

**UNIVERSITÉ PARIS 1 PANTHÉON-SORBONNE
U.F.R. ÉCONOMIE**

THÈSE

Pour obtenir le grade de
DOCTEUR ÈS-SCIENCE ÉCONOMIQUE

Présentée et soutenue publiquement par :
Judith FAVEREAU
Le 14 février 2014

Sous le titre :

**L'APPROCHE EXPÉRIMENTALE DU J-PAL EN ÉCONOMIE DU
DÉVELOPPEMENT : UN TOURNANT ÉPISTEMOLOGIQUE ?**

Membres du Jury :

Monsieur Jean-Louis Arcand, professeur au Graduate Institute, Genève (rapporteur)
Madame Annie L. Cot, professeur à l'Université Paris 1 Panthéon-Sorbonne (directrice)
Madame Agnès Gramain, professeur à l'Université Paris 1 Panthéon-Sorbonne (présidente)
Monsieur Emmanuel Picavet, professeur à l'Université Paris 1 Panthéon-Sorbonne
Monsieur Daniel Serra, professeur à l'université Montpellier 1 (rapporteur)
Monsieur David Teira, professeur à l'UNED, Madrid

L'université Paris 1 Panthéon-Sorbonne n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses, ces opinions doivent être considérées comme propres à leurs auteurs.

A Claude et Pierre,
et à Michel

Introduction Générale

« Il vient une heure où protester ne suffit plus ;
après la philosophie il faut l'action ; la vive force
achève ce que l'idée a ébauché. » (Victor Hugo,
Les Misérables, 1862, 532)

« Il faut penser la pauvreté autrement »¹. Cette urgente nécessité a été affirmée par Esther Duflo, il y a maintenant deux ans dans *L'Express*, et c'est autour de cette nécessité que se sont orientées l'ensemble de ses recherches. Beaucoup de politiques de développement auraient échoué, selon Esther Duflo et Abhijit Banerjee, en raison de ce qu'ils nomment les trois « i » (Banerjee et Duflo, 2011, 38) : ignorance, idéologie et inertie. Le premier « i », l'ignorance renvoie au fait que les agences internationales, en ne se déplaçant pas sur le terrain, ignorent la nature des besoins des populations touchées par la pauvreté. L'idéologie, ensuite, traduit l'idée que les politiques de développement sont pensées au travers d'une idéologie soit pro-marché soit anti-marché. L'inertie, enfin, réside dans le fait que les agences internationales continuent à instaurer les mêmes politiques. Ces trois « i » résultent d'une appréhension toujours trop générale et abstraite de la pauvreté. A l'inverse, Banerjee et Duflo entendent montrer que la pauvreté² n'est pas abstraite, mais qu'elle se constitue d'« une série de problèmes concrets qui, une fois correctement identifiés et compris, peuvent être résolus. » (Banerjee et Duflo, 2011, 19) Afin d'identifier cette série de problèmes et de contrecarrer les trois « i » Banerjee et Duflo préconisent de mener des expériences de terrain évaluant les programmes de développement. En ce sens, le « terrain » occupe une place centrale dans leur approche, c'est à travers lui que l'action est placée au cœur de la lutte contre la pauvreté. L'objectif de cette thèse sera d'examiner l'approche de Banerjee et de Duflo en économie du développement. Afin de définir et de délimiter cet objectif, je la présenterai, tout d'abord, en détail. Dans un deuxième temps, je définirai précisément l'objectif de cette thèse, puis la problématique qui y préside. Dans un troisième temps, je spécifierai la méthode que j'utilise

¹ Propos d'Esther Duflo recueillis par Valérie Lion, *L'Express*, numéro 3106, semaine du 12 au 18 janvier 2011.

² Tout au long de ce travail lorsque je parlerai de pauvreté je me référerai à la définition d'Angus Deaton et d'Olivier Dupriez (2011). Ces derniers fixent le seuil de pauvreté à 99 centimes de dollars en parité de pouvoir d'achat. C'est ce seuil qui est retenu par Banerjee et Duflo. La banque mondiale dans son rapport en 1990 définissait le seuil de pauvreté à un dollar par jour (voir, Banque mondiale, 1990). Depuis 2009, à la suite des travaux de Martin Ravallion, Chen Shaoua et Prem Sangraula (2009), la Banque mondiale définit le seuil de pauvreté à un dollar et vingt cinq centimes en parité de pouvoir d'achat.

ainsi que la perspective que j'adopte afin de répondre à la problématique posée. Je développerai, enfin, la structure de la thèse.

1. Rendre la lutte contre la pauvreté efficace

Pour Banerjee et Duflo³, la lutte contre la pauvreté, prise dans sa généralité, se heurte fréquemment à un important pessimisme, la tâche apparaissant presque insoluble. Les acteurs politiques, les citoyens, les chercheurs se sentent alors impuissants et submergés par l'ampleur du problème. C'est pourquoi, selon Banerjee et Duflo, il importe de fournir des preuves d'efficacité de programmes de développement locaux et ciblés, l'objectif étant de produire des changements à la marge, qui ajoutés les uns aux autres permettront de lutter avec efficacité contre la pauvreté.

