entre sciences, industrie et publications et les intérêts financiers qui en découlent est souvent mis à l'index. Scheffer pointe ces conflits comme suit : « Les journaux de référence dont les quatre majeurs sont le Lancet, le New England Journal of Medicine (NEJM), le Journal of the American Medical Association (JAMA) et le British Medical Journal (BMJ) pourraient jouer un rôle de contre-pouvoir face à l'influence de l'industrie pharmaceutique dans les essais cliniques par le biais de leur sélection et de leurs relectures critiques. Cependant, ces journaux sont, eux aussi, concernés par la problématique des conflits d'intérêts. Richard Smith, qui a été rédacteur en chef du British Medical Journal y a, notamment, consacré tout un ouvrage. On y apprend que l'industrie paye les copies, appelées « reprints » des articles publiés dans ces journaux qui peuvent avoir un intérêt pour elle. L'aura de sérieux et de scientificité que ces revues confèrent est un atout dont savent se servir les visiteurs médicaux face aux médecins notamment. Les revenus issus de ces copies constituent une part non négligeable des recettes de journaux. Ces achats représentent 41 % du budget du Lancet (mais seulement 3 % du British Medical Journal qui s'avère moins dépendant), ce chiffre est sans doute encore plus important pour le New England Journal of Medicine, le journal le plus coté en termes de facteur d'impact et par conséquent le plus ciblé et le préféré des industriels, mais il n'a pas été communiqué à ce jour aux investigateurs. [Ce qui] entraîne un risque de publier des articles susceptibles d'avoir la faveur de l'industrie, en vue de générer des profits des reprints, même si leur qualité peut être douteuse » (Scheffer, 2017, p. 31). Le *gold standard*, revendiqué par la science médicale et la pharmacopée, demeure marqué par les stratégies et l'influence des industriels pharmaceutiques déployées en direction des parties- prenantes (décideurs publics, régulateurs, acheteurs et gestionnaires) sans oublier le travail mené auprès des chercheurs et des médias dont l'influence structure la formation des opinions publiques.



Nous reviendrons au chapitre V sur les publications et la rhétorique qui les accompagne pour la promotion des RCT en microéconomie du développement.

### **Section 3 : Expériences de programmes de lutte contre la pauvreté et évaluation**

Relevant d'une véritable quantophrénie, les évaluations aléatoires contrôlées embrassent un vaste champ de programmes liés à cette microéconomie du développement, principalement en matière de lutte contre la pauvreté. Le site du J-PAL en recense les principaux domaines suivants : agriculture, éducation, santé, genre, crimes, violences et conflits, firmes, marché du travail. Ainsi, plus d'un millier d'évaluations ont-elles été menées dans 84 pays, toutes réalisées par des chercheurs affiliés au laboratoire ou financés par son entremise.

Ces expérimentations reposent sur un certain nombre de postulats visant à mettre au jour des *causalités insoupçonnées* à partir de modèles d'explication des comportements et des processus de pauvreté. Pour ses promoteurs, l'efficacité des programmes sociaux serait nettement améliorée si les décisions politiques étaient fondées sur des preuves scientifiques rigoureuses, si bien, que la théorie économique s'avère être, parfois, un « mauvais guide » de choix entre des programmes concurrents visant un même objectif (Banerjee et Duflo, 2009, p. 693). Les évaluations doivent, dès lors, reposer sur les expériences de terrain car celles-ci ont « un pouvoir subversif que n'ont ni les évaluations rétrospectives, ni les expériences de terrain... leur force principale, au-delà même de la force de l'identification, [elles] sont une opportunité de faire avancer la science et la lutte contre la pauvreté (Duflo, leçon inaugurale au Collège de France, 8/01/2009).

Les deux idées simples qui gouvernent cette microéconomie du développement s'insèrent dans le paradigme consistant à considérer que la théorie peut être un « mauvais guide » et qu'il faille, par conséquent, mener des études empiriques susceptibles d'aider à concevoir les politiques visant à réduire efficacement la pauvreté. Par ailleurs, la meilleure façon de tirer des conclusions précises sur le cheminement des causes aux effets est de mener un essai randomisé contrôlé sur le terrain à même d'évaluer l'impact causal d'un programme en répondant à des questions contrefactuelles qui cherchent à savoir comment les personnes non



exposées à l'intervention se seraient comportées si elles n'avaient pas eu la possibilité d'y participer. Les mécanismes de découverte inductive proviennent du pouvoir de manipuler les traitements d'intérêt pour créer des événements qui n'ont pas été observés.

En vue de saisir la manière dont les RCT cherchent à inférer des causalités en vue d'identifier ce qui fonctionne, on s'attardera sur un certain nombre d'expériences phares, très souvent invoquées par ses promoteurs en matière de lutte contre la pauvreté (Glennerster et Takavarasha, 2013, Banerjee et Duflo 2011, Duflo, 2010).

Le guide des expérimentations aléatoires contrôlées, établi par Glennerster et Takavarasha, évoqué plus haut, donne les détails des étapes de la planification et de la réalisation de toute évaluation aléatoire contrôlée. S'appuyant sur une expérience phare menée en Inde en matière d'éducation (Banerjee et al., 2010) les détails en question explicitent les différentes phases de cette technologie, mise au service d'une recherche de nouvelles connaissances scientifiques d'utilisation pratique, selon le sens que lui confère Mario Bunge (1983). Ces phases se déroulent selon des séquences qui démarrent par l'affinement de la conception du programme à évaluer *ex-ante* et ses itérations à tester. En accord avec les partenaires, cette première phase doit déterminer les objectifs du programme et l'examen des conditions de sa généralisation dans des conditions où, souvent, les ressources allouées sont limitées. Viennent, ensuite, l'intervention à tester et son pilotage. Durant cette seconde phase, il est généralement défini un pilote pour les interventions qui visent à réaliser plusieurs objectifs : vérifier, en final, si le programme tel que conçu est réalisable et s'il a une chance raisonnable de succès ; renseigner, ensuite, les détails des voies par lesquelles le programme pourrait produire des effets, affiner les mesures des résultats intermédiaires et finaux ; et, enfin, décider de la taille de l'échantillon nécessaire à l'évaluation à grande échelle. A l'issue de ce travail, l'assignation aléatoire intervient proprement avec son lot de questionnements pour savoir, en particulier, si l'assignation se fait au niveau de l'individu ou du groupe, voir comment traiter les retombées potentielles dès l'étape de la conception, et apprécier comment stratifier et couvrir les mécanismes de la randomisation. Le plan de collecte des données, l'analyse de puissance, la mise en

œuvre du programme, la surveillance de sa bonne exécution, l'analyse des données, ce que les résultats de l'étude signifient pour les politiques, les résultats de la politique, et ce que les résultats impliquent, tous éléments faisant partie intégrante des mécanismes à l'œuvre.

Cette cinématique méthodologique, ainsi, mise au service des expérimentations aléatoires contrôlées, et variées, vise l'obtention d'effets à partir du traitement de problèmes concrets.

Ainsi, l'éducation, la santé, la gouvernance ou la finance sont-elles abordées sous l'unique primat des comportements des agents. Nous examinerons dans ce qui suit quelques exemples d'expérimentations conduites par le J-PAL ou ses partenaires/associés.

Comme nous l'avions indiqué, l'utilisation des RCT dans des contextes de politique sociale remonte au début des années soixante pour diverses expériences sur les politiques sociales américaines. Concernant les évaluations appliquées en matière de développement, la base de données 3IE recense plus de 130 RCT publiés au cours de la période 1981-1999. La première RCT de cette base de données provient d'un projet de recherche de la Banque mondiale sur les interventions éducatives (manuels scolaires et leçons radiophoniques) visant l'amélioration des résultats en mathématiques des élèves au Nicaragua (Ravallion, 2018). En la matière, parmi les RCT antérieures à 2000, on peut citer celle réalisée par le Gouvernement mexicain pour l'évaluation de Progreso (1997). Cette expérimentation a eu, par la suite, une influence considérable sur l'expansion des transferts monétaires conditionnels dans plus de 50 pays (Fiszbein et al., 2009).

Avec la création du J-PAL au tournant du nouveau millénaire, la production annuelle d'expérimentations aléatoires contrôlées a littéralement explosé. En matière d'éducation (plus de 240 selon le site du J-PAL), les RCT, telles que pratiquées par les randomistes, prennent le contre-pied de la théorie du capital humain pourtant parfaitement en ligne avec l'école néoclassique pour ce qui concerne l'arbitrage rationnel des individus en situation d'information parfaite. A l'équilibre, le salaire est égal à la productivité marginale et les différences de revenus sont expliquées par la formation. La méthodologie du J-PAL, à la fois

descriptive et prescriptive, s'écarte aussi bien des fondements théoriques du concept définissant le capital humain comme l'ensemble des compétences et de l'expérience accumulées visant à rendre les salariés plus productifs (Barro, 1991 ; Becker, 1964 ; Romer, 1986 ; Aghion et al., 2004 ; Stiglitz, 2007) que des apports théoriques et empiriques utilisant les régressions transversales (Mincer, 1974), utilisées pour établir les liens entre l'investissement en capital humain et les gains.

Les partisans de la randomisation considèrent que les rendements transversaux estimés de l'éducation mesurés à partir d'années de scolarité sont souvent biaisés en raison de la variation sous-jacente qui reflète une sélection systématique.

A l'instar de ce qui est considéré pour l'aide au développement et la critique qu'ils y adressent (chapitre I), Banerjee et Duflo (2011) opposent, en matière d'éducation, les « fournisseurs de l'offre » appelés par eux « Wallah » en Hindi, au nombre desquels on trouve Sachs, jugés interventionnistes en matière d'éducation et les « Wallah de la demande », en référence à William Easterley, qui « pensent qu'il ne sert à rien de développer l'offre de l'éducation, tant que celle-ci ne fait pas l'objet d'une demande claire » (Banerjee et al., 2011, pp. 124-128). Face à cette « tentation de l'absolu » (Duflo, 2009, p. 26), les randomistas questionnent des modalités pratiques, fixent des objectifs modestes, manipulent des objets concrets et se concentrent sur l'amélioration de la pédagogie (notamment par le recours à des enseignements complémentaires au profit d'enfants en retard par rapport au niveau scolaire considéré). Ils s'attachent, également, à améliorer la gestion des écoles et à responsabiliser les enseignants. Le but ultime est d'introduire les réformes pédagogiques qui adaptent l'enseignement aux niveaux des élèves et d'utiliser le levier des incitations comme l'embauche d'enseignants sous contrats à court terme.

Nous examinerons un certain nombre d'expérimentations en la matière qui ont été rapportées par le guide écrit par Glennerster et son co-auteur, nous ayant parues à la fois assez emblématiques et fortement médiatisées. Ainsi, en matière de cours de rattrapage, il sera question en particulier de l'expérience ayant concerné 122 écoles primaires à Vadodara et 77 à Mumbai, en Inde, menée entre 2001 et 2004 par Duflo, Cole et Linden (2007).

Prenant le contre-pied des méthodes axées sur l'amélioration de l'enseignement par l'accroissement ou l'ajout de ressources (distribution de manuels scolaires, réduction de la taille des classes, etc.), l'expérimentation consiste en l'élaboration d'un programme de rattrapage scolaire où des jeunes femmes sont embauchées pour accompagner les élèves en retard en matière d'acquisition de compétences de base en lecture, écriture et calcul. Ce programme a pour visée l'étude d'impact de deux interventions menées en collaboration avec une ONG indienne (Pratham) en évaluant l'apprentissage des élèves les plus faibles. La première intervention consistait à recruter des paras professionnels pour accompagner des élèves de troisième et de quatrième années dont le rendement scolaire est médiocre à l'extérieur de leur classe régulière. La seconde visait l'apprentissage, assisté par ordinateur, des mathématiques en y consacrant deux heures par semaine au profit d'élèves de quatrième année. Le rationnel qui a guidé ces deux expériences randomisées part du constat, établi par enquête, que 44 % des enfants indiens, âgés de sept à douze ans, n'arrivent pas à lire correctement un paragraphe basique et que, par ailleurs, un élève sur deux est incapable d'effectuer une simple soustraction.

D'après les études, les évaluations menées dans deux villes, sur une période de deux ans, concluent à l'efficacité des deux programmes : les résultats aux tests des enfants dont les écoles ont bénéficié du programme de rattrapage se sont améliorés de 0,14 écart-type en première année et de 0,28 en deuxième année. Les enfants directement touchés par le programme ont amélioré leurs résultats de 0,6 écart-type en deuxième année, alors que les enfants qui sont restés dans la classe ordinaire n'en ont pas profité. Le programme d'apprentissage assisté par ordinateur semble, lui aussi, avoir été efficace, augmentant les résultats en mathématiques de 0,36 écart-type la première année, et de 0,54 écart-type la deuxième année. L'explication des différences d'efficacité entre les deux programmes est imputée aux retombées du programme des paras professionnels qui, pourtant, sont moins bien formés que les enseignants titulaires, lesquels sont contraints par le programme officiel et manquent de temps pour aider les élèves en difficulté. L'autre raison explicative avancée est que les attitudes sociales et les préjugés communautaires peuvent limiter l'efficacité des professeurs, ces derniers ayant le sentiment de « rendre un grand service en enseignant à des enfants de communautés autrefois intouchables

ou à des migrants très pauvres ». Ces facteurs peuvent également contribuer à expliquer l'efficacité du programme d'apprentissage assisté par ordinateur, par la stimulation individuelle de chaque enfant, abstraction faite du niveau de réussite.

Enfin une évaluation économique est donnée pour justifier les deux programmes. Le programme dit Balsakhi (amies des enfants en hindi pour désigner les assistantes non titulaires) est remarquablement bon marché (les enseignantes auxiliaires sont payées 10 à 15 dollars par mois correspondant à 2,25 dollars par élève et par an). Le coût du programme d'apprentissage assisté par ordinateur est, quant à lui, de 15,18 dollars par élève et par an selon les auteurs, y compris le coût amorti sur cinq ans des ordinateurs.

L'autre étude, menée entre 2005 et 2007 par Duflo, Dupas et Kremer et intitulée « School Governance, Teacher Incentives, and Pupil-Teacher Ratios », a concerné 210 écoles primaires de l'Ouest du Kenya. Cette expérimentation intervient après la décision du gouvernement kenyan de décréter la gratuité de l'enseignement primaire avec pour conséquence l'accroissement de 30 % de la fréquentation scolaire. Beaucoup de ces nouveaux élèves étaient des primo-inscrits et n'avaient pas fréquenté auparavant l'école maternelle. Par conséquent, dans la même classe, les élèves avaient des capacités initiales très variées en lecture, en écriture et en mathématiques. Menée en collaboration avec l'International Child Support Africa, l'expérience a consisté à évaluer trois interventions dont l'objectif a été de remédier à la problématique de l'encombrement des classes et d'atténuer les différences de résultats scolaires initiaux entre élèves d'un même niveau de classe. Sur 201 écoles primaires, 140 ont été choisies au hasard pour bénéficier du Programme d'Enseignants Supplémentaires (ETP : Extra Teacher Provision) qui a permis de financer l'embauche d'un enseignant contractuel local afin de pallier la surpopulation des classes. Les comités scolaires étaient chargés d'embaucher les enseignants contractuels en s'octroyant la liberté de remplacer ou de garder l'enseignant contractuel initial en fonction de son rendement. Les enseignants contractuels recevaient environ un quart du salaire des enseignants civils réguliers tout en disposant des mêmes qualifications professionnelles. Les 70 écoles restantes servaient de groupe de comparaison.

Dans la moitié des écoles de l'ETP, les élèves ont été affectés au hasard à la classe prise en charge par l'enseignant contractuel ou à la classe assurée par l'enseignant fonctionnaire. Dans l'autre moitié des écoles de l'ETP, les élèves ont été répartis en classes plus petites sur la base de leurs résultats aux examens standardisés, par un processus appelé « tracking ». La moitié des écoles de l'ETP ont également été sélectionnées au hasard pour recevoir une formation à la gestion scolaire (School-Based Management : SBM). Cette formation a été conçue pour permettre aux parents membres du comité scolaire de s'assurer que le processus de recrutement de l'enseignant contractuel local était équitable et objectif, et assurer le suivi des performances des enseignants.

Dans l'ensemble, le programme ETP a permis d'améliorer les résultats des tests de 0,22 SD (standard deviation ou écart-type) ; bien que tous les élèves n'en aient pas profité de la même manière. L'impact sur les résultats des élèves a été beaucoup plus important pour les élèves affectés aux enseignants contractuels que pour les élèves affectés aux enseignants de la fonction publique. Les enseignants contractuels avaient 28 points de pourcentage de plus de chance d'enseigner lors d'une visite aléatoire que les enseignants du gouvernement dans les écoles non-ETP, le programme ETP ayant réduit l'effort des enseignants de la fonction publique.

La combinaison du programme ETP et de la formation (SBM) a permis d'augmenter les résultats des élèves aux tests plus que ne l'a fait le programme ETP seul. Cette formation a permis aux parents de suivre plus efficacement le processus d'embauche des enseignants contractuels et a permis de réduire de moitié, environ, le nombre d'enseignants contractuels, embauchés simplement parce qu'ils étaient liés à un enseignant existant. Le tri des élèves par classe en fonction de leurs résultats aux tests standardisés a, également, permis d'améliorer l'apprentissage. Enseigner à des élèves ayant un niveau scolaire similaire a été bénéfique en partie ; ce qui, d'après Glennerster et al., a permis de mieux adapter les matériaux d'enseignement au profit de leurs élèves et d'augmenter la présence des enseignants dans les classes (pp. 425-426). Les auteurs en déduisent que les enseignants



fonctionnaires des écoles qui ont reçu des fonds pour embaucher un enseignant contractuel ont à la fois augmenté leurs absences et influencé négativement le processus d'embauche des enseignants contractuels. En revanche, les élèves affectés à des enseignants contractuels recrutés localement, dont le taux d'absentéisme était plus faible, ont augmenté considérablement leurs résultats aux examens. Par ailleurs, le programme se révéla être un outil de gouvernance ayant permis aux parents de participer aux comités d'école améliorant, ainsi, les résultats des élèves quel que soit le statut de l'enseignant, contractuel ou fonctionnaire.

Tout comme l'éducation, la santé ou la microfinance (micro-crédit ou micro-assurance) n'ont pas échappé à cette multiplicité expérimentale. Les mêmes ingrédients déclencheurs s'y retrouvent : accompagnement d'une ONG locale, disponibilité de financement ou bien réponse à appel d'évaluation. A la lisière entre la santé et l'éducation, l'on retrouve immanquablement le programme phare, très largement médiatisé et souvent invoqué dans la littérature expérimentale, transcendant la polarisation au sein de la communauté de la recherche et de l'évaluation entre partisans et pourfendeurs des expérimentations aléatoires contrôlées. Il s'agit de l'expérimentation menée par Edward Miguel et Michael Kremer au sujet du traitement des vers intestinaux. L'expérience recouvre, au demeurant, aussi bien des préoccupations sanitaires qu'éducatives puisqu'une des vertus du traitement administré (vermifugation) est de faire baisser l'absentéisme scolaire.

Le fait, rarement signalé dans la littérature spécialisée, est que cette étude, menée entre 1997 et 2001, et intitulée *Worms : Identifying Impacts On Education and Health in the presence of Treatment externalities* (Miguel et Kremer, 2004), n'est pas le produit d'une expérience randomisée. Les écoles incluses dans les groupes de traitement et de contrôle ont été assignées sur la base de l'ordre alphabétique. L'expérimentation échelonnée sur trois années a été menée en collaboration avec le partenaire habituel dans cette région de l'ouest du Kenya (district de Busia près du Lac Victoria) : ICS Africa (International Child Support). Elle a touché soixante-quinze établissements scolaires répartis en trois groupes. Les enfants infectés par les helminthes (vers intestinaux) ont été traités deux fois par an de la manière



suivante : pour 25 des 50 écoles des deux premiers groupes, les frais du vermifuge l'ont été à la charge des familles. Les deux tiers des écoles participantes ont reçu un type de médicaments et un tiers s'est vu administré deux types de médication. Une campagne de sensibilisation, avec cours et affiches, a accompagné le traitement pour éviter, notamment, les contacts avec des déchets dangereux que constituent les excréta humains ou l'eau polluée du lac.

La littérature issue de cette étude (site du J-PAL et guide pratique de la randomisation déjà cités) indique que, d'une part, la prise du médicament chute de manière importante dès que le prix devient positif, montrant qu'il est contre-productif de faire payer pour traiter des maladies infectieuses et, parallèlement la fréquentation de l'école s'est accrue de 7,5 % de points équivalent à 25 % de chute de l'absentéisme scolaire pour les écoles s'étendant à trois kilomètres. Dix années plus tard, on a mesuré que les bénéficiaires du traitement vermifuge ont gagné 0,5 à 0,8 années de scolarité supplémentaire et réussi à travailler 3,5 heures de plus par semaine !

Ozier (2018) montre, au contraire, que malgré la réduction de l'anémie et l'amélioration de la fréquentation scolaire, l'intervention initiale ne semble pas avoir amélioré les mesures de l'intelligence générale à long terme, ni les résultats des tests scolaires à court terme pour les bénéficiaires directs du vermifuge. Cela s'explique en partie du fait que, pour certains types de résultats, l'intervention vermifuge arrivait trop tard pour les participants à l'étude initiale. Ozier constate que le fait de vivre dans une communauté traitée par vermifuge dès le début de la vie (de 0 à 2 ans) entraîne une amélioration des performances cognitives et scolaires dix ans plus tard. En outre, les chocs nutritionnels et les modifications des stimuli environnementaux sont déterminants au cours de cette période de croissance.

L'autre étude de Baird, Hicks, Kremer et Miguel (2015, p. 36) visant à estimer l'ampleur et la portée spatiale du traitement vermifuge constate, par ailleurs, que les calculs fiscaux et de TRI sont spéculatifs pour plusieurs raisons, notamment la projection des revenus futurs, et que les enfants de l'école primaire sont probablement trop âgés pour que le vermifuge ait un impact majeur sur le développement du cerveau et rien n'indique un tel impact.

Au sujet des expérimentations relatives aux finances, Banerjee et Duflo reviennent longuement (2011) sur les « entrepreneurs malgré eux », loués par la microfinance et l'entrepreneuriat social fondés sur le présupposé que les pauvres sont des entrepreneurs nés et qu'il suffit de leur donner l'environnement adéquat et l'aide nécessaire pour se lancer. Leur conclusion, quant à la capacité du microcrédit à sortir les pauvres de l'ornière du dénuement, est sans appel : le microcrédit peut aider à soutenir les toutes petites entreprises pour le rôle qu'ils peuvent jouer en faveur des pauvres mais il serait vain de penser que cette modalité a le pouvoir d'ouvrir la voie à une sortie massive de la pauvreté (p. 352). Tout au plus, considèrent-ils, la microfinance comme un outil qui fait apparaître les entreprises des pauvres comme un moyen « d'acheter un emploi... [plutôt] que l'expression d'une authentique pulsion entrepreneuriale » (p. 341).

En dépit de cette profession de foi relevant à la fois de l'ordre positif et normatif, le nombre d'expérimentations aléatoires en matière de finances, microfinances et micro-assurances recensé par le site du J-PAL dépasse les 290.

L'expérience menée aux Philippines par Dean Karlan et Jonathan Zinman en est une illustration paradoxale entre l'intérêt pour ce levier de lutte contre la pauvreté et les limites qu'il suggère. Cette expérimentation est également intéressante par le profil bien singulier de Karlan dont les directeurs de thèse furent, précisément, Duflo et Banerjee. Il est producteur prolifique d'expérimentations randomisées tout comme le sont les principaux acteurs du J-PAL dont il préside le secteur finances. Comme indiqué plus haut, il est également président-fondateur de l'IPA (Innovation Poverty Action), ONG ayant la charge de mobiliser les fonds nécessaires aux RCT.

L'étude en question (2011) part de deux constats : le premier est que les institutions de microcrédit distribuent des montants se comptant en milliards de dollars sous forme de petits prêts, principalement à des femmes entrepreneures, et en second lieu ses partisans pensent que cela atténue les défaillances du marché, stimule la croissance des microentreprises et accroît le bien-être des emprunteurs. En collaboration avec la First Macro Bank, les auteurs ont cherché à tester ces hypothèses à l'aide d'un modèle expérimental qui a permis d'attribuer de manière

aléatoire des microcrédits de responsabilité individuelle (de 200 à 1000 dollars, 225 en moyenne) à 1601 personnes aux Philippines grâce à un score de crédit. 1272 ont été acceptées et ont servi comme groupe de traitement et 327 n'ont pas réussi au test de score et ont servi comme groupe de contrôle. Au bout de 11 à 22 mois, il a été constaté que les éléments de preuve sont compatibles avec la demande non satisfaite au prix actuel (le taux d'intérêt annualisé étant d'environ 60 %). Ainsi, l'emprunt net a-t-il augmenté de 9,4 % de plus pour le groupe de traitement par rapport au groupe témoin en notant, toutefois, que le nombre d'activités commerciales et d'employés dans le groupe de traitement a diminué par rapport aux témoins, entraînant une baisse du bien-être subjectif. De même, a-t-il été constaté le peu de preuves que les effets du traitement aient été plus prononcés chez les femmes. Il a également été prouvé que les microcrédits augmentent la capacité à faire face aux risques, renforcent les liens communautaires et augmentent l'accès au crédit informel. En conclusion, argue-t-on, le microcrédit peut fonctionner mais plutôt à travers des canaux différents que ceux supposés par ses partisans.

### CHAPITRE III : Limites du *Gold Standard*

A examiner de près ce modèle, placé au rang de gold standard, on s'aperçoit qu'il 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.

Comme évoqué plus haut, ce qui distingue les expériences contrôlées de toutes les autres méthodes de recherche est l'attribution aléatoire de traitements différents aux unités d'observation de l'étude. Ainsi, le chercheur manipule-t-il délibérément l'environnement pour une population déterminée en introduisant un changement et en mesurant ensuite les conséquences. Si le contrôle s'exerce sur les variables étudiées et sur l'environnement dans lequel ces variables sont observées, la mesure de l'effet nécessite, quant à elle, une base de comparaison.

Les expérimentations aléatoires menées par les chercheurs du J-PAL et leurs associés se distinguent de celles menées dans les années soixante et soixante-dix en Amérique du Nord par un aspect fondamental. Si les dernières visaient à estimer les paramètres sous-jacents d'une fonction de réponse de la population comme ce fut le cas pour l'expérience de l'impôt négatif sur le revenu (Negative Income Tax) testant diverses combinaisons de garanties de revenu, les autres ont cherché à mesurer des effets globaux d'un ou plusieurs traitements distincts. Ainsi, les premières expériences sont de type structurel dans le sens où les traitements individuels sont définis comme des points dans un espace de paramètres politiques continus alors que pour l'autre type d'expérience, chaque traitement testé représente une intervention unique. Les expériences qui testent plusieurs traitements ne trouvent aucun moyen naturel d'être définis comme des points sur un continuum politique (Burtless, 1995).

Le gold standard attribué aux RCT repose sur trois postulats. D'abord, la randomisation fournit un test équitable en égalisant tout sauf le traitement et permet donc une estimation rigoureuse de ce dernier. En second lieu, elle est nécessaire pour résoudre la sélection sans biais qui isole l'inférence causale recherchée. Enfin cette inférence statistique ne nécessite que la comparaison de deux moyennes (différence des moyennes dégagées pour les groupes de traitement et de contrôle).

Pour ses promoteurs, ces postulats érigés en principe fondateurs, confèrent aux RCT des avantages jugés incomparables. Contrairement à d'autres méthodes d'analyse, ces avantages, constituant les principaux arguments en leur faveur, sont faciles à décrire puisque l'application au hasard de traitements alternatifs permet d'appréhender avec une grande fiabilité les effets sur le comportement (Burtless, 1995 ; Allegre, 2004). La procédure d'assignation assure également la saisie de la direction de la causalité entre le traitement et le résultat : les différences de résultats moyens entre les différents groupes de traitement sont dues aux différences de traitement, et les différences de résultats moyens ne sont pas la cause des différences de traitement observées.

L'attribution aléatoire éliminerait par ailleurs toute corrélation systématique entre le statut de traitement et les caractéristiques observées et non observées des participants. Les effets estimés sont donc exempts du biais de sélection susceptible d'entacher, potentiellement, toutes les estimations basées sur des sources d'information non expérimentales. Ainsi, une expérience soigneusement conçue et bien administrée permettra-elle de convaincre que les données expérimentales peuvent produire une estimation interne valable de l'effet moyen du traitement. Cette orientation normative rattache les programmes expérimentaux à la définition des politiques fondées sur les preuves permettant aux analystes, aux décideurs politiques et donateurs de mesurer - et d'observer - les effets de stimuli économiques ou de nouveaux types de traitement qui n'auraient pas été observés auparavant. Par exemple, la variation naturelle des prix relatifs ou des traitements politiques, est bien trop faible pour permettre aux économistes de déduire de manière fiable la promesse de nouvelles politiques : « ...De nombreux hommes

politiques pensent, par exemple, que les employeurs privés engageraient un plus grand nombre de travailleurs défavorisés si les salaires versés à ces travailleurs étaient généreusement subventionnés. Si un programme de ce type n'a jamais été testé, il serait difficile, voire impossible, de prévoir la réaction des employeurs à un niveau de subvention salariale particulier. Bien entendu, de nouvelles politiques peuvent également être testées dans le cadre de programmes de démonstration. Par rapport à la plupart des sources d'information non expérimentales, les expériences permettent aux économistes de connaître les effets d'un éventail beaucoup plus large de prix et de politiques » (Burtless 1995).

Rompant avec le pluralisme méthodologique présidant à l'activité des évaluations expérimentales et se présentant comme les meilleures, les ECR réduisent ainsi la question de la conduite des politiques publiques à ce « qui marche » et ce « qui ne marche pas ». Or, sans « remonter les phénomènes à leurs principes, et découvrir les structures qui les règlent » (Robert Frank, 2007), l'absence de lien entre les ECR et d'autres connaissances scientifiques rend difficile leur utilisation en dehors du contexte exact dans lequel elles sont menées (A. Deaton, N. Cartwright, 2016).

Dans les sections qui suivent, nous montrerons que les données qui en sont issues sont loin d'être idéales pour estimer les paramètres structurels des modèles comportementaux. Leurs limites rendent difficile la généralisation des résultats d'une expérience à l'autre ou l'utilisation d'expériences pour identifier les paramètres structurels invariants, programme nécessaire à l'évaluation de la politique (Heckman, 2010). La section (1) s'attachera à mettre en exergue les apories issues de l'identification causale issues de l'absence lacunaire de modèle théorique sous-jacent. Les RCT s'attachent à révéler ce qui marche en étant incapables d'expliquer comment ça marche, la description statistique bute alors sur le problème de la généralisation inductive. La culture économique exige la combinaison de la vision historique avec celle issue de la maîtrise des techniques d'observation et des modèles théoriques (Schumpeter, 2004). Les controverses qui en résultent, comme nous le verrons, sont anciennes.

Au niveau de la section (2), nous nous attardons sur les nombreux biais, soit d'ordre statistique, liés à la taille des groupes d'expérimentation, à la puissance, voire à l'hétérogénéité des résultats, soit encore au niveau de biais cognitifs ou psychologiques, produisant des effets liés davantage à la motivation des sujets qu'aux programmes expérimentaux à l'instar de l'effet Hawthorne ou l'effet Henry. Les biais de publication, tout comme les dilemmes éthiques résultant de l'asymétrie induite par les rapports de pouvoirs établis entre chercheurs et sujets d'expérimentation, affectent la rigueur réclamée par les promoteurs de cette microéconomie du développement. Enfin, la section (3) s'attachera, tout particulièrement, à mettre en exergue la faiblesse de la chaîne argumentative des RCT au regard de la déconnexion méthodologique par rapport à tout rattachement théorique.

### **Section 1 : Les apories de l'identification causale**

Les controverses portées sur les limites de ces méthodes trouvent leur origine dans ce que Deaton (2009) et Dani 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 et le problème n'est pas tant qu'est-ce qui marche mais comment ça marche. Rodrik ne dit pas autre chose lorsqu'il écrit : on expérimente, mais qu'apprend-on ?

Alors qu'il était inenvisageable de publier un article utilisant les méthodes économétriques empiriques sans qu'il ne fût adossé à un modèle théorique, cette 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 des conditions d'interaction des phénomènes en vue de dégager des causes ou des déterminants d'effets.

Cette transformation fondamentale pose la question de la mise à l'épreuve des théories et du délicat passage de la description statistique à la généralisation inductive d'autant que l'absorption de l'économie du développement par l'approche néo-classique coïncide, comme il fut mentionné, 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.

Mookherjee (2005) constatant que l'économie du développement devient de plus en plus une discipline empirique, considère que ses principales préoccupations au cours de la dernière décennie sont de plus en plus motivées par des questions d'observation des données et de politique économique et se demande si le pendule ne s'était pas trop éloigné devant, ainsi, imposer la nécessité de réorienter les programmes de recherche et les méthodes à l'œuvre. Aussi, propose-t-il, en conséquence, de revenir sur les débats concernant le lien entre théorie et empirie ayant scandé l'histoire de l'économétrie dès son émergence et de situer l'évolution de la méthodologie en économie du développement dans ce cadre. Convoquant les débats entre Tinbergen et Keynes, dont on connaît l'attitude profondément méfiante à l'égard des statistiques d'une part, et ceux entre Koopmans et Vining d'autre part, Mookherjee rappelle que l'équilibre requis entre théorie et propositions observationnelles avait bien donné lieu à controverse tout le long de la première moitié du XX<sup>e</sup> siècle. Pour Schumpeter, évoqué plus haut, on ne comprendra rien aux phénomènes économiques sans maîtrise des données historiques ; maîtrise qui passe par la combinaison de la vision historique avec l'appropriation des techniques d'observation et des modèles théoriques.

Mookherjee relate la résistance de Keynes vis-à-vis du travail empirique de Tinbergen où ce dernier énumère les sources possibles de biais qui, dans le langage économétrique d'aujourd'hui se traduiraient par des problèmes de variables omises, d'erreurs même dans les variables, de forme fonctionnelle incorrecte, de décalage temporel, de multi colinéarité, de simultanéité, d'instabilité de paramètres et de problèmes d'extrapolation, tous éléments qui justifient son scepticisme quant à l'utilité des régressions dans le processus d'inférence concernant la validité des théories. Sur un plan méthodologique, Keynes va jusqu'à questionner la pertinence d'appliquer les corrélations multiples à des données économiques non analysées en supposant constants les coefficients adoptés sur de longues périodes et pose la question de la détermination du futur sur la seule base de statistiques passées (De

Vroey, et al., 2012, p. 22). Critiques à propos desquels, Tinbergen a considéré, dans sa réplique, que de nombreuses influences pourraient être exclues de la régression estimée et collectées sous formes d'hypothèses appropriées et permettre, ainsi, l'inférence ; les problèmes de multi colinéarité étant, selon lui, des problèmes de degré et non de nature.

Sur ce même registre historique, Koopmans, dans sa critique du recueil statistique des cycles économiques de Burns et Mitchell, pointe les défauts de la mesure sans théorie et considère que sans le recours à la théorie, il est impossible de tirer des conclusions pertinentes pour orienter les politiques économiques. Ainsi, pour cet auteur la question primordiale de la théorie économique est celle de la vérité logique et de la clarté des modèles et non pas la vérité empirique (Vachet, 1975 ; Koopmans, 1970, p. 146). La réponse de Vining, dans le débat rapporté dans 'Methodological Issues in Quantitative Economics' (1949), cité par Mookherjee, s'articule autour de quatre volets de la recherche quantitative initiée par Haavelmo et consistant à construire des modèles théoriques, tester les théories, procéder à l'estimation et enfin établir la prédiction. La réplique de Vining rappelle, ici, l'importance de l'analyse exploratoire en tant que phase préliminaire de la recherche sur les phénomènes économiques concernés et c'est ce qui permet de comprendre l'importance de la théorie pour expliquer les dérivations d'implications potentiellement observables.

L'économie du développement qui est apparue comme une discipline distincte n'a pas échappé à ce découpage où les principaux protagonistes ont combiné à la fois les empiristes tels Kuznets ou Myrdal et les théoriciens comme Rosenstein-Rodan, Hirschman, Leibenstein ou plus récemment Amartya Sen. La préoccupation politique de ces économistes était focalisée sur la conception appropriée de stratégies de planification, du développement et de réglementation, toutes tâches dévolues à la puissance publique. Plus récemment, s'écartant du paradigme walrasien, des travaux théoriques importants ont abordé les questions essentielles telles les défaillances de la coordination, la croissance endogène, les marchés incomplets ou bien les problématiques de gouvernance et le rôle et fonctionnement des normes sociales.

Ce rappel permet de souligner les difficultés qu'a l'empirie à saisir les composantes observationnelles aussi bien à partir de séries temporelles économiques que de données expérimentales pour se livrer à un travail de prédiction. En dépit de tous les perfectionnements apportés aux techniques statistiques dans l'approche consistant à faire parler les données sans guide théorique, les expériences aléatoires contrôlées n'échappent pas aux cycles du genre, que les données soient issues de séries historiques ou construites en utilisant un modèle expérimental d'évaluation randomisée. La revendication de supériorité intrinsèque de ces évaluations viendrait de sa capacité à répondre à des problèmes généralisés d'identification faible en économie empirique non seulement dans le développement mais également dans d'autres domaines appliqués comme l'économie du travail, la santé ou l'économie publique (Lalonde, 1986 ; Angrist, 1990 ; Angrist, Imbens et Rubin, 1996 ; Imbens et Wooldridge, 2009). Cette prétendue supériorité s'appuie également sur la capacité à résoudre les problèmes économétriques associés aux effets de placement et de sélection. La variable randomisée est alors un instrument strictement exogène (Barrett et al., 2010). L'absence de biais sélection est avancée comme le gage de l'identification causale rigoureuse. Elle tend, cependant, soit à ignorer délibérément soit à passer sous silence les biais dits de randomisation (Heckman, 1992 ; Fougère, 2010). Par ailleurs, de nombreuses contributions et travaux rapportent les cas où les influences politiques et les interactions chercheurs/expérimentateurs/partenaires sont pointées comme des « bricolages » affectant négativement le crédit attribué aux RCT. C'est, par exemple, le cas de l'expérimentation au Cambodge relative à l'évaluation d'un programme de micro-assurance (Aurélie Quentin et Isabelle Guérin, 2013) ou, encore, pour l'expérience tant vantée et citée sur les vermifuges où, comme indiqué au Chapitre II, les partenaires locaux avaient refusé l'utilisation du tirage au sort pour affecter les écoles objet de l'expérimentation et imposer l'assignation par ordre alphabétique. L'expérimentation randomisée a également servi comme « transposition d'une méthode scientifique hors de son cadre d'origine » (Massieu, 2011) où la notion d'expérimentation introduite dans la révision de la Constitution en 2003 fut utilisée alors même que le dispositif d'évaluation était en cours ; la loi généralisant le RSA (Revenu de Solidarité Active)

a en effet été promulguée sans attendre les résultats de cette expérimentation (Massieu, 2011 ; Allègre, 2012). Un autre cas concerne le programme Progres de l'aide ciblée au Mexique (Faulkner, 2012). A l'aide d'une série d'entretiens menés auprès d'assistants de recherche, maillon indispensable des expérimentations aléatoires qui joue un rôle primordial d'interface entre le terrain et le chercheur, Jatteau (2015) a étudié « les expérimentations aléatoires telles qu'elles se font et non telles qu'elles devraient se faire idéalement » (Jatteau, 2017, p. 336). Ses entretiens ont donné à voir un certain nombre de bricolages et arrangements qui caractérisent la façon dont certains travaux sont menés. À titre d'exemple, Jatteau cite l'expérimentation menée en Inde visant à réduire l'absentéisme des instituteurs et contrôler leur présence en les photographiant lors de la prise de service en classe puis pouvoir, ainsi, lier la rémunération à la présence effective. Pour cet auteur, cet exemple illustre la manière dont on imagine des artifices : on s'efforce de tester les variables jusqu'à obtenir des résultats, et l'on procède ensuite à leur interprétation avec pour objectif de les présenter comme généralisables, sans en fournir la moindre preuve (*sociologie économique* du 06 Mars 2017).