Afin d'évaluer l'efficacité des programmes de développement et d'identifier des problèmes précis, Banerjee et Duflo importent de la médecine la méthode des essais cliniques. Ces derniers consistent en des expériences randomisées⁴. Cette méthode expérimentale permet de comparer deux situations, une première où le programme est mis en place et une seconde dont il est absent. Il serait alors possible de déterminer l'impact du programme en comparant les deux situations. La constitution de ces deux groupes se réalise aléatoirement. Par la dimension aléatoire qui les caractérise, les expériences randomisées permettent d'annuler les biais de sélection⁵, ce que les méthodes non expérimentales ne parviennent pas à faire. En matière d'évaluation, cela offre à la randomisation son statut de *gold standard* méthodologique, en annulant les biais de sélection, la randomisation offre des résultats fiables et solides, permettant de produire des preuves. Cette méthode possède alors deux principaux avantages, à savoir, d'une part, la possibilité de concevoir des preuves d'efficacité des programmes de lutte contre la pauvreté ; et par ailleurs, de par sa dimension expérimentale la rencontre avec le terrain. Lors de la mise en place de l'expérience, les chercheurs sont directement confrontés au terrain et donc aux manifestations de la pauvreté, rompant avec l'ignorance que Banerjee et Duflo critiquaient. En se fondant sur ces deux avantages, Banerjee et Duflo se font les principaux promoteurs de la randomisation en économie du développement. Ils créent, en 2003, avec Sendhil Mullainathan⁶, le *Abdul Latif Jameel*

³ Voir (Banerjee et Duflo, 2011) et Duflo (2009).

⁴ Tout au long de ce travail, je prendrai pour synonyme expériences randomisées, randomisation, expérience par assignation aléatoire, essais randomisés.

⁵ L'attribution aléatoire au sein des deux groupes rend parfaitement comparables ces derniers ; autrement dit seul l'effet du programme est rendu visible, et l'intervention d'autres facteurs est annulée.

⁶ Sendhil Muallainathan a depuis quitté la direction du J-PAL, il a été remplacé par Rachel Glennester.

Poverty Action Lab (J-PAL) au *Massachusetts Institute of Technology*. L'objectif principal de ce laboratoire est de « transformer la recherche en action »⁷. Il existe donc une double volonté. La première - caractérisée par la recherche - est de parvenir à fournir via une méthode expérimentale des preuves d'efficacité concrètes des programmes de développement. La seconde - caractérisée par l'action - vise à ce que ces preuves soient utilisées par les décideurs politiques. En somme, l'action politique doit être fondée sur des preuves. La randomisation représente pour les chercheurs du J-PAL le meilleur outil pour réaliser ces objectifs. Tous les chercheurs du J-PAL ont en commun de réaliser des expériences randomisées.

L'une des premières évaluations de ce type en économie du développement fut menée par Edward Miguel et Michael Kremer (2004) de 1998 à 2001, Miguel au Kenya. C'est l'une des expériences randomisées les plus fameuses⁸. Elle cherche à évaluer l'impact de la distribution de médicaments contre les vers intestinaux sur l'assiduité des élèves dans le district de Busia, à l'ouest du Kenya. Pour cela, Kremer et Miguel (2004) évaluent le *Primary School Deworming Project* (PSDP). Ce programme se fixe pour objectif d'offrir des médicaments contre les maux intestinaux en vue de favoriser l'éducation des enfants. Il est mené par l'organisation non gouvernementale (ONG) *Internationaal Christelijk Steunfonds Africa* (ICS), en partenariat avec le ministère de la santé Kenyan. Les soixante-quinze écoles de la région sont assignées, aléatoirement⁹, à trois groupes distincts de vingt-six écoles. Le premier groupe bénéficie gratuitement de médicaments contre les maux intestinaux en 1998 et en 1999 et constitue durant ces deux années le groupe de traitement. Le deuxième groupe bénéficie gratuitement des médicaments en 1999, et cette année là il constitue avec le premier groupe le groupe de traitement. Le troisième groupe reçoit le traitement en 2001. En 1998 et

⁷ Voir : <http://www.povertyactionlab.org/fr/j-pal>

⁸ Deaton (2010a) qualifie cette expérience de « produit phare » du J-PAL et insiste sur le fait qu'elle a été très souvent reprise par les différents chercheurs du J-PAL : « *I turn to the flagship study of the new movement in development economics – Miguel and Kremer's (2004) study of intestinal worms in Kenya. This paper is repeatedly cited in DGK's manual and it is one of the exemplary studies cited by Duflo (2004) and by Banerjee and He (2008). It was written by two senior authors at leading research universities and published in the most prestigious technical journal in economics.* » (Deaton, 2010a, 446)

⁹ Les différentes écoles ont, tout d'abord, été listées alphabétiquement, et c'est à partir de cette liste que l'assignation aux différents groupes s'est réalisée. Angus Deaton (2010a) critique très vivement cette expérimentation, considérant que cette expérience n'est pas une expérience randomisée : « *As befits a paper that aims to change method, there is emphasis on the virtues of randomization, and the word « randomization » or its derivatives appears some sixty times in the paper. However, the « randomization » in the study is actually an assignment of schools to three groups by their order in the alphabet (...). Alphabetization may be a reasonable solution when randomization is impossible but we are then in the world of quasi or natural experiments, not randomized experiments.* » (Deaton, 2010a, 446)

1999, ce dernier groupe constitue le groupe de contrôle. En plus des médicaments, lorsque les écoles font partie du groupe de traitement, elles bénéficient de conférences de prévention.

L'évaluation a montré que la distribution de médicaments contre les maux intestinaux réduit de moitié les infections intestinales au sein des groupes de traitements. Les élèves qui ont bénéficié du traitement sont moins fréquemment malades, leur niveau d'anémie est plus faible, et un grand nombre d'entre eux ont pris du poids et grandi d'un demi centimètre. Finalement, le traitement augmente l'assiduité scolaire de plus de sept points de pourcentage. Autrement dit, le traitement a un effet positif sur l'assiduité scolaire. Par contre, aucun effet positif sur les résultats scolaires n'a été démontré. Miguel et Kremer ont ensuite mené une analyse coût / bénéfices, au terme de laquelle ils concluent que le programme de distribution des médicaments contre les vers intestinaux s'avère tout à fait rentable :

« Deworming was by far the most cost-effective method of improving school participation among a series of educational intervention implemented by ICS in this region of Kenya that were subject to randomized evaluations. » (Kremer et Miguel, 2004, 205)

Les résultats de cette expérience ont donc été jugés extrêmement positifs. Qui plus est, cette expérience va dans le sens des recommandations de l'Organisation Mondiale de la Santé (OMS, 1987) en promouvant les programmes de distribution de médicaments contre les maux intestinaux dans les écoles des pays en développement. De plus, les résultats positifs de cette expérience ont donné lieu à la création par Kremer d'une organisation non gouvernementale *Deworm the World*¹⁰. Cette ONG a pour but de collecter des fonds afin de financer des programmes de distribution de médicaments contre les vers intestinaux. Au travers de cette expérience, on trouve la volonté du J-PAL d'identifier un problème précis - l'absentéisme des élèves. Une solution concrète est évaluée - offrir des médicaments contre les maux intestinaux ; mais aucune solution d'envergure n'est envisagée. Les deux objectifs du J-PAL sont alors saillants : il s'agit bien d'évaluer des programmes de développement en vue de guider la décision politique. L'expérience par assignation aléatoire a permis de déterminer des preuves d'efficacité d'un programme de développement. Le rôle de l'ONG ICS et du ministère Kenyan permet de pointer le lien étroit que le J-PAL entretient avec le terrain et la sphère politique. En s'adressant à un problème précis et en lui proposant une solution concrète, l'action politique a été rendue possible.

¹⁰ Pour plus d'information sur cette organisation, voir le lien ci-dessous : <http://www.dewormtheworld.org/>. On peut noter que Kremer est le président de cette association et que Duflo fait partie de son conseil d'administration.

Ce lien entre le terrain, les chercheurs et le politique a grandement participé au succès du J-PAL au cours de ces dix dernières années. Depuis cette période, cette approche a rencontré un succès croissant en économie du développement¹¹. Les expérimentations du J-PAL constituent l'un des outils actuellement les plus utilisés pour évaluer des politiques de développement. Angrist et Pischke (2010) parlent même de « révolution empirique »¹². Une telle approche tend à orienter l'économie du développement dans une nouvelle direction en remplaçant en son centre les évaluations par assignation aléatoire. Cette nouvelle perspective offerte par les chercheurs du J-PAL vise à montrer qu'il n'est pas nécessaire d'implanter des politiques d'envergure afin de lutter contre la pauvreté. Face aux problèmes concrets, des solutions concrètes impliquant des changements à la marge peuvent avoir d'importants effets. Dans cette optique, l'approche du J-PAL, à travers les expériences randomisées, évalue des solutions concrètes de lutte contre la pauvreté, dont l'accumulation pourrait conduire à une « révolution douce » :

« Ce n'est pas là « ravalier nos ambitions » : nous croyons que des progrès graduels et l'accumulation de petits changements peuvent parfois aboutir à une révolution douce. » (Banerjee et Duflo, 2011, 356-357)

Même si l'objet du J-PAL se résume à évaluer des programmes locaux, l'ambition n'est pas pour autant mineure : l'accumulation de solutions concrètes pourrait contribuer à transformer la situation des pauvres, et en ce sens à lutter efficacement contre la pauvreté. En tout cas, cela n'implique pas d'imposer la violence d'une révolution politique. Bien au contraire : la randomisation en économie du développement, institutionnalisée par Banerjee et Duflo, n'a pas uniquement pour vocation d'être une révolution empirique, elle souhaite aller