Devaux-Spatarakis dresse de façon, à peu près similaire, le même constat :

« Une expérimentation randomisée, entre la très bonne idée sur le papier, d'une mesure ou même d'un protocole de randomisation très bien *designé* [anglicisme] et, trois ans après, la récupération des données et des résultats, il s'est passé potentiellement beaucoup de choses. Il y a plein de vraies gens qui interagissent dans une expérimentation, que ce soient les traités ou les intermédiaires qui dispensent le traitement : les opérateurs, les gens qui travaillent concrètement à cela, il y a 1000 raisons qui font qu'à un moment donné la randomisation peut avoir sauté ou ne pas avoir été appliquée, que les règles n'aient pas été parfaitement appliquées et, aux quels cas, les données que l'on récupère ne valent rien » (2014, p. 268). Elle cite, en appui de son propos, le constat d'un évaluateur du Crest (Centre de recherche économique et statistique) pour qui la mise en œuvre de l'expérimentation constitue la phase où se déroule le travail institutionnel entre parties prenantes. Cette phase qui peut durer de deux à trois ans, nécessite de manière permanente l'interaction entre porteurs de projets et équipes d'évaluation.

Ces dernières, au-delà de leurs ambitions initiales, doivent composer avec les réalités du terrain. S'agissant des opérateurs du terrain, l'apprentissage scientifique expérimental auquel ils sont soumis, peut se révéler si difficile qu'il menace l'intégrité du protocole (p. 266). Cette politique technocratique de la statistique, consistant « à produire conjointement de la connaissance et de la contrainte sur les comportements les plus intimes des individus » (Bardet et al., 2012, p. 186), se veut exempte de biais d'ordre statistique. Elle se heurte, cependant, à des difficultés inhérentes à cette technologie expérimentale liées aux biais de randomisation admis, même, par ses propres partisans.

## **Section 2 : Les biais de randomisation entre rigueur méthodologique et remède miracle**

Il est généralement admis que l'estimation des relations empiriques est souvent entachée de biais. Dans les approches structurelles, il s'agit d'identifier les mécanismes économiques à prendre en compte à partir des contributions théoriques mobilisées et les données disponibles utilisées. La démarche non structurelle, quant à elle, se fixe pour seul objectif l'identification causale de la politique à évaluer en focalisant l'analyse sur la variabilité exogène des données qui permettent d'isoler l'effet causal (Roux, 2015). Utilisant une approche non structurelle, les promoteurs des RCT ont pour visée première la constitution, à partir du terrain, des données adaptées à l'évaluation des programmes qu'ils se sont fixés de mener. De ce fait, l'expérimentation contrôlée demeure enchâssée entre le Charybde d'une rigueur méthodologique et le Scylla du remède miracle selon l'expression de Vincent Guillin (2013). Or de nombreux biais menacent la supériorité revendiquée tant ces expérimentations présentent souvent des problèmes de conformité (compliance) par le présupposé que tous les individus alloués au traitement y participent et que les individus du groupe de contrôle n'y participent pas. En pratique, la réalité, comme nous le verrons après, est tout autre. Dans ce cas fréquent, l'effet moyen du traitement dégagé n'est plus que l'effet moyen de l'intention à traiter. La conformité est alors imparfaite et l'effet mesuré ne concerne donc que ceux ayant participé au programme que l'on appelle TOT (Treatment On the Treated) dont l'effet estimé

s'obtient en divisant l'estimateur de l'intention à traiter par la proportion du nombre de ceux qui participent effectivement, corrigé de la proportion de ceux qui y participent mais censés être dans le groupe témoin. C'est précisément ce TOT, agissant comme variable instrumentale, qui capte l'effet sur les individus qui respectent l'assignation (Bouguen et al., 2014, p. 124). Cette démarche inductive agit, alors, selon le post hoc, ergo hoc (après cela donc à cause de cela). Elle pose, en conséquence, des problèmes lorsque les caractéristiques qui influencent la non-conformité sont, par ailleurs, reliées à l'efficacité du traitement éliminant, ainsi, les bénéfices de la randomisation. La littérature rapporte abondamment les cas des individus *always-takers* (toujours participants) et les *never-takers* (jamais participants) quelle que soit leur assignation. Le choix du protocole d'évaluation est déterminé par le niveau d'assignation qui préside à l'application du traitement considéré. Il est guidé, aussi bien, par les besoins de précision statistique que par les compromis issus des arbitrages aux niveaux des différentes stratégies d'assignation ou du niveau où s'opère celle-ci. Les arbitrages sont, également, liés aux déterminants de la puissance statistique aux problèmes déterminés par l'acceptabilité, ou encore par les considérations d'ordre éthique. Les arrangements susmentionnés, les considérations précitées et les limites à la conformité que nous venons d'évoquer font peser de sérieuses menaces sur la randomisation et sa validité interne, génèrent plusieurs apories statistiques et affectent négativement la revendication d'absence de biais que la randomisation est censée éliminer (Heckman, 1992 ; Heckman et Smith, 1995 ; Manski, 1997 ; Fougère, 2010). Elles limitent la capacité de la randomisation à fournir un test équitable en égalisant tout sauf le traitement censé permettre une estimation rigoureuse de ce dernier, et peinent, ainsi, à résoudre la sélection sans biais censée isoler l'inférence causale recherchée par la seule comparaison de deux moyennes (différence des moyennes dégagées pour les groupes de traitement et de contrôle).

Un autre problème, souvent rapporté, a partie liée avec la qualité de la collecte des données tant au niveau qualitatif que quantitatif. Cette qualité est entachée par la méthode de collecte souvent confiée à des assistants de recherche ou ayant recours aux données d'administrations, critiquées par ailleurs pour la défaillance de leurs



dispositifs statistiques, voire par des agents engagés par les ONG auprès desquels les données sont collectées (Jatteau, 2015). Le crédit accordé est, également, entaché par les problèmes d'acceptabilité du protocole expérimental (Aurélien Quentin et Isabelle Guérin, 2013). Bien souvent, les données sont collectées « pour l'évaluation elle-même ou dans son seul but » (Bouguen et al., 2014). Les menaces sur la qualité des données constituent alors l'un des principaux paramètres explicatifs de l'attrition. Du fait de cette attrition, la validité interne de l'estimateur de l'effet du traitement pourrait être sujette à caution du fait même de la question de la probabilité d'y répondre ou de la qualité de la réponse induisant aussi bien une difficulté à isoler ce qui est causé par les problèmes de collecte que de ce qui correspond réellement à l'effet du traitement.

Si les effets d'attrition concernent les membres qui décident de quitter l'expérience, qu'ils soient du groupe du traitement ou du contrôle, l'effet de substitution concerne, généralement, les membres des groupes qui se soumettent au traitement ou plus encore à d'autres programmes ; phénomène que l'on observe particulièrement pour ceux destinés au retour à l'emploi. S'y ajoutent les problèmes de perméabilité, les individus des groupes témoins peuvent s'arranger pour avoir accès au traitement, et qu'à l'inverse, les individus du groupe de test y échappent. Ce phénomène a été rapporté dans l'enquête au sujet des RCT conduites par le fonds d'expérimentation de la jeunesse, appuyée par des témoignages des chercheurs qui y ont été impliqués, menée par Devaux-Spatarakis (p. 291). En mettant à mal la puissance statistique, ces deux effets combinés minent sur le terrain la validité interne revendiquée. C'est ce qui constitue un des principaux obstacles rencontrés dans le processus de production des résultats d'autant que les effets de diffusion d'un programme sur tout ou partie du groupe témoin pourraient être affectés indirectement par le programme lui-même.

Il n'est pas toujours acquis que ceux qui gèrent l'expérience soient capables ou disposés à protéger la conception des expériences de ces menaces à la validité interne. Les biais de sélection et de diffusion du traitement sont souvent aggravés par les réactions potentielles liées souvent aux comportements involontaires parmi



les participants ou les groupes témoins. La randomisation peut susciter de la rancune entre les différents groupes participant à l'expérience (Deaton, 2009 ; Cohen et Easterley, 2009), les sujets n'acceptant pas leur affectation à un groupe en refusant de participer ou bien en finissant par obtenir le traitement alors qu'ils en étaient exclus.

La littérature critique des RCT fournit moult exemples où nombre de participants refusent l'expérimentation et, plus gênant encore, où d'autres cessent de participer à l'expérience avant que celle-ci ne parvienne à son terme. Dans ces conditions, la difficulté provient de l'incapacité d'isoler ce qui est dû aux effets du traitement de ce qui vient de la qualité de la collecte des données. L'estimateur final obtenu sera, alors, celui produit par une randomisation altérée.

Les efforts déployés dans le domaine de l'économie pour traiter le risque de biais dans les estimations du RCT sont de portée limitée et les altérations correspondantes, régulièrement, pointées par la littérature critique sont souvent passées sous silence (Eble et al., 2014). A partir de leçons des sources de biais d'estimation identifiées dans la littérature médicale, ces chercheurs montrent que les ECR appliquées à l'économie ne fournissent pas d'informations essentielles ni sur les décisions relatives à la conception ni dans les comptes rendus liés aux sources de biais ainsi mises en évidence. De ce fait, ils estiment que les ECR en économie présentent un risque bien plus élevé de rapporter des effets de traitement exagérés. Ils concluent en suggérant à la fois des normes contribuant à la qualité de ces expériences et des outils pour aider les chercheurs et les décideurs à évaluer le risque de biais et de l'éviter. Selon Bédécarrats et al. (2015), les défaillances relèvent plus d'une « désinvolture liée à de moindres exigences de qualité scientifique qu'à des procédures de contrôle plus lâches par les pairs ».

Aux limites précitées s'ajoutent ce que l'on appelle l'effet Hawthorne lié au processus psychologique qui fait référence davantage à la motivation des sujets observés dans l'étude qu'au programme lui-même. En effet, la motivation résulterait de l'attention accordée aux participants et qui leur confère la conscience de participer à une expérience dans laquelle ils sont testés. L'effet Hawthorne, découvert par suite des expériences menées par Elton Mayo à Chicago entre 1924

et 1932 au sein de l'usine « *Western Electric de Cicero* », induit des observations biaisées en provoquant des modifications dans le comportement et les attitudes des sujets lorsqu'ils ont conscience qu'ils ont été sélectionnés pour devenir un objet d'étude. Ainsi, lorsque les individus se savent observés ils adaptent leurs comportements.

Un autre biais similaire à l'effet Hawthorne, mais différent par ses implications est l'effet qui porte le nom de John Henry (Saretsky, 1975) conducteur de train qui s'était mis en tête de défier une foreuse à vapeur en enfonçant rageusement pointes et forets et de faire mieux que la machine. Son triomphe lui coûta la vie par excès d'effort. Saretsky attribue cette mort à l'effet de l'effort extraordinaire de ceux qui exécutent le traitement de contrôle. L'auteur décrit l'effet John Henry comme un facteur de confusion potentiel entre les approches expérimentales et les approches de groupe de contrôle que l'on relève dans les évaluations des innovations éducatives.

Ce phénomène est généralement observé dans les classes de soutien supplémentaire où les élèves qui font partie du groupe de contrôle peuvent être incités à fournir plus d'efforts pour surmonter le désavantage de n'être pas choisis pour le groupe de traitement. Saretsky estime que, généralement, les membres du groupe de contrôle perçoivent les innovations comme une menace sur leur emploi, salaire ou statut et à cause de ce sentiment, ils peuvent faire montre de performances atypiques et altèrent, en conséquence, les résultats de l'évaluation. L'autre effet de cette nature est ce que l'on appelle l'effet Pygmalion qui présuppose que les préjugés des enseignants peuvent influencer les résultats des élèves. La réussite des élèves résulterait parfois de l'intérêt qu'on leur accorde.

L'autre limite statistique, relative aux RCT, est liée à des questions de coûts ou de temps alloués à l'expérience amenant les chercheurs expérimentateurs à recourir à des effectifs réduits (Glennerster et al., 2013). La taille réduite des échantillons peut déboucher sur des estimateurs biaisés en raison de l'écart type qui peut être très large.

Par ailleurs, les études sociales tout comme les études médicales peuvent être entachées de biais liés à la publication des travaux de recherche, phénomène bien

connu sous le nom de « biais de publication ». Lesquels biais apparaissent lorsque la publication du travail va dépendre soit de la direction de la signification de la recherche soit des résultats obtenus par le chercheur. Les revues académiques y contribuent elles-mêmes en sélectionnant davantage les expériences à résultats positifs plutôt que celles qui aboutissent à des résultats négatifs : « plusieurs études ont montré que certaines revues académiques avaient tendance à un certain degré de publication sélective en ne retenant à la publication qu'un certain type d'études, notamment, celles qui présentent des résultats statistiquement significatifs. Dès lors, une synthèse de la littérature qui s'appuierait uniquement sur les résultats publiés pourrait s'avérer biaisée en faveur d'études présentant des effets empiriques plus larges qu'ils ne sont réellement » (Laroche, 2007, p. 29). Le même constat est rapporté par Franco et ses co-auteurs (2014) lors de l'examen de 221 études financées par la National Science Foundation où il ressort que les résultats positifs ont 40 % de plus de chance d'être publiés que les résultats nuls et même 60 % de chance d'être rédigés.

Ce phénomène est bien connu en matière d'expérimentation médicale et particulièrement thérapeutique où la tendance est, souvent, à la publication de résultats statistiquement significatifs. Entendu, comme nous le montrions plus haut, que la significativité statistique ne dit absolument rien de la validité méthodologique d'une étude.

A travers les enquêtes qu'il a menées au sujet des ECR conduites par les chercheurs liés au J-PAL durant son enquête socio-économique, Jatteau (2013 et 2015) montre que la pression à la publication pousse à l'obtention de résultats tangibles au détriment des questions de développement ou de politiques publiques. Même du côté des partisans des RCT, l'on voit pointer les limites dues aux coûts et au temps engendrés par les expérimentations qui, souvent, peuvent les rendre plus coûteuses que les recherches non expérimentales sur le même sujet. Elles consomment beaucoup de ressources réelles et sont presque toujours coûteuses en temps et si les décisions politiques concernant une politique publique particulière ne peuvent être reportées, l'utilité d'une expérience peut être mise en doute. Plus généralement, on peut plus facilement persuader les responsables d'affecter de petites sommes à la

recherche pure ou des sommes moyennes à une nouvelle étude que de les convaincre que certaines personnes devraient se voir systématiquement refuser une intervention potentiellement bénéfique dans le cadre d'une nouvelle étude coûteuse (Burtless, 1995). Cet auteur s'inquiétait, par ailleurs, de la fortune moins solide de l'industrie des expérimentations sociales dont la tradition remonte à plus d'un quart de siècle et dont les dépenses consacrées aux essais aléatoires sur le terrain aux États-Unis ont consommé plus d'un milliard de dollars (de 1994). Il relève que les ressources réelles consacrées à l'expérimentation dans les années 1970 et au début des années 1980 ont été plus importantes que celles investies depuis lors et les considère comme des tentatives ambitieuses et coûteuses d'estimer les paramètres comportementaux de base - le revenu et l'élasticité des prix - des fonctions d'offre de travail et de demande de logement et l'élasticité de la demande de soins de santé en réponse à des régimes d'assurance alternatifs. Ces somptueuses expériences (*lavish experiments*) ont donné lieu, d'après lui, à des centaines de rapports de recherche et à de nombreux articles dans les principales revues scientifiques. Et qu'au moment où les expériences sociales ont acquis une nouvelle influence sur l'élaboration des politiques, certains économistes éminents ont été désenchantés par cet outil de recherche et ont remis en question la valeur des expériences pour répondre aux questions centrales sur le comportement humain et l'efficacité des politiques. Ce faisant les politiciens peuvent être amenés à se méfier, naturellement, d'une méthode de recherche impliquant l'expérimentation sur des sujets humains et des dommages que cela pourrait causer à des individus qui sont surtout des électeurs, « Also Known as Voters », écrit-il (p. 64).

Outre les biais susmentionnés, les objectifs de réduction de la pauvreté par des interventions locales placent les RCT dans le champ du questionnement déontologique. Un certain nombre d'expérimentations dénotant un manque de réflexion sur leur aspect moral et, engageant de surcroît des sujets humains, provoquent un certain malaise éthique. C'est d'autant plus problématique que les expérimentations créent de fait des rapports de pouvoir entre expérimentateurs et sujets d'expérimentations marqués par des asymétries de statuts qui soulèvent des enjeux éthiques importants (Ryoa Chung, 2019). Cet auteur dresse un panorama historique qui permet d'illustrer ces rapports de pouvoir qui ont caractérisé les

pratiques de la recherche expérimentale en sciences sociales et plus particulièrement en biomédecine et montre comment ces « rapports de pouvoir ont été déterminés par les inégalités socio-économiques, genrées et racialisées [entre autres] donnant lieu à des injustices structurelles » du fait de manquements éthiques. L'asymétrie dans la relation entre le chercheur qui possède un savoir et le sujet de l'expérimentation qui, dans la situation de passivité où il est confiné, ne peut maîtriser le savoir mobilisé. Le peu d'attention et le manque de réflexion sur les inégalités opèrent un déplacement du cadre conceptuel de l'épistémologie comme cadre de justification et de validité des connaissances scientifiques produites, d'une hiérarchie des positions sociales et des savoirs déterminants, alors, la détention de la connaissance. Stéphane Baele (2013) s'est également consacré aux problématiques d'ordre éthiques soulevées par les approches expérimentales en développement économique et pointe sans ambiguïté les dilemmes éthiques auxquelles font face les ECR qui, parfois, franchissent allègrement les limites morales et leur mise en œuvre est alors prise à la légère. Deaton (2009, p. 40) note que les ECR en médecine - l'étalon or - auquel les ECR en développement se comparent souvent, ne sont pas exempts de difficultés pratiques, et leur primauté n'est pas sans poser un problème.

Les questions éthiques portant sur les sujets humains et entourant les ECR en médecine, sont devenues suffisamment graves pour limiter sérieusement le champ de ce qui pourrait être entrepris. D'autant que l'on nourrit de craintes par les enjeux financiers importants induits et qui, parfois, jettent le doute sur la véracité ou la sincérité des résultats. Ce n'est pas [encore] un problème en économie, mais il le deviendrait certainement si, comme le soutiennent ses défenseurs, les ECR devenaient une condition préalable au financement des projets. Baele cite Concato et al. (2000) qui considèrent qu'en pratique les ECR ne fournissent pas d'informations utiles au-delà de ce que l'on peut apprendre d'études d'observation bien conçues et soigneusement interprétées.

La philosophie morale occidentale a distingué deux approches en matière d'éthique : les approches déontologiques, les approches conséquentialistes (Baele 2013) et celles issues du raisonnement moral : « Grosso modo, dans une perspective déontologique, on se concentre sur le fait de connaître et de suivre

intentionnellement des règles appropriées et non sur les conséquences des actions régies par ces règles. Il suffit d'avoir compris et intentionnellement suivi les règles pour être libéré de la culpabilité morale. Dans un cadre conséquentialiste, en revanche, on juge les actes par leurs résultats, et non par leur conformité à une règle donnée. Dans de tels systèmes, il est possible de suivre une règle avec de bonnes intentions et d'effectuer une action qui, dans le contexte, s'avère avoir des conséquences néfastes et ainsi être coupable de ces conséquences », écrit Robbins dans *La philosophie morale* (Joël Robbins, 2013).

Donc si l'approche déontologique juge répréhensible ce qui va à l'encontre d'une règle qui interdit un certain type d'actions, le conséquentialisme considère qu'aucune action n'est répréhensible en soi et l'éthique de la pratique ne peut être évaluée que par rapport à ses conséquences.

Les promoteurs des expériences aléatoires randomisées se situent clairement dans la catégorie conséquentialiste. Un chercheur comme Robert Dingwall (2006 et 2008) considère que les évaluations éthiques sont un contrôle bureaucratique supplémentaire et que les sciences sociales ne devraient pas être soumises à ces évaluations préjudiciables à la qualité du travail des chercheurs. Le système de régulation éthique préventive, développé dans les sciences biomédicales « est devenu une menace majeure pour la recherche en sciences humaines et sociales » (Dingwall, 2008). Dans son allocution plénière à la conférence annuelle du groupe de sociologie médicale de la *British Sociological Association* à Édimbourg (septembre 2006), il soutient qu'au contraire des sciences biomédicales, les spécialistes des sciences sociales ne peuvent nuire aux sujets humains de manière comparable à la médecine puisque aucun risque inhérent de mort ou de dommages physiques graves n'est encouru. Les chercheurs en sciences sociales ne sont que des invités dans la vie d'autrui (*guests in other people's life*), leur seule obligation est le devoir de vigilance vis-à-vis des détenteurs d'informations.

Ce point de vue, rétif aux contrôles d'ordre éthique, n'est pas isolé dans la communauté des chercheurs en sciences sociales. Vassy et Keller (2008) versent dans le même sens que Robert Dingwall en pointant la propension de certains comités éthiques à transposer les principes, certes utiles, en recherche biomédicale



mais qui, parfois, est contre-productive dans la recherche en sciences sociales : « Ces comités font peser sur les chercheurs des exigences qui révèlent une grande méconnaissance des pratiques de recherches en sciences sociales, donc nuisent au bon déroulement des projets, voire empêchent le chercheur de travailler » (p. 130). Duflo balaye ces préventions d'un revers de main ; pour elle, lorsqu'il s'agit de la recherche, les questions auxquelles s'attaquent les chercheurs sont encadrées par des comités d'éthiques déplorant leur inexistence en France au contraire des États-Unis. Elle déclare soumettre les projets auxquels elle participe au comité éthique du MIT (*La vie des idées*, 05 mai 2009). Elle invoque l'exemple du Kenya, où il incombe à un comité d'éthique de se prononcer sur la rigueur morale en relation avec les expériences qu'elle-même et ses collègues mènent dans ce pays et toujours en veillant à limiter les risques auxquels les personnes peuvent être exposées. Cette figure de proue de la randomisation mentionne, toutefois, certaines limites que son laboratoire veille à gérer avec soin : les ruptures de confidentialité où des informations détenues par les chercheurs peuvent nuire aux sujets et la limitation des budgets dont résulterait la privation des bénéfices utiles d'un programme pour les populations. Toutes ces préventions peuvent paraître dérisoires si l'on garde à l'esprit les risques auxquels peuvent être exposés les sujets humains en matière de recherche expérimentale, en particulier, en sciences médicales et biomédicales. Que l'on se rappelle le rapport Belmont qui a fait suite au scandale de la tristement célèbre affaire Tuskegee en Alabama, où près de 400 sujets noirs ont fait l'objet sur quarante ans (1932 à 1972) d'un suivi thérapeutique sur la syphilis sans que personne de la cohorte ait bénéficié du traitement à la pénicilline, connu pourtant depuis 1943 (Benmarhnia et al., p. 50). Révélé par le New York Times, le scandale fut d'autant plus retentissant que l'étude en question a bénéficié d'un financement public américain, par suite de quoi le rapport Belmont fut établi. Il repose sur trois principes : respect des personnes, bienfaisance, un principe plus proche du conséquentialisme évoqué précédemment et justice dans « la distribution équitable du fardeau et des avantages de la recherche expérimentale » de manière à éviter d'exploiter les catégories vulnérables ou susceptibles de subir les formes insidieuses de discrimination (p. 51). Lorsque Duflo appelle à ne pas exagérer l'importance des barrières éthiques, elle ne semble pas saisir la dimension profondément sociale de



la recherche scientifique où se créent des relations asymétriques entre chercheur et participant comme le signale Chung (2019, p. 57).

Ce type de conflit entre la méthode et son objet débouche à la fois sur les deux questions liées au contrôle de l'interaction et à la qualité de l'objet. Les résultats produits sont alors contingents dans la mesure où l'artefact pourrait déboucher sur un résultat qui disparaît lorsqu'il est retiré ou bien aggrave la situation des bénéficiaires dès lors qu'il n'est pas intégré dans une politique publique appropriée. C'est précisément le cas pour deux études menées dans le cadre d'expérimentations aléatoires contrôlées, à l'instar de celle menée à Tanger au Maroc (Devoto et al., 2011) pour aider des ménages pauvres à obtenir l'obtention de branchements d'eau à domicile au lieu de se rendre aux robinets publics. Le résultat débouche sur le constat que les populations concernées ont gagné du temps pour leurs loisirs alors même qu'elles vivent dans des quartiers où le chômage et la déscolarisation frappent durement, subissant de surcroît des augmentations du budget de l'eau en dépit de l'état de pauvreté dans lequel elles sont plongées. S'y ajoute également le coût des branchements couverts par des prêts à taux zéro assuré par le fournisseur d'eau local ! L'on se trouve ici face à un double dilemme moral : les problèmes éventuels (augmentation du budget de l'eau, impactant une partie du budget) « étaient prévisibles et n'ont pas été clairement surpassés par les gains attendus de la science, qui n'étaient pas particulièrement significatifs » (Baele, 2013). Cette expérimentation fait, par ailleurs, peu cas du conflit d'intérêt né de la situation où le financement est assuré par l'entreprise Véolia qui fournit le service, accorde le prêt pour le branchement (remboursable sur sept ans par échéances mensuelles de l'équivalent de dix euros) et, du coup, élargit sa base clientèle.

Une autre étude symptomatique des dilemmes provoqués par les questions éthiques concerne la tolérance vis-à-vis de la corruption à l'effet d'obtention du permis de conduire en Inde. L'étude en question a été menée par la chercheuse Bertrand, associée au J-PAL, et ses co-auteurs à l'aide d'une expérience randomisée visant l'expérimentation d'incitations à soudoyer des fonctionnaires par des sujets ne possédant pas de permis de conduire. Les sujets en question n'ont pas suivi avec succès la formation requise et obligatoire et font ainsi prendre des risques pour eux-

mêmes et pour des innocents. Barrett et Carter (2010, p. 520) jugent sévèrement cette expérience en pointant l'irresponsabilité de la conception de la recherche, publiée pourtant dans l'une des revues les plus prestigieuses de la profession, alors qu'en tant que telle elle n'a pas accordé l'attention suffisante aux contraintes éthiques de l'expérimentation sur le terrain de l'économie.

Ces auteurs recensent quatre types de dilemmes éthiques (pp. 519-522) : une première catégorie, la plus évidente, concerne les conséquences négatives imprévues mais prévisibles de certains modèles expérimentaux. Le principe « *ne pas nuire* » est sans doute l'obligation morale la plus fondamentale à laquelle doivent se conformer les chercheurs. La plupart des universités et des organismes de recherche sérieux ont mis en place des comités d'examen institutionnels pour se prémunir contre de telles éventualités. Pourtant, les résultats issus de conceptions très discutables sont publiés dans des revues réputées. La seconde catégorie de problèmes éthiques qui émergent dans de nombreuses expériences de développement tourne autour de la suspension du principe fondamental du consentement éclairé. Ce principe sous-tend l'obligation de traiter les sujets humains comme agents volontaires disposant du droit de participer ou non et comme ils l'entendent. Or souvent, il apparaît plus évident de les traiter comme des sujets à manipuler entraînant les biais exposés plus haut et susceptibles de saper la validité interne (effets Hawthorne et assimilés). Les auteurs évoquent les expériences randomisées en grappes maintenant l'exogénéité de l'intervention et dispensant le chercheur d'obtenir ce principe de consentement éclairé. La troisième catégorie concerne le traitement expérimental en aveugle, bien en cours en médecine mais impossible dans les ECR en microéconomie du développement. Ceux qui se savent dans un groupe témoin sont en réalité soumis à un traitement différentiel susceptible de générer des troubles émotionnels (sentiment d'être exclu d'un bénéfice) pouvant générer des conséquences négatives qui altèrent les différences entre les groupes témoin et de traitement ; les dilemmes éthiques deviennent potentiellement des problèmes statistiques.

Enfin, la rareté des ressources est à l'origine de la quatrième catégorie. Elle implique d'exploiter l'information asymétrique que, précisément, la randomisation

est censée surmonter, du fait d'une sélection de personnes qui n'ont pas besoin de l'intervention au détriment de ceux qui en auraient le plus besoin, entraînant ainsi un gaspillage de ressources rares. Ce biais réintroduit de fait l'hétérogénéité non observée que la randomisation vise à surmonter. Le fait d'offrir un programme d'aide à une communauté et pas à une autre peut susciter la jalousie et la différence de résultats est présentée comme un effet positif du programme en dépit, parfois, des effets négatifs infligés à la communauté témoin.

Cette difficulté intervient lorsque les unités traitées et de contrôle s'engagent dans des actions stratégiques les unes par rapport aux autres comme pour les expériences évaluant la participation à la démobilisation militaire (Humphreys et al., 2009). La pureté statistique dessert, alors, les objectifs de l'expérimentation.

S'ajoutent encore deux problèmes fondamentaux : le calcul dangereux et la randomisation elle-même, selon Baele. Le calcul dangereux (*hazardous calculus*) provient du principe conséquentialiste évoqué plus haut qui, tout en admettant les dommages, franchit les limites déontologiques pour mettre en évidence les effets positifs (scientifiquement parlant) ouvrant, ainsi, la question de la responsabilité des chercheurs vis-à-vis des résultats qui résultent de manipulations mises en œuvre par des tiers. Citant Humphreys et ses co-auteurs (2009), Baele reformule encore plus directement le questionnement : les chercheurs sont-ils responsables des effets néfastes de leurs ECR ? Quant au problème de la randomisation, l'auteur pointe son caractère moralement litigieux car, tout en semblant une procédure équitable, elle produit des résultats inéquitables. En ce sens, elle contrevient au principe de la bienfaisance selon lequel on doit d'abord assurer un certain niveau de bien-être à la population la plus défavorisée (p. 17).

### **Section 3 : Pléthore de l'empirie, déficit de la théorie**

Ian Hacking (1989, p. 260) considère que les travaux expérimentaux sont portés par la théorie sans quoi les faits expérimentaux sont ininterprétables. Cependant, pour cet auteur il existe des circonstances heureuses où travail expérimental et théorie peuvent s'harmoniser sans prétendre que l'un puisse exister sans l'autre et qu'il arrive que des recherches fondamentales peuvent précéder des théories pertinentes. L'économie a, de ce fait, longtemps été marquée par l'interdit millien consistant à

évoquer l'impossibilité de procéder à des expérimentations en sciences sociales (que J. S. Mill décrit comme sciences morales) par l'impossibilité de contrôler les variables économiques clés et les conditions de fond nécessaires à la vérification de l'effet de la manipulation de chaque cause isolément. Le long passage d'une économie politique discursive à une économie expérimentale a mobilisé les mathématiques, les statistiques et les modèles non seulement comme instruments mais également comme outils d'investigation.

Dès les années d'après-guerre, l'usage expérimental a pris naissance en économie alimentant, cependant, de fortes controverses d'ordre méthodologique (Serra, 2012). Dupouy (2001) et Soler (2009), chacun selon son angle d'attaque, interrogent le statut épistémologique des expériences et définissent l'expérimentation comme l'intervention volontaire, systématique et contrôlée sur le cours des phénomènes ou événements visant à déterminer les paramètres qui participent de façon effective à produire un effet donné. En médecine, l'intervention se fonde sur la connaissance physiologique et pathologique assurée de son objet ; la science expérimentale étant attachée à une motivation morale, ajoute Dupouy (2010, p. 249). Les contrôles agissent aux niveaux de la conception et de la manipulation et concernent, aussi bien, la variable à modifier ou à manipuler que les conditions de fonds fixées par l'expérimentateur (Guala, 2008). C'est précisément le contrôle des conditions de fond, essence même de la méthodologie expérimentale qui permet d'attribuer le changement attendu à la variable manipulée (Smith, 1976).

Issue de ce que l'on a nommé l'économie comportementale et distinguée quatre fois en moins de vingt ans par le prix de la banque de Suède, l'expérimentation en économie s'est érigée en sous discipline dont le parcours est daté, bien que cette histoire reste à écrire, selon le propos de Guala (2008, p. 152).

Serra (2012) fait remonter l'histoire de l'économie expérimentale au paradoxe dit de Saint-Pétersbourg dévoilé par Bernoulli (neveu) explicitant l'aversion au risque lorsque la variable aléatoire est faible mais son espérance infinie. Guala, Serra et Annie Cot et ses co-auteurs (2016) retracent à grands frais cette histoire entrelacée et situent dans la théorie des jeux ses premiers fondements comme « tournant

expérimental ». Guala rattache la naissance de l'économie expérimentale à la théorie des jeux et le comportement économique à Von Neumann et Morgenstern (1944) ainsi qu'aux développements ultérieurs de la théorie de la décision. En effet, les travaux de Von Neumann et son co-auteur ont eu pour visée le dépassement des frontières disciplinaires pour résoudre des problèmes issus de conflits inhérents aux relations intergroupes, voire dans les relations internationales ou au sein d'entreprises (2008).

Ainsi, deux méthodes sont-elles privilégiées : la coopération pour partager coûts et bénéfices (branche coopérative) ou le désaccord, où chaque partie se satisfait des conséquences de sa stratégie (branche non coopérative). Au sein de cette histoire entrelacée évoquée plus haut, la Rand Corporation a, également, joué un rôle central dans l'orientation modélisatrice et la recherche économique dans l'immédiat après-guerre (Serra, p. 753).

Morgan (2005) va plus loin en soutenant que les expériences, comme moyen d'investigation du monde économique, offrent un plus grand pouvoir épistémologique que les modèles et estime inutilement limitée la restriction du rôle des expériences à la seule fonction de tester les théories. Il faudra, cependant, attendre les années soixante-dix pour voir émerger l'intérêt croissant pour cette sous discipline que constitue l'économie expérimentale. Plott, à l'origine spécialiste de l'économie politique, et Smith ont fondé, au sein de l'université d'Arizona, le « Economic Science Laboratory » pour y développer leurs travaux sur les enchères secrètes en termes de jeux et l'analyse des données issues de leurs expériences (Serra, p. 761). Pour Guala (2009), les économistes expérimentaux ont cherché, dans les années soixante-dix, à tirer profit du lent épuisement de la théorie de l'équilibre général et la désillusion croissante envers l'économétrie.

La méthodologie de laboratoire permet de passer de la focalisation sur des économies particulières telles qu'on les rencontre naturellement à une focalisation sur les théories, modèles et principes généraux. Pour ce faire, le laboratoire permet d'utiliser simplement, de tester et d'évaluer la capacité de prédiction des théories générales lorsqu'elles sont appliquées aux cas particuliers. Cela n'a pas empêché toutes les controverses et les débats portés sur les faits produits en dehors du

laboratoire et sur la signification de leurs résultats empiriques en plein essor dans les années quatre-vingt attachés à cette économie expérimentale (Cot et al., p. 30). Le lancement de la revue *Experimental Economics* en 1998, débouchera, , en 2002, sur la reconnaissance de l'apport de cette économie par l'attribution du prix de la banque de Suède discerné à Smith et Kahneman. Smith définit le mode opératoire par la fondation d'un système microéconomique défini par un environnement dont on fait subir des variations, d'institutions différenciées et de mesure de la performance résultante (Smith 1982). Smith conclut en épilogue (p. 952) : "Above all, we need to develop a body of knowledge which clarifies the difference between what we have created (theory as hypothesis) and what we have discovered (hypothesis that, to date, is or is not falsified by observation)." Assez naturellement, ces expériences ont été étendues au terrain (pour plus de réalisme) dans le but de faire correspondre ce qui se passe en laboratoire (garantissant le contrôle sur les variables testées) aux contextes réels en prenant, bien entendu, le risque de perdre le contrôle sur ces variables testées avec les menaces pesant sur la validité externe des expérimentations afférentes.

Si les expériences en laboratoire se fondent sur la théorie ou s'assignent pour objectif le test des théories, les promoteurs des expériences aléatoires de terrains s'affranchissent de toute référence à la théorie « un mauvais guide de choix entre programmes concurrents » (Duflo et al., 2009). Ils se proposent d'aborder chaque problème de façon particulière et spécifique en se défiant de « toute approche théorique ou englobante » (Servet, 2018, p. 162), l'unique objectif étant de produire les preuves d'efficacité sans chercher à explorer les mécanismes sous-jacents au risque de s'exposer aux problèmes de validité aussi bien interne qu'externe (nous y reviendrons).

Le *Gold Standard*, dont ils se revendiquent, a fait l'objet de critiques systématiques et appuyées (Leamer, 1983 ; Mookherjee, 2005 ; Basu, 2005 ; Ravallion, 2009 ; Deaton, 2010), Cartwright considérant, quant à elle, que sans ontologie particulière, la valeur probatoire des ECR n'est pas seulement affaiblie mais bien réduite à néant (Guillin, 2013, p. 102).



Nous ne revenons pas sur les limites d'ordre statistique et éthique déjà évoquées plus haut ni sur les nombreux bricolages que les expérimentateurs s'autorisent et encore moins sur tous les biais de publications où ne sont révélées que celles susceptibles de présenter des résultats positifs. Tous ces biais ont fait l'objet d'une très large littérature évoquée pour grande partie dans notre bibliographie qui couvre aussi bien des champs économiques, statistiques, qu'épistémologiques.

A l'examen de l'expérience tant commentée menée sur les vermifuges au Kenya, Barrett et Carter (2010) s'arrêtent sur les problèmes des fausses exogénéités qui altèrent les conclusions. Ils reprennent l'étude séminale sur les vermifuges réalisée par Miguel et Kremer (2004) dont le traitement est basé sur une pharmacopée dans laquelle il était possible de savoir exactement ce qui avait été donné et reçu par le sujet traité.

Laquelle étude avait concerné soixante-quinze écoles primaires rurales à l'Ouest du Kenya où les élèves ont bénéficié d'un traitement vermifuge, les vers intestinaux et la schistosomiase qui infectent une personne sur quatre dans le monde affectant, en particulier, les enfants dans les pays en développement en âge d'être scolarisés. Le programme a réduit l'absentéisme scolaire d'au moins un quart, avec des gains de participation particulièrement importants chez les plus jeunes enfants, faisant du déparasitage un moyen très efficace pour stimuler la participation scolaire des jeunes enfants. S'agissant des externalités interscolaires mesurées par l'impact du vermifuge pour les élèves des écoles situées à proximité des écoles de traitement, lorsque l'on utilise la variation exogène de la densité locale des élèves des écoles de traitement générée par la randomisation au niveau de l'école, on constate que la recherche a pour résultat supplémentaire la réduction de la charge des vers intestinaux et une plus grande présence scolaire des enfants des écoles primaires voisines (Miguel et Kremer 2004). Bien que les auteurs revendiquent des preuves d'externalités de traitement au sein de l'école, et bien que la randomisation ait eu lieu entre les écoles plutôt qu'entre les élèves d'une même école, il a été difficile d'utiliser ces résultats pour décomposer l'effet global sur les écoles de traitement en un effet direct et un effet d'externalité au sein de l'école, et dès lors l'usage de méthodes nécessairement plus provisoires et non expérimentales est nécessaire.