¹¹ Comme le souligne d'ailleurs Banerjee et Duflo : « Au cours des dix dernières années, on a assisté à une véritable explosion des expérimentations par assignation aléatoire dans le champ de l'économie du développement. A l'automne 2008, à l'occasion de la conférence NEUCD (*New England Universities Conference on Development Economics*), sur les 112 articles présentés utilisant des données microéconomiques (non comprises les expériences de laboratoire), 24 rendaient compte d'expériences randomisées. En 2004, il y en avait 4. Toujours à l'automne 2008, à l'occasion de la conférence BREAD, la conférence la plus prestigieuse en matière d'économie du développement, 4 des 8 orateurs invités ont présenté des expériences randomisées. Sur les six articles publiés ou à paraître en 2008 dans le *Quarterly Journal of Economics* qui utilisent des données microéconomiques de pays en développement, 3 sont des expériences randomisées. Et cet enthousiasme n'est pas l'apanage des seuls universitaires. A la Banque Mondiale, pour le seul continent africain, on pouvait recenser en 2008 67 évaluations par assignation aléatoire sur les 89 programmes d'évaluation en cours. » (Banerjee et Duflo, 2009, 692)

¹² C'est d'ailleurs le titre d'un de leur article : « The Credibility Revolution in Empirical Economics : How Better Research Design is taking the Con out Economics », voir (Angrist et Pischke, 2010). Cet article est une réponse à un article d'Edward Leamer (1983) nuancant l'importance de l'utilisation de l'économétrie en économie.

au-delà et conduire à « une révolution douce ». Toute la question est de savoir si ces changements à la marge peuvent donner lieu à une transformation aussi profonde?

2. Mener une analyse épistémologique de l'approche du J-PAL

Cette thèse a pour objectif de mener une analyse épistémologique¹³ de l'approche du J-PAL. Cette analyse comportera une double dimension, méthodologique et théorique. La dimension méthodologique vise à interroger la méthode utilisée par le J-PAL. Elle questionne les modes d'inférence que la randomisation permet de produire (comment les résultats sont-ils produits ?), la fiabilité des résultats obtenus par la randomisation (ces résultats sont-ils fiables ?), ainsi que la possible utilisation par la sphère politique de ces résultats (ces résultats sont-ils utilisables par les décideurs politiques ou sont-ils cantonnés à leur contexte – celui de l'expérience ?), tout comme la possible généralisation des résultats (est-il envisageable d'étendre un programme à une échelle plus large, à partir des résultats obtenus ?). Pour résumer, deux principales questions président à mon analyse méthodologique : la nature de *gold standard* méthodologique de la randomisation (1), ainsi que la possible transposition des résultats obtenus par le J-PAL dans la sphère politique (2).

La seconde dimension, théorique, questionne l'apport du J-PAL aux débats théoriques qui ont traversé l'économie du développement au cours de ces dix dernières années. Cette dimension s'interroge sur le tournant que le J-PAL aurait opéré sur l'économie du développement : s'agit-il uniquement d'un tournant empirique ou est-il aussi question d'un tournant théorique et politique ?

S'intéresser à ces deux dimensions me permet d'examiner l'approche du J-PAL dans son ensemble, cette approche ne se limitant pas à la méthode qui la fonde – la randomisation. Elle se fixe aussi deux principaux objectifs : (1) produire des preuves, afin (2) de guider la décision politique en matière de lutte contre la pauvreté. Etudier uniquement la dimension méthodologique de cette approche ne me permettrait que de traiter d'un de ses objectifs. C'est pourquoi, il est nécessaire de s'intéresser aussi à sa dimension théorique, afin de pouvoir questionner les deux objectifs indissolublement liés, que s'est donné le J-PAL.

2.1 La dimension méthodologique de l'approche du J-PAL : les critères de validité interne et externe

¹³ Pour une analyse des différentes manières de faire de l'épistémologie, voir, par exemple, Picavet (1995 ; 1996).

La dimension méthodologique de cette recherche s'attache à la validité interne et externe des résultats obtenus par la randomisation. Questionner une méthode particulière revient communément à proposer deux critiques : une critique interne et une critique externe. Alors que la première critique s'interroge sur la cohérence logique permettant d'aboutir à un résultat particulier, la seconde critique vise à confronter les résultats obtenus à la réalité afin d'attester ou non de leur véracité. La manière dont je choisis de questionner la randomisation s'éloigne sensiblement de cette distinction commune, mon intention étant de m'interroger sur la validité interne et externe d'une telle méthode et non d'en proposer une critique interne et externe¹⁴.