Ainsi, « lorsque la randomisation est utilisée pour des sujets plus vastes, à caractère économique (par exemple, pour modifier les attentes des agents en leur proposant de nouvelles conditions contractuelles ou de nouvelles technologies), le véritable traitement reçu par les sujets devient plus difficile à discerner. Les traitements perçus subjectivement sont probablement répartis de manière non aléatoire entre les sujets expérimentaux dont la capacité à comprendre et à agir varie de façon subtile mais substantielle. Les perceptions inobservables d'un nouveau produit, contrat, arrangement institutionnel, technologie ou autre intervention varient selon les participants et de manière presque certainement corrélée avec d'autres attributs pertinents et les bénéfices attendus du traitement. Le problème de l'hétérogénéité non observée auquel on cherche à remédier peut revenir ».

Il convient de noter par ailleurs, que l'effet de traitement moyen est identifié uniquement pour la sous-population qui se conforme au traitement. Avec la notion *Intention de traiter*, mentionnée plus haut, la randomisation ne permet pas de surmonter le problème. Ce dernier est lié à la question de la puissance statistique où l'expérimentation a, précisément, pour objectif de repérer l'association réellement existante entre le traitement évalué et la variable de résultat considérée, la taille de l'échantillon  $y$  est un facteur déterminant (Fougère, 2012). Les essais thérapeutiques utilisant les ECR n'échappent pas à ce problème. Répondant à Fisher (le père de la randomisation) qui avait écrit que « la procédure complète de randomisation [est la méthode] par laquelle le test de signification peut être garanti contre la corruption par les causes de perturbation qui n'ont pas été éliminées (c'est-à-dire qui n'ont pas été contrôlées avant la randomisation), John Worrall (2002) considère cette affirmation trivialement insoutenable (trivially unsustainable). Pour Worrall, la randomisation contrôle tous les facteurs connus ou inconnus « dans un certain sens probabiliste » car la randomisation peut produire « par hasard » une division entre les groupes de contrôles et groupes expérimentaux qui soit significativement biaisée par rapport à un facteur pronostic non contrôlé qui, cependant, joue dans le résultat thérapeutique (p. S322).

Alors que la littérature épidémiologique et de santé publique fait bien la distinction entre l'impact d'un traitement imposé de manière exogène et le traitement

permettant une endogénéité, les processus endogènes qui guident l'allocation des ressources par les agents humains, que ce soit par les décideurs politiques ou les individus, peuvent en fin de compte altérer la quête d'élimination de l'endogénéité par la conception de la recherche. Lorsque nous imposons des allocations exogènes, nous ne reproduisons pas, en fait, le comportement humain réel, nous violons la proposition la plus fondamentale de la microéconomie : l'allocation des ressources est endogène. (Barrett et al., cité).

En conséquence, les impacts et les comportements obtenus par l'expérience sont généralement endogènes à des conditions environnementales et structurelles qui varient de manière inconnue dans le cadre d'un plan d'expérience nécessairement très stylisé. Cette fausse exogénéité mine les prétentions d'identification propre en raison de la randomisation.

Dans les expérimentations biomédicales, auxquelles s'apparentent les ECR en microéconomie du développement, le fonctionnement est déterminé par des lois physico-chimiques de base qui assurent une approche ascendante intégrant les meilleures preuves externes et les mécanismes impliqués sont plus mécaniques que dans les sciences comportementales et sociales. C'est la raison pour laquelle on pourrait s'attendre à ce qu'elles confèrent un certain degré d'homogénéité de la réponse à une expérience. En économie, il y a peu de raisons de croire à l'homogénéité de la réponse à un changement des conditions environnementales et l'hétérogénéité des micro-environnements est telle qu'il faut être très prudent avec les erreurs de spécification des modèles. Ces préoccupations s'appliquent à toutes les recherches qui, pourtant, semblent particulièrement négligées au sein de la pratique expérimentale telle que pratiquée par le J-PAL.

Nombreux ont été les chercheurs à avoir attiré l'attention sur l'importance du contexte social pour juger de la pertinence des conclusions tirées des programmes de développement (Pawson et Tilley, 1997 ; Rodrik, 2008 ; Deaton, 2010). La signification du concept de « contexte social » reste souvent vague lorsque les chercheurs affirment que le contexte est important pour l'efficacité des programmes de développement. Par exemple, Deaton (2010a, p. 426) plaide pour un recentrage

sur « l'étude des mécanismes potentiellement généralisables qui expliquent pourquoi et dans quels contextes on peut s'attendre à ce que les projets fonctionnent ». Pawson et Tilley (1997) affirment que c'est la combinaison de mécanisme et de contexte qui génère des résultats et, par conséquent, sans comprendre ces combinaisons le progrès scientifique n'est pas possible.

Deaton d'une part, et Pawson et son co-auteur d'autre part n'offrent pas de suggestions quant à l'importance du contexte pour l'efficacité du développement des programmes. A l'inverse, Posner (2011) suggère d'intégrer les caractéristiques de la communauté telles que la densité des réseaux, sa périodicité, son homogénéité ethnique ainsi que ses normes sociales et culturelles dans la conception d'évaluations de programmes de développement.

Les dimensions humaines, historiques, sociales et anthropologiques ne se laissent point dissocier des dimensions qui déterminent les contextes sociaux mais également « la multitude indéfinie des variables qui constituent un contexte social, la non-réurrence de leur configuration (leur historicité ?) et la complexité de leurs interactions » (Allamel-Raffin et al., 2019, p. 11).

Les expériences aléatoires sont, souvent, mises en œuvre dans des contextes singuliers et même artificiels comme pour le cas de l'expérience menée en Inde avec le soutien de l'ONG locale pour contrôler le temps de présence des enseignants par prise de photos qui indiquent date et heure de prise de service. Imagine-t-on une telle expérience en France ou ailleurs dans tout pays développé ? Imagine-t-on les réactions syndicales face à ce qui serait considéré comme affront à des éducateurs ! A titre de commentaire à propos de cette expérimentation, le site du J-PAL, lui-même, considère difficile de tester un tel dispositif de surveillance quotidienne sur des enseignants de l'école publique sans penser, toutefois, que les obstacles seraient insurmontables. Contrairement à l'affirmation que les ECR ont participé à « l'effort visant à développer de manière cumulative notre compréhension des questions économiques concernant les pays et les populations pauvres » (Banerjee et Duflo 2012, p. 61) ; les expériences susmentionnées, et bien d'autres, montrent clairement qu'elles sont « peu susceptibles de découvrir les clés insaisissables du

développement, ni de servir de base à un programme de recherche cumulatif qui pourrait progressivement conduire à une meilleure compréhension du développement » (Deaton, 2009, p. 3).

A examiner de près, cette dissonance épistémologique découle de tensions entre un haut niveau de contrôle, créant des situations artificielles où le chercheur construit une réalité qui n'est pas la réalité et amène ainsi les participants aux expériences à avoir des comportements qu'ils n'auraient peut-être jamais eus en situation naturelle d'une part, et la validité externe des résultats, où la question de la « transposabilité » des résultats pose des problèmes en dehors du contexte de l'expérience d'une part ; l'expérimentateur n'est pas infallible et les sujets ne sont pas que des pourvoyeurs de données, d'autre part. Ces tensions permanentes naissent du déséquilibre entre recherche d'efficacité et découverte de vérités sous-tendues par les visées pratiques, techniques, instrumentales d'un côté, et les soubassements théoriques, explicatives et proprement scientifiques de l'autre. Pour saisir encore la difficulté de ces tensions, il y a lieu de relever que les énoncés de l'observation sont occasionnels et tributaires de données et de procédures construites impliquant la précarité des résultats renvoyant, au demeurant, à l'articulation en observables et inobservables qui mine les capacités d'inférences causales par la contrefactualité.

Nombre de contributions académiques critiques des ECR ont mis en exergue la forte validité interne de leurs résultats, tout en considérant la transférabilité des résultats, associée à la validité externe, problématique (Bloom, 2006 ; Labrousse, 2010 ; Jatteau, 2015 ; Favereau, 2014 ; Bouguen et al., 2014). Pour Rodrik (2008), les évaluations aléatoires sont fortes sur la validité interne, mais produisent souvent des résultats qui peuvent être contestés sur des bases de validité externe. Il considère qu'en revanche, les approches économétriques et qualitatives standard qu'il a appliquées au cas des moustiquaires, objet de l'expérience menée par Dupas et Cohen (2010) à propos de la lutte contre le paludisme par l'utilisation de ces moustiquaires, posent moins de problèmes de validité externe quand, bien même, elles sont jugées plus faibles sur le plan de la validité interne.

Seules les preuves tangibles générées par les RCT concernent des questions dont la portée et l'application sont si limitées qu'elles sont inintéressantes (Rodrik, 2008, p. 5). L'évaluation type est réalisée dans un lieu spécifique, sur un groupe spécifique dans des conditions expérimentales spécifiques et sa généralisabilité à d'autres contextes n'est jamais assurée du fait que sa validité externe n'est pas établie par l'évaluation elle-même. Et les preuves « dures » établies par l'évaluation aléatoire doivent être complétées par des preuves tangibles avant de pouvoir être utilisées (p. 4).

Les conditions dans lesquelles certaines expérimentations ont été menées sont sujettes à caution. Considérons d'abord l'expérience menée au Maroc sur le sujet du microcrédit. Cette étude parue dans une revue prestigieuse (*American Economic Journal : Applied Economics*) et censée démontrer la supériorité intrinsèque des RCT (Bédécarrats et al., 2019). Ayant eu accès aux micro-données de l'étude initiale menée au Maroc par quatre chercheurs affiliés au J-PAL au sujet de l'évaluation de l'impact de l'accès au microcrédit en milieu rural marocain, Bédécarrats et ses co-auteurs ont procédé à la réplication de cette expérience par suite d'une commande passée par l'AFD. Forte des résultats de ce travail, l'équipe du J-Pal avait conclu que le microcrédit peut être un puissant instrument financier pour sortir de la pauvreté grâce à un meilleur investissement dans l'auto-emploi, au moins à moyen terme (deux ans après l'introduction du programme) (Crépon et alii, 2015, p.148). Signalons d'abord, la différence entre l'analyse nuancée de Banerjee et Duflo sur les bienfaits de la microfinance, mentionnée au chapitre II de ce travail, et les conclusions auxquelles aboutit l'expérimentation en question.

Ensuite, Bédécarrats et ses co-auteurs montrent que le « protocole de la RCT et les données collectées sont tellement entachées de problèmes et d'erreurs que les conclusions qui en sont tirées questionnent aussi bien la validité interne qu'externe » (p. 5). La réplication de l'expérience menée met en évidence la multiplicité d'erreurs qui minent profondément sa validité. La qualité des données est jugée douteuse, l'identification du groupe de traitement est sujette à caution et la zone d'enquête prétendue vierge ne l'était pas. Le diagnostic de la réplication est jugé si alarmant que ses auteurs se demandent comment de telles défaillances



n'aient pas conduit à un rejet de l'article par la revue ou au minimum à relativiser la portée des résultats par des mises en garde.

L'autre article mentionné également dans le Journal précité, concerne une expérience indienne de microcrédit. Elle y est présentée, par ses auteurs, comme la plus ancienne évaluation du prêt de groupe à faire connaître la microfinance. L'évaluation aboutit au constat que la demande d'accès aux produits du microcrédit n'est pas universelle contrairement aux prétentions des institutions qui y sont engagées (IMF et ONG). Seuls un tiers des ménages sélectionnés en fonction de la propension à recourir à l'emprunt le font auprès d'une IMF. Une explication avancée serait liée au fait que la plupart des emprunteurs n'ont pas de projets réalisant un taux de rendement supérieur à 24 % - taux effectif appliqué par l'IMF Spandana, partenaire de l'étude -. L'autre constat est l'inadéquation entre l'amélioration du bien-être et l'accès précoce au microcrédit. De plus, il n'y a pas eu d'effets sur l'autonomisation des femmes ni de résultats tangibles pour le développement au bénéfice des pauvres (Banerjee et al., 2015). Ainsi, le microcrédit ne serait pas le miracle que l'on prétend être quand bien même il permettrait à certains d'entreprendre. Les adeptes du microcrédit surestiment son potentiel d'être une source de revenus et moyen d'autonomisation des femmes (pp. 51-52).

Dans cette expérimentation, l'absence de références académiques qualitatives ou économétriques menées au sujet de la microfinance en Inde, interroge. L'Inde représente, tout de même, un des plus grands marchés du monde avec son potentiel de trois cent vingt millions de clients potentiels où l'on estime les clients non bancarisés à près de six cent vingt millions (Guerin et al., 2008). Ce pays a été un champ très large d'investigation sur le sujet ces trente dernières années et, pourtant, l'ostracisme attesté par le seul recours à l'auto-référencement prive ses promoteurs de connaissances et savoirs disponibles et nécessaires aux heuristiques interprétatives et explicatives. Michel Servet (2018) impute l'engouement pour les expérimentations à l'extension du néolibéralisme qui, au nom du progrès et de l'efficacité économique, est responsable de la « chosification » de la personne humaine. A l'instar de la société liquide bien décrite par Zygmunt Bauman (2013), qu'il cite, la finance façonnant du global au local a pour effet de se protéger contre

les risques inhérents à la fragilité et a, fortement, contribué à ancrer une culture de l'endettement en Inde où « l'être humain peut y être pensé comme un tissu de dettes aux multiples connotations » (p. 82). Comme mentionné plus haut, les promoteurs des RCT, forts de la croyance en la supériorité des preuves fournies par leurs tests, font peu de cas des contextes sociaux ou anthropologiques et se préoccupent moins des enquêtes et des observations qui donnent à voir les déterminants et les logiques qui gouvernent les habitudes des populations comme, par exemple, la détention de bien physiques, tels l'or ou l'épargne en nature, souvent considérés risqués (risky assets) et cette détention n'est, généralement, pas prise en compte. S'y ajoute, également, le fait que les autres modalités de conservation de la monnaie sont ignorées. La prépondérance du cash s'explique, souvent, par l'informalité qui donne le sentiment que les projets ne peuvent s'évaluer qu'à l'aune des solidarités qui affranchissent des soucis du rendement (pp. 93-94).

De même, est ignoré le rôle déterminant que jouent les usuriers par la minimisation des coûts de structures formelles et de collecte des remboursements, l'abolition de cette structure d'agents intermédiaires bute sur l'extraordinaire souplesse qu'offre l'informalité face aux rigidités des réseaux formels et ses coûts cachés. Guerin et ses co-auteurs (p. 134) considèrent que « les dettes s'inscrivent dans un tissu de relations sociales qu'[ils] actualisent et reproduisent, si bien que toutes tentatives d'abolition demeurent vaines ».

Tous aspects non pris en compte par les randomisateurs permettent d'approfondir l'intelligibilité des situations qui semblent leur échapper et leur font voir l'envers des événements que l'histoire ne montre pas selon les bons mots de Chateaubriand dans *Mémoires d'outre-tombe*. C'est cette non prise en compte des savoirs établis qui sape aussi bien la validité interne qu'externe des ECR enserrés dans ce tournant pratique qui s'opposant, selon Lena Soler, aux produits constitués et stabilisés de la science a déplacé la tonalité des discours de cette même science et n'a finalement produit que des résultats nécessairement contingents. En se focalisant sur le calcul des moyennes, les expérimentateurs randomisateurs saisissent bien l'hétérogénéité des déterminants mais ne peuvent rendre compte du caractère hétérogène des résultats. Quand ils appellent à répliquer les expériences pour pallier les difficultés de la validité interne comme externe, ils s'exposent alors aux disparités normatives

en menant à des résultats susceptibles d'être peu conciliables voire contradictoires. La réplication de l'expérience de microcrédit menée au Maroc, montre bien les limites de cet exercice de réplication censé affranchir des difficultés à valider la généralisation des expériences. Soler (2011, p. 67) estime que l'exigence de reproductibilité des expériences afin de vérifier et de revérifier à volonté les résultats est un impératif scientifique difficilement atteignable en sciences humaines.

Le poulet de Bertrand Russell (Russell, 1912) fournit, à cet égard, un excellent exemple des limites de l'extrapolation à partir de répétitions réussies. A partir de preuves répétées, l'oiseau déduit que chaque matin le fermier ne vient que pour le nourrir. Cette induction lui est utile jusqu'au matin où on l'égorge pour le dîner de Noël. Ce sont les concepts théoriques qui concourent à asseoir les validités interne et externe, les tactiques déployées pour s'en assurer convergent pour leur renforcement mutuel (Ayerbe et Missonier, 2007) et le passage du descriptif à l'interprétatif ou l'explicatif se paie de ce prix.

Dans « *Critique de la raison pure* », Kant (1846) écrit que « la raison ne voit que ce qu'elle produit elle-même d'après ses propres plans [...], elle doit obliger la nature à répondre à ses questions et ne pas se laisser conduire, pour ainsi dire, en laisse par elle » (Kant, 2012, p. 17). A cette aune, considérer que les études menées par le J-PAL « font clairement apparaître que [notre] intuition (ou la théorie économique en tant que telle) est un bien mauvais guide du choix des programmes concurrents qui visent le même objectif : comment alors [aurions- nous] deviner à priori que le traitement contre les parasites intestinaux est beaucoup plus efficace que le recrutement d'un professeur supplémentaire ? » (Banerjee et Duflo, 2009, p. 693) est heuristiquement surprenant et ontologiquement sans fondement épistémologique. On pourrait, en effet être surpris par la comparaison entre traitement contre un parasite intestinal et le recrutement d'un professeur alors que la visée de l'étude mentionnée était de démontrer l'amélioration de l'assiduité scolaire corrélativement au traitement dont les fondements biologiques sont établis par la théorie scientifique et à la base des remèdes adéquats à la pharmacopée utile contre les parasites intestinaux (Abdelghafour, 2017). Se prévaloir de la rigueur des

RCT pour cette expérience menée par Kremer et Miguel, désignée comme la « flagship study of the new movement in development economics » (Deaton, 2010, p. 446), alors même que la sélection des écoles, opérée sur base d'une sélection par ordre alphabétique et non par randomisation, est un fait surprenant, s'agissant de recherche académique. D'ailleurs, la méthode réelle utilisée n'est pas décrite dans les documents, où pourtant, le mot aléatoire et ses dérivés « apparaissent une soixantaine de fois ». Deaton rapporte, également, que l'autre étude menée par Glewwe et al. (2004) est, également, issue d'une expérimentation non randomisée.

Ce réel construit, et saisi comme donné, heurte la raison qui porte en elle-même les principes d'objectivité et d'universalité et pour paraphraser Kant ne « doit [s'y] laisser conduire en laisse ». Ainsi la portée universelle au cœur du mouvement des RCT se focalise-t-elle « sur ce qui marche » pour constituer un réel destiné à produire des résultats qu'à justifier ce qui pourrait en être fait. Duflo et Kremer (2004, p. 32) n'hésitent pas à ranger ce réel dans la catégorie des biens publics dont la raison n'en est pas moins que de constituer la source des connaissances universelles et nécessaires. Selon ces auteurs, les biens publics internationaux ainsi constitués offrent l'avantage de connaître, au-delà de l'organisation du pays concerné, si un programme fonctionne ou pas. Toujours selon ces auteurs, les programmes efficaces peuvent être adaptés à d'autres pays ou étendus à l'intérieur des pays et ceux qui échouent sont abandonnés. Aussi, plaident-ils pour que la promotion, l'encouragement et le financement de ces évaluations rigoureuses (entendez les RCT) puissent servir aux organisations internationales portant les programmes d'évaluation et indiquer ainsi aux bailleurs, aux ONG et aux gouvernements la voie qui permettrait de contrecarrer le scepticisme qui entoure l'aide au développement. Cette profession de foi vise, à assurer les financements des RCT et, peut-être autant, la production de « bien publics » dont aucun sous-jacent n'est fourni à son appui (aspect qui sera traité au chapitre IV). Elle tranche, en tout cas avec toute la littérature consacrée en soixante ans au développement. En dépit des différences d'approches, les économistes du développement se sont accordés pour considérer que travail, capital, ressources et innovation doivent se combiner, sans quoi, ces composantes du progrès se renforcent les unes les autres et plongent les pays pauvres dans le cercle vicieux du sous-développement. Le

franchissement des barrières de la pauvreté demande un effort concerté et coordonné sur de nombreux fronts pour provoquer la forte poussée nécessaire au développement et la généralisation du bien-être au profit du plus grand nombre.

Le pragmatisme théorique, dont se réclament les promoteurs des ECR, ne les empêche pas de recourir aux modèles (Neyman, Rubin et Holland) utilisés pour inférer des relations et d'en mesurer les effets attendus. Pour accroître leur légitimité, les ECR se focalisent sur une empirie dont l'inspiration méthodologique se nourrit plus de l'analyse statistique que de la théorie économique (de Curraize et al., 2018). Par la variété des thématiques abordées par les ECR, Labrousse (2010, p. 27) trouve leurs fondements théoriques « flous, disparates et éclatés en une multitude de propositions à géométrie variable » et, par ailleurs, distendus sur le plan épistémologique. Faute de guide théorique, cette méthodologie se confine dans l'impasse et se couvre d'un voile d'ignorance (*Veil of Ignorance*) selon l'expression de Heckman (1992) attribuée à Vickrey (1945, 1960).

La thèse de Duhem (1981) pointe l'impossibilité de tester une hypothèse isolée du fait que le verdict d'une expérience ne s'exerce pas sur l'hypothèse isolée mais sur tout un ensemble d'hypothèses. On peut conclure à la validité d'une hypothèse, mais, l'expérience n'indique point celle qui doit être changée ou non. Quine (1951), dont l'essence de la thèse est la sous-détermination de la théorie par l'expérience, considère que des réajustements sont nécessairement entraînés par quelque expérience à la périphérie et que les valeurs de vérité doivent être redistribuées sur certains énoncés dont la réévaluation entraîne, en raison de leurs interconnexions logiques, celle d'autres. Selon lui, aucune expérience n'est liée à un énoncé particulier à l'intérieur du champ sauf indirectement par des considérations d'équilibre affectant le champ dans son ensemble.

Guala (2012, p. 606) exprime la thèse Duhem-Quine autrement : « les résultats expérimentaux n'indiquent pas, en général, de manière déductive la véracité/fausseté d'une hypothèse théorique isolée, mais, plutôt tout un ensemble de connaissances ou d'un groupe d'hypothèses théoriques et auxiliaires à la fois ». La thèse vaut pour les domaines qui traitent aussi bien les sciences dites de la nature que les sciences humaines. L'homme apparaît alors comme objet d'investigation et

d'étude, à la fois être naturel déployant des mécanismes de vie dans le corps et réalité culturelle interagissant avec ses dimensions biologiques et culturelles. Qu'il concerne l'homme, son environnement ou son milieu naturel, le débat expérience/théorie est très ancien. L'histoire de la philosophie des connaissances fait remonter leur confrontation aux temps immémoriaux. Pour Platon et son allégorie de la caverne, la connaissance qui relève de l'expérience est une connaissance imparfaite et décevante (Kahn, 1988, p. 8). La connaissance vraie n'est accessible que par « la vue de l'intelligence » qui élève à la vérité et que l'on appelle *theoria*. Dans cette vision platonicienne, la théorie est vue comme la traduction immédiate des idées, opposée à l'empirie du rapport sensible, mais la valorisation de celle-ci « ne conduit pas à une forme dévoyée d'idéalisme, qui verrait dans la seule capacité d'abstraction du sujet la source privilégiée de la connaissance » (Dekens, 2005, p. 11). La théorie est alors, à minima, le médium à partir duquel se construit toute expérience objective sans que cela puisse impliquer nécessairement que la théorie soit première dans l'ordre épistémologique. Sans quoi, dire le réel sans conceptualisation relèverait du solipsisme.

Sans enjamber les siècles, on ne peut ignorer la position de l'élève et non moins alter-ego de Platon, Aristote, qui défend une position plus proche des empiristes et considéré, par conséquent, comme le philosophe de l'expérience. Dekens alerte, cependant que ce considéré n'est pas à prendre comme « le primat ontologique de la sensibilité, mais bien au nom d'une détermination de la connaissance comme élaboration théorique prenant source dans l'expérience de la singularité » (p. 13), la singularité n'empêche pas Aristote de considérer qu'un savoir n'est scientifique que s'il a une portée générale.

Plus tard, à partir du 17<sup>e</sup> siècle, bon nombre de philosophes ont cherché à faire face au poids de la tradition et instituer la raison comme mode de gouvernement de tous les domaines de l'esprit. Cette institution de la raison et de la rationalité s'appuie en particulier sur l'appréhension et le contrôle des forces et ressources de la nature. Initié, en particulier par Descartes et John Locke, ce mouvement se modifie en profondeur par les contributions de Kant qui, invitant à penser par soi-même, distingue la raison pratique comme règle morale et la raison pure comme produit de



la pensée et non de l'expérience, lorsqu'il s'agit de la connaissance. Contrairement à Descartes, Locke dans *Essais sur l'entendement humain* considère que l'esprit de l'homme étant vierge à sa naissance (« tabula rasa »), les idées ne sont pas innées mais proviennent de l'expérience fondement de toutes connaissances. Les encyclopédistes étendent le pouvoir de la raison en rassemblant systématiquement les connaissances humaines à l'objet d'améliorer concrètement le sort de leurs semblables en s'autodéterminant dans le cadre d'une société éclairée. D'Alembert et les encyclopédistes invitent à s'appuyer sur l'expérimentation dans le progrès de la science en s'écartant du rationalisme cartésien sans dénier à la théorie son rôle fondamental. Précisons que ce rationalisme cartésien ne rejette point l'expérience qui n'est utile, par la régularité de ses observations et son inférence généralisatrice, que déductive et dérivée d'une théorie (p. 17).

Signalons, enfin, pour ce rétrospectif historique sur la tension entre théorie et expérience, Francis Bacon que l'on considère comme le promoteur de l'expérimentation conçue comme modalité de la recherche de la vérité. Pour l'auteur du *Novum organum* l'induction, pour permettre le passage du particulier à l'universel, doit avoir la capacité à trancher entre théories concurrentes, esquissant ainsi la théorie de l'expérience cruciale (*instantia crucis*).

Sur cette période, à travers un prisme original et un intérêt particulier à la mécanique classique, Alexandre Koyré (1973) apporte un éclairage intéressant sur l'histoire des sciences, où, le « vrai savant » doit se soumettre à l'épreuve de l'expérimentation, le premier temps de la recherche appartient à la théorie et non à la praxis (Rochot, 1966, p. 390).

Pour l'illustrer, Bernard Rochot utilise une métaphore expressive : « on ne règle pas le baromètre en pleine tempête ; mais une fois bien réglé, il aide à prévoir la tempête... Bref, les questions auxquelles la science répond sont celles qui sont bien posées, et c'est parce qu'on s'est mis à bien les poser à la lumière des mathématiques, et non seulement selon la hiérarchie des besoins ou des intérêts économiques, que la période sur laquelle portent les présentes études a été féconde » (p. 391).

Les sciences sociales ne peuvent échapper à ce questionnement épistémologique. Elles recouvrent différentes disciplines (économie, sociologie, ethnologie, anthropologie, histoire, psychologie...) au sein desquelles s'opère la production des connaissances avec des interprétations et croisements ayant, cependant, les mêmes objets d'étude : la société, l'individu et leurs interactions. Toutes ces disciplines sont marquées par la tension entre l'induction et la déduction d'une part, et par l'analytique et le synthétique d'autre part à l'intérieur d'une tension holisme/individualisme et où « l'individu et la société, le sujet et la structure font figure de référents déterminants, exposés à des accentuations et à des interprétations diverses qui, d'emblée, semblent faire sens » (Valade 2012, p. 357). Valade fait remonter le rôle de cette antinomie holisme/individualisme dans l'accumulation de la connaissance des faits sociaux aux traitements qu'en ont faits Montesquieu et Rousseau. A grands traits, la ligne de démarcation s'établit comme suit : pour l'auteur de « L'esprit des lois », il faudrait s'attacher uniquement aux faits humains d'où l'on déduit l'observation de ces faits, de l'activité sociale qui les engendre et des vicissitudes passées et présentes de ces mêmes sociétés. Pour l'auteur du *Contrat social*, il faudrait s'en tenir au raisonnement pur et la tâche de gouverner s'assimile alors à résoudre un problème d'algèbre. Montesquieu vise une explication globale et Rousseau part de l'individu pour recomposer la totalité sociale (pp. 361-362).

Qu'elles se situent dans des perspectives holistes ou individualistes, les sciences sociales s'attachent à répondre efficacement à la question du comment vivre ensemble selon l'injonction de Maurice Allais. L'écart entre leurs visées compréhensives et leurs capacités explicatives est au cœur de la réflexion épistémologique, théorique et leurs pratiques interprétatives. Ainsi, « tant l'étude historique minutieuse des controverses ou des travaux scientifiques que la description méticuleuse du travail au jour le jour des chercheurs ne font qu'accroître la béance entre la science telle qu'elle se fait et telle qu'elle se dit, entre la réalité de son inscription concrète et l'idéalisation de sa représentation normative » (Berthelot, 1998, p. 28).

Au sein des sciences sociales, l'économie a pris une trajectoire autonome ; on situe celle-ci d'Adam Smith à sa forme aboutie que représente l'équilibre d'Arrow-

Debreu (Walliser, 2012). Bien que d'essence holiste, où le tout (enrichissement des nations) est supérieur à ses différentes parties (agents, institutions, production, consommation, échange, revenu, monnaie, emploi, etc.), ses visées ont pour objet la coordination des actions d'agents rationnels coordonnés par des institutions diversifiées, dont en particulier le marché, guidée en cela par l'observation brute de données empiriques et la proposition d'abstractions explicatives ou interprétatives, voire prédictives des phénomènes mis en évidence. Pour ce faire, elle saisit des processus sociaux complexes par le biais d'équations ou de causalités claires comme preuves d'engagements scientifiques supérieurs justifiant sa distance par rapport aux disciplines plus discursives (Fourcade et alii, 2015). Elle est à la fois positive du fait qu'elle nous renseigne sur ce qui est, normative car elle fournit les connaissances sur ce qui doit être. Elle est également un art, qui selon John Neville Keynes, permet d'utiliser les leçons tirées de l'économie positive pour atteindre des buts normatifs (Cozic, 2011).

Contrairement aux sciences naturelles, étant donné la manière dont elle s'exprime, les questions qu'elle sélectionne, les questions qu'elle pose et les réponses qu'elle y apporte, l'économie n'est pas axiologiquement neutre. Elle est toujours imprégnée de jugements de valeur y compris pour ses dimensions évaluatives. La discipline fut longtemps marquée d'une part par ce que Black, cité par Blaug (1994, p. 113), a nommé la « *guillotine de Hume* » énonçant qu'on ne « peut déduire ce qui doit être de ce qui est », consacrant cet art de Keynes (John Neville) à propos de l'incapacité du descriptif à amener le prescriptif. En réalité, Keynes isole trois activités distinctes : une économie positive qui cherche à découvrir des lois (what is), une éthique économique qui fait émerger des idéaux économiques (what ought to be) et enfin un art ou une technique pour y arriver (Mongin, 2007). D'autre part, et nous l'évoquions plus haut, l'économie est longtemps restée gouvernée par l'interdit millien édicté au milieu du 19<sup>e</sup> siècle où John Stuart Mill très critique, par ailleurs, envers la logique déductive considère pourtant, que pour les sciences morales (comprendre les sciences sociales), les méthodes inductives sont inapplicables en raison de « la fréquence des effets composites de nombreuses forces » (p. 62). Il conclut ainsi à l'impossibilité de procéder à des expériences contrôlées par le jeu des causes variées et de l'interdépendance d'effets séparés.

Nous ne revenons pas sur les méthodologies néo-positives nées des réticences vis-à-vis du caractère défensif du déductivisme millien ayant trait aux particularités épistémologiques de l'économie (p. 498). Dans cette perspective, on se contente de signaler, ici, l'opérationnalisme basé sur les « théorèmes opérationnellement significatifs » (Blaug, 1994, p. 87) ou bien « la théorie des préférences révélées » de Samuelson (Cozic, 2011, pp. 498-501). D'autre part, on évoquera également « la thèse de la non-pertinence des hypothèses » où Friedmann considère que le seul critère pertinent à la validité est à jauger à la capacité prédictive des théories qui le portent (Blaug, p. 91), celles-ci ne sont pas à jauger par le réalisme de ses hypothèses, les théories ayant des visées prédictives et non explicatives.

A l'appui de sa thèse, Samuelson cité par Blaug (p. 88), soutient que : « L'utilité d'une théorie vient du fait que, par notre analyse, il est souvent possible de déterminer la nature des variations de nos variables inconnues résultant d'une variation déterminée d'un ou plusieurs paramètres. En fait, notre théorie n'a pas de signification au sens opérationnel si elle n'implique pas des restrictions sur les quantités observables, par lesquelles on peut concevoir qu'elle soit réfutée ».

Selon Cozic, Friedmann et Savage (1948) explicitent la règle épistémologique de l'irréalisme des hypothèses comme suit : « l'hypothèse n'affirme pas que les individus calculent et comparent les utilités espérées explicitement ou consciemment... l'hypothèse affirme plutôt que, dans une classe particulière de décisions, les individus se comportent comme s'ils avaient calculé et comparé l'utilité espérée et comme s'ils connaissaient les chances. La validité de cette affirmation ne dépend pas de la question de savoir si les individus connaissent les chances précises, encore de celles de savoir s'ils disent qu'ils calculent et comparent les utilités espérées ou s'ils pensent qu'ils le font... mais seulement de celle de savoir si elle engendre des prédictions suffisamment précises concernant la classe de décision sur laquelle porte l'hypothèse » (Cozic, 2010, p. 30).

De l'examen de la méthodologie économique traitée dans son ouvrage hyponyme, Blaug (p. 5) écrit que : « les lois universelles sur lesquelles reposent les explications ne sont pas extraites grâce à des généralisations inductives des cas particuliers ; ce sont de pures hypothèses, des conjectures faites par des auteurs... ». Contrairement

au vérificationnisme issu de l'idée que l'économie repose sur des propositions donnant lieu à des conclusions qui sont vraies tant qu'elles ne sont pas exposées à des « causes perturbatrices particulières » (Blaug, p. 54), les néopositivistes évoqués, ci-dessus, relèvent de la catégorie des « infirmationistes » dans la tradition poppérienne.

Signalons enfin la notion de capacité développée par la philosophie des sciences où pour Nancy Cartwright (2007 et 2009), « beaucoup d'affirmations causales, scientifiques ou non, sont des attributions de capacité, les lois de la répulsion et de l'attraction électromagnétique, comme la loi de la gravité, et, un grand nombre d'autres lois sont des lois qui portent sur des tendances persistantes ou capacités » (Cozic, 2010, p. 43). Cartwright (2009, p. 47) soutient qu'il n'est pas possible d'apprendre à propos des capacités à partir de modèles cherchant à les isoler... Or, la petite quantité de principes économiques acceptable ainsi que le fonctionnement spécial des capacités économiques sont les raisons pour lesquelles il ne faudrait pas espérer beaucoup de la modélisation concernant les capacités à l'œuvre dans le cas de phénomènes économiques. Ainsi en est-il des modèles économiques qui, comme c'est le cas pour les expériences galiléennes, l'objectif est d'isoler une cause particulière qui puisse opérer correctement et permettre à la capacité de se manifester en enlevant les contraintes entravant son action.

Ce panorama concis de l'histoire de la méthodologie économique et ses implications d'ordre épistémologique nous amènent à la question de la portée et la validité de l'économie expérimentale, comme sous discipline de l'économie, à dire la vérité scientifique. Longtemps et suivant en cela le constat de Samuelson (1985) que pour l'économiste il n'est pas possible de se livrer à des expériences contrôlées, à l'instar du chimiste ou du biologiste, il devrait se contenter de « observer ». Or, depuis au moins quatre décennies et, plus particulièrement encore, depuis l'attribution en 2002 par la banque de Suède du prix Nobel en sciences économiques à Daniel Kahneman et Vernon Smith, l'économie expérimentale, utilisant l'expérimentation comme modalité d'investigation, s'est imposée en courant majeur de l'économie. Les expérimentations peuvent être de terrain et sont « naturelles » ou se dérouler en laboratoire et, alors, elles sont « artificielles ». Lorsque l'expérimentation en

laboratoire développe un environnement et des institutions suffisamment artificiels pour être facilement contrôlés, elle s'expose au risque de perdre en représentativité ; alors que l'expérimentation de terrain cherche à prendre en compte le plus possible les contextes réels pour assurer la signification statique des sujets expérimentaux, elle s'expose au risque de perdre le contrôle des différentes variables mises en jeu (Cot et al., 2016).

Cette économie expérimentale se donne pour objectifs : le test des théories, la production de nouvelles connaissances et l'aide à la décision (Eber et Wellinger, 2005, p. 13). Par rapport aux autres méthodes empiriques, l'expérimentation vise à contrôler l'environnement et assurer la répliquabilité. Pour Plott, les expériences ont eu tendance à être rejetées, soit parce qu'elles étaient incomplètes en tant que simulations, soit trop compliquées si bien que les tests des modèles n'ont pas convaincus. En d'autres termes, l'expérience était récusée soit parce qu'elle ne reflétait pas un processus naturel, soit parce qu'elle le faisait. Pourtant fort favorables à l'économie expérimentale, Eber et Wellinger précisent que le niveau de contrôle est loin d'être comparable à celui du physicien ou du chimiste et qu'« un tel degré de contrôle de l'environnement n'est tout simplement pas possible dans la mesure où chaque sujet vient dans le laboratoire avec sa propre histoire et constitue en soi un objet très complexe ». Il s'en suit inéluctablement une interaction entre l'histoire personnelle du sujet et les conditions créées par l'expérimentateur.

Quant au problème relatif à la répliquabilité, nous renvoyons simplement à ce que nous évoquions auparavant à propos de l'exemple rapporté par Bédécarrats et ses co-auteurs, sur le microcrédit au Maroc pour ce qui concerne les expériences aléatoires contrôlées.

Nous opérons, cependant, la distinction entre les expériences de laboratoire plus tournées vers la théorie dans des perspectives de réfutation et les expériences aléatoires contrôlées (EAC ou RCT) tournées vers l'évaluation *ex-ante* des politiques publiques et la lutte contre la pauvreté dans des perspectives de confirmation. Nous nous concentrons sur le cas de ces dernières pour souligner l'impasse épistémologique dans laquelle elles se trouvent.



Rappelons que l'intention de la méthodologie des EAC est de nature inférentielle cherchant à déterminer de façon précise les conséquences d'une intervention ou d'un traitement selon la terminologie consacrée. Le modèle causal, rappelons-le encore, permet la détermination de la variable d'intérêt et la comparaison entre le groupe traitement et le groupe témoin, les deux groupes étant tirés aléatoirement en mobilisant la notion de contrefactuel, et ce, à l'effet d'éliminer tout biais de sélection.