Les notions centrales, lorsque l'on s'intéresse à la méthodologie des expérimentations, sont, en effet, celles de validité interne et externe. Floris Heukelom (2009) retrace l'histoire de ces notions et de leur développement au sein des différentes disciplines où les expériences ont pris une place importante. Selon lui, l'intérêt croissant pour ces notions commence par les travaux du statisticien Ronald Fisher et du psychologue Louis Leon Thurstone à la fin des années vingt. Leur objectif est de déterminer quel procédé expérimental permettra de rendre compte d'inférences valides. Autrement dit, l'expérience permet-elle de définir des inférences valides ? La question de la validité définit alors un critère permettant d'évaluer la portée des procédés expérimentaux. Au milieu des années cinquante, un questionnement similaire est repris par l'*American Psychological Association* (APA), association dont le but est de qualifier les standards méthodologiques de la recherche en psychologie. A l'issue de ce questionnement, un manuel définissant les recommandations techniques en matière d'expériences en psychologie a été établi. Donald Campbell (1957) critiquera vivement les recommandations provenant de ce manuel tout en proposant une distinction précise des notions de validité interne et externe¹⁵. La discussion sur les termes validité interne et externe de Campbell et de l'*American Psychological Association* définit le vocabulaire qui prévaut, toujours aujourd'hui, en psychologie expérimentale¹⁶. Une fois cette définition solidement établie en psychologie, elle se dissémine au sein des autres sciences sociales¹⁷. Comme

¹⁴ Même en se centrant sur les notions de validité interne et externe on ne s'éloigne, cependant, jamais complètement des critiques internes et externes. Ces dernières se retrouvent, d'une certaine manière, au sein de mon analyse, mais ne constituent pas mon analyse.

¹⁵ Je développerai précisément les notions de validité interne et de validité externe chez Campbell (1957 ; 2002) lors des trois prochaines sous-section (2.1.1 ; 2.1.2 ; 2.1.3).

¹⁶ « *The discussions of validity in the Technical Recommendations and by Campbell between the early 1950s and early 1960s defined the experimental vocabulary in the psychology for subsequent decades, up into the twenty first century.* » (Heukelom, 2009, 6)

¹⁷ Heukelom ne l'évoque pas mais ce critère devient, avec le développement de la médecine expérimentale autour des essais cliniques médicaux, le principal critère pour évaluer les inférences produites par les essais

Heukelom le souligne, il n'est pas surprenant que les économistes s'emparant des méthodes expérimentales utilisées en psychologie s'emparent aussi du principal critère permettant de définir et de délimiter un bon procédé expérimental.

Aujourd'hui, ce critère est celui qui prédomine lorsque l'on s'intéresse à la méthodologie des expérimentations et je structurerai mon analyse méthodologique de la randomisation autour de ces deux critères. En premier lieu, car ils sont les critères de référence dans le domaine ; et en second lieu, en ce qu'ils me permettent d'éclairer et de questionner une telle méthode. Afin de présenter le plus clairement possible l'analyse méthodologique que je souhaite mener, je développerai, tout d'abord la notion de validité définie par Campbell (2.1.1) puisque sa définition fait aujourd'hui office de référence. A partir de la typologie qu'il établit, je définirai les notions de validité interne (2.1.2) et que celles de validité externe (2.1.3).

2.1.1 La notion de « validité »

La notion de validité au sein d'un procédé expérimental est définie par Campbell de la façon suivante : « *We use the term validity to refer to the approximate truth of an inference. When we say something is valid, we make a judgment about the extent to which relevant evidence supports that inference as being true or correct.* » (Campbell, 2002, 34). La validité est donc comprise comme ce qui permet d'évaluer la fiabilité d'une inférence obtenue à partir d'un procédé expérimental particulier¹⁸. De cette définition de la validité, Campbell extrait une typologie organisée autour de quatre types de validité : (i) celle des conclusions statistiques ; (ii) celle qualifiée d'interne ; (iii) celle comprise comme construite ; et enfin (iv) la validité externe. La validité des conclusions statistiques porte, en premier lieu, sur l'utilisation des statistiques afin d'inférer que la mise en place du traitement ou du programme

cliniques. Le rôle important de ce critère dans le domaine médical peut aussi expliquer sa dissémination au sein des sciences sociales.

¹⁸ Campbell (2002) relie la validité d'une inférence à la vérité d'une telle proposition. Cependant, il explique qu'il est impossible d'être certains du caractère vrai d'une inférence, ce dernier ne pouvant qu'être falsifié. Campbell s'inscrit alors dans une tradition Poppérienne. Il relie tout de même la notion de validité à celle de vérité. Pour cela, il distingue trois théories de la vérité : la théorie de la correspondance, la théorie de la cohérence et la théorie pragmatique. La première théorie (celle de la correspondance) considère qu'une affirmation est vraie si cette dernière est conforme à la réalité observée. La deuxième (celle de la cohérence) juge qu'en affirmation est vraie si elle fait partie d'un ensemble d'affirmations cohérentes entres elles. La troisième (celle du pragmatisme) déclare qu'une affirmation est vraie si cette dernière est utile. Campbell associe Bruno Latour à la théorie pragmatique, Eihenhart et Howecal à la théorie de la cohérence, et il associe l'ensemble des scientifiques dont le travail est de collecter des données à la théorie de la correspondance. La notion de validité que définit Campbell souhaite défendre une utilisation de chacune de ces théories dans l'appréhension de la vérité. Une inférence valide est donc considérée comme vraie si elle est conforme à la réalité, et si elle est cohérente au sein d'autres inférences, puis si elle est jugée utile.

entraîne une variation. L'évaluation de cette validité relève de deux types de questionnements : « *(1) whether the presumed cause and effect covary and (2) how strongly they covary.* » (Campbell, 2002, 42). La validité des conclusions statistiques concerne l'existence d'une variation statistique, ainsi que le degré de celle-ci. La validité interne renvoie au fait que la variation statistique observée est bien le fruit d'une relation causale et non une simple corrélation : « *We use the term internal validity to refer to inferences about whether observed covariation between A and B reflects a causal relationship from A to B in the form in which the variables were manipulated or measured. To support such an inference, the researcher must show that A preceded B in time, that A covaries with B (already covered under statistical conclusion validity) and that no other explanations for the relation are plausible.* » (Campbell, 2002, 53) La validité interne garantit le fait que la variation statistique observée est bien causée par le traitement - ou par le programme - et non par d'autres facteurs¹⁹. Afin d'évaluer et de mesurer un phénomène particulier, il est nécessaire de construire ce phénomène ; à titre d'exemple, pour mesurer la taille d'un individu, il faut construire la notion de taille, définir à quoi renvoie cette taille, ainsi que l'unité qui va en permettre la mesure comme les centimètres ou les mètres. Les deux enjeux de la validité construite sont donc de comprendre les constructions nécessaires à l'appréhension de phénomènes ainsi que leurs évaluations. Ces constructions sont jugées valides si elles permettent de passer d'un échantillon particulier à des constructions plus larges, dans lesquelles elles s'inscrivent : « *Construct validity involves making inferences from the sampling particulars of a study to the higher-order constructs they represent.* » (Campbell, 2002, 65) La validité construite permet aussi de passer d'un échantillon spécifique à un échantillon plus large, cela à l'aide de constructions conceptuelles qui permettent d'envisager ce passage. La validité externe, enfin, a trait à la possibilité de généraliser la relation-causale que l'expérience a déterminée : « *External validity concerns inferences about the extent to which a causal relationship holds over variations in persons, settings, treatment, and outcomes.* » (Campbell, 2002, 83) La validité externe exprime la généralisation d'une inférence causale d'un contexte à un autre, d'une population à une autre, ou encore d'un traitement à un autre traitement.

2.1.2 La validité interne

¹⁹ Campbell (2002) appelle aussi cette validité interne « validité locale d'ensemble » (*local molar validity*). Le terme local traduit le fait que la conclusion causale est limitée au cadre de l'expérience, et l'idée d'ensemble montre que même limitée à un cadre la conclusion causale s'inscrit dans un ensemble de plusieurs composantes.

Ces quatre types de validité entretiennent des liens très étroits entre eux. La validité interne et la validité des conclusions statistiques s'intéressent toutes deux au lien entre le traitement – le programme – et le résultat : « *These two validity types are closely related. Both are primarily concerned with study operations (rather than with the constructs those operations reflect) and with the relationship between treatment and outcome.* » (Campbell, 2002, 63) Elles sont, dès lors, intimement liées et ne se séparent que parce qu'elles représentent deux étapes distinctes du processus. La validité de la conclusion statistique permet de mettre en évidence une variation, la validité interne prouve que cette variation est bien le fruit d'une causalité. Compte tenu de ces similitudes, lorsque je me référerai à la notion de validité interne, j'entendrai simultanément ce que Campbell appelle validité des conclusions statistiques et validité interne. Je justifie cela par le fait que ces deux types de validité concernent le lien du traitement et du résultat. De plus, cette distinction établie par Campbell en 2002, vise uniquement à affiner la notion de validité interne telle qu'il l'avait définie en 1957²⁰.

Lorsque Francesco Guala s'intéresse à la notion de validité en économie expérimentale, il retient la seule distinction entre la validité interne et externe. L'intérêt qu'il porte à ce critère montre bien l'importance qu'il revêt lorsqu'on s'attache à une analyse méthodologique des procédés expérimentaux²¹. La validité interne concerne, selon Guala, la validité des inférences à l'intérieur du domaine de l'expérience - au sein du laboratoire, puisque l'analyse qu'il propose se concentre principalement sur des expériences de laboratoire. A l'opposé, la validité externe désigne le lien de ces inférences avec le monde réel, en dehors du laboratoire. Selon Heukelom, avec l'analyse de Guala les notions de validité interne et externe se trouvent au cœur des analyses méthodologiques en économie expérimentale²².