Nous avons examiné le long de ce chapitre les nombreux et variés biais auxquels fait face la randomisation, qu'ils relèvent des modes d'échantillonnage, de puissance statistique, de comportement des sujets, des biais de publication ou simplement des bricolages rapportés très largement, aussi bien, par la littérature critique que par celle qui lui est favorable. Nous avons susmentionné les limites éthiques provoquées par l'approche conséquentialiste qui considère qu'aucune action n'est répréhensible en soi, l'éthique étant une pratique qui ne peut être évaluée que par rapport à ses conséquences. De même, avons-nous souligné ce que Ryoa Chung appelle l'asymétrie épistémologique inhérente à la relation inégale de pouvoir et de savoir entre le sujet de la recherche expérimentale et le chercheur. A ces asymétries s'agrègent les injustices structurelles que l'apolitisme, revendiqué par les promoteurs des EAC, tend délibérément à ignorer l'amont responsable des causes à traiter.

Ces limites interrogent également le positionnement du chercheur face aux attentes des partenaires des expérimentations et des bailleurs de fonds qui font de cette méthode un prérequis pour le financement de projets.

Ces manquements aux canons de la validité statistique entament à eux seuls la crédibilité de la validité interne conférée par la randomisation puisque très souvent l'inférence qui en est faite ne concerne que la population échantillonnée et non la population cible. L'idéal du *gold standard* n'est jamais atteint. La multiplicité de facteurs sociaux, politiques ou historiques, faisant rencontrer divers mondes sociaux, provoque des phénomènes susceptibles d'inférer avec l'indicateur d'intérêt. Les ECR sont alors des artefacts qui produisent des résultats contingents et les

changements observés disparaissent dès que la procédure qui les a fait naître disparaît.

Plus fondamentalement encore, les critiques de la communauté scientifique se concentrent sur la question de la validité externe et les difficultés de généralisation. Nous pensons que les problèmes de validité externe sont consubstantiels à ceux induits par les apories affectant la validité interne.

Les cadres analytiques et méthodologiques, empruntés à l'épistémologie et à la philosophie des sciences, déjà évoqués, situent clairement l'économie comme science fondamentalement sociale qui poursuit une démarche hypothético-déductive. A l'observation se succède la déduction, le matériau empirique s'analysant dans le cadre d'un enchâssement théorique.

Après avoir considéré que la théorie est un mauvais guide, Duflo estime que « la théorie inspire les chercheurs (mais les expériences) peuvent ensuite la contredire ; la validité de l'expérience ne repose pas sur la validité de la théorie » (Duflo, 2009, pp. 66-67). Pour elle et son courant, chaque expérience apporte un nouvel éclairage et pose de nouvelles questions et c'est, ainsi, que se justifie le nombre élevé d'expériences menées par chacun des principaux sociétaires du J- PAL.

René Thom (1986, 1988) auteur d'un article séminal sur la problématique de la rencontre théorie-expérience, considère qu'il n'y a pas de méthode expérimentale et qu'il n'y a méthode qu'attachée à une théorie (explicite ou implicite) préexistante qui prescrit les procédures protocolaires d'expérimentation : « sauf heureuse surprise, tant vaut la théorie, vaut l'expérience ». Thom rappelle à cet égard les préceptes, qu'il qualifie de quasi-cathéchique, de Francis Bacon et Claude Bernard, deux théoriciens de l'expérimentation : possibilité de reproduire les faits, se défier des idées préconçues et honnêteté de la pratique. A examiner de près et à partir des nombreuses expériences aléatoires contrôlées que nous avons étudiées, et partant des controverses qu'elles ont suscitées et rapportées ou présentées dans ce travail, il s'avère que ces préceptes n'ont pas toujours été observés.

Lorsque Duflo observe que : « Les expériences du terrain ont un pouvoir subversif que n'ont ni les études rétrospectives ni les expériences de laboratoire. Elles forcent

à la fois scientifiques et acteurs de terrain à accepter d’être contredits et surpris », la surprise invoquée n’étant pas exprimée par rapport à un corpus préexistant de connaissances ou de théories. René Thom constate, au demeurant, que les découvertes par les faits de sérendipité ont été courantes à l’instar de la pénicilline de Flemming et, dans ce cas, se pose alors la question de l’explication qui régit le fonctionnement de l’agent invisible (ici, l’arrêt de la croissance bactérienne dû aux sécrétions de *Penicillium notatum*).

Le succès du J-PAL est à rechercher du côté de « la multinationale scientifique » développée par Pollak (1979) à propos de Lazarsfeld, symbole de l’empirisme apolitique où ce dernier « ... s’appuyant sur la politique des fondations, a contribué à exporter à l’échelle de l’Europe de l’Est et de l’Ouest la sociologie empiriste américaine en même temps que le modèle des relations entre recherche sociale et agents économiques et politiques qui s’était élaboré aux États-Unis pendant les années vingt et trente du siècle dernier, par une stratégie de conquête de nouveaux marchés intellectuels qui coïncidait avec les stratégies de modernisation et de reconquête idéologique de l’Europe » (exergue de l’article).

Jatteau (2013) impute le succès de l’entreprise J-PAL à la promotion d’une cause : la randomisation en s’appuyant sur un fonctionnement hiérarchique, mobilisant des financements favorables et organisés en vue de mener une offensive méthodologique par une élite composée de professionnels aux affiliations institutionnelles prestigieuses.

En plus des apories liées à l’identification causale, à la pléthore d’empirie dépourvue de soubassement théorique et aux biais qui affectent la méthodologie en question, nous concluons ce chapitre en évoquant quelques appréciations provenant aussi bien d’auteurs critiques que de partisans des RCT, et formulées pour souligner la faiblesse de la chaîne argumentative des expérimentations aléatoires. Howard White (2009, abstract), directeur exécutif de « l’international Initiative for impact evaluation », organisme engagé dans le soutien et le financement des évaluations de politiques publiques considère que les approches ancrées dans la théorie sont faibles : « application of a theory-based approach remains weak ».

Deaton (2010) et Guillin (2013) signalent que la validité des preuves produites par l'*evidence based-policy* dépend du maillon de la chaîne argumentative le plus faible et relativisent son avantage revendiqué : “As Cartwright notes, the validity of evidence-based depends on the weakest link in the chain of argument and evidence, so that by we seek the experimental results, the advantage of RCTs over matching or other econometric methods has evaporated” (Deaton, 2010, p. 450). En complément, Vincent Guillin ajoute : « Dans bien des cas, nos difficultés ne sont pas dues au fait que nous n’avons pas encore trouvé la théorie appropriée (des capacités), mais plutôt qu’il n’y a pas de théorie à trouver. (...) Il n’y a pas de nécessité que ce qui est produit par une cause dans une situation donnée ait une quelconque relation systématique avec ce qu’elle fait dans une autre situation. Une bonne part des phénomènes sociaux pourrait bien être trop holistique pour que la méthode analytique puisse y être employée ». Et de préciser : « en l’absence d’une théorie des capacités des causes que nous étudions, la valeur probatoire d’un ECR n’est pas seulement affaiblie, mais bien (...) réduite à rien » (Guillin, 2013, p. 102).

En fondant leur légitimité sur les méthodes statistiques au détriment de la théorie économique, les RCT débouchent sur un « gigantesque bric-à-brac d’effets » selon London (2009) cité par Labrousse (2010, p. 13). Cette auteure impute la réduction des références théoriques à « la circulation réticulaire entre un nombre réduit de chercheurs (...), les chaînons théoriques manquants dans l’argumentation du J-PAL limitent à ce jour la portée de la recherche-action : sans théorisation fouillée et systématique des mécanismes à l’œuvre, les effets de connaissances sont réduits ».

Les limites méthodologiques du modèle causal sur lequel s’appuient les expérimentations aléatoires contrôlées présentées comme le gold standard proviennent notamment, du déficit théorique du fait que les promoteurs de la méthodologie, assise sur le modèle d’inférence causale, considèrent la théorie comme mauvais guide. Elles découlent, également, de l’identification de nombreux biais d’ordre statistique qui altèrent l’utilisation de la randomisation et sont, entre autres, entachés par les nombreux bricolages et arrangements induits par les interactions chercheurs/expérimentateurs/partenaires auxquelles les expérimentations donnent lieu. Enfin, le modèle causal tel qu’utilisé par le J-PAL est enserré

entre deux tensions : holisme/individualisme et sujet/société, mal appréhendées sociologiquement et anthropologiquement et sur le plan épistémologique, elles donnent lieu à des heuristiques argumentatives faibles.

Les deux derniers chapitres de notre thèse développeront les intrications entre l'expertise, l'académique et le politique servies par une performativité dont les énoncés échappent aux problématiques du vrai et du faux mais également à l'aide d'une rhétorique permissive au « anything goes » de Feyerabend où, paraphrasant Robert Nadeau, l'économique agit comme littérature.

## **Chapitre IV : Expérimentation, pauvreté et reconfiguration du Consensus de Washington**

Les retournements idéologiques opérés à la fin des années soixante-dix après les révolutions thatchérienne et reaganienne qu'accompagnent, par ailleurs, la crise de la dette issue des chocs pétroliers de 1973 et 1979 et le ralentissement économique qui s'en est suivi, ont marqué le retour aux premiers plans des institutions de Bretton Woods. Ces institutions, remises en selle par le virage libéral en vogue, appliqueront aux pays en développement la potion amère des plans d'ajustements structurels destinés prioritairement à assurer le remboursement des dettes extérieures et la transition vers le marché. Encouragées de surcroît par la chute de l'Union soviétique, ces institutions imposeront des conditions visant la limitation du champ d'intervention de l'État, la recherche des équilibres monétaires et budgétaires et la compression des dépenses publiques et des charges fiscales. C'est le triomphe du monétarisme, comme pensée dominante, qui va servir d'instrument scientifique et guide méthodologique. L'important, dès lors, est la protection des prix des actifs et les fortunes accumulées par leurs détenteurs. En raison de leur caractère normatif simplificateur et apolitique, les politiques d'ajustement ont moins « pour effet de modifier les politiques économiques et le monde d'insertion internationale que de changer les comportements et les opportunités internes d'enrichissement ; autrement dit, [ses] effets sont moins économiques que socio-politiques » (Hibou, 2014, p. 3). Le matériau théorique, puisant dans les fondements de l'économie libérale, est mobilisé pour justifier ces politiques se caractérisant par la parcimonie et l'extrême simplicité dans l'utilisation avec les concepts sous-jacents où « seuls certaines conclusions sont retenues » par un appel plus grand aux « empirical evidence » (p. 103). Pour combattre la pauvreté, la Banque mondiale recommande la croissance élargie et l'amélioration de l'accès aux services sociaux (BM, 1990). L'amélioration des services sociaux exige « le renforcement entre les décideurs publics, les prestataires et les citoyens » (BM, 2004) sous-tendant la libéralisation des échanges et des marchés et l'amélioration de la fiscalité. Ces visées reposent sur trois principes : norme libérale de l'État minimal, libre échangisme et apolitisme



(Hibou, 2014). C'est sur ces fondements, construits à l'aide d'arguments scientifiques et apolitiques, que naît le consensus de Washington.

Dans le présent chapitre, notre intention sera d'explicitier les fondements prescriptifs du CW et de montrer comment, sous des variantes et appellations différentes, les politiques sociales prescrites demeurent dépendantes de ce schéma de pensée imposé, se muant, même, en orthodoxie nouvelle et améliorée selon l'expression de Maxwell (2005). L'octroi des aides, soumis à des impératifs d'évaluation, est assorti de conditionnalités d'extension du marché et de retrait de l'État. De ce fait, l'extension de cette doxa libérale en appelle à la responsabilité et la rationalité individuelle et les exclus du marché sont sommés d'y participer ! (Section 1).

On s'attachera à décrire comment les expérimentations aléatoires contrôlées sont mobilisées par la transformation de la recherche académique, censée expliquer et prévoir, en une ingénierie de la décision au service de ce nouvel ordre. Ainsi l'éducation, la santé, l'agriculture, les finances, les questions du genre et le marché du travail sont-ils soumis à des expérimentations randomisées. Ces dernières recouvrent les préoccupations référentielles du consensus de Washington aspirant à faire de la généralisation de ces évaluations l'instrument d'une gestion apolitique mise en œuvre dans une logique de réduction des dépenses sociales et de rationalisation de l'intervention publique (section 2). Enfin en section 3, seront traités les liens entre la lutte contre la pauvreté et les politiques de ciblage liées par une nouvelle gouvernance du social et de nouveaux systèmes d'incitation.

### **Section 1 : Consensus et marché des évaluations des politiques publiques**

Issu d'un glissement de la notion de gouvernement vers celle de la *gouvernance*, le consensus de Washington est présenté comme une perspective normative et prescriptive d'organisation du retrait de l'État au profit du marché en raison de son incapacité à assurer l'efficience en matière de maximisation du bien-être social sous contraintes budgétaires. La version originale portée par John Williamson et ses contributeurs associés visait principalement la macroéconomie et la gestion financière en incitant à prendre des mesures de stabilisation et à opérer des réformes

structurelles. Bâti autour de 10 axes principaux, le dit consensus visait la discipline budgétaire par la maîtrise du niveau de la dette publique, la mobilisation de l'épargne nationale et le maintien de la sincérité des comptes publics. Par conséquent, les dépenses publiques doivent assurer la croissance et les missions d'intérêt général en incitant à mieux dépenser. La réforme fiscale doit se construire autour de l'élargissement de l'assiette avec des taux d'imposition plus faibles. Pour faire face à la pénurie du capital, talon d'Achille des économies en voie de développement, la libéralisation financière est prônée comme instrument d'allocation des ressources qu'accompagne un contrôle prudentiel strict. Dans ce sillage, il est également recommandé d'abandonner, même progressivement, le système des changes fixes, d'adopter des changes flottants et des parités glissantes à même d'assurer l'insertion dans le marché international. Parmi les prescriptions du CW, la libéralisation du commerce international en est un des piliers qui agit, dans la tradition de l'analyse ricardienne de spécialisation, comme moteur du progrès technique et des économies d'échelle, et levier qui favorise la croissance par les gains de l'échange. L'élimination des barrières à l'investissement direct étranger est censée permettre le décollage et l'incorporation du progrès technique. Alors même que ces investissements directs étrangers se concentrent au Nord à plus de 65 % et à 35 % en direction de peu de pays (Chine, Inde, Brésil, ...), ailleurs les IDE se réalisent davantage sous forme d'actifs de portefeuille que d'investissements nouveaux ; d'où la recommandation visant à inciter à la privatisation des entreprises publiques présentées souvent comme des canards boiteux. Cette prescription est généralement présentée comme un levier visant la création d'un environnement économique plus concurrentiel et plus transparent. Elle s'inscrit également dans la conception d'outils susceptibles d'alléger les finances publiques alors même que la réalité est souvent plus nuancée se traduisant parfois par la substitution de monopoles privés aux monopoles publics. Au demeurant, les conditions de concurrence plus vive sont recommandées comme source d'innovation, de bien-être pour le consommateur, donc de croissance et, *in fine*, de compétitivité pour l'économie nationale. Cette préoccupation est portée par la recommandation de réformer la réglementation des marchés et l'élimination des principales barrières

d'entrée. Enfin, et dans la pure tradition de l'économie mainstream, il est fortement recommandé de garantir les droits de propriété jusqu'à en faire une condition de l'état de droit et un droit naturel et imprescriptible de l'homme.

Williamson avait lui-même pris ses distances avec la rigidité idéologique avec laquelle sont prescrites et appliquées sous forme de conditionnalités imposées par les institutions dominantes en matière de financement du développement ou de soutien aux finances publiques des États. Dans un papier que lui avait commandé le CIDOB (2004), il affirmait qu'« à l'époque, je pensais que les opinions s'étaient rapprochées sur le type des politiques qui étaient préconisées depuis longtemps par l'OCDE. Je ne croyais pas, en particulier, que la plupart des innovations « néolibérales » de l'administration Reagan aux États-Unis ou du gouvernement Thatcher en Grande-Bretagne avaient survécu à la disparition de la première (le gouvernement de M<sup>me</sup> Thatcher était alors encore en train de mourir). L'exception était la privatisation, qui était le cadeau personnel de Madame Thatcher au programme de politique économique du monde, et qui, en 1989, avait fait ses preuves. Mais je pensais que toutes les autres idées nouvelles avec lesquelles Reagan et Thatcher étaient entrés en fonction, notamment le monétarisme, l'économie de l'offre et le gouvernement minimal, avaient été rejetées comme des modes peu pratiques ou indésirables, de sorte qu'on n'en trouve aucune trace dans ce que j'ai appelé le « consensus de Washington ». Pour Williamson, l'acceptation comme pertinentes pour le monde en développement d'idées longtemps décrites « maternité et tarte aux pommes » (motherhood and apple pie) dans le monde développé a été un changement capital alors que, tout au long de la guerre froide, le monde était resté figé dans la classification des années cinquante entre Premier, Deuxième et Tiers Monde, chacun étant supposé avoir son propre ensemble distinct de lois économiques ». Williamson conclut son propos en constatant que la mondialisation des connaissances n'a jamais signifié l'acceptation général du néolibéralisme, quelle que soit la définition qu'on en donne.

Cette inflexion est à apprécier dans le cadre du mouvement plus global de critiques adressées au CW (Stiglitz, 2002 ; Rodrik, 2008). Ces critiques se concentrent sur l'échec du consensus attribué au constat que ces recommandations en question ont davantage profité aux plus riches qu'aux plus défavorisés et qu'elle se sont appliquées au détriment de la démocratie et de la justice sociale (Stiglitz, 2002). Elles ont également mis en cause la propension à disqualifier l'État en raison des vertus supposées du marché et de l'application indifférenciée des prescriptions incriminées sans tenir compte des spécificités de chaque pays et de ses institutions et qu'« il n'est de bonne économie que dans un contexte donné » (Rodrik, 2008, p. 6). Dans ce sillage, nombre de digressions ont été déclinées pour en atténuer les aspérités issues des difficultés d'intégration économique en profondeur dans un environnement de fragmentation de la souveraineté politique. Ainsi en 2005, James Wolfensohn, ancien président de la Banque mondiale, observait que le consensus de Washington est mort depuis des années et est remplacé par toutes sortes de consensus » (Maxwell, 2005). Pour ce dernier, le Consensus de Washington se mue en une orthodoxie nouvelle et améliorée qui accorderait plus d'attention à l'équité et la justice sociale en prenant en compte les questions politiques liées à l'architecture de l'aide. L'accent est alors mis sur les objectifs du millénaire pour le développement (OMD). Relevons, au passage, que les OMD adoptés en 2000 n'avaient pas atteint les résultats escomptés pour 2015 ; ils ont été transformés en objectifs à atteindre pour 2030 sous l'appellation d'objectifs de développement durable (ODD). La sémantique évolue, le fond demeure en prenant la forme d'un méta-narratif qui, selon l'observation de Maxwell, doit allier l'orientation vers le marché avec des impératifs en relation avec la qualité des institutions et le capital humain. Ce faisant, les sujets d'inclusion sociale, de trajectoire, de développement et d'aide sont inscrits dans un vaste agenda de recherches traçant la voie pour un véritable marché de recherches académiques et d'expertises qui légitime l'ouverture macroéconomique, la réduction des dépenses, l'aide dite harmonisée en faveur du marché et de la bonne gouvernance dans le cadre plus global de l'agenda international du commerce et de la finance. La bonne gouvernance et l'environnement institutionnel sont censés assurer le passage d'une économie administrée par les règles vers une économie « pilotée » par les prix.

Le « tout marché », appliqué sous la pression des institutions financières internationales, s'est accompagné par une montée en puissance du marché des évaluations des politiques publiques conditionnant l'octroi de prêts et d'aide. Ainsi, tout un espace fut aménagé pour l'évaluation et l'expertise. Les RCT sont à analyser à travers cette grille de lecture. Contrairement à d'autres contributions (Labrousse, 2010 ; Jatteau, 2013 et 2017), fort éclairantes au demeurant, nous pensons que les RCT ont prospéré dans le sillage du consensus de Washington. Ce fait est daté et le signal en avait été donné par l'éditorial de la revue *The Lancet* (2004) apprécié comme plaidoyer décisif en faveur de ces expérimentations et de leurs promoteurs. L'éditorial en question cherchait, d'emblée, à savoir si, l'aide, quelle qu'elle fût, n'était pas préférable à son annulation (*debt forgiveness*). Question qui dépasse le cadre de la Banque mondiale tant les programmes dits d'ajustement structurel ont été, entre autres, construits à l'effet d'assurer le remboursement des dettes que les pays pauvres avaient accumulées en raison de diverses restrictions commerciales, du protectionnisme et de politiques financières changeantes. La Banque - et le public - devraient se demander, premièrement, si l'aide fonctionne et, deuxièmement, si les projets d'aide sont le meilleur moyen d'améliorer la vie des personnes qui ont le plus besoin du soutien des donateurs. *The Lancet* constate alors que : « ... des signes récents montrent que la Banque prend des mesures nécessaires depuis longtemps pour répondre à la première question, du moins en partie. Au cours de l'été 2004, elle lance une série d'essais aléatoires pour déterminer si ses projets d'aide sont utiles. Alors que le « succès » de la Banque a parfois été déterminé en fonction du nombre de prêts accordés, des méthodes plus rigoureuses devraient désormais remplacer cette batterie de mesures simplistes ». Cette revue n'hésite pas, d'ailleurs, à classer le J-PAL parmi les collaborateurs de la Banque mondiale, relevant que seuls 2 % des projets qu'elle a financés ces dernières années ont fait l'objet d'une évaluation critique, c'est-à-dire d'évaluations randomisées, une nouveauté pour la banque selon elle : « c'est une statistique effroyable : de telles évaluations sont des biens publics, et la responsabilité publique les exige certainement », déplore le *Lancet* se demandant alors comment, sans preuve, peut-on savoir s'il faut modifier, supprimer ou étendre un programme existant ? Reprenant les propos de Duflo, le *Lancet* poursuit : « La création d'une culture dans

laquelle des évaluations randomisées rigoureuses sont promues, encouragées et financées a le potentiel de révolutionner la politique sociale au cours du 21<sup>e</sup> siècle, tout comme les essais randomisés ont révolutionné la médecine au cours du 20<sup>e</sup> siècle ».

Cet éditorial fait, aussi, écho à la charge de Banerjee contre la Banque mondiale, ses évaluateurs, les concepteurs de projets étant traités « d’armées ignorantes qui s’affrontent la nuit » (*ignorant armies clashing by night*), un propos qu’il attribue à Lant Pritchett (2002). Les donateurs n’échappent pas à cette critique. L’auteur considère qu’ils ne savent pas très bien ce qu’ils doivent faire valoir, conduisant à la corruption et à la mauvaise utilisation des fonds et impute ces tares à la pensée paresseuse – *lazy thinking* – (Banerjee, 2006). La charge des randomisateurs, tout comme celle du Lancet, semblent avoir eu de l’impact comme le relèvent Bédécarrats et ses co-auteurs (2013, p. 111) : « si la Banque mondiale semble avoir une politique plus nuancée en matière de méthodologie d’évaluation d’impact, en diversifiant les méthodes (néanmoins toujours quantitatives), les RCT en représentent près des deux-tiers, avec 64 % des 368 évaluations engagées (jusqu’en 2010). Non seulement les RCT tendent à occuper une position de plus en plus dominante, mais elles exercent un effet d’éviction sur les autres approches, [...]. Au cours de la période 2000-2004, à peine 20 % des évaluations étaient des RCT ».

Cette conjonction a, également, partie liée avec les effets influents du rapport du Center For Global Development « *When will we ever learn? Improving lives through impact evaluation* », évoqué plus haut, auquel avait participé E. Duflo avec d’autres, à sa rédaction, en bénéficiant de l’appui des fondations Bill & Melinda Gates et William & Flora Hewlett. Ce rapport indiquait : “Yet after decades in which development agencies have disbursed billions of dollars for social programs and developing governments and non-governmental organizations (NGOs) have spent hundreds of billions more, it is deeply disappointing to recognize that we know relatively little about the net impact of these social programs” (Center For Global Development, 2006).

Cette doxa, érigée en consensus, reflète la prédominance des grilles d’analyse issue de l’économie néoclassique et libérale qui considère que l’économie du



développement n'existe pas à proprement parler, mais, qu'elle serait à rechercher dans la transposition des théories économiques issues de l'économie standard applicables aux pays en développement. L'accent est, désormais, mis sur la croissance et les mécanismes qui y sont sous-tendus (information, coordination, externalités, rationalité des agents, anticipation...). La distinction faite par les pionniers de l'économie du développement entre croissance et développement est vouée à l'abandon. François Perroux (1966) considérait que le développement « est la combinaison des changements mentaux et sociaux d'une population qui la rendent apte à faire croître cumulativement et durablement son produit réel et global » (p. 240). Pour Hirschman (1984), Assidon lui impute le constat que l'on « ne saurait aborder l'étude des économies sous-développées sans modifier profondément, sous un certain nombre de rapports importants, les données de l'analyse économique traditionnelle, axée sur les pays industriels » (Assidon, 2002).

Krugman (1993, 1996), tout en signalant l'intérêt que revêtent ses découvertes, - entendez celles de l'économie du développement -, considère que les travaux des économistes du développement sont des « écrits merveilleux et source d'inspiration » mais l'incapacité à formaliser les a éloignés de l'économie du développement du courant mainstream. Et c'est cette incapacité à les modéliser qui aurait conduit à la mort intellectuelle de la discipline (Herrera, 2007). On pourrait dire que le courant mainstream va plus loin, comme avec Lucas qui considère que si la croissance est ce que l'on connaît, l'économie du développement est ce que l'on ne sait pas : « It seems to be universally agreed that the model I have [just reviewed] is not a theory of economic development. Indeed, I suppose this is why we think of 'growth' and 'development' as distinct fields, with growth theory defined as those aspects of economic growth we have some understanding of, and development defined as those we don't » (Lucas, 1988, p. 13).

Ces coups de boutoir de la part de l'économie dominante ont poussé à l'utilisation systématisée des méthodes quantitatives et mathématisées dans l'intégration de l'économie du développement dans l'approche standard (Raffinot, 2015). Un phénomène nouveau ces trois dernières décennies, couplé à l'influence quasi monopolistique des organismes internationaux du développement (Banque mondiale principalement) ont fait qu'il est devenu, par ailleurs, presque impossible

de publier dans les meilleures revues sans présenter une analyse empirique appuyée. Dès lors, l'attention est de plus en plus portée sur la qualité des données, l'inférence causale et la batterie d'outils économétriques et statistiques. C'est ce que d'aucuns n'ont pas hésité à appeler le « tournant empirique » en économie du développement la faisant évoluer d'une discipline académique à une sorte « d'ingénierie de la décision ». Ainsi les spécificités de l'économie du développement se diluent peu à peu dans des fondements microéconomiques renouvelés (Assidon, 2012). Les expérimentations aléatoires appliquées à la microéconomie du développement et à la lutte contre la pauvreté sont nées dans ce contexte porté par la vague néo-libérale dans la pratique et la domination néoclassique et qu'Angrist et Pischke (2010) ont appelé « la révolution de la crédibilité ».

Instaurée au forceps en vue de maintenir la discipline budgétaire imposant l'austérité, la doxa en question se traduit le plus souvent par des réductions drastiques des dépenses sociales. Elle accorde une trop grande confiance au fonctionnement des marchés pour remplir les fonctions sociales concernées (éducation, santé, électrification, accès à l'eau, logement...), tout en reconnaissant que ces services « fourmillent d'exemples de dysfonctionnements du marché... le secteur privé, livré à lui-même, ne procurerait donc pas à la société le niveau de santé et d'instruction qu'elle souhaite » (BM, 2004). La Banque mondiale recommande, en conséquence, aux pouvoirs publics d'agir avec les communautés et les partenaires extérieurs en coopération avec le secteur privé pour assumer la responsabilité fondamentale de fournir les services de santé et d'éducation en faveur des pauvres. Pour ce qui concerne l'Argentine par exemple, Bruyas (2010, p. 36) note que : « Certaines dépenses sociales y sont encouragées, ou tout au moins considérées comme acceptables, alors que l'équilibre budgétaire y est présenté comme l'un des objectifs majeurs de la politique macroéconomique. La pratique encouragée par les institutions financières internationales valorise d'ailleurs clairement le second objectif par rapport à celui du soutien à l'éducation et à la santé » soulignant, ainsi, les contradictions discursives du consensus.

Les expériences aléatoires s'inscrivent et se placent précisément dans cette construction à l'effet de la légitimer académiquement. Dans le rapport de la Banque mondiale, cité en amont, les experts de l'institution pointent, qu'en dépit de moyens

considérables (le tiers du budget allant à l'éducation et à la santé) les défauts qui caractérisent les services aux pauvres de quatre manières : premièrement, les dépenses engagées profitent aux non pauvres. En second lieu, lorsque les dépenses sont recentrées sur les pauvres, à part les salaires, les crédits alloués n'atteignent pas les services en première ligne. Troisièmement, les agents affectés à fournir ces services ne sont pas incités à remplir leurs missions et font rarement l'objet de contrôle ou de sanction pour absentéisme. En quatrième lieu, le défaut des services aux pauvres souffre d'absence de demande ! Les pauvres s'abstiennent d'envoyer leurs enfants aux écoles et ne les amènent pas à l'hôpital. Cette faiblesse de la demande est imputée à l'extrême pauvreté, aux longues distances à parcourir, et même à des raisons de type essentialiste comme la culture (par exemple les discriminations à l'égard des femmes ou les distances sociales entre pauvres et prestataires de services) ou, plus encore, aux considérations qui relèvent du clientélisme politique. Ces constats viennent à l'appui de la revendication, appuyée par maintes expériences relatées, de la nécessité de sous-traiter ces services au secteur privé, aux ONG ou à des organismes publics dotés d'outils de gouvernamentalité tirés du *new public management* et potentiellement privatisables.

A examiner de près, le découpage qu'opère le J-PAL des secteurs d'interventions visant l'évaluation des politiques publiques croise ces préoccupations libérales. L'examen minutieux de nombre d'expérimentations menées et sous-tendues par un apolitisme revendiqué montre clairement cette confiance attribuée aux marchés pour la fourniture des services essentiels aux populations. Les différents thèmes (éducation, santé, agriculture, finances, genre, marché du travail reposant sur les expérimentations randomisées), recouvrent les préoccupations référentielles du consensus de Washington érigeant la généralisation de ces évaluations comme doctrine de gestion apolitique mise en œuvre dans la logique de réduction des dépenses, notamment sociales, et de rationalisation de l'intervention publique.

Le capital savant fourni par cette expertise dote les organisations internationales (FMI et BM) d'outils et de capacité à mettre en œuvre et à imposer les programmes liés au rôle de prêteur conditionnel renforçant leur pouvoir normatif et de sanction (Delmas, 2011, p. 81). La justification de l'État minimal, de la dissolution du social dans le marché et le démantèlement des barrières protectrices reposent sur les

modèles économiques orthodoxes de concurrence parfaite, de rationalité des agents, de flexibilité des prix et des salaires et enfin aux hypothèses relatives aux élasticités.

Nous spécifierons dans les deux sections suivantes formant corps de ce chapitre comment le J-PAL contribue à la reconfiguration du consensus de Washington et donnons-nous à voir comment ses prescriptions sociales y afférentes relèvent du pouvoir normatif des politiques promues par les institutions de Bretton Woods.

## **Section 2 : Rationalité individuelle et accès aux ressources**

On oublie souvent que la Banque mondiale ou le Fonds Monétaire International sont avant tout des institutions bancaires qui lient l'octroi de prêts à la capacité des emprunteurs à être solvables. Toutes les opérations d'investissements et de prêts des banques relèvent du principe de marché et de la règle de concurrence. Une posture qui explique très largement les recommandations de ces institutions relatives au démantèlement des banques de développement et la privatisation des banques publiques. Dans un bulletin d'information du bureau de la BM au Maghreb et consacré au nécessaire redéploiement du rôle de l'État et de promotion d'un environnement favorisant le renforcement de l'industrie financière, son directeur de l'époque, aujourd'hui vice-président pour la région Moyen-Orient et Afrique du Nord écrivait assez poétiquement : « ... à une époque où la compétition devient loi, où les marchés s'internationalisent et où le monde s'aplanit (pour reprendre le titre de l'excellent ouvrage de Tom Friedman « The World Is Flat ») je dirai que le système financier marocain est tel le navire bien arrimé à son port. Stable et sûr. Le défi consiste pour ce navire à sortir en haute mer et à affronter sans trop tanguer, la houle du large et de naviguer sur les vagues hautes de la compétition » (Belhadj, 2006).

Ce tropisme du tout marché, de la concurrence, de la réduction des budgets publics, de l'appel à la responsabilité et à la rationalité individuelles se retrouvent dans l'utilisation des expérimentations au service des orientations relatives à la manière de mener et de conduire les politiques publiques. Ces orientations s'appuient sur les théories du courant d'économie libérale (néo-institutionnalisme, théories de la société civile, individualisme méthodologique et postulats individualistes) « qui tendent à dépolitiser les perceptions et les interprétations » (Hibou, 2014, p. 12). En

évoquant les théories de la société civile et en s'appuyant notamment sur les travaux de Hernando de Soto en particulier, Béatrice Hibou établit, de manière judicieuse, un lien intéressant entre les analyses de la Banque mondiale et l'approche consistant à s'appuyer systématiquement sur des ONG (très souvent la même association pour chaque secteur et pays concernés), ce que le J-PAL fait systématiquement pour mener ses expérimentations. Cette société civile est perçue comme la forme appropriée de résistance à l'État et capable d'apporter des réponses créatives à son incapacité (p. 15), tout en occultant les différentes imbrications liant cette société civile, les bailleurs et le croisement des appartenances. L'expérimentation créative est d'ailleurs une expression/argument qu'emploie Duflo (2009, p. 54) lorsqu'elle évoque les co-expérimentations réunissant jusqu'aux profanes qui s'investissent dans des sujets réservés aux experts (Labrousse, 2010). Ainsi, lorsque les expériences aboutissent à des résultats « provoquants » et contredisant la théorie, il conviendrait de considérer que si la théorie peut inspirer les chercheurs, l'expérience ne repose pas sur la validité de la théorie (Duflo, 2009). Dans son ouvrage « *Le développement humain, lutter contre la pauvreté (I)* », elle appuie encore son argumentation en concluant par : « il est donc possible – et apparemment ce n'est pas si difficile d'améliorer l'accès des plus pauvres à des systèmes éducatifs et médicaux de qualité. De nombreuses initiatives montrent que des interventions peu coûteuses (stages de lecture, traitement contre les vers intestinaux, distribution de lentilles pour encourager la vaccination, diffusion d'informations claires sur les taux d'infection par le VIH) peuvent avoir des effets spectaculaires contre l'analphabétisme et la prévalence de certaines maladies » (Duflo, 2010, p. 101). Elle ajoute, non sans emphase, : « Ces succès isolés forment un contraste étonnant avec les piètres performances de l'ensemble des institutions scolaires et de santé. S'il est possible à un jeune tout juste sorti du lycée d'apprendre à un enfant à déchiffrer les lettres en quelques semaines, pourquoi tant d'enfants passent-ils cinq ans à l'école, sous la houlette d'enseignants formés, sans apprendre à lire ? Et si un kilo de lentilles multiplie par 7 les taux de vaccination, comment se fait-il que 20 millions d'enfants ne soient pas vaccinés chaque année ? ».

Duflo et ses co-promoteurs des ECR imputent les échecs des politiques sociales à la frilosité politique et au manque d'imagination et d'innovation des gouvernements

ou des organisations internationales dont le seul levier d'intervention, généralement inutile, est l'augmentation des budgets (plus d'enseignants, plus de tableaux noirs, plus d'infirmières, plus de médicaments). On considère que les politiques traditionnelles reposent sur deux principes : les parents sont l'obstacle principal à l'instruction universelle (*sic*) d'une part, et des coûts de scolarité associés aux coûts d'opportunité (travail à la ferme parentale par exemple) constituent une charge trop lourde pour ces parents, d'autre part. La santé n'est pas en reste du fait que « les comportements, l'ignorance, les préjugés jouent également un rôle essentiel » (pp. 22-26).

Ce qui n'est pas le cas des ONG, véritable source d'innovation. Les aides ciblées conditionnent, à leur tour, les comportements désirables (le programme Progreso au Mexique est donné comme l'exemple à suivre malgré les tripatouillages politiques qui l'ont accompagné et dont nous avons évoqué les bricolages institutionnels plus haut (Faulkner, 2014). Tous constats qui finissent par considérer que les problèmes auxquels les secteurs de l'éducation et de la santé font face « n'ont rien avoir avec l'aide internationale » et que celle-ci « n'est pas une bonne recette pour les bureaucrates des pays en développement » (pp. 103-104).

Lorsque l'on rapproche le rappel de cette doxa à ce que nous mentionnions à propos des services aux pauvres tels que conçus par les experts de la Banque mondiale, nous retrouvons tous les présupposés véhiculés par le consensus de Washington et le substrat académique qui en constitue le support idéologique. Ainsi, trouve-t-on tous les ingrédients positifs et explicatifs de l'économie standard relative à la place du marché, aux comportements, à la responsabilité individuelle, aux instructions, à la place des incitations et à la conception de la pauvreté vue comme une « maladie à traiter » et non une construction sociale bâtie sur le socle des inégalités.