En économie du développement, Duflo et Banerjee prennent aussi pour critères les notions de validités interne et externe. La validité interne renvoie à la fiabilité des résultats trouvés à l'aide des expériences randomisées, ou comme la définit Duflo : « *whether we can conclude that the measured impact is indeed caused by the intervention in the sample.* » (Duflo, Glennerster et Kremer, 2007, 3950). Autrement dit, la validité interne rend compte de

²⁰ En 1957, Campbell ne distingue que la validité interne de la validité externe : « *Validity will be evaluated in terms of two major criteria. First, and as a basic minimum, is what we can be called internal validity (...) the second criterion is that of external validity* ». (Campbell, 1957, 297)

²¹ Voir (Guala, 1999, 2003, 2005) et (Guala et Mittone, 2005).

²² « *With the analytical treatment by Guala and others, internal and external validity in economics came to the center of a logic of experimentation in economics.* » (Heukelom, 2009, 14)

la fiabilité de l'évaluation. C'est à travers elle qu'il est possible de questionner le premier objectif du J-PAL : produire des preuves.

2.1.3 La validité externe

La validité construite et la validité externe sont, elles aussi, très similaires. Elles portent sur un questionnement identique, se rapportant à la question de la généralisation. Malgré quelques légères différences²³, ces deux types de validité peuvent être pensées simultanément. C'est pourquoi, lorsque je parlerai de la validité externe, tout au long de ce travail, j'entendrai indistinctement la validité construite et la validité externe telles qu'elles sont définies par Campbell. Appliquée à la méthodologie du J-PAL, la validité externe renvoie à la portée générale de cette approche en économie du développement. Duflo la définit de la façon suivante : « *whether the impact we measure would carry to other samples or populations. In other words, whether the results are generalizable and replicable.* » (Ibid.) La validité externe a donc une double caractérisation. Je choisis dans ce travail de distinguer ces deux niveaux au sein de la validité externe. Le premier niveau traduit la possible utilisation d'un résultat dans un contexte autre que celui de l'expérience dans lequel il a été établi. Lorsque ce premier niveau de validité externe est garanti, on parle d'indépendance du résultat par rapport à l'environnement ; à l'inverse, lorsqu'il n'est pas assuré, on parle de dépendance à l'environnement :

« La dépendance à l'environnement est un élément clé de la possibilité de généralisation (ou de la validité externe). La question posée est la suivante : obtiendrons-nous le même résultat si nous menions la même expérimentation dans un contexte différent ? Ou plus exactement, le programme qui est en cours d'évaluation aurait-il les mêmes effets s'il était mis en œuvre ailleurs (et pas dans le cadre d'une expérimentation) ? » (Banerjee et Duflo, 2009a, 702)

Le premier niveau exprime la possible transposition d'un résultat donné à un autre contexte. Le second niveau, quant à lui, envisage la généralisation d'un programme à une échelle plus large, le passage d'un programme à une échelle plus large pouvant, en effet, transformer le résultat obtenu par l'expérience :

²³ Les inférences provenant de la validité construite sont par définition des constructions et peuvent être détachées de la relation causale mise en évidence ; alors que tel n'est pas le cas pour la validité externe. De plus les méthodes qui visent à améliorer ces deux types de validité sont différentes. La validité construite afin d'être améliorée a besoin d'une explication et d'une bonne évaluation des éléments de l'étude. La validité externe peut être améliorée par une réflexion sur la relation causale mise en évidence.

« Un problème connexe est celui posé par ce que l'on appelle couramment les effets d'équilibre général (bien que ce terme puisse créer une confusion ; nous préférons donc parler d'effets d'équilibre car le concept d'équilibre général est essentiellement un concept multimarché). Les effets d'un programme identifiés dans le cadre d'une petite étude peuvent être différents des effets du programme s'il était généralisé au niveau national ». (Banerjee et Duflo, 2009, 712)

L'analyse méthodologique du J-PAL que je propose s'organise autour de ces deux critères, celui de validité interne et celui de validité externe. Le premier me permet de questionner la nature des inférences produites par une telle approche. Le second, au sein duquel je distingue deux niveaux, vise à interroger la possible utilisation par la sphère politique des résultats expérimentaux du J-PAL ainsi que leur généralisation à une échelle plus large.

2.2 La dimension théorique (et politique) de l'approche du J-PAL

La dimension théorique de mon analyse épistémologique s'attache à questionner l'apport du J-PAL à l'économie du développement. Cette dimension théorique ne peut être détachée de sa portée politique. D'une part, le J-PAL entend explicitement guider la décision. En ce sens, ce laboratoire possède une importante vocation politique²⁴. D'autre part, les débats théoriques auxquels le J-PAL souhaite offrir une réponse sont en eux-mêmes éminemment politiques. A titre d'illustration, le principal débat auquel le J-PAL se réfère est celui de l'aide au développement. Ce débat oppose les tenants d'une aide massive destiné à éradiquer la pauvreté, aux critiques qui pointent le rôle désastreux de l'aide dans le processus de développement et la rendent fautive des échecs de développement. Le J-PAL cherche à offrir des preuves d'efficacité de l'aide. Il entend trancher ce débat à travers des preuves, qui doivent ensuite constituer l'un des principaux outils des décideurs politiques. La question est alors de savoir si le J-PAL parvient, réellement, à fournir ces preuves d'efficacité et donc à guider la décision politique.