Les auteurs de « *Repenser la pauvreté* » refusent la « pensée paresseuse et stéréotypée » et invitent à s'attaquer aux problèmes de la pauvreté comme une série de problèmes concrets qui peuvent être résolus un à un. Cette approche intervient dans ce que l'on appelle la gouvernance sociale où les politiques sociales sont encadrées dans les politiques économiques du marché (*market friendly*), l'État se cantonnant au rôle de protecteur contre les risques excessifs du marché. Ses auteurs



n'invoquent qu'indirectement les présupposés de l'économie standard sur la rationalité, le comportement et le capital humain. L'éducation et la santé sont analysées, sans jamais être évoquées explicitement, sous l'angle de la théorie moderne du capital humain avec Theodore Schultz (1970) ou Gary Becker (1964) armés d'analyses théoriques et empiriques portant sur les liens entre l'investissement en capital humain et la rémunération, ou encore aux nouvelles théories de la croissance économique, auxquelles sont associés les noms de Romer (1986) et Lucas (1988), le capital humain étant considéré comme un facteur déterminant de la croissance économique. Selon ces canons d'analyse, le capital humain doit faire l'objet d'un investissement consacré à l'éducation ou à la santé et les revenus additionnels sont à apprécier à l'aune de ce que peut permettre cet investissement et les rendements que l'on peut en tirer. Il s'agit donc du choix d'un individu rationnel qui maximise les gains futurs nets des coûts à supporter aujourd'hui. L'intensité de l'effort à consentir est à apprécier au regard du coût d'opportunité qui croît avec le montant investi. Les inégalités d'investissement entre individus ne sont pas considérées comme résultant des conditions sociales mais plutôt comme la rencontre entre la capacité à tirer un rendement de l'investissement à supporter et les coûts auxquels les familles sont confrontées pour investir. S'en suivent, également et fort logiquement, les considérations liées à la santé et la nécessaire limitation des naissances chez les pauvres. Nos auteurs estiment que les pauvres y consacrent une bonne partie de leurs maigres ressources qui sont mal affectées soit par ignorance soit par défauts d'informations. Les politiques de prévention, le bon accès aux soins ou encore l'accès à l'eau potable sont expérimentés avec des recommandations consistant à limiter l'accès des pauvres aux médicaments dont ils n'ont pas besoin. Un fléau comme le paludisme n'est « traité » qu'à travers les moustiquaires subventionnées sans jamais évoquer les intérêts importants et croisés qui s'accommodent bien du décès de millions de personnes à cause de la malaria. Cette maladie relève de questions de santé publique aux conséquences épidémiologiques désastreuses et entretient de façon endémique dans son sillage les poches de pauvreté. Le récit d'Émile Roubaud sur la manière dont les américains en sont venus à bout, après qu'ils eurent à prendre la relève des français pour le parachèvement du percement du canal de Panama, en dit long sur

la volonté et les moyens que l'on met pour traiter les questions épineuses : « La lutte contre la fièvre jaune ou fièvre amarile, ce fléau des continents américains et africain tropicaux devait naturellement marcher de pair avec la campagne précédente. Les mesures adoptées dans l'isthme pour combattre l'affection amarilique découlent de celles qui avaient été prises à la Havane en 1901 sous la direction du général Gorgas. La campagne de la Havane basée sur la destruction du moustique transmetteur *Stegomyia fasciata* et de ses gîtes avait abouti en peu de temps à délivrer complètement la ville cubaine de la désastreuse endémie. Les mêmes rigoureuses méthodes furent employées dans l'isthme. Les gîtes de ponte du *Stegomyia* qui sont essentiellement urbains furent impitoyablement surveillés. Des équipes sanitaires chargées de l'inspection rigoureuse des locaux au point de vue de l'existence des larves de moustiques contrôlent toutes les habitations dans la zone du canal. Les maisons sont protégées contre l'introduction des moustiques ailés dans les chambres de malades, des fumigations insecticides sont pratiquées. L'autorité américaine assume la direction et la surveillance de mesures préventives sévères (quarantaines) aux escales maritimes de Panama et de Colon. Grâce à cet ensemble de mesures, la fièvre jaune a disparu de la zone du canal dès la fin de 1905 » (Roubaud, 1918, p. 68). Au-delà de l'enjeu géostratégique de l'époque et du scandale qui s'en était suivi, Roubaud impute l'échec de l'entreprise française à l'incapacité à traiter cette question sanitaire majeure. La Compagnie française antérieure enregistrait en effet, pour 1886, une mortalité de 4 pour 100 des employés. Ainsi, l'entreprise tout entière était condamnée par suite de l'effet moral produit qui eût entravé le recrutement des ouvriers alors que « l'œuvre d'assainissement accomplie par les Américains dans la zone du canal représente en fait une magistrale expérience et un admirable enseignement »... et [qu'il] n'est pas de contrée tropicale où un européen ne puisse vivre et travailler dans de bonnes conditions si l'on sait efficacement protéger son existence et si l'on veut y mettre le prix (pp. 68-69). L'œuvre de développement et l'action sur la nature ont été, en effet, à ce prix.

Le contraste est saisissant avec les prescriptions des auteurs susmentionnés, qui sur un autre registre, mêlant santé, sécurité, microfinance et marché constatent que : « ... le microcrédit s'est banalisé, assurer les pauvres semble être un champ d'action

tout désigné pour les capitalistes créatifs et pleins de bons sentiments (un éditorialiste de Forbes a ainsi parlé de marché naturel encore non pénétré).

Les pauvres subissent des risques considérables et devraient sans doute être prêts à payer des primes raisonnables pour assurer leur vie, leur santé, leur bétail ou leurs récoltes. Avec ces milliards de pauvres attendant de souscrire une police, même un profit minime par contrat constituerait un projet commercial fantastique et, en même temps, une aide décisive pour tous les pays en développement. Il ne manque plus, semble-t-il, que quelqu'un pour organiser le marché : c'est ce qui a incité des organisations internationales (comme la Banque mondiale) ou des fondations importantes (comme la fondation Gates) à investir des centaines de millions de dollars pour encourager le développement de solutions de prévoyance pour les pauvres » (Banerjee et Duflo, 2011, pp. 252-253).

De manière générale, les pauvres sont invités à « construire leur épargne brique par brique » pour faire face aux risques qu'ils encourent et aussi pour créer leurs entreprises d'autant qu'ils ont un accès restreint aux crédits de financement. Ainsi, « s'ils avaient une petite réserve, les conséquences d'une mauvaise récolte ou d'une maladie seraient moins dramatiques. Et l'épargne pourrait également leur permettre de créer une entreprise » (p. 307). Ils nous renseignent, par ailleurs, que la communauté internationale (sans la définir précisément), sous l'impulsion en particulier de la fondation Bill et Mélinna Gates, déploie des efforts considérables pour faire accéder les pauvres à des comptes d'épargne.

S'agissant des entreprises des pauvres, Banerjee et Duflo expliquent pourquoi les entreprises des pauvres peuvent avoir des rendements très élevés pour leurs investissements, mais avertissent, à la manière dont Gary Becker explique les rendements de l'investissement individuel en capital humain, qu'il est nécessaire d'éviter la confusion entre le rendement marginal d'un dollar et le rendement global de l'entreprise. En d'autres termes, pour le rendement marginal, il faudrait évaluer si un dollar d'investissement en plus ou en moins rapporte plus ou moins que le remboursement du capital et les intérêts. Le rendement global, quant à lui, détermine les revenus totaux une fois déduits les frais de fonctionnement. Ces auteurs constatent, cependant, et paradoxalement que « le taux marginal peut être

élevé alors même que le rendement global est faible » (p. 354). Ce paradoxe trouve son explication dans le mimétisme qui fait que les pauvres entreprennent de manière identique sans niches différenciatrices. Sans cela, ils achèteraient de l'emploi par un moyen détourné lorsque des choix professionnels plus classiques font défaut. Pour créer de bons emplois, le cadre réglementaire régissant le travail est important mais si trop restrictives, les lois seront contre-productives freinant ainsi l'embauche ou de potentiels employeurs. Le rôle utile de l'État est de favoriser le développement d'entreprises importantes en garantissant ou en cédant, comme en Chine ou en Corée, les entreprises publiques :

« ... du moins une partie de leurs équipements ou de leurs bâtiments ont été discrètement cédés à leurs employés » (p. 381). A noter que le terme « discrètement » surprend dans cette pièce rapportée.

Cette approche des comportements et des motivations liées aux subventions publiques se retrouve également traitée comme fait stylisé pour la question de la nutrition, besoin primaire s'il en est. Les auteurs constatent que les pauvres ne sont pas affamés, ils consacrent leurs ressources à acheter des choses inutiles au lieu de s'acheter à manger et même lorsqu'ils se procurent des aliments, à cause du manque d'instruction et d'information, ils préfèrent des aliments plus chers mais ayant un meilleur goût plutôt que d'acheter des aliments nutritifs. Ils arguent, par ailleurs, que les pauvres sacrifient la nourriture au profit de l'organisation de grands événements comme les mariages ou les funérailles. En effet, « manger n'est pas le plus important » pour les pauvres, il y a « des choses plus importantes dans leur vie que la nourriture » (p. 72). L'attitude des pauvres vis-à-vis de la nutrition tranche pourtant, d'après les auteurs, avec les grands efforts consacrés par les États à aider ces catégories sociales à se procurer de quoi manger (2,8 % du PIB pour un pays comme l'Égypte) alors qu'une partie de la distribution est perdue, notamment, par les rats (p. 46) !

L'apolitisme de cette démarche se manifeste dans l'opposition à la vision, partagée par de nombreux spécialistes du développement, selon laquelle la politique prime sur l'économie. Pour les randomisateurs, il est en effet nécessaire d'interroger les politiques les mieux intentionnées et les mieux conçues en dépit des défaillances des

gouvernements, souvent invoquées, pour s'opposer à l'aide internationale et aux influences que l'on peut mener en faveur des politiques sociales. Les auteurs arguent qu'il est possible de « faire progresser la transparence et de lutter contre la corruption même dans les institutions globalement « mauvaises », il n'y a, à l'inverse, aucune garantie que de « bonnes » institutions fonctionnent bien » (p. 402). En appui, ils convoquent le rapport de la Banque mondiale, mentionné ci-haut, où l'on peut lire : « lorsque les enjeux sont suffisamment importants, les communautés sont capables d'affronter le problème ». Duflo et Banerjee invoquent également Acemoglu et Robinson (cités abondamment) pour qui la raison du sous-développement est à chercher du côté des institutions héritées des périodes coloniales, que les dirigeants, arrivés au pouvoir après les indépendances, ont utilisées et utilisent encore à leur seul profit, omettant, au passage, de signaler l'opinion des auteurs concernés à propos de la microéconomie du développement qu'ils promeuvent et les méthodologies qu'ils emploient. Ainsi, Daron Acemoglu (2010, p. 26) avertit que lorsque les institutions et politiques ne sont pas mises en place exclusivement, ou même principalement, ce n'est guère en raison d'un manque de compréhension des principes économiques de la part des décideurs politiques. Au contraire ces mêmes décideurs introduisent ou maintiennent de telles orientations pour rester au pouvoir, pour s'enrichir, ou parce que des élites politiquement puissantes s'opposent au changement : rivalité, concurrence, introduction de nouvelles technologies ou amélioration des droits de propriétés de leurs travailleurs, et qu'ainsi, la perspective implique que la théorie redevienne particulièrement importante pour évaluer les effets possibles des interventions politiques à grande échelle. Pour cet auteur, les analyses contrefactuelles qui ignorent les facteurs d'économie politique, comme celles qui ne tiennent pas compte des effets d'équilibre général, peuvent donner des réponses trompeuses. Les preuves micro ou même macro (équilibre général) convaincantes concernent les effets d'un changement particulier de politique sur les résultats économiques et ne sont pas en elles-mêmes suffisantes pour évaluer ce que seront les implications lorsqu'une telle politique sera encouragée, voire mise en œuvre.

Pour les promoteurs des ECR, il n'y a pas de formule magique pour éradiquer la pauvreté mais il est possible d'améliorer la vie des pauvres. Pour ce faire, les auteurs

s'appuient sur ce que Kako Nubukpo (2012, p. 112) appelle « la normativité déconcertante » consistant à informer les pauvres, les soulager du trop-plein de leurs responsabilités, leur rendre le marché accessible, croire en leur capacité à s'en sortir et à éradiquer les préjugés à leur encontre.

### **Section 3 : Pauvreté, ciblage et gouvernement du social**

Les schémas normatifs s'appuient et se fondent, à coup d'expérimentations, sur des aspects explicatifs qui, souvent, relèvent de l'anecdotique ou du lieu commun pour inférer des généralisations contestables. Les schémas de financement de ces expérimentations assurés par des bailleurs de fonds (organismes internationaux ou grandes fondations) et l'accompagnement sur le terrain par des associations de ce que l'on appelle la société civile (serait-on face à des sociétés militaires ?) relèvent de cette doxa qui veut que le traitement des politiques sociales soit confié au marché et à « la charité organisée » nationale et internationale. Alors que les États du tiers monde fonctionnent comme rhizome, selon l'expression de Jean- François Bayart, où les tiges apparentes constituent les institutions incarnant la modernité et les tiges souterraines portent la trace des mémoires historiques et anthropologiques et incarnent les forces de la tradition. Pour Bayart (entretien Alternatives Économiques N 152 du 01/10/1997), les forces portant la modernité et celles se drapant sous la couverture de la tradition sont généralement incarnées par les mêmes acteurs qui interviennent dans ces deux dimensions.

S'agissant des approches microéconomiques du développement face aux carences des États, Bayart estime, non sans raison, que cette propension à réduire les champs des interventions de l'État repose sur le postulat qu'il y aurait d'un côté une société civile et, de l'autre, une autorité souveraine minimum pour laisser la société civile prospérer et produire un gouvernement efficace du social. Très souvent, cette société civile est indissociable de l'État et nombre d'organisations non gouvernementales, partenaires privilégiés des bailleurs de fonds internationaux, « servent en fait de cache-sexe aux réels détenteurs du pouvoir » selon cet auteur.

A titre d'illustration, au Maroc, souvent décrit comme le « bon élève » des institutions de Bretton Woods, l'initiative nationale pour le développement humain annoncée par le Roi en Mai 2005 a été instituée comme un lieu symbolique



d'énonciation du politique « déchargeant » l'État de ses responsabilités par un appel à la société civile pour enclencher la technicisation et la formalisation des dispositifs d'intervention publics créant une opportunité nouvelle pour les bailleurs de fonds en matière de politiques sociales et participer à leur co-production (Hibou et Tozy, 2015). Cette construction participe de la consécration de la « légitimité gouvernementale des associations, considérées comme particulièrement aptes à comprendre les questions de justice sociale et mieux armées que les administrations pour agir en faveur des pauvres et des démunis. Parfois liées aux partis politiques mais le plus souvent issues de ladite « société civile », légitimées par l'hégémonie internationale de la rhétorique néolibérale et relayées par les bailleurs de fonds, les associations sont en passe de devenir des acteurs incontournables du gouvernement du social » (p. 382).

Les lectures du J-PAL, par ses principales figures, donnent à voir des manières très spécifiques par lesquelles le social, anhistorique, est saisissable à travers une traduction des recommandations du consensus de Washington où des acteurs précis, incarnant la figure du pauvre, « font face à des problèmes concrets qui une fois bien identifiés sont résolus un à un ». L'échec des politiques destinées à lutter contre la pauvreté serait dû « aux failles des théories qui sous-tendent ces politiques plutôt qu'au caractère écrasant de la tâche » (Duflo et al., 2012, quatrième de couverture).

Or en dépit d'une rhétorique (développée dans le cinquième chapitre) savamment entretenue, et en dépit de la réduction des références théoriques due à « la circulation réticulaire entre un nombre réduit de chercheurs », selon les bons mots d'Agnès Labrousse (2010) signalés plus haut, les expérimentations randomisées s'inscrivent dans un schéma théorique qui puise dans les fondements des théories néo-classiques, qu'elles concernent la place du marché, des agents individuels, la rationalité des pauvres ou encore la place des incitations et, en somme, tout ce qui est de nature à participer à l'analyse des aspects positifs et explicatifs : capital humain, inclusion, participation et rationalité qui met l'accent sur les individus et leur propre responsabilité.

La rationalité, entendue ici, relève des développements de la théorie qui sous-tend que l'esprit est dans des dispositions où l'on cherche à « ne pas gaspiller les

ressources et à vouloir obtenir un résultat à moindre coût » en tenant compte de « l'aspect de la vie sociale marquée par la lutte contre la pauvreté et l'enchaînement de la production » (Berthoud, 2012, p. 9). Berthoud poursuit, non sans malice d'ailleurs, qu'il y a des économes, des économies et des économistes. Il s'interroge sur cette distinction : est-elle valable en tous lieux et pour tous les temps ? Les hommes n'ayant jamais pu vivre dans des situations d'abondance ni dans des sociétés où l'on ne doit éviter les gaspillages, définissant les institutions et les règles, rattachant, ainsi, les actes économiques à toutes les autres dimensions de la vie sociale. L'hypothèse de la rationalité individuelle est appréhendée par l'économie néo-classique comme une sortie de « l'état d'enfance des sociétés » vers « l'état civilisé », s'analysant comme un processus économique garantissant les libertés individuelles, et évitant la confusion des passions et les désirs humains de semer la guerre (Mardellat, 2006, p. 28). Pour cette école de pensée, la forme d'économie qui réalise l'essence même de l'économie comme discipline est « l'économie en équilibre général et que l'activité de l'homo-economicus, lorsqu'elle est rationnelle et que nulle entrave ne s'oppose à la rationalité, produit cette économie » (p. 20).

Dès lors, les notions d'individu et la rationalité qui lui est attribuée est à définir dans l'organisation dans laquelle il est inséré (Benetti et Cartelier, 1994, p. 19). Ces deux auteurs notent que « la rationalité dont le théoricien dote les individus est en fait le reflet de la rationalité attribuée à la société qu'ils forment. Dès lors, le problème central n'est pas de savoir si les individus sont ou ne sont pas maximisateurs mais celui de la possibilité de cohérence (ou de la rationalité) de la société de marché ». L'hypothèse de rationalité est définie alors comme le calcul individuel de la maximisation de la fonction d'objectif sans tenir compte, curieusement, de la contrainte budgétaire. Il est vrai que pour la pensée orthodoxe, la réponse est à trouver du côté de la détermination de l'équilibre général concurrentiel. C'est, du reste, ce qui rend cette hypothèse secondaire pour l'aspect positif mais central pour l'aspect normatif.

Pour ce courant, la pauvreté et sa persistance sont dues aux imperfections du marché. Ces imperfections sont, en particulier, de trois ordres : crédit, externalités et pièges découlant des risques. D'autres analyses y ajoutent des raisons plutôt de

type essentialiste découlant de l'histoire, de la géographie ou de la culture. Nous comprenons alors combien ces grilles explicatives occupent une place importante dans les approches du J-PAL et les thématiques qu'il aborde : accès aux services financiers avec le crédit pour la génération de revenus et l'épargne pour se prémunir contre les chocs, protection contre les risques par des produits alternatifs de micro-assurance (santé, récoltes et bétail) ou bien les transferts ciblés.

Le débat, dès lors, situé entre ciblage et universalité des prestations sociales est encadré dans le dilemme efficacité-équité. Cet emboîtement n'est pas étranger à celui des évaluateurs « embarqués » des politiques publiques, partagés entre indépendance et connivence, entre science et ingénierie sociale, selon l'expression de Dubois et Podevin (2011).

Ce qui n'apparaît guère dans la démarche des randomistas a trait aux fondements théoriques sous-tendant la philosophie qui motive leurs interventions. Ignorant systématiquement les théories de la justice, même lorsque Amartya Sen est évoqué pour les questions des libertés et des capacités, les fondements théoriques reliés aux imperfections du marché relèvent plutôt des théories du bien-être où les considérations d'équité doivent être traitées séparément des questions d'efficacité. Ainsi, lorsque le marché fonctionne dans une configuration où les agents se comportent de façon concurrentielle et que chacun dispose de toute l'information nécessaire sur les caractéristiques de tous les biens, alors l'équilibre est optimum et l'efficacité produit spontanément de l'équité. Ce n'est que lorsque la distribution de dotations initiales est très inégale et lorsque les droits sont mal spécifiés ou bien mal respectés que l'État peut jouer un rôle de redistribution pour satisfaire aux impératifs de l'équité. Sans s'attarder sur cette dernière, notion équivoque que l'on oppose souvent à l'égalité ou à la justice, l'allocation optimale peut être réalisée en tant qu'équilibre concurrentiel avec des transferts appropriés.

Les transferts ciblés promus comme un levier de lutte contre la pauvreté sont à apprécier à l'aune de cette justification du rôle que peut jouer l'état dans « la répartition » des ressources sans affecter négativement l'efficacité. Les aides ciblées relèvent de cette catégorie de traitement à l'instar du programme Progresa ou

Opportunidades au Mexique, Tayssir (facilitation) au Maroc. Le programme des bons de scolarité en Colombie est encore plus intéressant à saisir, en ce sens que les pays ont été incités à expérimenter des programmes donnant aux pauvres un accès aux écoles privées jugées meilleures que les écoles publiques par le biais de bons de réduction. Ces incitations sont à rapprocher de celles utilisées à encourager les mères indiennes à vacciner leurs enfants. Ce n'est que lorsque les lacunes du marché sont difficiles à combler qu'il est envisagé de combiner conjointement les fonctions allocatives, par des biens collectifs, et redistributives, par des transferts appropriés. De cette manière, on peut procéder aux corrections des défauts du marché pour atteindre l'équilibre concurrentiel sans déroger aux prescriptions du Consensus de Washington tant pour ses aspects conjoncturels que structurels.

Le rapport de la Banque mondiale de 2006, consacré aux thèmes de l'équité et du développement, imbrique égalité des chances et absence de privation absolue comme modalité nécessaire à la croissance et au développement et, qu'avec les imperfections des marchés, « inégalités de pouvoir et de richesse se traduisent en inégalité des chances, source de gaspillage du potentiel productif et d'inefficacité dans l'allocation des ressources » (p. 8) et que le fonctionnement imparfait du marché est dû soit à des « défaillances intrinsèques », soit à « l'asymétrie de l'information » ou encore aux distorsions imputables à l'État (pp. 8-9). Ce qui frappe, au demeurant, dans cette étude par rapport à cette micro-économie du développement, est la réduction de toute la problématique du développement à l'unique question de la lutte contre la pauvreté. C'est un étrange déplacement de l'aspiration légitime au développement qui range les approches structurelles issues de plus de soixante-dix ans d'économie du développement dans les oubliettes de l'histoire. Ce déplacement opère un glissement sémantique du développement vers le rôle que devrait jouer l'État (de la revendication de moins d'État à mieux d'État) mais aussi à la notion de « gouvernance » et ce qu'elle charrie au niveau du discours au sujet de la lutte contre la corruption. L'aide au développement et l'accompagnement par les institutions financières internationales et les grandes fondations sont à apprécier à l'aune de ce glissement et les programmes d'évaluations qui y sont associés. La montée de la randomisation des programmes

du J-PAL soutenus par la Banque mondiale et d'autres organismes internationaux entre dans ce cadre paradigmatique de l'évaluation par la preuve. Comme nous le verrons, un des instruments importants qu'utilise la Banque mondiale est le document appelé DSRP (documents de stratégie de réduction de la pauvreté) où les bénéficiaires de l'aide doivent expliciter leurs stratégies de lutte contre la pauvreté en lien avec une stratégie de développement global. Temple (2010) indique que les DSRP examinés semblent souvent et remarquablement similaires au Consensus de Washington, et n'hésite pas à qualifier cette caractérisation de ventriloque (*Ventriloquism*) qui ne permet pas d'appropriation par les bénéficiaires (*Ownership Capacity*). Temple impute l'incapacité d'appropriation aux pressions des donateurs qui restent trop prescripteurs et finalement les engagements en faveur de l'autonomie ne serait, d'après cet auteur, que poudre aux yeux (*Window Dressing*) ; pressions que Rodrik (2007, p. 973) appelle contraintes contraignantes (*Binding constraints*).

L'histoire de l'enrichissement généralisé au sein d'une nation qui s'était opérée à travers toute une série de transformations structurelles et mentales a permis les processus d'accumulation que l'on connaît et qui n'effacent pas pour autant les inégalités et encore moins les raisons fondamentales de leur persistance. Ces inégalités, n'épargnent ni les pays développés ni les pays en développement.

La montée du marché de l'évaluation a partie liée avec la fabrique du développement et la lutte contre la pauvreté axée sur la promotion de conditions institutionnelles, juridiques et sociales impliquant les pauvres eux-mêmes dans cette lutte contre la pauvreté. Les vertus du marché sont avancées comme le moyen de sortir de la pauvreté, structurant le débat international et la recherche « évaluative ». Dans ce sillage, la question de l'efficacité de l'aide, depuis au moins le début des années 2000, a provoqué une multiplication d'expérimentations appelant à l'usage de nouveaux mécanismes et de nouvelles approches telles que les évaluations *ex-ante* mobilisatrices de financements et privilégiant les stratégies de réformes globales au détriment du financement des projets. Les nouvelles orientations prises s'articulent autour de trois axes : cadre stratégique unique, financement intégré comprenant le coût des évaluations et mécanismes de contrôle et d'évaluation. Les

pays qui souhaitent bénéficier de ces mécanismes se doivent de préparer leurs programmes de lutte contre la pauvreté, consignés dans un document appelé « Document stratégies de réduction de la pauvreté (DSRP) » (Cling et al., 2002). Les initiatives, qui en ont résulté, ont été entreprises par suite des mises en cause adressées à l'encontre des institutions de Bretton Woods pour pointer les effets négatifs des politiques restrictives qu'elles imposent aux pays en développement.

Cling et ses co-auteurs soulignent d'une part les critiques, dits de droite, qui cherchent à dépolitiser le contenu des interventions en soulignant l'absence de preuves quant à l'efficacité de la participation dans les processus d'amélioration des politiques publiques, en revendiquent des approches strictement technocratiques s'appuyant sur des études « scientifiques ». D'autre part, les critiques, dits de gauche, qui considèrent, de leur côté, que le contrôle citoyen et la responsabilité démocratique des autorités et du personnel politique sont déficients du fait de la captation des ressources publiques par ces dernières et assurée de la nature prédatrice des États concernés. Pour ces derniers critiques fort suspects, par ailleurs, à l'égard des véritables intentions de ces institutions, la lutte contre la pauvreté ne serait qu'un slogan (pp. 165-167).

Les auteurs convoquent fort judicieusement Hirschman pour souligner la rhétorique des « effets pervers », mobilisée par les deux camps pour « disqualifier les nouvelles stratégies de lutte contre la pauvreté ». Hirschman dans *Deux siècles de rhétorique réactionnaire* (Charle, 1992) part du schéma évolutionniste de Marshall selon lequel les deux derniers siècles furent marqués successivement par la société civile et la déclaration des droits de l'homme, la politique avec le suffrage universel et le social avec l'organisation de l'État-providence.

Hirschman, cité par Adler, souligne que ces différentes étapes ont vu émerger à chaque fois des opposants aux réformes pour mettre des contre-feux avec une argumentation se matérialisant par trois figures rhétoriques réactionnaires : l'effet pervers, l'inanité et la mise en péril, de ce que l'on a appelé la carte des rhétoriques de l'intransigeance (1992, pp. 510-512). Cling et ses co-auteurs pensent infondés les griefs de « duplicité systématique des institutions de Bretton Woods dans



l'adoption du principe fondateur de la lutte contre la pauvreté » mais évoquent, à juste titre, Bruno Lautier qui relève que cette lutte se focalise sur l'imposition de l'incidence de la pauvreté et voit dans « l'indicateur central de mesure de la pauvreté, à l'aune duquel seront jugées ses interventions, le résultat d'un calcul politique ». En effet, la Banque mondiale et les arguments scientifiques mobilisés à cet égard, cherchent à privilégier « les moins pauvres des pauvres » et ainsi, afficher des résultats rapides et à moindre frais (p. 168). Le problème de l'identification des pauvres ne va pas sans soulever d'énormes problèmes étant donné le caractère multidimensionnel de la pauvreté. Ainsi, d'après Ehrhart (2006), il convient de prendre en compte certaines dimensions non monétaires de sorte que la mesure de la pauvreté fondée sur la consommation ou le revenu puisse être complétée par d'autres éléments d'appréciation comme la nutrition, l'espérance de vie, la mortalité des moins de cinq ans et les taux de scolarisation. C'est ce que fait la Banque Mondiale dans son Rapport de 1990 en étendant la définition de la pauvreté à la privation de besoins fondamentaux (nutrition, santé, éducation, logement, etc.) (Ehrhart, 2006, p. 622). Cependant, ses analyses demeurent prisonnières d'une approche monétaire traitant l'ensemble des pauvres comme une catégorie où l'on ne tient compte ni des effets d'hétérogénéité ni des différences d'un pays à l'autre, ni même au sein d'un même pays, s'autorisant, par conséquent, des généralisations hasardeuses qui ne tiennent pas compte des processus d'appauvrissement à l'œuvre dans chacun des pays. Bruno Lautier et Pierre Salama (1995, p. 246) relèvent, à ce propos, que « les débats sur la pauvreté menés dans les organisations internationales évacuent généralement la question de sa définition, au profit de son repérage et de sa mesure, préalables à l'invention de solutions techniques ». Ainsi, mettre un seuil identique et même en vision unidimensionnelle (un dollar par jour en parité de pouvoir d'achat de 1985) pour le pauvre de la lointaine banlieue de Sao Paulo et le pauvre paysan du Nord du Niger, est impossible pour des raisons techniques (indices des prix) et surtout à cause des désaccords, sur le panier de consommations non alimentaires (Lautier, 2002, pp. 145-146). Le seuil unique, que les chercheurs du J-PAL utilisent, indistinctement et à l'identique, ne sert pas à comparer mais à additionner les pauvres, il entraîne l'adhésion des médias et évacue la question de

l'inégalité. La Banque mondiale (1990, propos rapportés par Lautier) écrit : « si un grand nombre de pauvres se situent juste au-dessous du seuil, une hausse du revenu fait passer de nombreux individus au-dessus du seuil et l'incidence de la pauvreté diminue considérablement. A l'inverse, s'il y a une plus grande inégalité aux environs du seuil de la pauvreté la même hausse du revenu ne fait passer qu'un petit nombre de pauvres au-dessus du seuil et la réduction de l'incidence de la pauvreté est beaucoup moindre » (p. 148). Ainsi, ne mesure-t-on plus les inégalités entre riches et pauvres mais entre pauvres et pauvres.

Dans le cadre de la stratégie de réduction de la pauvreté menée par la Banque mondiale, il est requis de recueillir les opinions des différents acteurs de la société dont, en particulier, celles des pauvres et à les impliquer dans le processus participatif, en recommandant de donner à ceux-ci « l'opportunité de s'exprimer et d'influer sur les décisions publiques qui touchent à leurs conditions de vie » (Cling et al. 2002, p. 141). A l'effet d'assurer cette participation des pauvres, l'approche dite « évaluation participative de la pauvreté » connue sous le terme PPA (*Participatory Poverty Assessment*) est utilisée pour recueillir le point de vue des participants. Sans être formalisée de la sorte, on retrouve systématiquement cette approche dans toutes les approches du J-PAL ; aussi bien à travers ses articles que dans les ouvrages commis par ses figures emblématiques. Les nombreuses histoires rapportées à l'appui de la revendication de la solidité des méthodes expérimentales randomisées, confinant parfois à l'anecdotique, émaillent le récit de pauvres « experts en pauvreté » ; contrairement aux experts qui ont pris l'habitude de décider à la place des pauvres de ce qui est bon pour eux sans prendre la peine de les consulter, où la vision de la pauvreté est dominée par les caricatures et les clichés : « le pauvre paresseux, le pauvre entrepreneur, le pauvre affamé ». Duflo, dans un entretien accordé à l'AFP (14/09/2019), promeut une démarche inverse : « Si on veut comprendre les problèmes liés à la pauvreté, il faut dépasser ces caricatures et comprendre pourquoi le fait même d'être pauvre change certaines habitudes dans les comportements, et d'autres non ! ». Aussi, ces expérimentateurs entreprennent des démarches visant l'éradication de la pauvreté en s'appuyant sur une approche plus pragmatique et plus « clinique » mettant à l'épreuve les faits

et le terrain en s'inspirant de la vie réelle des pauvres. Elles interviennent comme un processus de naturalisation de la décision politique (Robert, 2008) dépassionnée et dissociée des intérêts partisans.

Cette vision n'est pas lointaine de l'esprit de l'évaluation participative de la pauvreté prônée par la Banque mondiale, mentionnée plus haut, reconnaissant que « les pauvres sont des « experts » en matière de pauvreté et qu'ils sont les mieux placés pour définir la nature du phénomène, ses origines et les moyens de « s'en sortir ». Par ailleurs, la reconnaissance des multiples dimensions de la pauvreté sont difficiles à appréhender dans les enquêtes qualitatives classiques et sont, de surcroît, irréductibles au concept couramment retenu basé sur le seul critère monétaire. La démarche participative s'inscrit de manière générale dans une optique plus large qu'un simple recueil d'information. Elle vise à inclure différents acteurs clefs, notamment des représentants des pauvres, dans le processus de suivi des politiques mises en œuvre (Razafindrakoto et Roubaud, p. 143). Duflo précise plus explicitement, encore, le rôle des expérimentations dans l'accompagnement des organisations internationales en la matière : « ... Nous travaillons beaucoup avec la Banque mondiale qui, sous l'impulsion de François Bourguignon lorsqu'il en était l'économiste en chef, s'est de plus en plus intéressée à cette approche. La Banque mène aujourd'hui 80 expérimentations en Afrique. Entre autres au Maroc où le gouvernement prépare un projet d'allocation scolaire dont nous testons plusieurs schémas et que finance la Banque mondiale. Nous faisons, ainsi, des allers-retours avec la théorie, comme en Inde sur le microcrédit, ce qui nous a permis de valider des questions profondes sur l'aléa moral et l'antisélection » (les Échos du 01/03/2009). Au demeurant, cette participation s'appuie sur l'alignement aux mesures structurelles portées par le consensus de Washington comme les droits de propriété sur les biens que possèdent les pauvres pour les solvabiliser, suivant en cela les approches de Hernando de Soto ou encore l'accès au marché financier par la microfinance telle que pratiquée par la Grameen Bank et plus encore, et de manière différente, ce que Duflo appelle « les institutions de marketing social et toute une partie de la Banque mondiale qui pousse au contrôle des enseignants et des infirmières par les usagers de ces services » (Les Échos, 01/03/2009). Cette ingénierie relève de ce qu'elle-même appelle « l'empowerment formé », qui part de

l'idée de donner aux pauvres la responsabilité de lutter, par eux-mêmes, contre la pauvreté. Se réclamant une filiation avec Amartya Sen ou encore avec Joseph Stiglitz, elle prend ses distances avec les alternatives formalisées par Théodore Schultz pour qui les paysans traditionnels sont pauvres mais efficaces, en ce sens qu'ils utilisent efficacement les ressources rares auxquelles ils ont accès, le caractère traditionnel venant surtout du fait qu'ils manquent de ressources qui leur permettent d'acheter de nouvelles machines et que, souvent, ces machines sont inappropriées à leur situation (Petit, 2011, p. 101) alors que pour elle (2009, 2012), une des raisons de l'inefficience est due à la procrastination des paysans. Il conviendrait, cependant de rétablir l'explication à l'endroit : « les pays pauvres dépendent beaucoup de leur agriculture comme principale source de leurs exportations, du paiement de leurs importations et du service de la dette, d'alimentation de leur population, etc. Cependant, d'une part, ces agriculteurs sont sous-équipés et ne disposent pas d'infrastructures appropriées. D'autre part, ils sont mis à mal par la vague de libéralisation néolibérale imposée par les politiques d'ajustements structurels du Consensus de Washington alors qu'ils n'étaient guère en mesure de supporter les distorsions de concurrence qui leur était imposées. La mondialisation néolibérale aggrave en effet le problème de compétitivité de ces agriculteurs qui subissent de plein fouet la détérioration des termes de l'échange, laquelle rend leur avenir de plus en plus difficile à cerner » (Sergot, 2009, pp. 26-27).

Nous nous trouvons, en conséquence, face à des stratégies appuyées par des études scientifiques, *ex-ante* ou *ex-post*, s'inscrivant dans des trajectoires d'économie de marché favorables aux pauvres et élaborées selon des procédures participatives à l'effet de leur donner une légitimité « démocratique » sans savoir si l'expression issue de la participation est représentative des pauvres ni si elle découle de conflits d'intérêts ou de perceptions multiples et variées. Razafindrakoto et Roubaud notent que le recueil de la « voix des pauvres » pose « le double problème de la représentativité des groupes interrogés et de la transcription des résultats en information utilisable pour mettre en œuvre des mesures concrètes au niveau national » (p. 145) ; ce qui amène « à s'interroger sur la gestion des conflits d'intérêts au sein de la société civile et sur la manière dont la définition des politiques prend en compte les résultats des processus participatifs » (p. 190).

Comme nous l'indiquions plus haut, les déterminants relatifs à la responsabilité et à l'individuation des communautés concernées se réduisent généralement à la parentèle élargie selon l'expression de Cogneau (2003, p. 63). Ce dernier signale, d'ailleurs, que la faiblesse des États subsahariens leur a fait accepter formellement et sans conditions les *policy-packages* qu'on leur a proposés non sans fortes résistances au sein de ces sociétés (p. 65). De telles « résistances » ont souvent été mises sur le compte du culturalisme africain comme « si ces sociétés ne souhaitent pas aussi plus d'emplois industriels, plus de concurrence et plus de démocratie » signalait, fort justement, l'auteur.

Précisons enfin que le lien J-PAL/Banque mondiale et ses affiliations institutionnelles ont une histoire que Jatteau (2018) retrace dans une sociographie très fouillée portant sur le fonctionnement du J-PAL et sa structure pyramidale et fortement hiérarchisée. Ainsi, lorsque Duflo (2004) écrit que la Banque mondiale ne joue pas le rôle de promoteur de nouvelles idées alors qu'elle le pourrait, Jatteau estime, quant à lui, que « la conversion de la Banque mondiale s'est alors opérée rapidement », ajoutant que « dès 2005, elle crée un département entièrement consacré aux questions d'évaluation, le DIME (*Development Impact Evaluation*) et, deux ans plus tard, un fonds spécifiquement consacré aux évaluations d'impact, le SIEF (*Strategic Impact Evaluation Fund*) qui, s'il n'est pas entièrement dévolu à la randomisation, lui est très favorable », données recoupant celles fournies par l'enquête menée par Pamies-Sumner (2013, pp. 11-25).

Jatteau (2018) évoque le lien qu'entretiennent les chercheurs enseignants du J-PAL avec les grandes fondations caritatives américaines en signalant une pratique inconnue ailleurs. Ainsi, Duflo est “Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics”, Banerjee “Ford Foundation International Professor of Economics” et, enfin, Kremer porte le titre de “Gates Professor of developing Societies”.

Comme déjà mentionné, le rôle et l'intérêt porté par François Bourguignon aux approches microéconomiques du développement et de la lutte contre la pauvreté qui, à partir de ses responsabilités comme vice-président en charge de l'économie du développement et chef économiste, a donné une impulsion décisive à l'usage des

expérimentations aléatoires comme méthodologie d'évaluation en usage par la Banque mondiale. Devaux Spatarakis évoque l'accusation portée par Banerjee contre la Banque comme poche de « résistance à la connaissance ».

Le cadrage cognitif, évoqué ci-haut, montre comment les institutions du développement mobilisent des groupes de travail à l'effet de donner des gages de scientificité à leurs prescriptions. Devaux-Spatarakis (p. 69) revient sur « *The Evaluation Gap Working Group* », groupe de réflexion ayant compté parmi ses membres J. Gueron, du MDRC (Manpower Demonstration Research Corporation) et fervente partisane des expérimentations aléatoires contrôlées au sein de la « *coalition for evidence based policy* », Bourguignon et autres auteurs du rapport « *When will we ever learn? Improving lives through impact evaluation* » pour évoquer la « pièce manquante » dans la conduite rigoureuse d'évaluations d'impact. L'auteure précise, encore, que le message en question est relayé aussi par des financeurs nationaux qui exigent de la Banque mondiale le développement de la pratique évaluative. Suivant les instigations de cette « Coalition », les États-Unis exigent, à leur tour, que certains projets de la Banque mondiale soient évalués rigoureusement avec une préférence pour les expérimentations aléatoires. L'Espagne, de son côté, a alloué un fonds dédié spécifiquement au le développement de la conduite d'évaluations d'impact, fortement encouragé par le DIME (*Development Impact Evaluation*) créé en 2005 (pp. 68-69).

Après avoir quitté la BM, Bourguignon se démarque de la revendication du meilleur standard. Ainsi, à propos de l'approche qui a dominé cette économie du développement qui ne tient pas compte des considérations d'ordre macroéconomique, il écrit : « Cette approche est-elle seule pour autant susceptible de donner à elle seule la clé du décollage économique ? ». On peut en douter répond-il : « une mesure expérimentée avec succès dans un contexte culturel, environnemental ou économique n'est pas nécessairement transposable dans un autre contexte. En outre, ses résultats peuvent être différents si la mesure s'applique à toute une population plutôt qu'à de petits échantillons... D'autre part, même si leur efficacité est prouvée pour réduire la pauvreté, des interventions de nature essentiellement microéconomiques restent insuffisantes pour lancer et entretenir le développement » (2015, pp. 40-42).



Gaston Bachelard, qui considère que les problèmes ne se posent pas d'eux-mêmes, que rien n'est donné et que tout est construit, estime que les vérités ne sont pas à chercher dans l'expérience, elle-même objet à correction par l'abstraction des concepts. Et il soutient que pour son achèvement, la science s'oppose à l'opinion qui, pour être légitimée, doit se fonder sur d'autres raisons que l'opinion (1934, p. 17). Pour l'auteur de la *formation de l'esprit scientifique*, « le commerce n'hésite pas à faire glisser l'emploi des comprimés sur les malaises les plus variés » (p. 131). C'est ainsi que la question que pose Bruno Lautier (2002) à propos du « *pourquoi faut-il aider les pauvres ?* » apparaît, dès lors, pertinente à examiner à l'aune de cet ordre qui publicise la pauvreté dans des contextes de privatisation des missions publiques. Lautier situe les réponses de la Banque mondiale aussi bien dans le rôle de bailleur de fonds que dans sa fonction de « mentor idéologique de nombreuses institutions et experts ». Usant d'une rhétorique assez coutumière dans son discours qui « organise le dicible et assure la division du discours discursif » (Corten, 1998, p. 6), la Banque mondiale déclare que sa raison d'être est d'aider les pays à réduire la pauvreté et, qu'à partir de ce travail discursif et de ses implications positives, il convient d'analyser et de situer les travaux d'expertise destinés à « euphémiser » la dimension normative et prescriptive des recommandations des institutions de Bretton Woods.

A ce propos, Cécile Robert (p. 316), évoquant les experts médicaux et les membres des comités d'éthique, montre comment les procédés discursifs sont utilisés pour légitimer les diagnostics et les préconisations, non en tant que tels, mais amenés « discrètement » et « subrepticement » aux détours de démonstrations scientifiques ou de propos philosophiques. En somme, pour emprunter la voie d'une science morale, Corten va plus loin pour parler de « pauvreté » comme objet sémiotique voué à disparaître, en opérant un tournant marquant le passage de la raréfaction à l'incitation, deux notions empruntées par l'auteur à Michel Foucault et utilisées comme procédures du discours social. L'assignation aléatoire et le discours qui l'entoure dans la conception et le développement des expérimentations semblent, à travers toute la littérature du J-PAL examinée, relever de cette division du travail discursif. Le soin mis pour objectiver le travail mené et la méthode utilisée expliquent l'évitement des pièges de l'économie politique inapte, selon ces

expérimentateurs, à saisir les mécanismes et ressorts de la pauvreté et les moyens de son éradication. Ainsi, pour les bailleurs de fonds, donateurs et ONG attentifs aux besoins des pauvres, André Corten nous invite, malicieusement, à le comprendre ainsi : « résolu à appliquer la stratégie ». De la politique, de l'économie politique et des rapports de pouvoir, le glissement, qui n'est pas que sémantique, verse dans le registre du débat commandé par la morale (Lautier, p. 138). Retraçant l'histoire de la pauvreté, Lautier rappelle que la question sociale était au cœur de la responsabilité des États, quatre siècles durant et que ce sont leurs dirigeants qui sont jugés lorsque la pauvreté est présentée comme affligeante et infamante : « On pourrait, bien sûr, incriminer un système, capitalisme, impérialisme, néo-colonialisme, néo-libéralisme ou autre ; cela fut fait, dans un passé pas si lointain, mais ce n'est pas fait, ni par la Banque mondiale, ni par aucune des organisations de L'ONU. Si la pauvreté n'est pas le fait du système, il ne reste que deux possibilités : les populations elles-mêmes (et avant tout les populations pauvres) ou les dirigeants des États. S'agissant des populations pauvres, on verra plus loin que la rhétorique du « mauvais pauvre » n'a en rien disparu ; mais par construction, elle ne peut désigner tous les pauvres : « certes, il y a toujours eu de nombreux simulateurs, fainéants, drogués, ivrognes et délinquants ; la technologie du repérage et de la classification des pauvres est là pour les débusquer mais le vrai problème, ce sont les pauvres malgré eux » (p. 139).

Comme mentionné précédemment, Duflo et Banerjee (2011, p. 404) vont jusqu'à considérer « qu'il est possible de modifier la gouvernance ainsi que les politiques sociales et économiques sans changer les structures existantes ». Ces positions morales construites sont à appréhender à travers deux questionnements. Primo : *quelle cohérence peut-il y avoir entre principes moraux et techniques mises en œuvre pour satisfaire l'argument moral ?* Secundo : *le discours fondé sur la morale produit-il ses effets uniquement dans le champ de la pauvreté ou cherche-t-il à créer un nouveau « type de citoyenneté et nouveau type de société » insérés dans l'ordre dominant conforme aux prescriptions d'organismes internationaux ?* (p. 141).

Cette préoccupation d'ordre moral se rattache en réalité à une lignée qui remonte à Walras, distinguant justice productive et question morale, où les rapports marchands ont pour rôle de « produire la justice ». L'ordre moral dont il est question

ici se trouve légitimé par les apports d'une société civile (grandes fondations) engagée dans la perpétuation de l'ordre libéral dominant d'une part, et l'expertise académique « embarquée » d'autre part, mettant le marché dans son rôle de producteur du maximum d'effets « justes ». L'action ne doit plus se porter sur les structures mais sur les individus dont il convient de renforcer les capacités en les dotant de droits de propriétés, par exemple, ou en faisant bénéficier les personnes en extrême pauvreté d'aides ciblées. Ainsi la focalisation sur la question de la pauvreté réduit-elle la notion du développement à la « recherche de compensations aux effets de la mondialisation marchande » (Lautier, 1999, p. 789).

A propos des aides ciblées, Rocha (2004) signale, à travers un travail très minutieux mené au Brésil, pays où les politiques sociales se sont largement appuyées sur des mesures pro-pauvres menées lors de la présidence Lula et largement saluées, que les programmes d'allocations sociales à destination des pauvres au Brésil démarrés dans les années 1970 avaient connu un point d'inflexion décisif au milieu des années 1990. C'est ainsi qu'à partir de 2003, la Bourse-École et les différents nouveaux programmes fédéraux d'allocations sociales ont été réunis sous l'appellation Bourse-Famille (Bolsa Família). L'auteure estime, cependant, que c'est grâce au plan Real de 2004, qui a mis fin à l'hyperinflation chronique, que la vie des Brésiliens et la gestion publique ont changé, permettant l'amélioration significative des conditions de vie des familles (en matière de santé, d'éducation et d'assainissement, de baisse de la mortalité infantile et d'allongement de l'espérance de vie à la naissance) dans un pays pourtant très inégalitaire. Rocha impute ces améliorations davantage au rôle tenu par le revenu du travail qu'à la contribution des allocations sociales. En dépit de l'expansion importante de ces allocations sociales, celles-ci ne représentaient, en 2011, que 1,4 % du revenu des familles – avec des contributions quasiment identiques de la BPC (Prestação Continuada) et de la Bourse-Famille.

Cela rend la question du ciblage inextricable selon ce que Rocha considère comme difficulté à définir des « cibles » sûres. Il est souvent difficile de s'assurer que les fonds alloués soient bien redistribués à ceux qui en ont réellement besoin. Au demeurant, il se trouve que la place laissée à l'évaluation censée produire des preuves est utilisée plutôt comme un faire-valoir (Progresá au Mexique ou réforme

constitutionnelle en France pour autoriser les expérimentations sociales avant généralisation), les décisions ayant été prises avant la fin des expérimentations censées les légitimer. Cet usage politique de la preuve destiné à l'objectivation des options politiques choisies et à la mise en apparence scientifique de l'action publique n'échappent pas aux postures de « manipulations des ressources et des modes de vie » selon Chelle (2012, p. 261).

Campbell (1991, p. 223) semble avoir raison lorsqu'il parle de sociétés d'expérimentations qui n'existent pas encore : « Une société expérimentatrice sera celle qui expérimentera vigoureusement des solutions à des problèmes récurrents, qui fera des évaluations déterminées et multidimensionnelles des résultats et qui s'orienteront vers de nouvelles alternatives lorsque les évaluations démontrent qu'une réforme a été inefficace voire dangereuse ». La scientificité est, en conséquence, mobilisée pour organiser alors le dicible que nous mentionnions plus haut. Ainsi l'objet des politiques du développement se trouve-t-il détourné de son objet visant à fonder l'espoir de milliards d'individus à sortir des trappes du sous-développement et qui se trouve dissous dans la lutte contre la pauvreté déplaçant l'aide vers l'humanitaire et les situations d'urgence selon Lautier (p. 790), les transferts ciblés jouant, au mieux, le rôle de revenus primaires.

Un autre avatar de cette doxa libérale, avancé un moment, est à voir dans ce que l'on appelle la « redistribution à l'envers » ou « les effets de ruissellement » (*trickle down effect*), consistant à donner davantage de moyens aux plus fortunés par l'abaissement des niveaux d'imposition et nourrir, ainsi, le tissu productif et, partant, le reste de la société. Arnaud Parienty (2018), commentant le propos du Président Macron (15 octobre 2017) à propos des premiers de cordée, signale qu'aucun scientifique n'a jamais démontré cet effet de ruissellement.

Un autre résumé du *trickle-down effect* est donné par le directeur du budget de Reagan (1981) : « Donner les réductions d'impôts aux tranches supérieures, aux individus les plus riches et aux plus grandes entreprises, et laisser les bons effets « ruisseler » à travers l'économie pour atteindre tout le monde ». La courbe de Laffer démontre, a contrario, que la pression fiscale responsable de réduction des recettes de l'État est démentie dans les faits et, en tous cas, ces mesures se sont

plutôt accompagnées, corrélativement dans les pays développés, de baisses du niveau des prestations sociales (aide au logement et indemnisation du chômage en France à titre d'exemples).

Ce qui frappe dans ce renouvellement de l'économie du développement prôné par ses promoteurs est qu'il fait sciemment l'impasse sur toute la littérature qui fonde cette discipline depuis soixante-dix ans, ce qui en relativise la prétention à la nouveauté. Le renouvellement en question postule que seuls les résultats sont à confronter aux faits, révélant ainsi une perspective qui se caractérise par la fixation de « bornes conceptuelles de la lutte contre la pauvreté en envisageant la richesse de quelques-uns et la pauvreté de nombreux autres comme les conséquences nécessaires du fonctionnement ordinaire de la société » (Lagneau - Ymonet 2013, p. 75). Servet (2018) estime que la multiplication des expérimentations, dans divers pays, engendre un « vagabondage planétaire, difficilement compatible avec une investigation en profondeur de chacun des terrains abordés » et, faute de recherches fouillées en collaboration avec des spécialistes locaux, on ne débouche que sur le constat de certains résultats peu éloquents quant à leurs causes et aux processus ayant engendré les faits (p. 126).

Ceux que l'on désigne comme pauvres, eux-mêmes ne se désignent que rarement comme tels et, pour s'en convaincre, il n'y a qu'à consulter une partie des nombreux travaux consacrés aux thèmes de la pauvreté qui tous soulignent, très largement, la variabilité de cette notion et son évolution à travers les époques. Beaucoup de ces travaux, dont certains ont été évoqués plus haut, sont allés au-delà des questions strictement techniques avec pour seul objectif le dénombrement des pauvres ou leur identification, et se sont intéressés au caractère multidimensionnel de la pauvreté. Dans cet ordre, Paugam (1998, p. 140) met en exergue la difficulté pour les chercheurs qui ont pour objet d'étude la pauvreté et l'exclusion de maintenir la distanciation conceptuelle nécessaire entre les modes de pensées qui nourrissent le débat social et enrichir simultanément ce débat. La Banque mondiale et les experts se rangeant du côté de ses modes de pensée et d'action se focalisent sur la nécessité de faire accéder les pauvres au marché sans questionner les relations de pouvoir qui génèrent leur exclusion du marché et les processus de disqualification sociale qui les frappent.

A cet égard, il convient de mentionner l'ouvrage de l'auteur iranien Majid Rahnema, *Quand la misère chasse la pauvreté*, où il distingue « *pauvreté-misère* » d'aujourd'hui (privations, frustrations et indignités) et « *pauvreté digne* » dans certaines cultures (vie frugale, sentiment d'appartenance à une culture, fierté de cette différence et de dimensions spirituelles fortes). S'y ajoute également l'ouvrage de Gilbert Rist (2013), *Le développement, une croyance occidentale*, pour qui la pauvreté n'est pas le signe de la maladie du capitalisme mais, qu'au contraire, la preuve de sa bonne santé. Pour Rist, dans cet élan, lutter contre la pauvreté revient à chercher à en atténuer les effets les plus désastreux et non à transformer le système qui la produit. Il considère par ailleurs que les objectifs du millénaire, portés par l'Organisation des Nations Unies (ONU) et son corollaire, l'aide au développement, sont des avatars qui remplissent cette fonction et contribuent, par conséquent, à soutenir le système de pensée qui les porte. Pour sortir l'étude des inégalités des revenus en économie de la marginalisation, Anthony Atkinson (2016) appelait à faire sortir du froid - du désert - la répartition des revenus. *Bringing Income Distribution in from the Cold* était l'intitulé d'un discours qu'il avait prononcé en sa qualité de président de la Royal Economic Society.

Pour la Banque mondiale (Rapport 2000/2001, p. 93) et les experts mobilisés pour fonder scientifiquement ses prescriptions « le pouvoir de coercition de l'État peut être un puissant atout pour les autorités nationales, décidées à favoriser l'accumulation d'actifs par les pauvres, mais [...] une ingérence induite de l'État peut entraver le fonctionnement des marchés et éroder les incitations à l'investissement privé – détruisant les possibilités d'emploi au lieu d'en créer ». Duflo considère, à son tour, que « As economists increasingly help governments design new policies and regulations, they take on an added responsibility to engage with the details of policy making and, in doing so, to adopt the mindset of a plumber. Plumbers try to predict as well as possible what may work in the real world, mindful that tinkering and adjusting will be necessary since our models give us very little theoretical guidance on what (and how) details will matter... Economists should seriously engage with plumbing, the interest of both society and our discipline » (2017, exergue). Cette vision contraste avec les notions où les inégalités économiques ont



pour origine essentielle les inégalités salariales, de revenus et de patrimoine, les deux premières, mesurées en termes de flux, alimentent la constitution de stocks de patrimoines qui génèrent à leur tour des revenus de capital ou de propriété, aggravant les inégalités de distribution que ce soit au niveau de la distribution primaire (par les droits et protections) ou au niveau de la répartition secondaire que ce soit par l'impôt progressif ou par les transferts sociaux sous conditions de ressources (Biasutti et Braquet, 2016).

Sur chacun des aspects analysés dans ce chapitre, se dégage une remarquable convergence entre l'agenda des politiques promues par la Banque mondiale, notamment celles s'attachant aux politiques dites d'empowerment, et le succès des RCT dans la production des données quantitatives (Bardet et al., 2012, p. 178).

## CHAPITRE V : Expérimentation, rhétorique, performativité et fabrication des politiques publiques

Le positionnement scientifique des expérimentations aléatoires contrôlées s'appuie sur des logiques microéconomiques et micropolitiques qui opèrent des renversements inédits dans les rapports entre savoirs et pouvoirs. Alors que son impuissance théorique et sa limitation à un champ fort restreint dans les modalités mises en avant pour lutter contre la pauvreté sont largement admis, le discours porté par leurs promoteurs sature les champs académique et médiatique et leurs performances dépassent leurs intentions. Les politiques fondées sur les preuves (*evidence based policies*) mises en scène, sont « performées » par des déplacements qui transforment les peuples en sociétés civiles, les territoires en espaces, la justice en efficacité, le jugement en évaluation, la règle en objectif et la liberté en flexibilité selon une catégorisation sémantique qui, selon Alain Supiot (2015), « congédie le vocabulaire de la démocratie politique au profit de celui de la gestion » (pp. 47-48). Ce déplacement paradigmatique qui relève de ce que l'on a appelé le « courant pratique » se fonde dans le mouvement général porté par la circulation internationale du modèle de la preuve. Importé de la médecine (*evidence based medicine*), il s'étend à tous les champs couverts par ce que Foucault appelle la gouvernementalité libérale entendue comme déplacement des débats sur la nature et la légitimité de l'État vers ses pratiques définies comme exercice du pouvoir. Les nombreux liens et dépendances de réseaux nationaux et internationaux, voire supranationaux, les insèrent dans des réseaux qui consacrent de nouveaux rapports entre la recherche, l'expertise et la politique que l'on cherche paradoxalement à « dépolitiser ». Les standards et les bonnes pratiques tiennent lieu de règles de droit échappant ainsi aux contrôles démocratiques par la substitution des marchés aux responsabilités dévolues à l'État social.

Cette *empirical evidence* est portée par une rhétorique savamment entretenue, elle « incarne une épistémologie positiviste établissant une hiérarchie dans l'établissement des données probantes où les essais contrôlés randomisés occupent le premier rang aux côtés de dispositifs quasi-expérimentaux, des revues systématiques de la

littérature scientifique et des méta-analyses des données (Normand, 2017, p. 21). Les référentiels sur lesquels s'appuient cette rhétorique comprennent trois dimensions définies par Bruno Jobert (1992) pour décrire les controverses et débats qui traversent les représentations sociales dans la conduite des politiques publiques : dimension cognitive fournissant les éléments de l'interprétation causale, dimension normative définissant les valeurs à respecter pour traiter les problèmes et, enfin, dimension instrumentale pour orienter l'action en fonction des savoirs et des valeurs.

Selon cette grille, nous retrouvons ces trois dimensions dans les travaux du J-PAL : le contrefactuel pour inférer les causalités, le marché comme norme et modalité de substitution à l'État social et la randomisation comme instrument. La portée de cette économie par les preuves ne saurait se saisir en dehors du cadre d'études empiriques précises qui donnent à voir « ce qui marche » sans jamais se soucier du comment ou du pourquoi « ça marche ». La performativité est alors à appréhender dans le cadre de cette épistémologie particulière dont les limites et les apories ont été évoquées plus haut.

En construisant le socle de leurs autonomisations politiques par rapport à tous pouvoirs transcendants, les États ont été conduits à renforcer leurs pouvoirs dans la lignée de l'œuvre de Machiavel (le prince doit désormais moins renforcer son pouvoir que la puissance de l'État), par les sciences camérales d'abord et la statistique ensuite et enfin l'économie politique, signale Cécile Robert (2008, p. 309). Au fil du temps, les États se sont dotés d'instruments et de dispositifs de production de politiques publiques, armés de connaissances et d'outils de compréhension où se nichent les savoirs et les techniques spécifiques nécessaires au développement des richesses matérielles et humaines. Au tournant du vingtième siècle, et plus particulièrement après les années cinquante, la figure de l'expert scientifique ou du spécialiste, auquel est reconnue une autorité singulière, se généralise pour contribuer à la fabrication des politiques publiques ou pour conseiller « le prince ». L'expertise judiciaire (médecine légale, arpentage, horlogerie, comptabilité, architecture, informatique, etc.) était déjà admise pour éclairer les décisions des tribunaux (Dumoulin, 2000). Comme suggéré par C.

Robert, la source de cette autorité singulière se matérialise par le travail rhétorique qui permet d'attester de « l'universalité et du désintéressement du jugement expertal » et se fonde sur « une compétence spécifique des compétences nécessaires à la conduite de l'action publique » (p. 311). S'opère, ainsi, le transfert de légitimité qui s'appuie sur l'objectivité et l'absence d'arbitraire. Le développement du marché de l'expertise a multiplié les espaces et les opportunités où se déploie ce savoir-faire particulier : comités d'éthiques, clubs de réflexion ou think tank, comités de normalisation et groupes d'évaluations relevant aussi bien de centres privés que de laboratoires de recherche. L'influence anglo-saxonne est, bien entendu, dominante par le jeu d'acteurs, où recherche de performances et culture de l'évaluation se conjuguent consubstantiellement à la réédition des comptes.

C'est dans cet ordre qu'il convient d'apprécier le développement des techniques basées sur la preuve et l'élaboration des politiques fondées sur celle-ci. Le J-PAL totalement inscrit dans ce cadre a, dès le départ, étendu son influence dans une perspective partenariale de construction et de fourniture de preuves en s'appuyant sur un objet massif de réflexion, d'action et d'articulation avec les circuits de prise de décision. Aussi, est-il intégré à ce que Jobert appelle les « *forums* », instances mobilisées au service de ce que l'on appelle le *tournant libéral*. Cet auteur distingue trois espaces parmi ces forums : le scientifique, celui de la communication politique et enfin ce qu'il appelle le forum des communautés politiques. Boussaguet (2014) y ajoute judicieusement le forum professionnel auquel nous n'hésiterons pas à associer le J-PAL tant cette entité, par son organisation et sa structure (on y reviendra), s'apparente davantage à une entreprise professionnelle fortement hiérarchisée qu'à un strict laboratoire académique, eût-il été porté par un prestigieux établissement d'enseignement et de recherche, le MIT en l'occurrence.

Dans ce chapitre, nous approchons deux notions rarement traitées dans la recherche critique au sujet des essais randomisés appliqués à la microéconomie du développement : la mobilisation de la rhétorique, entendue au sens aristotélicien, investissant les espaces appropriés et mobilisant les discours adaptés pour emporter la conviction. La force de l'argumentation (*gold standard*) et les formes mises en

œuvre s'appuient sur une performativité entendue dans le sens que lui confère Michel Callon où il ne s'agit pas simplement de représenter le monde mais d'aspirer à le réaliser, voire le constituer et finalement instaurer ce dont on parle (Section 1). La Seconde section s'attachera principalement à décrire les mécanismes par lesquels le J-PAL s'insère dans les forums aussi bien académiques, politiques et médiatiques et où, chercheurs, réseaux d'action locale et bailleurs conjuguent leurs efforts au service du tournant libéral » évoqué ci-dessus.

## **Section 1 : Expertise, rhétorique et performativité**

On doit à Angus Deaton (2009 et 2019) l'usage de la rhétorique pour qualifier la prétention des RCT à dire la vérité. Deaton (2019) estime que les promoteurs des expérimentations aléatoires randomisées ont tendance à présenter les estimations comme si elles étaient des faits établis. La rhétorique des RCT est que les essais peuvent établir la vérité, ce qui n'est pas le cas (*They cannot*, écrivit-il).

Il faut, cependant, remonter jusqu'à Aristote et son ouvrage *la rhétorique* (édition 1991) pour saisir les principes de la rhétorique contemporaine utilisée dans le traitement de l'usage du discours « pour plaire comme pour convaincre, pour plaider comme pour délibérer, pour raisonner comme pour séduire » souligne Michel Meyer dans l'introduction du livre éponyme (p. 5). Elle se laisse définir, poursuit Meyer « comme l'art de persuader et elle est, plus généralement, le lieu de rencontre de l'homme et du discours, ce qui vérifie la définition classique de l'homme comme animal raisonnable, où la passion issue de l'animalité vient échouer sur les rives du logos » (p. 6). Dans l'usage de la rhétorique que pratique les promoteurs des expériences randomisées comme outils d'élaboration de la preuve par des données probantes, on détecte les figures et les catégories mentionnées déjà par Aristote pour soutenir une raison, défendre une méthode ou attaquer les concurrents. Ainsi sont mobilisées les techniques ou l'enthymème. Dans un article fondateur sur le rôle des évaluations aléatoires dans l'évaluation de l'impact des programmes d'aide au développement, la figure emblématique du J-PAL écrit que « les bénéfices de l'évaluation de l'impact des programmes de développement et de la connaissance des programmes qui fonctionnent et ceux qui ne fonctionnent pas ne dépassent pas

beaucoup un programme ou une agence quelconque » (Duflo, 2005). Pour cette auteure, les évaluations d'impact crédibles sont des biens publics et le manque d'évaluations crédibles est perçu comme signe de la défaillance des programmes eux-mêmes, alors même que la crédibilité participe de la construction du support du développement. Elle souligne, encore, que « les méthodes traditionnelles [...] sont sujettes à des biais sérieux dus à des variables omises » considérant que si l'indépendance de la banque centrale requiert d'autres méthodes d'évaluation, tous les programmes visant les individus (hygiène, gouvernement local, éducation ou santé) peuvent être candidats aux évaluations par assignation aléatoire. Aussi, constate-elle, qu'il y a un déséquilibre dans les méthodologies d'évaluation au détriment des RCT. Cet article datant de 2005, l'on ne peut que constater sa puissance rhétorique et sa capacité à faire effet par la conjonction d'éléments de langage (dirions-nous aujourd'hui) et évoqués par Aristote près de 24 siècles auparavant ! La force de persuasion et de plaidoyer associée au reproche du cofondateur du J-PAL Abhijit Banerjee à l'égard de la Banque mondiale, de « résister à la science », semblent avoir produit leurs effets. Le constat de Bédécarrats et ses co-auteurs est sans appel : « Si la Banque mondiale semble avoir une politique plus nuancée en matière de méthodologie d'évaluation d'impact, en diversifiant les méthodes (toujours quantitatives), les RCT en représentent près des deux tiers, avec 64 % des 368 évaluations engagées (jusqu'en 2010). Non seulement les RCT tendent à occuper une position de plus en plus dominante, mais elles exercent un effet d'éviction sur les autres approches. Au cours de la période 2000 - 2004, à peine 20 % des évaluations étaient des RCT » (p. 111).

Conscients des limites de l'utilisation des RCT à embrasser les questions liées à la macroéconomie en matière d'aide au développement, matérialisée par les déboursements en tant que support au budget général de certains PED, les promoteurs des RCT, pour des motifs politiques et d'efficacité, n'en demandent pas moins de les fractionner en des programmes qui ciblent les individus ou communautés pour des objectifs bien définis. Pour faire face aux oppositions exprimées à l'encontre des expériences aléatoires, ils suggèrent, par ailleurs, d'exiger que tout programme à financer doit être nécessairement être évalué avant



son déploiement à grande échelle (le programme Progresa au Mexique par exemple) et qu'il faudrait envisager de financer les évaluations par subventions et non par prêts (Duflo 2005). Le rôle des grands bailleurs de fonds (agences créées à cet effet et grandes fondations) répond précisément à cette requête et rend ainsi viable le cadre partenarial et performatif au sein duquel prospèrent les RCT.

Aristote nous apprend que la rhétorique s'attache à la dialectique, Plutarque mentionnant qu'elle n'y est pas subordonnée mais elle lui est coordonnée. Ainsi en va-t-il des preuves. Celles qui sont inhérentes au discours, nous dit Aristote, sont de trois catégories : le caractère moral de l'orateur qui amène la persuasion, la disposition de l'auditoire quand ses passions sont excitées par le discours et enfin le discours lui-même lorsqu'on persuade qu'il porte la vérité ou paraît l'être. Les équipes de l'évaluation par assignation aléatoire usent de l'ensemble des registres de ces techniques à longueur d'articles où les uns citant les autres dans l'exercice réticulaire d'auto-référencement déjà évoqué. Face aux critiques frappant les limites de la randomisation, Duflo avance des considérations morales pour dire, en conclusion de sa leçon inaugurale au Collège de France, sa passion pour la lutte contre la pauvreté : « Je souhaite pratiquer l'économie comme une vraie science *humaine*. Une science, rigoureuse, impartiale, sérieuse. Une science *de l'homme*, dans toute sa richesse et sa complexité. Mais une science humaine dans sa fragilité et sa modestie. Une science humaine finalement : généreuse, ambitieuse, engagée. Prendre ma part dans la construction laborieuse d'un savoir contre la pauvreté est mon travail et ma raison de vivre ». Cette profession de foi relève du même ordre moral que la déclaration de la Banque mondiale « notre raison d'être est de réduire la pauvreté » ou encore « notre rêve, un monde sans pauvreté ». S'ensuivent les attaques en règle contre les autres méthodologies d'évaluation sur les registres mêmes où les limites de leur méthode sont pointées (biais y compris de publication, généralisation, coûts, etc.). La force de persuasion se manifeste par une saturation de l'espace médiatique et la présence assidue dans les forums et espaces où sont définies les modalités d'évaluations et leurs financements. Le discours médiatique participe à la fabrication du consentement et tend à structurer les champs de la réflexion et à définir les positionnements. Même lorsque les savoirs sont imprécis, leur circulation en direction des sphères profanes peut les rendre audibles et aptes

à emporter l'adhésion. Nous avons déjà mentionné le Center of Global Development et son rapport (déjà cité) : « *When will we ever learn? Improving lives through impact evaluation* » auquel il faudrait ajouter le groupe consultatif d'aide aux populations les plus pauvres (Consultative Group to Assist the Poor : CGAP), l'Innovation for Poverty Action (IPA) et le Laboratoire Abdul Latif Jameel d'action contre la pauvreté (J-PAL). Ces entités interviennent au sein de forums tels que les définit Jobert réunissant des experts et des spécialistes de renom à l'objet « d'examiner comment l'innovation dans la mise au point et la distribution de produits financiers peut permettre aux 2,5 milliards d'adultes, qui à l'échelle mondiale ne sont toujours pas bancarisés, de bénéficier de services financiers formels » peut-on lire dans une brève de la Banque mondiale du 27 Février 2015.

L'auditoire dont il est question fait partie intégrante du forum évoqué par Bruno Jobert. Quant à la vérité, les Randomistas suivent les préceptes aristotéliens professant de voir à la racine de la connaissance des simulations proprement intellectuelles, un désir de la connaissance pour la connaissance, par-delà les fins utilitaires, le projet de l'étonnement... » (Jouary, 1996, p. 14). Le même auteur cite Rousseau et son discours sur l'inégalité : « nous ne cherchons à connaître que parce que nous désirons de jouir » et qu'un être qui ne connaît rien, ne s'étonne de rien. Pascal dans *De l'esprit géométrique, De l'art de persuader* cité par Boudon, parle de « puissances qui nous portent à consentir ». Ce dernier considère, par ailleurs, que la question n'est pas tant la subjectivité des raisons (*ni arbitraires ni liées à la singularité du sujet*) qui pose un problème mais le fait que ces raisons soient objectivement non fondées. Il ajoute que les idées fragiles ou fausses proviennent souvent, après détours et médiations plus ou moins complexes, des sciences humaines, les idées « les plus viables sont peut-être celles qui reposent sur une argumentation explicite valide et dont les conclusions sont contaminées par l'implicite », et ainsi tenons-nous pour démontrées des idées douteuses (Boudon, 1990, pp. 45, 61 et 68). Lorsque, par exemple, Duflo évoque la filiation de la randomisation dans les sciences sociales avec celle appliquée à la médecine, au moment même où des ténors et des pères de cette discipline s'inquiètent des dérives causées, soit par la manipulation des résultats cliniques soit par l'influence des grands laboratoires pharmaceutiques ou encore par la parcimonie avec laquelle sont

dispensés les crédits publics à la recherche, elle feint d'ignorer cet avertissement quant au pouvoir de cette filiation à conférer à la méthode revendiquée un transfert de scientificité.

De même, lorsqu'on reproche aux travaux menés sous la bannière du J-PAL le déficit de théorie, elle parle du pouvoir des expériences aléatoires randomisées à tester les théories : « L'expérience de terrain présente des avantages pour tester la théorie, à la fois sur l'évaluation rétrospective des programmes et sur les expériences de laboratoire » (Duflo, 2009, leçon inaugurale au Collège de France). Ainsi tour à tour plaide-t-elle pour, à la fois réfuter et valider la théorie comme l'expérimentation des moustiquaires menée par Dupas et Cohen qui permet de tester la théorie de l'élasticité des prix à l'aide de l'expérience menée au Kenya : « ...Cohen et Dupas ont découvert que le fait d'augmenter le prix des moustiquaires imprégnées d'insecticide de 0 à 60 cents réduit de 60 % le nombre de ceux qui achètent les moustiquaires ». En complément, elle ajoute : « En Zambie, le fait d'augmenter le prix d'un désinfectant de l'eau de 9 à 24 cents a réduit de 30 % la part des personnes acceptant l'offre » (Banerjee et al., 2009, page 719). Ici, les tests de la théorie se veulent relever à la fois du réfutationisme et de la confirmation sans rapport avec l'idéal poppérien.

Nadeau (1995, p. 86) rapporte que McCloskey caractérise la science économique de « processus conversationnel » où la rhétorique est mobilisée pour faire valoir son point de vue sur tout autre. Facchini (La vie des idées, 03/2017) explique qu'en participant à la critique de la significativité statistique qui fait confondre « significativité substantielle et significativité statistique », McCloskey dans « *Les péchés secrets de la science économique* » révèle les insuffisances et les fautes de l'économie dominante en pointant les trois péchés des économistes qui pourraient être des vertus : obsession de la mesure, goût pour les mathématiques et biais en faveur du libéralisme économique. Les péchés dits véniels, pour ce qui concerne cette catégorie se matérialisent par l'usage immodéré de la figure de l'homo œconomicus. McCloskey recense, enfin, les péchés impardonnables marqués par le désintérêt pour les enquêtes, l'histoire économique, l'histoire des idées et l'arrogance en matière d'ingénierie sociale (Facchini, 2017). Dans la traduction

que l'on doit à Frobert (2004) de l'article de McCloskey *The Rhetoric of Economics* (1983), cette auteure considère que « ...La faconde des économistes est la preuve que les multiples méthodologies officielles dont ils se réclament ne sous-tendent jamais leurs recherches ». Et c'est ainsi, qu'elle en appelle à étudier la rhétorique des sciences économiques entendue comme un jeu verbal creux, une « rhétorique vide », ou une « rhétorique pure », qui trahit l'étroitesse d'un esprit imperméable aux nuances du débat. Elle s'appuie sur Wayne Booth (1974) et son *Modern Dogma and the Rhetoric of Assent* pour décliner la rhétorique en différentes significations : art d'étudier ce en quoi les hommes croient devoir croire et non d'établir la vérité au moyen de méthodes abstraites ; évaluent précisément les plus ou moins bonnes raisons permettant de parvenir à de plus ou moins probables ou plausibles conclusions fragiles mais préférables à celles obtenues par le hasard ou l'intuition aveugle ou encore l'art de découvrir des croyances non justifiables et de renforcer celles-ci au sein d'un discours partagé. L'objectif n'étant pas « de convertir une autre personne à une idée préconçue, mais de favoriser une reconnaissance mutuelle » comme cela se conclut, généralement, dans le marché des idées. Le même Booth (1967, p. 13) écrit : « Nous croyons à la persuasion réciproque comme mode de vie ; nous passons notre temps à aller de colloque en colloque ». En définitive, McCloskey considère que la rhétorique est l'exploration de la pensée au moyen de la communication. De manière générale lorsqu'il s'agit de politiques sociales, les politiques publiques et plus particulièrement celles qui ont rapport avec les questions économiques, mobilisent davantage la rhétorique argumentative. L'analyse n'est pas seulement d'ordre normatif ; ses énoncés sont aussi positifs pour dire ce qui est, ce qui était et ce qui sera, selon les propos d'André Gosselin (1998). Dans la littérature des partisans de la randomisation en micro-économie du développement, les procédés et dispositifs mobilisés au service de la rhétorique et évoqués plus haut sont légion. Que ce soit pour louer la méthodologie où l'on voit fleurir les expressions : gold standard, révolutionner les sciences sociales, expériences créatives, sensibilité atavique à la pauvreté, etc. Ils sont mobilisés également pour « attaquer » les méthodologies d'évaluations concurrentes sur leurs défauts putatifs : biais de toutes sortes, incohérences

temporelles dues aux jugements *ex-post*, coûts, incapacité « de deux millions de régressions à percer le secret de la croissance », etc.

La rhétorique en question ici, s'appuie également sur ce que l'on pourrait désigner par l'argument de communauté en ce sens que les expériences randomisées menées sous le patronage du J-PAL sont utilisées de manière réticulaire et agissent comme facteur décisif de persuasion mise en œuvre au service d'une stratégie-réseau. Cette stratégie explique largement la reprise répétitive d'expériences venant illustrer les effets démonstratifs de l'argumentaire, déployés à travers divers médias largement repris ci-haut (articles, ouvrages, interviews, conférences et leçons) en faveur des RCT.

Concernant la fonction performative du discours de la Banque mondiale sur la lutte contre la pauvreté, nous reprenons le propos d'André Corten au sujet du dispositif discursif qui fonctionne par la construction d'un objet sémiotique nouveau et qui, par ses conditions de production, est voué à disparaître. Comme modalité appréciative, cette description s'applique, à notre avis, aux dispositifs discursifs développés par le J-PAL et la rhétorique qui les soutient.

L'emploi de cette rhétorique est mis au service d'une performativité qui explique largement le succès de « l'entreprise » J-PAL. Cette performativité puise ses ressorts et se déploie avec vitalité aussi bien dans le fonctionnement, les structures que dans les moyens mobilisés pour sa prospérité. Pour saisir ce que cette notion donne à voir, nous l'examinerons, tour à tour, telle qu'elle est traitée dans ce que les Anglo-saxons appellent les « Sciences Studies » ou par la nouvelle sociologie des sciences développée, particulièrement, en France par Michel Callon et Bruno Latour.

Le concept de performativité dérive des énoncés performatifs dus au langage développé par John Austin, particulièrement dans son ouvrage posthume *Quand dire, c'est faire* (1962). Les conditions à remplir, selon Austin, pour que les énoncés performatifs puissent s'étendre à la performativité sont de deux ordres : premièrement, l'énoncé postule l'existence d'une procédure qui rende le discours légitime et intelligible. C'est ce que permettent la randomisation et tout

l'argumentaire construit autour de la méthode usitée par les expérimentateurs concernés ; deuxièmement, lors de l'énoncé, la procédure doit être exécutée par tous les participants de façon correcte et dans son intégralité.

L'Institution J-PAL évoquée tout au long de ce travail est érigée en gardien du temple garantissant l'unité du discours et des conditions de son exécution, nul ne pouvant soliloquer. Ainsi, une assertion comme « je vous fais mari et femme » prononcée par un Maire sonne comme un oracle qui scelle un destin et produit des effets irréversibles sur le cours d'une vie. Cet énoncé performatif est à comparer à l'énoncé « les RCT sont un gold standard » seul à même d'inférer une causalité. La performativité qui en résulte relève alors de la catégorie des actes « perlocutoires » selon la catégorisation qu'en fait Austin distinguant, ainsi, les actes de paroles en « locutoire » (acte de dire quelque chose), « illocutoire » (acte effectué en disant quelque chose) et « perlocutoire » (acte provoqué par le fait de dire quelque chose) d'après la lecture de Régis Meyran (2017, p. 13) de l'œuvre d'Austin. Ravallion (2018, Abstract) démontre par ailleurs que cette affirmation est discutable « While RCTs have a place in the toolkit for impact evaluation, an unconditional preference for RCTs as the “gold standard” is questionable ».

La performativité des dispositifs « perlocutoires » des Randomistas repose précisément sur des instruments construits autour d'un fonctionnement qui s'apparente davantage aux moyens dont se doterait une entreprise que ceux d'un laboratoire d'enseignement et de recherche. Comme évoqué plus haut, la très précise sociographie établie par Arthur Jatteau (Op. cit.) au sujet du J-PAL décrit un fonctionnement fortement hiérarchisé piloté par un conseil d'administration de 24 membres et un comité exécutif composé de huit chercheurs dont ses figures tutélaires, Esther Duflo et Abhijit Banerjee. La marche quotidienne est prise en charge par une administration (une bureaucratie selon Jatteau) occupée par de nombreux non chercheurs au sein du laboratoire central à Boston dont le site officiel indique: « J-PAL Global is our head office and coordinates J-PAL's worldwide research, policy, education, and capacity-building work to promote our mission of ensuring that policy is informed by scientific evidence » mais également au niveau des antennes régionales disséminées à travers le monde (Asie du Sud-



Est, Amérique Latine et Caraïbes, Afrique, Asie du Sud et Europe). Chacun des bureaux régionaux est dirigé par un Directeur-pays, personne clé pour contrôler l'unité méthodologique du laboratoire. Citant Colander et Coats (1989), Jatteau (2018) rapporte que « les compétences bureaucratiques peuvent être plus importantes que les compétences académiques... Les contextes organisationnels sont en effet primordiaux pour saisir le fonctionnement de la recherche ». La structure J-PAL est organisée également au niveau thématique avec des directeurs pour chacun des thèmes d'intervention du Laboratoire (agriculture, crimes violence et conflits, entreprises, environnement et énergie, finance, genre, marché du travail, santé, économie politique et gouvernance, éducation).

Les intitulés des thèmes (économie politique et gouvernance ou environnement et énergie, par exemple) peuvent surprendre dans la mesure où l'unité de la méthode et la focalisation sur la lutte contre la pauvreté restreignent le champ des investigations généralement abordées par les sciences économiques. Les financements favorables et conséquents ont été pensés et menés de manière à permettre une taille et un rayonnement à la hauteur de la notoriété et à l'influence acquises. Ces financements se montent à des millions de dollars et les rémunérations sont à l'image de ce que permet le Job Market américain. Toujours selon Jatteau, on dénombre plus de 700 partenaires du J-PAL se répartissant entre gouvernements, organismes de financement internationaux, ONG locales et très grandes fondations américaines. Ainsi a-t-il été recensé entre 2003 et 2015, près de 50 donations pour un montant légèrement supérieur à 60 millions de dollars par les seules fondations Hewlett, Gates, MacArthur et Ford. La consultation par nos soins des états financiers de l'IPA (Innovation for Poverty Action), et en particulier, le formulaire 990 de la déclaration d'impôts déposée au département américain du trésor (accessible au public) révèle que ce bras financier du J-PAL a reçu 41,1 millions de dollars en 2018 sous forme de donations et contributions contre 42,5 millions de dollars une année auparavant.

Les dépenses de l'association en question se sont établies à près de 40 millions de dollars en 2018 et 38,3 millions en 2017. Le total bilan de l'association a évolué de 19,4 millions de dollars en 2017 à 22,8 millions en 2018.

Il est intéressant d'examiner la nature des dépenses déclarées (chiffres arrondis aux milliers ou aux centaines) :

Dons et assistance aux organisations domestiques et gouvernements locaux	1.016 KUSD
Dons pour autres assistances, aux gouvernements et personnes étrangères	618,7 KUSD
Compensations et primes aux agents, aux directeurs, administrateurs et employés principaux	651,1 KUSD
Salaires et autres rétributions	20.504 KUSD
Voyages	4.730 KUSD
Voyages et divertissements officiels étatiques et locaux, conférences, conventions et réunions	673 KUSD

Source: Form 990. IPA, Department of the Treasury, Internal Revenue Service.

Pour assurer le succès de l'entreprise J-PAL, il ressort clairement de cette comptabilité, simplifiée à l'extrême, que ses promoteurs ont doté leur laboratoire et ses démembrements de moyens dignes du fonctionnement d'une firme orientée vers la maximisation de ses ressources, l'extension de son influence et se pourvoyant de moyens conséquents pour motiver son « personnel ». L'argent est alors mobilisé comme « argument persuasif afin de surmonter les réticences de certains » avertit Arthur Jatteau (2018). L'aspect Marketing, dont la nature de certaines dépenses, montre clairement qu'il est mis au service du primat scientifique de la méthode promue et de ses énoncés performatifs. Cette débauche de moyens, sans comparaison avec ce qui a généralement cours dans les milieux de la recherche académique, explique les conditions d'énonciation du discours et son épaisseur performative. La validité ou non, l'objectivité ou non de l'énoncé a, dès lors, moins d'importance que les modalités de sa construction. Nous devons à Nicolas Brisset (2015, p. 36) cette citation de Bruno Latour extraite de « *La science en action* » : « Pour déterminer l'objectivité ou la subjectivité d'un énoncé, l'efficacité ou l'inefficacité d'un procédé, nous ne chercherons pas ses qualités intrinsèques mais nous reconstituerons l'ensemble des transformations qu'il subit plus tard entre les

maines des autres ». Brisset convoque l'étude des relations entre les économistes et l'économie comme pratique sociale en lien avec leur champ d'investigation pour constater que lorsque l'économiste se penche sur la réalité sociale afin de la comprendre et la décrire, il modèle cette dernière en un sens particulier. Brisset cite encore pour appuyer ce constat Muniesa et Callon (2009) : « Cette intuition permet de qualifier les situations dans lesquelles l'objet sur lequel porte un travail scientifique n'est pas simplement constaté ou décrit, mais modifié, voire appelé à exister » considérant, ainsi, que la science économique est sujette à un effet de circularité d'où découlerait une grande relativité de ses propositions et conclusions. La discipline finit par rendre l'économie comme représentation de la réalité engendrant des comportements sociaux qui agissent de sorte qu'elle advienne « vraie ».

Un autre instrument essentiel complète la panoplie des outils de promotion, de dissémination de la méthode et du récit « randomisateur » : la revue *American Economic Journal: Applied Economics* participe à ce “*Scientific Business Model*”. D'après la formule de Bédécarrats et al., « l'effet d'éviction [d'autres articles basés sur d'autres approches] est aussi lié au fait que les promoteurs les plus influents des RCT sont souvent membres des comités de rédaction des plus grandes revues d'économie et d'économie du développement ». Esther Duflo est le rédacteur fondateur (Editor) de la revue en question.

Nous avons mentionné plus haut comment cette revue a été mobilisée pour louer le miracle de la microfinance telle qu'elle est mise en avant par ses promoteurs pour faire bénéficier les pauvres des vertus du marché. Ce levier supplémentaire n'est pas sans rappeler le rôle joué par les publications au service de la randomisation dans l'*evidence based medicine* utilisée par l'industrie pharmaceutique pour accroître ses profits, et il n'est pas inutile de rappeler ici les péripéties nées des polémiques consécutives à la pandémie récente et des rétractions d'articles qui s'en étaient suivies.

L'ensemble de ces dispositifs réunis peut s'analyser sous le prisme de la sociologie de la traduction ou celle de l'acteur-réseau développée par Michel Callon, Bruno

Latour et Madeleine Akrich (2006) du Centre de sociologie de l'innovation de l'école des mines de Paris. Cette grille d'analyse invite à prendre en compte l'abatement des cloisonnements et à reconsidérer le fait scientifique en fonction de la multiplicité des relations qui le constituent. Se trame, ainsi, l'entrelacs qui mobilise une structure fortement hiérarchisée, des financements favorables, des institutions partenaires politiques, financières, caritatives et académiques, qui ensemble forment ce que Jobert appelle *le forum* réunissant dans cette sociologie de la traduction : scientifiques, communicants et communautés politiques.

## **Section 2 : Forum scientifique, politique et médiatique**

Quel que ce soit le régime économique étudié, la puissance publique joue un rôle prépondérant en matière de politiques économiques. Soit directement, soit en laissant au marché une large part pour le développement d'échanges marchands, soit encore par la régulation à l'effet de pallier les dysfonctionnements du marché et faire, ainsi, face aux causes qui sont à l'origine de ces dysfonctionnements : externalités, asymétries de l'information et imperfections de la concurrence. Cela peut aussi être pour produire des biens publics présentant généralement les caractéristiques de non-rivalité et de non-exclusivité. Souvent, les trois leviers sont utilisés pour déployer les interventions par la puissance publique investie de la légitimité nécessaire qu'elle exerce dans un domaine spécifique ou à l'échelle de tout le territoire. En ce sens, l'État ou la puissance publique habilitée usent d'une approche normative pour définir les choix des politiques publiques les plus efficaces au regard des objectifs fixés (Bozio et Grenet, 2010, p. 5). Bien entendu, cette puissance publique n'est pas toujours l'agent bienveillant capable d'arbitrer entre l'intérêt général et les préférences individuelles organisées en groupes de pression constituant des réseaux de clientèles homogénéisés par des appartenances diverses et très actives dans l'orientation des interventions publiques. Les transformations des contextes internationaux induits par l'ouverture, les regroupements régionaux ou encore la décentralisation des décisions politiques et administratives opèrent également dans la fabrique de la décision publique et, partant, au niveau de l'élaboration des politiques associées. Prenons l'exemple du phénomène de la

corruption qui n'est pas seulement l'apanage des pays sous-développés biaisant, ainsi, l'intervention publique par le comportement de ses élites administratives et/ou politiques et, de ce fait, pervertit le marché tant loué. La puissance publique et ses interventions sont également à appréhender comme dispensateurs de situations de rentes par le levier de la dépense publique et son financement que ce soit par les incitations ou en substituant des monopoles privés à des monopoles publics à l'occasion des privatisations d'entreprises publiques générant de nouvelles « élites programmatiques » selon l'expression chère à William Genieys (2006). C'est à Domhoff (1967) que l'on doit l'émergence du concept de *Policy Planning Networks* agrégeant forums, think tanks et fondations pour influencer en amont la fabrication de la décision publique.

C'est dans ce contexte qu'intervient la déconstruction des systèmes de décision conduisant à « analyser des politiques publiques sous la forme de réseaux d'acteurs » opérant, de cette manière, la convergence des « paradigmes étatistes et pluralistes » (Muller, 2018). Il convient de signaler, cependant, que c'est à la sociologie économique que l'on doit ces changements paradigmatiques et de perspectives d'analyse. La sociologie économique introduit les dimensions de relations sociales et la dimension intéressée que l'on doit à Swedberg (2003) cité par Philippe Steiner (2007). Ainsi Steiner écrit-il en introduction de l'ouvrage susmentionné : « à rebours du mouvement qui porte les sciences sociales à diviser le travail intellectuel, à spécialiser les recherches, la sociologie économique suggère qu'il est nécessaire et utile de faire se rencontrer les théories économiques et sociologiques de manière à fournir de meilleures explications des faits économiques que ne le font l'un et l'autre de ces deux savoirs lorsqu'on les met en œuvre d'une manière isolée ou, pire, contradictoire » (p. 3) et que la résurgence de la sociologie économique dans la compréhension des phénomènes économiques, voire de la théorie, a « partie liée avec l'insatisfaction ressentie vis-à-vis de la théorie économique lorsqu'elle s'isole des autres sciences sociales » (p. 4). En ne tenant pas compte de cette interaction nécessaire avec les autres sciences sociales et de l'encastrement social des relations marchandes, le J-PAL se prive d'outils indispensables à la compréhension des phénomènes de la pauvreté et des modes de

sa construction dans les sociétés fortement inégalitaires. Pour saisir correctement cet encastrement social dans l'économie, il faudrait remonter à Karl Polanyi et l'opposition qu'il dresse entre deux forces sociales : celles qui travaillent à l'extension des échanges marchands et celles qui luttent contre cette extension au nom de principes autres que ceux de l'efficacité économique (Steiner, p. 39). Toute une littérature est par ailleurs mobilisée et tour à tour, Granovetter (2000), Wayne Baker (1984) et enfin Laville et Nyssens (2001), auteurs cités par Steiner, s'attachent à saisir les apports de la sociologie économique à la compréhension de ces phénomènes d'encastrement entre l'économie et le social. Granovetter, pointant le défaut d'approche de Polanyi à propos du clivage jugé non pertinent entre les sociétés primitives dans lesquelles l'économie est pleinement enchâssée dans les relations sociales et les sociétés modernes, où l'économie est totalement désenclavée, considère que pour le marché du travail, par exemple, les explications retenues par les économistes sont insuffisantes du fait qu'ils ne tiennent pas compte des « médiations sociales par lesquelles passe l'encastrement structural » ainsi que de l'emprise des cadres relationnels et institutionnels dans lesquels prend place l'action économique. Quant à Baker, il utilise le fonctionnement de la bourse de Chicago pour souligner le rôle de cet enchâssement des traders, au sein d'un réseau d'interactions, pour constituer une forme de coordination importante qui permet d'en saisir le fonctionnement. Enfin, Laville et Nyssens évoquent l'encastrement politique en prenant l'exemple du rôle des associations dans le secteur de l'aide aux personnes âgées selon que les décisions publiques favorisent ou non l'existence d'une économie plurielle articulant marché, État et associations.

Il en ressort que les activités du J-PAL en termes d'organisation, d'instruments et de discours tendent davantage à performer ses interventions qu'à produire de l'amélioration de la connaissance économique nécessaire à la compréhension des phénomènes étudiés ou à la conceptualisation de cette nouvelle économie du développement. Selon le langage de Callon, la performativité, ici, passe par l'incorporation du savoir économique dans les technologies mises en œuvre par les acteurs.



L'approche des acteurs-réseaux relève de ce que Jacques de Maillard et Daniel Kubler appellent les « coalitions de cause » mieux connue en anglais par les termes « *Advocacy Coalition Framework* ». Attribuée aux auteurs Paul Sabatier et Hank Jenkins-Smith, l'expression est due à l'insatisfaction vis-à-vis de la séquentialité de l'analyse des politiques publiques, en ce sens qu'elle ne contient pas de modèle causal et ne permet pas de formuler des hypothèses empiriques testables (de Maillard et al., 2015, p. 185). Décrivant les logiques de pouvoir dans le récit des politiques publiques de l'Union Européenne, Radaelli constate que ces récits « représentent la forme que revêtent les éléments cognitifs attachés à des politiques publiques » (2011, p. 257) et qu'ils retracent « les histoires causales » liées à la définition et à l'énonciation de celles-ci. Ainsi le récit remplit-il à la fois une fonction performative et persuasive quant à l'acceptation des programmes d'action qui les sous-tendent (de Maillard et al., p. 174) ; de cette manière il permet d'éviter le surgissement des questions critiques (p. 175). Les prémisses de cette approche par les coalitions de cause sont de nature top-down. Elles ne se limitent pas aux acteurs étatiques, elles impliquent d'autres « partenaires », organisations multilatérales, groupes de pressions, experts et experts académiques, acteurs associatifs eux-mêmes liés à des sources de financement et, enfin, communicants politiques au sens large. C'est ce que Jobert désigne par la terminologie forums, au sein desquels les coalitions sont encore plus actives depuis le tournant libéral de la fin des années quatre-vingt. Sabatier et Jenkins, cités par de Maillard et Kubler, parlent de « *Policy Subsystem* » formé par une pluralité d'organisations publiques et privées pour influencer les politiques. Jobert (1992) définit, en effet, les variations d'environnement non pas en fonction de la réaction mécanique des individus mais à travers les faits susceptibles d'être nommés et interprétés impliquant, ainsi, la réalité sociale de l'acteur et prédétermine sa ligne de conduite. Pour ce faire, il dénombre les trois dimensions susmentionnées formant les référentiels des politiques publiques : dimension cognitive donnant les éléments d'interprétation causale, dimension normative spécifiant les valeurs à respecter et enfin dimension instrumentale fixant les principes d'action et d'orientation en fonction des savoirs et des valeurs (pp. 220-221). En projetant la place occupée par le J-PAL et ses

promoteurs dans le rôle joué au niveau des trois registres du référentiel susmentionné, on constate que les dimensions rhétoriques, performatives et cognitives, dont ils usent, expliquent le primat qui leur est aujourd'hui accordé, tant au niveau scientifique et médiatique que par les dispositifs financés d'évaluation auxquels ils contribuent. Ainsi, pouvoir à orienter les politiques se conjugue avec détention des ressources (cognitives et matérielles) pour influencer les politiques publiques dans l'ordre libéral établi.

Dans une note critique portant sur le gouvernement par l'expertise scientifique et technique, Jérôme Lamy (2019) cite Ursula Klein (2012) pour évoquer la figure de l'expert scientifique-artisan dont les portraits « passant des confinements académiques aux vents du mercantilisme fourmillent, donnant davantage à voir une figuration pointilliste d'acteurs particuliers qu'un véritable traitement analytique de ce qu'est l'expertise ». Lamy va plus loin en précisant que « les modalités de production, l'espace de validation et les formes de traçabilité respectives de production des énoncés distinguent nettement l'expertise des autres pratiques de savoir ». Il ajoute que lorsque le lien entre l'expert et celui qui a demandé l'expertise est exposé, la connaissance experte est utilisée pour « limiter la prolifération des débats » et le savoir produit ne pourrait plus être saisi comme un savoir scientifique (p. 5). Cette forme d'expertise « embarquée » dans ce que Pierre Muller appelle les réseaux de politiques publiques (*policy network*) s'insère dans ces nouvelles configurations d'acteurs qui transcendent les frontières habituelles entre États et société civile, caractérisées souvent par des combinaisons horizontales sans hiérarchie particulière et où les ressources techniques se mêlent aux ressources politiques et à celles liées aux sources de financement. Commentant le paradigme de Jobert à propos des forums, Muller constate que ceux-ci fonctionnent selon des règles et des temporalités propres et la mise en scène des acteurs est ainsi différenciée. Le forum académique investi dans la recherche de l'excellence est compatible avec le paradigme dominant, le forum de la communication politique cimente la coalition politique partisane et le politique quant à lui est affairé à minimiser les coûts de transaction entre global et sectoriel, entre l'administratif et le professionnel et enfin entre le Technique et le Politique. Il indique, par ailleurs,

que l'accès aux cercles de décisions offre à l'acteur concerné les ressources nécessaires pour faire avancer ses intérêts. A l'aptitude fondée sur le mandat politique légitimant traditionnellement l'action, s'ajoute l'aptitude fondée par la légitimité conférée par le potentiel de participation aux processus de fabrication et de développement des programmes politiques. Ainsi politics (vote et mobilisation) et policies se combinent pour réinterpréter les transformations ou le « formatage », selon l'expression de l'auteur, bousculant ainsi le système de représentation et d'interprétation qui opère en arrière-plan des politiques publiques. Ce système fonctionne alors de façon que les solutions paraissent inéluctables (de Maillard et al., p. 176).

Dans ce remodelage reconfiguré par le rôle que jouent les organismes internationaux de développement et appuyé par le support de la communication politique, le but vise l'agrégation des différents intérêts susceptibles de participer à la formation des coalitions de cause, dont il a été question plus haut, et à imposer, ainsi le « néo-libéralisme doctrinaire » alimenté par le discours de la contrainte économique internationale et la globalisation des échanges. L'action qui en résulte est alors, en paraphrasant Lascoumes et le Galès (2004, p. 12), cet « espace sociopolitique construit autant par des techniques et des instruments que par des finalités, des contenus et des projets d'acteur ». La figure de l'expert apparaît, dès lors, centrale dans le processus de justification et de légitimation des prescriptions formulées, en particulier, dans « les marchés » de recommandations de politiques publiques conditionnées par des aides ou des appuis aux projets de développement destinés aux pays dits en voie de développement. Le statut ambigu de l'expert provient, dans le cas qui nous importe, de sa dépendance vis-à-vis du commanditaire d'une part et l'indépendance que lui permet une neutralité conférée par sa compétence scientifique et technique comme fondement légitime de la décision, d'autre part. La rationalité instrumentale se heurte à la rationalité technique, Céline Granjou soutient pour sa part que, par voie de conséquence, « le savoir technique de l'expert, censé faire apparaître les contraintes objectives de la réalité, est le seul fondement légitime de l'action, la dimension pratique (choix des valeurs) se trouve absorbée par la dimension technique, ce qui correspond à une dépolitisation de la

société » (2013, p. 176). Cette instrumentation, où l'expert-scientifique apparaît comme la figure de proue, illustre en réalité l'ascension de l'expertise privée et internationale qui se traduit par l'importation, à tous les niveaux de l'action publique, des préceptes des politiques néolibérales soumettant les fonctions intellectuelles et la recherche à des intérêts économiques (Delmas, 2011, p. 88). L'action publique y est employée pour agir sur l'individu par son implication et sa participation ainsi que par l'entremise d'une subjectivation qui tend à confisquer la démocratie en faisant l'impasse sur l'État social par l'accroissement de l'autonomie et la responsabilité au détriment de l'accès aux droits sociaux (p. 91). Delmas pointe l'expérience profane « renforcée par l'émotion et l'intime tant dans les discussions publiques que dans des médias portés au compassionnel et au spectaculaire » (p. 92). La propension est alors à transformer le lien social en centres de profits soumis aux droits privés et sommés d'obtenir des résultats. L'État est appelé à tout mettre en œuvre pour garantir au marché de fonctionner et, le cas échéant, expertise et expérience sont invitées à tisser des liens « dans le cadre de politiques de ciblage de plus en plus individualisantes pour faire face aux souffrances » (p. 96).

Ce changement paradigmatique dans la conduite des politiques publiques n'est pas un résultat issu du débat technique et scientifique ; il obéit à une lutte politique et idéologique tendant à réaliser l'extension des marchés au sein d'un monde globalisé et dont les organismes internationaux, particulièrement financiers, se font le chantre et le gardien. L'expertise et les processus cognitifs mobilisés sont mis au service de ce projet où la pauvreté par exemple, qui écrase près de deux milliards de personnes dans le monde, n'est pas considérée comme un construit social produit par l'extension des inégalités entre pays et dans chacun des pays, mais plutôt comme le résultat d'imperfections de marchés faisant craindre l'implosion de l'État providence par la faute de structures administratives et productives en mal de réformes structurelles qui entravent, de surcroît, la liberté d'entreprendre et déresponsabilisent les individus par rapport à la prise en main de leurs propres destins. Selon de Maillard et Kubler ce changement s'opère à travers deux dimensions différentes : une dimension systémique qui opère au niveau des processus de coordination à l'échelon des actions gouvernementales, mais,

également au niveau des comportements sociaux, et une dimension qui a trait au travail intellectuel structurant les représentations sociales et véhiculant normes et valeurs. Ce cadre que l'on doit à Jobert et Muller dans *L'État en action* introduit le concept de « *référentiel de politiques publiques* » (pp. 179-180), par lequel, les acteurs s'orientent et donnent sens aux politiques publiques. Par la caution scientifique que les promoteurs des expérimentations aléatoires contrôlées apportent aux politiques de ciblage de la pauvreté et de ses données probantes ainsi qu'aux préceptes promues par les organisations internationales, ils participent à la consécration du référentiel du marché et de ses politiques publiques. En fournissant ce que Pawson (2006) appelle la monnaie vitale (*vital currency*), ils constituent, de cette manière, un rouage essentiel du dispositif du réseau d'acteurs qui y est impliqué. Le jugement de Pawson est, du reste, sans appel : si l'alliance entre les parties est intéressante, aucune d'entre elles n'est vraiment responsable du cumul des résultats de [l'évaluation] et le résultat est sans fin, avec beaucoup d'erreurs, mais peu d'efforts ou de résultats cumulés. Paraphrasant Paul Sabatier et Edella Schlager (2000, p. 211), nous pouvons conclure cette analyse par le constat suivant : si nous apprenons sur le plan politique, qui s'attache à identifier les ressources cruciales pour accroître le pouvoir du courant dominant, nous restons sur notre faim par rapport à l'apprentissage au niveau du contenu des politiques, de leurs causes et des bénéfices et coûts des solutions alternatives. On perd, ainsi, la distinction nécessaire entre le « *policy core belief* » qui définit une politique publique par « ses orientations normatives de base, les instruments d'action privilégiés, la perception de la gravité du problème, la distribution de l'autorité entre l'État et la société » et les « *secondary aspects* » qui concernent « la mise en œuvre des choix centraux de la politique dans des domaines relativement étroits, comme un territoire spécifique » selon Sabatier et Jenkins Smith.

Lorsque Banerjee et Duflo (2011) s'expriment clairement, contre l'économie politique à force d'exemples, aux caractères davantage anecdotiques et déficients analytiquement, on s'écarte du *policy core belief* pour mettre plutôt l'accent sur un récit qui vise à être persuasif en multipliant les *secondary aspects* étroits et spécifiques. De leur point de vue, on aurait tort de « se focaliser [à ce point] sur les

grandes institutions, présentées comme la condition nécessaire et suffisante », il est difficile de trouver de « grandes solutions » à de grands problèmes, soutiennent-ils (p. 432). En recommandant de mieux comprendre les motivations des pauvres, des fonctionnaires, des contribuables ou encore celles des élus pour que les institutions soient moins corrompues et moins déresponsabilisées, ils introduisent, en réalité, un clivage fondamental et sans fondement entre les vérités analytiques et les vérités synthétiques, selon l'expression de Quine (2003, p. 47). En estimant l'argument convaincant, selon lequel il faut rendre aux pauvres les politiques de lutte contre la pauvreté, prôné par la doxa des organisations internationales et justifiée par « l'échec manifeste de l'État » à fournir des services publics aux pauvres, « ... mieux placés que quiconque pour savoir ce qu'ils veulent et ce qui se passe sur le terrain ... et en leur donnant un pouvoir de contrôle sur les prestataires de ces services » (p. 404), la fonction rhétorique et performative de leur discours tend à justifier la place éminente qu'ils revendiquent au sein des forums et des réseaux d'acteurs décrits ci-dessus. Comme le souligne Agathe Devaux-Spatarakis (p. 322), évoquant particulièrement le cas en France, territoire de ses investigations sur le fonctionnement des ECR, la pratique des expérimentations aléatoires est l'objet permanent « de renégociations au gré autant des conjonctures, que des intérêts, des pratiques et de l'interaction des acteurs qui s'en saisissent » ; et ce, au niveau des tentatives de son institutionnalisation aussi bien comme outil légitime que lors de sa mise en œuvre effective par les acteurs. L'objectif vise, en définitive, la monopolisation de l'expertise et l'action au sein du forum comme puissants outils d'élimination de la pluralité de médiation et de construction de politiques publiques alternatives.



## Conclusion générale

En s'interrogeant sur le statut des expérimentations aléatoires contrôlées, au moyen d'une critique globale inspirée de la thèse Duhem-Quine, la présente enquête s'est attachée à montrer que la production de preuves à l'appui de relations causales en mesure d'évaluer l'impact des programmes et de fonder des interventions rencontre des difficultés majeures. La validité interne, souvent mise à l'actif de l'approche du J-PAL s'avère plus problématique qu'il n'y paraît. L'écart entre les conditions de la randomisation et les hypothèses sous-jacentes aux outils statistiques, l'évacuation de l'hétérogénéité consécutive à la détermination de l'effet moyen et le caractère contingent des résultats sont autant de limites qui mettent en doute les preuves de causalité. La duplication des expériences, justifiée par le principe qu'une base empirique étoffée garantit des résultats de plus en plus assurés est, par construction, inapte à éliminer les contingences. Les données obtenues à l'occasion de chaque intervention, restent toujours tributaires de la définition des programmes, des relations entre les parties prenantes, du dispositif expérimental, des procédures et attitudes des uns et des autres. L'élargissement du répertoire des résultats même convergents n'est pas en mesure d'apporter des preuves, non seulement parce que ces résultats ne sont pas contraignants, mais aussi parce qu'ils ne peuvent être tenus pour décisifs. Ce défaut de cumulativité est dû à la réduction de la théorie à la portion congrue. L'analyse des données ne contient pas en propre de critères de jugement de la pertinence des conclusions et des implications opérationnelles. En l'absence de ces critères, l'approche expérimentale ne peut résoudre le choix des énoncés à admettre comme acquis et à consolider sa base empirique. Ces énoncés ne peuvent être jugés significatifs qu'à l'issue d'interprétations informées par des propositions théoriques. Il y a loin entre le primat accordé à l'observation et l'ambition d'établir des preuves en faveur des propriétés révélées par l'expérimentation.

Au vu de ces limites, on ne saurait considérer que l'approche du J-PAL doit sa vogue à sa validité. Elle s'explique plutôt par la qualification qu'elle a acquise au sein de l'économie du développement en tant que microéconomie expérimentale au

service de l'expertise et la prescription. Est plus emblématique, l'imbrication des évaluations expérimentales avec les inflexions qu'opèrent les institutions financières internationales (IFI), notamment la Banque mondiale, aussi bien dans le discours qu'au niveau des projets visant à prendre en compte la situation des pauvres dans les pays en développement (*A world free of poverty* marqué sur le fronton de la Banque Mondiale). Le *mainstreaming* de la pauvreté, érigé en doxa, se traduit par des programmes et des prêts consacrés à la lutte contre la pauvreté organisée au moyen de dispositifs évaluatifs dans les domaines de la santé, l'éducation et l'inclusion financière. La prospérité des évaluations randomisées résulte de la conjonction entre la demande portée par les bailleurs et l'offre proposée par une ingénierie animée par des experts et des acteurs privés. Cette insertion dans le forum des politiques publiques s'appuie sur un mode opératoire soutenu par une rhétorique et une performativité dont la fonction discursive rend les performances plus grandes que les intentions et les effets. Ce faisant le J-PAL s'est doté de structures de type entrepreneurial et singulier, fait assez rare dans le monde académique, exerçant un quasi-monopole sur l'expertise et l'action qui conduit à restreindre la pluralité de médiation et de construction de politiques publiques alternatives.

« Elle [ l'économie du développement ] – écrit Lautier- défend l'idée selon laquelle la recherche de la justice passe par des politiques permettant au marché de produire le maximum d'effets justes. La nouveauté, depuis la fin des années 1980 (et la mise en scène du débat sur la pauvreté), est qu'on écarte a priori l'idée selon laquelle ces politiques devraient agir sur des « structures ». Il ne s'agit plus que d'agir sur des individus, en leur permettant d'acquérir par exemple, des droits de propriété, de la formation, une bonne santé.

La « nouvelle économie du développement » parle sans cesse de développer des sociétés sans jamais agir sur ces sociétés » (Lautier, 2002, p. 142).

L'approche du J-PAL conforte à maints égards ce jugement. D'abord, elle participe de ce cantonnement de l'étude à des objets locaux qu'illustrent les réaménagements de la figure de l'homo-economicus, qui consiste à examiner les conduites

individuelles et les incitations appropriées à l'amélioration de l'efficacité. Sous cet angle, elle confère, à l'instar de l'économie comportementale dont elle fait partie, un statut hégémonique en prenant en compte les biais vis-à-vis de la rationalité standard mais conserve la représentation standard du marché concurrentiel comme si cette représentation correspondait aux économies réelles. Cette ambivalence de la démarche, qui écarte les structures qui encadrent les choix des agents, est précisément consécutive au primat de l'individu. Ensuite, elle consacre le rétrécissement de l'objet de l'économie du développement en réduisant sa visée au champ de la seule micro-économie des décisions et des réactions à des interventions. La récusation de la macroéconomie par le J-PAL au nom de la modestie et de la nécessité d'études de cas n'a-t-elle pas poussé Easterley à remarquer que les expérimentations aléatoires procèdent à un abaissement systématique des ambitions de l'économie du développement ?

Enfin la primauté de l'expérimentation se traduit par une dévaluation de la théorie comme en témoigne la fameuse analogie de l'économiste avec le plombier. Très souvent, le terme de théorie est utilisé au sujet d'énoncés isolés, d'hypothèses ou des prédictions quantitatives ou de simples a priori. Dans le même temps, le terme, de fait, est tantôt tenu pour une réalité indépendante de l'observation tantôt assimilé à un énoncé empirique. Cette incertitude atteste un retrait par rapport à la thèse de l'imprégnation des faits par la théorie qui est un acquis de l'épistémologie générale. Dans ce contexte, la conception du développement comme faisceau de transformations des structures économiques et sociales permettant une extension du bien-être de la population disparaît sous les termes d'allocation des ressources et l'efficience des marchés.

# BIBLIOGRAPHIE

## 1) Livres

Acemoglu, D., Robinson J.A. [2012], *Why Nations Fail. The Origins of Power, Prosperity, and Poverty*, Profile Books.

Agostino S. et Duvert, N. [2008], *La pauvreté*, Bréal.

Akrich, M., Callon, M., Latour, B., [2016], *La sociologie de la traduction, textes fondateurs*, Presse des Mines.

Assidon, E. [2002], *les théories économiques du développement*, 3<sup>e</sup> édition, La Découverte, (Collection Repères).

Atkinson, A. [2016], *Inégalités*, Éditions du Seuil.

Austin, J. L. [1962], *Quand dire, c'est faire*, Éditions du Seuil, (Collection points), réédition 1991.

Bachelard, G. [1934], *la formation de l'esprit scientifique*, Bibliothèque des textes philosophiques, Vrin.

Banerjee, A., et Duflo E. [2012], *Repenser la pauvreté*, Éditions du Seuil.

Banerjee, A., et Duflo E. [2020], *Une économie utile pour des temps difficiles*, Éditions du Seuil.

Bauer, P.T. [1972], *Dissent on development: Studies and Debate in Development Economics*, Cambridge, Harvard University Press.

Bauman. Z. [2013], *La vie liquide*, Fayard.

Bayart, J. F.[1989], *L'État en Afrique. La politique du ventre*, Paris, Librairie Fayard, coll. L'espace du politique.

Benmarhnia, T., Davis, P.MM, Godrie, B. [2019], *Les sociétés de l'expérimentation, Enjeux épistémologiques, éthiques et politiques*, Presses de l'université du Québec, (Collection Problèmes sociaux et interventions sociales).

- Berthoud, A. [2012], *Essais de philosophie économique. Platon, Aristote, Hobbes, A. Smith, Marx*, Presses universitaires du Septentrion-Économie.
- Biasutti, J.P., Braquet L. [2016], *Comprendre les inégalités économiques*, Bréal.
- Booth, W. [1974], *Modern Dogma and the Rhetoric of Assent*, University of Chicago Press.
- Boudon, R. [1990], *L'art de se persuader*, Fayard.
- Bourguignon, F. [2015], *Pauvreté et développement dans un monde globalisé*, Collège de France/Fayard.
- Bozio A., Grenet J. [2010], « *Économie des politiques publiques* », Éditions la Découverte, (Collection Repères).
- Brisset, N. [2015], « *Comment (et pourquoi) repenser la performativité des énoncés théoriques* », L'Harmattan, L'Homme & la Société, 2015/3 n° 197, pp. 31-63.
- Cartwright, N. [2007], *Hunting Causes and Using Them. Approaches in Philosophy and Economics*, Cambridge: Cambridge UP.
- Chelle, E. [2012], *Gouverner les pauvres, Politiques sociales et administration du mérite*, Presses universitaires de Rennes, « Res Publica ».
- Chung, R. [2019], in « *Les sociétés de l'expérimentation, Enjeux épistémologiques, éthiques et politiques* », Presses de l'université du Québec, (Collection Problèmes sociaux et interventions sociales).
- Cling J.P., Razafindrakoto, M., Roubaud, F. [2008], in : « *Les Journées de Tam Dao : 'Nouvelles approches méthodologiques appliquées au développement (2)'* », Publications du Laboratoire d'économie Dauphine, pp. 16-47.
- Cling, J.P., Razafindrakoto, M., Roubaud, F. [2002], *Les nouvelles stratégies internationales de lutte contre la pauvreté*, Edition Economica.
- D'Agostino, S, Duvert, N. [2008], *La pauvreté*, Bréal.

Dawid, P. [2007], *Fundamentals of Statistical Causality*, RSS/EPSRC Graduate Training Program University of Sheffield.

Dickens, O. [2015], *Théorie et expérience*, Ellipses.

Delmas, C. [2011], *Sociologie politique de l'expertise*, (Collection repères), la découverte.

de Maillard, J., Kubler, D. [2015], *Analyser les politiques publiques*, Éditeur : Presses Universitaires de Grenoble - P.U.G. Collection : Le politique en plus 2<sup>e</sup> édition.

Duflo, E. [2009], *Expérience, science et lutte contre la pauvreté*, Fayard, Paris.

Duflo, E. [2010a], *Lutter contre la pauvreté, le développement humain*, tome I. Éditions du Seuil.

Duflo, E. [2010b], *Lutter contre la pauvreté, la politique de l'autonomie*. Tome II. Éditions du Seuil (Collection république des idées).

Easterley, W. [2006], *Les Pays pauvres sont-ils condamnés à le rester ?* Éditions d'Organisation, Groupe Eyrolles.

Easterley W. [2009], *Le Fardeau de l'homme blanc : l'échec des politiques occidentales d'aide aux pays pauvres*, éd. Markus Haller. 512 pages.

Eber, N., Wellinger, M. [2012], *L'économie expérimentale*, La découverte, (Collection Repères).

Ferguson, N. [1997], *Virtual History: Alternatives and Counterfactuals*, Basic Books.

Frobert, L. [2004], *Si vous êtes si malins... McCloskey et la rhétorique des sciences économiques, suivi de la traduction de : la rhétorique des sciences économiques* (Dreidre N. McCloskey), ENS Éditions.

Genieys, W. [2011], *Sociologie des élites*, Éditions Armand Colin (Collection U).



Glennerster, R., Takavarasha, K. [2013], *Running randomized evaluations*, Princeton University Press.

Hacking, I. [1989], : *Concevoir et expérimenter : thèmes introductifs à la philosophie des sciences*, édition C. Bourgois.

Hausman, D. [1992], *The inexact and Separate Science of Economics*, Cambridge University Press.

Hibou, B., Tozy, M [2015], *L'État d'injustice au Maghreb, Maroc et Tunisie*, Éditions Karthala.

JOBERT, Bruno. [1994], *Le tournant néo-libéral en Europe, Idées et recettes dans les pratiques gouvernementales*, Paris, l'Harmattan, coll. « Logiques politiques », 1994, 328 p.

Jouary, J.P. [1996], *Enseigner la vérité ? Essai sur les sciences et leurs représentations*, Stock.

Kant, E., [1846], *Critique de la raison pure*, J. Tissot, Librairie philosophique de Ladrange, Seconde édition française, Tome second.

Kant, E. [2012], *Critique de la raison pure [1781-1787]*, traduit de l'allemand par Tremesaygues et Pacaud, Éditions PUF, Collection « Quadrige ».

Laville, J-L., Nyssens, M. [2001], *Les services sociaux entre associations, État et marché*, La découverte.

Legay, J.M. [1997], *L'expérience et le modèle*, INRA éditions.

Lévy-Leblond, J.M. [2020], *La vitesse de l'ombre. Aux limites de la science*, Éditions le Seuil.

Lewis, D. [2007], *De la pluralité des mondes*, traduit par Caveribère et Cometti, Éditions de l'Éclat.

Lewis, D. [1973], *Counterfactuals*, Blackwell Publishers, Harvard University Press.

- Locke, J. [1972], *Essais sur l'entendement humain*, Vrin.
- Miles, M. B., Huberman, M. A. [2003], *Analyse des données qualitatives*, (2<sup>e</sup> éd.), de Boeck.
- Mincer, J. [1974], *Schooling, Experience and Earnings*, NBER, 152 pages.
- Mongin, P. [2000], « La méthodologie économique au XX<sup>e</sup> siècle. Les controverses en théorie de l'entreprise et la théorie des préférences révélées », in *La nouvelle histoire de la pensée économique*, sous la direction de A. Béraud et G. Faccarello, Éditions la Découverte, tome 3, Ch. 36
- Mounier, A. [2016], *Critique de l'économie du développement*, Bréal.
- Muller, P. [2018], « *Les politiques publiques* ». Presses Universitaires de France (Collection Que sais-je ?).
- Nadeau, R. [1999], *Vocabulaire technique et analytique de l'épistémologie*, Presses Universitaires de France, 904 pages.
- Parienty, A. [2018], *Le mythe de « la théorie du ruissellement »*, La découverte.
- Pawson, R. [2006], *Evidence-Based Policy: A Realist Perspective*, Thousand Oaks, Ca, Sage Publications.
- Pestre, D. [2013], *A contre-Science. Politiques et savoirs des sociétés contemporaines*, Éditions du Seuil.
- Petit, M. [2011], *Pour une agriculture mondiale productive et durable*, éditions Quae.
- Raffinot, M. [2015], *Économie du développement*, Dunod.
- Rahnema, M. [2003], *Quand la misère chasse la pauvreté*, éditions Fayard, Actes-Sud.
- Renouvier, C., [1876], *Uchronie, l'Utopie dans l'histoire*, PRNG Éditions (2013), Collection Uchronie.

- Rist, G. [2013], *Le développement. Histoire d'une croyance occidentale*, les presses Sciences Po, coll. « Monde et société ».
- Robbins, J. [2013], *La question morale*, in Didier Fassin et al., Presse Universitaire de France, présentation et traduction de Samuel Lézé, pp. 183-192.
- Rocha, S. [2014], *Allocations sociales et pauvreté au Brésil. Le Bon Combat ?* Éditions de la maison des sciences de l'Homme.
- Sachs, J. [2005], *The End of Poverty*, New York, The Penguin Press.
- Schultz, T.W. [1964], *transforming traditional agriculture*, New Haven and London, Yale University Press.
- Schumpeter, J. [2004], *Histoire de l'analyse économique, Tome I, l'âge des fondateurs (des origines à 1790)*, Gallimard.
- Sergot, J. [2009], *Pourquoi des agriculteurs rationnels sont-ils inefficients ? Capital humain, efficience productive et stratégies paysannes en Haïti*, Éditions Publibook (Collection Sciences Humaines et Sociales).
- Servet, J. M. [2018], *L'économie comportementale en question*, Paris, Éditions Charles- Léopold-Mayer.
- Smith, H. L. [2013], *la Causalité en sociologie et démographie : retour sur le principe de l'action humaine*, Penn Libraries, University of Pennsylvania.
- Soler, L. [2009], *Introduction à l'épistémologie*. Éditions Ellipses.
- Temple, J.W. [2010], *Handbook of Economic Development*, edited by D. Rodrik et M. Rosenzweig, Elsevier, chapter 67, pp. 4417-4523.
- Vedung, E. [2009], *Public Policy and Program Evaluation*, Transaction Publishers, New Brunswick (USA) and London (UK).
- Walliser, B., Prou, C. [1988], *La science économique*, Éditions du Seuil.

## 2) Thèses

Favereau, J. [2014], « L'approche expérimentale du J-PAL en économie du développement : un tournant épistémologique », Université Paris 1 Panthéon Sorbonne, UFR économie.

Jatteau, A., [2016], « Faire preuve par le chiffre ? Le cas des expérimentations aléatoires en économie », Université Paris-Saclay, École Normale Supérieure Cachan.

Devaux Spatarakis, A. [2014], « La méthode expérimentale par assignation aléatoire : un instrument de recomposition de l'interaction entre sciences sociales et action publique en France ? ». Thèse de Doctorat. Université de Bordeaux.

Scheffer, P. [2017], « Quelle formation à l'indépendance est-elle possible pour les étudiants en médecine, par rapport à l'influence de l'industrie pharmaceutique ? », thèse de Sciences de l'Éducation. Université Paris 8.

## 3) Articles

Abdelghafour, N. [2017], "Randomized Controlled Experiments to End Poverty. A sociotechnical Analysis", *Mythes sociotechniques et développement*, pp. 46-47, pp. 352-262.

Acemoglu, D., Johnson S., A. Robinson J. A. [2001], "The Colonial Origins of Comparative Development: An Empirical Investigation", *The American Economic Review*, Vol. 91, No. 5 (Dec. 2001), pp. 1369-1401.

Acemoglu. D. [2010], "Theory, General Equilibrium and Political Economy in Development Economics". *Journal of Economic Perspectives*, 24 N°3, pp. 17-32.

Allamel-Raffin, C., Dupouy, S., Gangloff, J.L. [2019], « Introduction. Philosophes sur l'expérimentation scientifique : bilan et perspectives », *Philosophia Scientiæ*, Éditions Kimé, 2019/2 23-2 | pp. 5-18.

Allègre G. [2012], « L'expérimentation du RSA et ses ambiguïtés », Caisse nationale d'allocations familiales, « *Informations sociales* » 2012/6 n° 174, pp. 51-60.

Angrist J. [1990], "Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records", *American Economic Review* 80(3), pp. 313-335.

Angrist J., Imbens G., Rubin D. [1996], "Identification of Causal Effects Using Instrumental Variables", *Journal of American Statistical Association* 91(434), pp. 444-472.

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*, pp. 3-30.

Angrist, J., Lavy, V. [1999], "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2), pp. 533-75.

Atler, N. ([1992], « Albert O. Hirschman, Deux siècles de rhétorique réactionnaire », *Sociologie du travail*, 34<sup>e</sup> année, N°4, pp. 510-512.

Ayerbe, C., Missonier, A. [2007], « Validité interne et validité externe de l'étude de cas : principes et mise en œuvre pour un renforcement mutuel », *Finance Contrôle Stratégie*, Volume 10, n° 2, juin 2007, pp. 37-62.

Baele, S. [2013], "The ethics of new development economics: is the Experimental approach to development economics morally wrong?" *The Journal of Philosophical Economics*, VII, 1, pp. 1-42.

- Baird, S., Hicks, H., Kremer, M., Miguel, E. [2015], “Worms at Work: Long-run impacts of a child health investment”, CEGA Working Paper Series N° WPS- 052, Center for effective global action, University of California, Berkeley.
- Baker, W. [1984], “The social structure of national securities market”, *American Journal of Sociology*, Vol. 89(4), pp. 775-811.
- Banerjee, A. [2006], “Making Aid work”, *Boston review*, 31 (4), pp. 7-9.
- Banerjee A., Cole, S., Duflo, E., Linden, L. [2007], “Remedying education: Evidence from two Randomized Experiment in India”, *The Quartely Journal of Economics*, 2007, Vol. 122, issue 3, pp 1235-1264.
- Banerjee, A., Duflo, E. [2006], « L’approche expérimentale en économie du développement », *Revue d’économie politique*, 2009/5, Vol. 119, pp. 691-726.
- Banerjee A., Duflo E., Glennerster R., and Kinnan, C. [2015], “The Miracle of Microfinance? Evidence from a Randomized Evaluation.” *American Economic Journal: Applied Economics*, 2015, 7(1), pp. 22-53.
- 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*, éditions La Découverte, 2012/2, n° 10, pp. 175-198.
- Barrett, C. B., Carter, M. R. [2010], “The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections”, *Applied Economic Perspectives and Policy*, Volume 32, Issue 4, pp. 515-548.
- Bédécarrats, F., Guerin, I., Roubaud F. [1989], « L’étalon-or des évaluations randomisées : du discours de la méthode à l’économie politique », *Sociologies pratiques*, 2013/2, n° 27, pp. 107-122.
- Bédécarrats, F., I. Guérin, et Roubaud, F. [2017], “L’étalon-or des évaluations randomisées: économie politique des expérimentations aléatoires dans le domaine du développement”, *Papiers de Recherche AFD*, n° 2017-44, Mai.



Belhaj, F. [2006], « Le Secteur Financier marocain : Quelles perspectives ? », Nawafid, lettre d'information trimestrielle du Groupe de la Banque mondiale au Maghreb. Mai 2006- N2. Page 44.

Berr, E., & Combarous, F. [2004], « *L'impact du consensus de Washington sur les pays en développement : une évaluation empirique* », Post-Print hal-00245919, HAL.

Bertrand, M., Djankov, S., Hanna, R., Mullainathan, S. [2007] “Obtaining a Driver’s License in India: An Experimental Approach to Studying Corruption”. *Quarterly Journal of Economics* 122 (4), pp. 1639-76.

Bertrand, M., Mullainathan, S., Shafir, [2006], “Behavioral Economics and Marketing in Aid of Decision Making Among the Poor”, *Journal of Public Policy and Marketing*, Vol. 25(1), pp. 8-23.

Benetti. C., Cartelier. J. [1994], « l'intérêt limité de l'hypothèse de rationalité individuelle », *Cahiers d'économie politique*, N 24/25, pp. 19-35.

Berthelot, J.-M. [1998], « les nouveaux défis épistémologiques de la sociologie », *Sociologie et société*, vol 30, n° 1, pp. 23-38.

Bertrand, M., Djankov S., Hanna, R., Mullainathan S. [2007], “Obtaining a Driver’s License in India: An Experimental Approach to Studying Corruption”. *Quarterly Journal of Economics*, 122 (4), pp. 1639-76.

Biddle, J. E., Hamermesh D. S. [2017], “Theory and Measurement: Emergence, Consolidation, and Erosion of a Consensus”, *History of Political Economy* (2017) 49 (Supplement), pp. 34-57.

Black, N. [1996], “Why we need observational studies to evaluate the effectiveness of health care”, *British Medical Journal*, Vol. 312, pp.1215-1218.

Bono, P.H., Trannoy, A. [2012], « Causalité et contrefactualité dans l'évaluation économique des politiques publiques », *Labyrinthe, atelier interdisciplinaire* 39/2012. pp. 35-53.

Boussaguet, L. [2014], « Forums » in Laurie Boussaguet éd., *Dictionnaire des politiques publiques*. 4<sup>e</sup> édition précédée d'un nouvel avant-propos. Paris, Presses de Sciences Po, 772 pages. « Références ».

Brisset, N. [2011], : « Les limites de la performativité des énoncés en économie. Les apports de John Searle », *Æconomia*, 1-4 | 2011, pp. 557-588.

Burtless, G. [1995], “The Case for Randomized Field Trials in Economic and Policy Research”, *Journal of Economic Perspectives*, —Volume 9, Number 2, spring 1995, pp. 63-84.

Campbell, D.T. [1969], “Reforms as experiments”, *American Psychologist*, 24(4), pp. 409-429.

Campbell, D.T. [1991], “Methods for the Experimenting Society”, Duplicated draft of speech delivered to American Psychological Association, 1971, Washington D.C., *Evaluation Practice*, VOL 12, N°3, pp. 223-260.

Cartwright, N. [2009], “If No Capacities, then No Credible Worlds, But Can Models Reveal Capacities?”, *Erkenntnis*, 70, pp. 45-58.

Charle, C. [1992], « Albert O. Hirschman, *Deux siècles de rhétorique réactionnaire* », *Annales. Économies, sociétés, civilisations*. 47<sup>e</sup> année, n° 6, pp. 1195-1197.

Clemens, M.A. [2017], “The Meaning of Failed Replications: A Review and Proposal”, *Journal of Economic Surveys*, 31(1), pp. 326-42.

Cling, J.P., Razafindrakoto, M., Roubaud F. [2002] « La Banque mondiale et la lutte contre la pauvreté : tout changer pour que tout reste pareil ? », *Politique africaine*, Éditions Karthala, n° 87, pp. 164-174.

Cohen, J., Dupas, P. [2010], “Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment”, *The Quarterly Journal of Economics*, Volume 125, Issue 1, pp. 1-45.

Cogneau, D. [2003], « pauvreté, inégalités des conditions et inégalités des chances », in Les nouvelles stratégies internationales de lutte contre la pauvreté, Éditions Economica, deuxième édition, Chapitre II, pp. 57 à 82.

Corten, A. [1998], « Le discours de la pauvreté de la Banque mondiale », *Langages et Sociétés*, n° 85. Maison des sciences de l'Homme.

Cot, A. L., Ferey, S. [2016], « La construction de « faits » économiques d'un nouveau type : éléments pour une histoire de l'économie expérimentale », *L'Actualité économique*, 92 (1-2), pp. 11-47.

Crépon B., Devoto F., Duflo E., Parienté W. [2015], "Estimating the Impact of Microcredit on Those Who Take It Up: Evidence from a Randomized Experiment in Morocco", *American Economic Journal: Applied Economics* 2015, 7(1), pp. 123-150.

David, P. [1969], « Transport Innovation and Economic Growth: Professor Fogel on and off the Rails », *The Economic History Review*, Vol. 22, N° 3, (Dec 1969), pp. 506-525.

Dawid, P. [2007], *Fundamentals of Statistical Causality*, RSS/EPSRC Graduate Training Program University of Sheffield.

Deaton, A. [2009], "Instruments of development: Randomization in the tropics, and the search for the elusive keys to economic development", NBER Working Paper No. 14690 January 2009, Revised March 2009.

Deaton, A. [2010], "Instruments, Randomization, and Learning about Development", *Journal of Economic Literature*, 48, pp. 424-455.

Deaton, A., Cartwright, N. [2018], "Understanding and Misunderstanding Randomized Controlled Trials", *Social Science & Medicine*, Elsevier. Volume 210, August 2018, pp. 2-21.

Deaton, A. [2019], “Randomization in the tropics revisited: a theme and eleven variations” Written for Florent Bédécarrats, Isabelle Guérin and François Roubaud, *Randomized controlled trials in the field of development: a critical perspective*, Oxford University Press.

Degeorges, O., Monnier, É., Spenlehauer, V. [1990], « L'évaluation comme outil de légitimation : le cas des grandes organisations internationales ». *Politiques et management public*, vol. 8, n° 4, 1990, pp. 1-23.

De Vroey, M., Malgrange, P. [2012], « Klein et l'émergence de la modélisation macroéconomique », in *Économie et Statistique*, pp. 451-453, La modélisation macroéconomique : continuités, tensions, pp. 21-30.

Dingwall, R. [2006], « Confronting the anti-democrats: the unethical nature of ethical regulation in social science », *Medical sociology*, online 1, pp. 51-58.

Dingwall, R. [2008], « The ethical case against ethical regulation in humanities and social research », *Journal of the academy of social sciences*, vol. 3, 2008, issue 1, pp. 1-12.

Drago, G. [2011], « Droits de propriété et liberté d'entreprendre dans la jurisprudence du conseil constitutionnel : une relecture », *CRDF*, n° 9, pp. 31-39.

Drouet, I., Dupouy, S., Jeanpierre, L., Nicodème, F. [2012], « Contrefactuels en histoire : du mot au mode d'emploi. Le moment de la new economic history », *Labyrinthe* 2012/2 (n° 39), pp. 81-112.

Dubois, J.M., Podevin, G. [2011], « Entre indépendance et connivence, la délicate posture de l'évaluateur embarqué », bulletin n° 294-1, emploi-formation du Cereq.

Duflo, E. [2004], « Scaling Up and Evaluation », In Bourguignon et Pleskovic, *Accelerating Development*, World Bank and Oxford University Press, pp. 342-367.

- Duflo, E. [2005], « Évaluer l'impact des programmes d'aide au développement : le rôle des évaluations par assignation aléatoire », *Revue d'économie du développement*, De Boeck Supérieur, 2005/2, vol. 13, pp. 185-226.
- Duflo, E. [2006], "Poor but Rational", In *Understanding Poverty* by Banerjee, A. Bénabou, R., Mookherjee, D., Oxford University Press, 443 pages.
- Duflo, E. [2007], "The economist as plumber», *American Economic Review*, vol. 107(5), pp. 1-26.
- Duflo, E. Hanna, R., Ryan, S. P. [2012], "Incentives Work: Getting Teachers to Come to School", *American Economic Review*, American Economic Association, pp. 1241-1278.
- Dumoulin, L. [2000], « L'expertise judiciaire dans la construction du jugement : de la ressource à la contrainte », *Droit et société*, n°44-45, 2000. *Justice et Politique* (III). *Les magistratures sociales*, pp. 199-223.
- Dupouy, S. [2011], « L'expérimentation », in : *Philosophie des sciences humaines*, Vrin, pp. 213-241.
- Dupuoy, S. [2012], « Philosophie de l'expérimentation biomédicale », *Revue philosophique*, n° 2010/2, pp. 249-256
- Eble, A., Boone, P., Elbourne, D. [2016], "On minimizing the risk of bias in Randomized Controlled Trials in Economics", *World Bank Policy Research*, Paper N. 7746.
- Ehrhart, C. [2006], « Croissance, redistribution et lutte contre la pauvreté : l'évolution non linéaire de l'approche de la Banque mondiale », *Revue d'Analyse économique*, Vol. 82, n° 4, pp. 597-641.
- Flanagin, A., Carey, L.A., Fontanarosa, P.B., Philips, S.G., Pace, B., Lindberg, G., Rennie, D. [1998], "Prevalence of Articles with Honorary Authors and Ghost Authors in Peer-Reviewed Medical Journals", *Journal of American Medical Association*, Vol. 280, N° 3, pp. 222-224.

Faulkner W.N., [2014], “A critical analysis of a randomized controlled trial evaluation in Mexico: Norm, mistake or exemplar?” *Evaluation*, 20(2) : pp. 230-243.

Favereau, J. [2016], Expérimentations, expériences randomisées en économie du développement et conventions, 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, pp. 121-124.

Ferrière, M. [2011], « Les principales critiques de la théorie du choix rationnel », *Idées économiques et sociales*, 2011/3, N° 165, pp. 37-45.

Fougère, D. [2010], « les méthodes économiques d'évaluation », *Revue française des affaires sociales*, 2010/1, pp. 105-128.

Fougère, D. [2012], « Les méthodes d'expérimentation en question », *Éducation et formation*, n° 81, pp. 41-47.

Franco, A. Malhotra, N. Simonovits, G. [2014] “Publication bias in the social sciences: unlocking the file drawer”, *Sciences*, Vol. 345, Issue 6203, pp. 1502-1505.

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.

Genieys, W. [2006], « Nouveaux regards sur les élites du politique », *Revue française de science politique*, 2006/1, vol. 56, pp. 121-147.

Glennerster, R., Takavarasha, K. [2013], *Running randomized evaluations*, Princeton University Press, 480 pages.

Gosselin, A. [1998], « la rhétorique économique dans les débats politiques » in *Communication, Information, Médias, Théories*, 1998/18-2, pp. 88-115.

Granjou, C. [2013], : « L'expertise scientifique à destination politique », Presses Universitaires de France, cahiers internationaux de sociologie, pp. 175-183.



Granovetter, M. [2006], « L'influence de la structure sociale sur les activités économiques », *Presses de Sciences Po, Sociologies pratiques*, 2006/2 n°13, pp. 9 à 36.

Guala, F. [2012], "Experimentation in Economics", in: *Philosophy of Economics*, Handbook of the philosophy of science, Edited by Uskali Maki, pp. 597 à 640.

Guala, F. [2018], "Experimental Economics, History of". In: Macmillan Publishers Ltd (Eds), the New Palgrave Dictionary of Economics. Palgrave Macmillan, London.

Guérin, I., Fouillet, C., Roesch, M. [2008], « La microfinance indienne : l'heure du bilan », *Transcontinentales : Sociétés, idéologies, système mondial* 6 | 2008, Mémoires et Nations.

Guillin, V. [2013], « De quoi les essais contrôlés randomisés sont-ils capables ? Évaluation, mécanismes et capacités en sciences sociales », *Cahiers philosophiques*, 2013/2, n° 133, pp. 79-102.

Hall, N. [2007], "R.A. Fisher and his advocacy of randomization", *Journal of History of biology* (2007) 40, pp. 295-325.

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., Ichimura, H., and Todd, P. E. [1997], "Matching as an econometric evaluation estimator; Evidence from evaluating a job-training program", *The review of economic studies*, 64(4), pp. 605-654.

Heckman, J. [2005], "The scientific Model of Causality", *Sociological Methodology*, 35 (1), pp.1-97.

Hibou, B. [1998], « Les méfaits du catéchisme économique : l'exemple de l'Afrique sub-saharienne », *Éditions Esprit*, pp. 98-140.

Holland, P.W., Rubin, D.B. [1983], “Causal inference in Retrospective studies”, Reprint From: *Evaluation Review*, edited by Howard Wainer and Samuel Messik, Routledge, pp. 3-25.

Holland, P.W. [1986], « statistics and causal inference », *Journal of The American Statistical Association*, Vol 81, N 396, pp. 945-960.

Holland, P.W. [1988], “Causal Mechanism or Causal Effect: Which Is Best for Statistical Science?”, *Statistical Science*, Vol. 3, N°2, pp. 186-188.

Humphreys, M., Weinstein, J. [2009], “Field experiments and the political economy of development”, *American Review of Political Science*, 7, pp. 367 à 378.

Imbens, G., Wooldridge, J. [2009], “Recent developments in the econometrics of program evaluation”. *Journal of Economic Literature* 47, N 1, pp. 5-86.

Ioannidis, J.P.A. [2007], “Why Most Published Research Findings Are False: Author’s Reply to Goodman and Greenland”, *PloS Med* 4(6): e215.

Jaillon, P. [2007], « L’essai clinique contrôlé randomisé ». *Bulletin de l’Académie Nationale de Médecine*, tome 191, 2007 avril-mai, N° 4 et 5, pp. 739-758.

Jatteau, A. [2018], « Comment expliquer le succès de la méthode des expérimentations aléatoires ? Une sociographie du J-PAL », in « Les professionnels de l’évaluation. Mise en visibilité d’un groupe professionnel ». *Sociologies* [En ligne], <http://journals.openedition.org/sociologies/6700>.

Jobert, B. [1992], « Représentations sociales, controverses et débats dans la conduite des politiques publiques », *Revue française de science politique*, 42<sup>e</sup> année, n° 2, pp. 219-234.

Karlan, D., Zinman, J. [2011], “Microcredit in theory and practice: using randomized credit scoring for impact evaluation”. *Science*, 2011, 332(6035), pp. 1278-1284.

- Kluve, J. [2004], “On the Role of Counterfactuals in Inferring Causal Effects”, *Foundations of Science*, 2004, Volume 9, n° 1, pp. 65-101.
- Kaufman, D. [2009], “Aid Effectiveness and Governance: The Good, The Bad and the Ugly”, in *Development Outreach*, Vol. 11, Issue 1 (Feb 2009), pp. 26-29.
- Klein, U. [2012], “Artisanal-Scientific Experts in Eighteenth-Century France and Germany”, *Annals of Science*, Vol.69, N°3, pp. 303-306.
- Lagneau-Ymonet, P. [2013], « Abhijit V. Banerjee, Esther Duflo, Repenser la pauvreté », *Travail et emploi*, 135/2013, pp. 73-77.
- Lalonde, R. [1986], “Evaluating the Econometric Evaluations of Training Programs with Experimental Data”, *The American Economic Review*, Vol. 76, No. 4, pp. 604-620.
- Lamy, J. [2015], « Des sciences pour et par le gouvernement. Sur le régime réglementaire des sciences contemporaines, *Sociologie et Sociétés*, Vol. 47, N°2, pp. 287-309.
- Leduc, G. [1963], « L’aide internationale au développement ». *Tiers-Monde*, tome 4, n°13-14, 1963, pp. 237-260.
- Laroche, P. [2007], « L’exploration statistique du biais de publication », *Journal de la société française de statistique*, tome 148, N 4, 2007, pp. 29-56.
- Lautier, B., Salama, P. [1995], « De l’histoire de la pauvreté en Europe à la pauvreté dans le Tiers Monde », *Tiers Monde*. 1995, tome 36 n°142. Pauvretés. pp. 245-255.
- Lautier, B. [1999], « L’étude de la pauvreté est-elle un substitut à l’étude du développement ? », *Revue du Tiers Monde*, tome 40, n° 160, 1999, pp. 789-791.
- Lautier, B. [2002], « Pourquoi faut-il aider les pauvres ? Une étude critique du discours de la Banque mondiale sur la pauvreté ». *Revue Tiers Monde*, n° 169, pp. 137-165.

Lecoutre, B. [2004], « Expérimentation, inférence statistique et analyse causale », *Intellectica, Revue de l'association pour la recherche cognitive*, N° 38/1, Raisonement causal, pp. 193-245.

Lind, J [1753], A treatise of the scurvy. In three parts. Printed by Sands, Murray and Cochran for a Kincaid and A. Donaldson. In James Lind Library.

L'Horty, Y., Petit, P. [2011], « Évaluations aléatoires et expérimentations sociales », *Revue française d'économie*, 2011/1 (Volume XXVI), pp. 13-48.

Lucas, R. [1988], "On the mechanics of economic development" *Journal of Monetary Economics*, Vol. 22, Issue 1, pp. 3-42.

Mardellat, P. [2006], « par-delà la notion de rationalité, l'économie comme sciences de l'esprit », *Cahiers d'économie politique* 2006/1 N 50, L'Harmattan, pp. 27-58.

Mendola, M. [2007], "Farm Household Production Theories", A Review of "Institutional" and "Behavioral" Responses. *Asian Development Review*, vol. 24, no. 1, pp. 49-68.

Meyran, R. [2017], « John Langshaw Austin. Quand dire, c'est faire », *Les Grands Dossiers des Sciences Humaines*, vol.46, n°3, pp.13-13.

Morgan, M. S. [2005], "Experiments versus models: New phenomena, influence and surprise", *Journal of Economic Methodology* 12:2, pp. 317-329.

Meyran, R. [2017], « John Lang Langshaw Austin- Quand dire, c'est faire », *Les grands dossiers des sciences humaines*, 2017/13, N 46, p. 13.

Mullainathan, S., Bertrand, M., Shafir, E. [2006], "Behavioral economics and marketing in aid of decision making among the poor". *American Marketing Association*, vol. 25 (1), pp. 8-23.

Nadeau, R. [1995], « L'économie comme littérature », in CNRS. Éditions « Hermès, La Revue », 1995/2 n° 16, pp. 85-111.

Normand, R. [2017], « La circulation internationale des techniques de la preuve dans les programmes d'intervention et de gestion publique en éducation », *Revue française d'administration publique*, 2017/n° 161, pp. 19-30.

Nubukpo, K. [2012], « Une approche simpliste de la pauvreté », in *l'Économie politique*, pp. 105-112.

Ozier, O. [2018], “Exploiting Externalities to Estimate the Long-Term Effects of Early Childhood Deworming”, *American Economic Journal: Applied Economics*, Vol 10, N3, pp. 235-62.

Pamis-Sumner, S. [2014], « Les évaluations d'impact dans le domaine du développement. État des lieux et nouveaux enjeux », *A savoir* 27 (Juin 2014), Département de la recherche, AFD.

Paulré, 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.

Paugam, S. [1998], « Les formes contemporaines de la pauvreté et de l'exclusion. Le point de vue sociologique », *Genèses* 31. *Femme, famille, individu*, pp. 138- 159.

Pawson, R., Tilly, N. [1997], “An Introduction to Scientific Realist Evaluation”, in *Chapter 29 of Evaluation for the 21<sup>st</sup> Century*, Edited by Eleanor Chelimsky and William R. Shadish.

Perroux, F. [1966], « Les blocages de la croissance et du développement : la croissance, le développement, les progrès, le progrès (définitions) », in *Revue du Tiers-Monde*, tome 7, n° 26, pp. 239-250.

Petit, M. [2011], « Pour une agriculture mondiale productive et durable », Éditions Quae, Collection Essais, 112 pages.

Plott, C. R. [1991], “Will Economics Become an Experimental Science?”  
*Southern Economic Journal*, Vol. 57, No. 4 (April 1991), pp. 901-919.

Plott, C. R. [1991], “Will Economics Become an Experimental Science?”  
*Southern Economic Journal*, Vol. 57, No. 4 (April 1991), pp. 901- 919.

Posner, R. A. [1999], “The Law and Economics of the Economic Expert Witness”,  
*Journal of Economic Perspectives*, Volume 13, Number 2, spring 1999, pp. 91-99.

Pollak, M. [1979], « Paul F. Lazarsfeld, fondateur d’une multinationale scientifique », in : *Actes de la recherche en sciences sociales*. Vol. 25, janvier 1979, pp. 45-59.

Pritchett, L. [2002], It pays to be ignorant: A simple rigorous evaluation, *The Journal of Policy Reform*, Rouledge 5:4, pp.251-269.

Quentin, A., Guérin, 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.

Quine, W. V. O. [1951], “Two Dogmas of Empiricism”, *The Philosophical Review* 60 (1951), pp. 20-43.

Ravallion, M. [2018], “Should the Randomistas (Continue to) Rule?” *CGD Working Paper 492*. Washington, DC : Center for Global Development.

Reiss, J., [2012], « Contrefactuels, expériences de pensée, et conception singulariste de la relation causale en histoire », *Labyrinthe* 39 : 2012, pp. 113-127.

Robert, C. [2008], « Expertise et action publique » in *Politiques publiques 1, la France dans les politiques européennes*, Presses Sciences Po, Olivier Borraz et Virginie Guiraudon, pp. 309-335.



Rochot, B. [1966], «Alexandre Koyré, Études d'histoire de la pensée scientifique», *Revue d'histoire des sciences et de leurs applications*, tome 19, n°4. pp. 390-391.

Rodrik, D. [2008], “the New Development Economics: We shall Experiment, but How Shall we learn”, John Kennedy School of Government, Harvard University, RWP08-055.

Romer, P. [1986], “Increasing Returns and Long-Run Growth”, *Journal of Political Economy*, Vol. 94, N° 5, pp. 1002-1037.

Roubaud, E. [1918], « L'assainissement du canal de Panama », *Annales de Géographie*, 1918, t. 27, N° 145, pp. 66-69.

Roux, S. [2015], « Les approches structurelles et non structurelles en micro-économétrie de l'évaluation des politiques publiques », *Revue française de l'économie*, Volume XXX, pp. 13-65.

Rubin, D.B. [1974], “Estimating Causal effects of treatments in Randomized and Nonrandomized Studies”, *Journal of Educational Psychology*, Vol. 66, n°5, pp. 688-701.

Rubin, D.B. [1977], “Assignment to treatment group on the basis of a covariate”, *Journal of Educational Statistics*, Vol.2, N°1, pp1-26.

Sabatier, P., Schlager, E. [2000], « Les approches cognitives des politiques publiques : perspectives américaines », *Revue française de science politique*, 50<sup>e</sup> année, n°2, pp. 209-234.

Sackett, D.L., Rosenberg, W.M.C., Gray, J.A.M., Haynes, R.B., Richardson, W.S. [1996], “Evidence based medicine: what it is and what it isn't”, *British Medical Journal*, 312 (1996), pp. 71-72.

Saretsky, G. [1975], “The John Henry Effect: Potential Confounder of Experimental vs. Control Group Approaches to the Evaluation of Educational Innovations”; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975), 32 pages.

Scheffer, P. [2016], « Préparer les étudiants en médecine à affronter l’influence de l’industrie pharmaceutique : un enjeu éthique, professionnel et éducationnel », *Pédagogie Médicale* 2016 ; 17(3), EDP Sciences / Société Internationale Francophone d’Éducation Médicale, pp. 189-197.

Schultz, T.W. [1966], “transforming traditional agriculture”, *Journal of Farm Economics*, Vol. 48, N°4, Part, Published by Oxford University Press, pp.1015-1018.

Schultz, T.W. [1970], “The reckoning of education as human capital”, in “*Education, Income and Human Capital*”, Lee Hanson, editor, pp. 297-306.

Serra, D. [2012], « Principes méthodologiques et pratiques de l’économie expérimentale : une vue d’ensemble », *Revue de philosophie économique*, 2012/1, vol. 13, pp. 21-78.

Simon, H. A. [1972], “Theories of bounded rationality”, *C.B. McGuire and Roy Radner, Decision and organization*. North-Holland Publishing Company. Chapitre 8, pp. 161-176.

Soler, L. [2011], « Les expérimentateurs sont-ils substituables les uns aux autres ? Opacité des pratiques expérimentales, fragmentation de l’expertise, désaccords irrésolus entre praticiens : quelles conséquences épistémologiques ? *Vrin « Le Philosophoire »*, 2011/1 n° 35, pp. 65-113.

Solow, R. [2001], « L’économie entre empirisme et mathématisation », *Critique économique*, N 4, pp. 150-152.

Soual, M., [2020], « L’économie du développement au prisme de la crédibilité empirique ». *Critique de l’économie*, n° 40. pp. 15-31.

Spencehauer, V., [2016], « La (f)utilité gouvernementale de l'évaluation des politiques publiques, quelques leçons américaines et françaises ». *Sciences Po, Laboratoire interdisciplinaire d'évaluation des politiques publiques*, N° 49, pp. 1–21.

Thaler, R. [1994], “Psychology and Saving Policies”, *American Economic Review*, pp. 186-192.

Taouil, R. [2018], « La macroéconomie est-elle une discipline empirique ? », *Revue d'histoire de la pensée économique*, n° 6, pp. 117-139.

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.

Thomas, Y. [2005], « Les artifices de la vérité en droit commun médiéval », *Revue française d'Anthropologie*, Éditions de l'EHEES, 2005/ 3-4, N° 175-176, pp. 113-130.

Vachet, A. [1975], « Review of [Koopmans, T.C., Trois essais sur la science économique contemporaine (Collection du Centre d'économétrie de la Faculté de droit et des sciences économiques de Paris), Éditions Dunod, Paris, 1970, 234 p] ». *Études internationales*, 6 (2), pp. 269- 271.

Vassy, C., Keller, R. [2008], « Faut-il contrôler les aspects éthiques de la recherche en sciences sociales, et comment ? », *La Découverte / « Mouvements*, 2008/3 n° 55-56, pp. 128-141.

Vedung, E. [2004], “Evaluation Research and Fundamental Research”, in Stockmann, Reinhard Edition, pp. 111-134.

Vining, R., Koopmans, T. C. [1949], “Methodological Issues in Quantitative Economics”, *Review of Economics and Statistics*, 31, pp. 77-94.

White, H. [2009], “Theory-based impact evaluation: principles and practice”, *Journal of Development Effectiveness*, 1/3, 271-284, DOI: 10.1080/19439340903114628. 15 pages.

Worrall, J. [2002], “What Evidence in Evidence-Based medicine”, *Philosophy of Science*, Vol. 69, N° S”, pp. S316-S330.

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.

#### **4) Conférences, colloques et congrès, symposium, rapports, articles internet**

##### **Symposium**

Banerjee, A, Bardhan, P., Basu, K, Mookherjee, D. [2005], “New Directions in Development Economics: Theory or Empirics?”, A Symposium in Economics a Political Weekly, pp. 1-40.

##### **Colloque**

Berthoud, A., Delmas, B., Demals, T. [2007], *Y a-t-il des lois en économie ?* 11<sup>e</sup> colloque international de l'Association Charles Gide pour l'Étude de la Pensée Économique (ACGEPE), Villeneuve d'Ascq, Presses Universitaire du Septentrion, 2007, 644 pages.

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 .

Duflo, E., Kremer, M. [2004], “Use of randomization in the Evaluation of development Effectiveness”, document préparé pour la conférence organisée par le département des évaluations de la Banque mondiale portant sur l'évaluation et l'effectivité du développement du 15-16 Juillet 2003.

Paulré, P. [2003], « La causalité en économie : mises en perspective et approches opérationnelles. École thématique CNRA de Philosophie Économique, Science économique et Philosophie des sciences. Aix en Provence, 9 septembre 2003.

### **Séminaire et Congrès**

Le Guennec, M.A., Mignemi, N., Plutniak S. [2017], « What if? Uchronie et raisonnement contrefactuel », in Séminaire des membres de l'École française de Rome, <https://semefr.hypotheses.org/2260>, [hdl:10670/1.gob0b6](https://nbn-resolving.org/urn:nbn:fr:semefr-2017-06-00001).

Massieu, V. [2011], « L'expérimentation comme exemple de transposition d'une méthode scientifique hors de son cadre d'origine », intervention 8<sup>e</sup> congrès de droit constitutionnel, 2011, Nancy.

<http://www.droitconstitutionnel.Org/congresnancy/comN3/massieurT3.pdf>.

Williamson, J. [2004], “From de Washington Consensus towards a new Global. Governance”, Conference CIDOB, Barcelona, September 2004, 14 pages.

### **Conférences**

Thom, R. [1988], « Un rapport problématique : la rencontre théorie-expérience », in « The role of Experience in Science», Proceedings of 1986 conference of the Académie Internationale de Philosophie des Sciences (Bruxelles). Éd. Walter de Gruyter & Co.

## Rapports

Banque mondiale [1990], *World Development Report 1990: Poverty*, New York: Oxford University,

<https://openknowledge.worldbank.org/handle/10986/5973>.

Banque mondiale [2003], *Rapport sur le développement dans le monde 2004 : Mettre les services de base à la portée des pauvres*, Banque mondiale, Washington

D.C. Co publication de la Banque mondiale et Éditions ESKA.

Banque mondiale [2006], « Rapport sur le développement dans le monde : Équité et développement », Oxford University Press.

Fiszbein, A. Schady, N., Ferreira, F., Grosh, M., Keleher, N., Olinto, P., Skoufias,

E. [2009], *Conditional Cash Transfers: Reducing Present and Future Poverty*. World Bank Policy Research Report. Washington, DC: World Bank.

## Articles internet

Blackhouse, R., Cherrier, B. [2014], *Becoming Applied: The transformation of economics after 1970*, “unpublished paper”, Birmingham University, <http://www.nber.org/papers/w22253>.

De Curraize, Y., Thoron, S. [2018], *A propos du statut épistémologique des expériences en économie*. Hal-01966636,

<https://hal.archives-ouvertes.fr/hal-01966636>, pp. 1-34.

David, J. [2015], « Une logique que rien n’arrête... Uchronie de Renouvier », URL : <http://journals.openedition.org/elh/585>.

DOI : <https://doi.org/10.4000/elh.585>.

Hibou, B. [2014], « Économie politique du discours de la Banque mondiale en Afrique sub-saharienne », <https://hal-sciencespo.archives-ouvertes.fr/hal-01010654>.



Juhel, J. [2015], « modèles structuraux et inférence causale », Université Rennes 2, pages 1-45.

**<https://www.researchgate.net/publication/324602998>**.

Lamy, J. [2019], « Gouverner par l'expertise scientifique et technique-Note critique », *Cahiers droits, sciences et technologies*, 8/2019, pp. 149-165.

**<https://doi.org/10.4000/cdst.789>**.

The Lancet [28 août 2004], "The World Bank is finally embracing science",

**[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(04\)16945-6/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(04)16945-6/fulltext)**.

Mawell, S. [2005], "the Washington Consensus is Dead! Long Live the Meta-narrative!", Working Paper N° 243, Overseas Development Institute, 14 pages:

**<https://cdn.odi.org/media/documents/2476.pdf>**.

Soltani, P., Patini, R. [2020], "Retracted COVID-19 articles: a side-effect of the hot race to publication". *Scientometrics* 125, 819–822 (2020).

**<https://doi.org/10.1007/s11192-020-03661-9>**.

Tröhler, U. [2003], "James Lind and scurvy: 1747 to 1795". *James Lind Library Bulletin*, Commentaries on the History of Treatment Evaluation,

**<https://www.jameslindlibrary.org/articles/james-lind-and-scurvy-1747-to-1795>**.

Frenkiel, E. [2009], « Nudge ou le paternalisme bienveillant », la vie des idées.