De plus, s'intéresser à la dimension théorique d'une telle approche revient aussi à se demander si les résultats obtenus à l'aide de la randomisation participent à la construction d'un nouveau cadre théorique en économie du développement. Beaucoup de résultats obtenus par les expérimentations du J-PAL sont décrits comme surprenants et contredisants les

²⁴ Le fait de mettre l'accent sur la lutte contre la pauvreté et non sur la définition de la pauvreté, comme le font Duflo et Banerjee, pointe aussi cette volonté d'action.

résultats de la théorie économique standard, notamment les travaux de Theodore Schultz (1964)²⁵ ; participent-ils, pour autant, à un nouveau cadre d'analyse ? Permettent-ils d'offrir une nouvelle appréhension des phénomènes de pauvreté ? Ont-ils une implication dans la compréhension de la pauvreté ?

Interroger la dimension théorique de l'approche du J-PAL, renvoie alors à deux questionnements : celui du cadre théorique potentiel offert par une telle approche et celui de sa portée au sein de l'économie du développement. Par théorie, j'entends la définition qu'en donne Marion Vorms (2009), c'est à dire un outil de représentation et d'inférences. En ce sens, une théorie exprime des hypothèses permettant de décrire et d'expliquer des inférences. Une théorie doit ainsi « dire quelque chose » des inférences qu'elle observe et vise à les expliquer. La question est alors de savoir si la randomisation du J-PAL permet de produire cette représentation tout en étant un outil d'inférence. Mais, il est aussi question de l'implication et des effets en économie du développement de cette représentation. L'objectif de cette thèse est donc de déterminer si la randomisation du J-PAL offre une révolution complète, c'est à dire à la fois empirique, théorique et politique, des modalités de lutte contre la pauvreté.

3. La problématique de la thèse

Le J-PAL a très souvent mis l'accent sur la validité interne de son approche²⁶. Mais, sa validité externe est beaucoup plus faible. Ces deux types de validités peuvent se compromettre l'une et l'autre. Autrement dit, favoriser la validité interne d'un procédé expérimental peut compromettre sa validité externe - et inversement²⁷. Concernant le J-PAL, l'insistance portée sur la validité interne tend à fragiliser sa validité externe. Le J-PAL, par sa volonté de produire des preuves, place au premier plan la fiabilité de ces résultats. Ses chercheurs entendent minimiser le nombre d'hypothèses *a priori* afin de favoriser au maximum la validité interne de leurs résultats. Ce faible nombre d'hypothèses *a priori*, voire

²⁵ Voir par exemple (Duflo, 2006c) où elle démontre explicitement ce point.

²⁶ Voir par exemple, (Duflo, 2006b) ; (Duflo, Glennerster, et Kremer, 2007) ; (Banerjee et Duflo, 2009) ; (Duflo et Kremer, 2005).

²⁷ Par exemple Guala (1999, 2003, 2005) et (Guala et Mittone, 2005) considèrent qu'il y a un compromis à trouver entre les deux types de validités. Heukelom (2009) tente de nuancer ce point de vue en montrant qu'il n'y a pas de lien de causalité entre la validité interne et externe. Je ne rentrerai pas dans le détail d'une telle discussion. Mon objectif n'est pas de discuter des notions de validité interne et externe, mais de les utiliser comme critère afin d'examiner la randomisation du J-PAL. Et il se trouve que dans ce procédé expérimental précis et dans les postulats des différents chercheurs qui constituent le J-PAL, les notions de validité interne et externe entrent clairement en opposition, et semblent être deux objectifs antagonistes. L'objectif du présent travail, on l'a vu, est de rendre cela explicite. Mais ce n'est pas à dire que la validité interne entre toujours en opposition avec la validité externe.

leur absence, ne permet pas aux chercheurs du J-PAL de mettre en évidence « pourquoi un programme fonctionne », mais leur permet uniquement de montrer que « le programme fonctionne ou ne fonctionne pas ». En d'autres termes, l'approche du J-PAL peine à mettre en évidence les mécanismes sous-jacents à ses résultats expérimentaux. Ces résultats fonctionnent ainsi comme une « boîte noire ». Ils mettent en évidence un lien causal fiable, mais n'arrivent pas à expliquer cette causalité. Cela se comprend en grande partie – et c'est l'une des thèses que je développerai dans ce travail – par l'ambiguïté du rôle donné à la théorie dans une telle approche. L'approche du J-PAL a souvent été critiquée pour sa dimension a-théorique. Pourtant, la place de la théorie ne s'y résume pas uniquement à une absence. Pour ne pas perdre en validité interne, les chercheurs du J-PAL se refusent à toute théorie pensée en amont. Mais ils entendent créer une théorie à partir de leurs résultats expérimentaux, ils souhaitent développer une théorie qui unifierait les différents résultats.

Il semble toutefois difficile, voire impossible, de créer cette théorie unificatrice sans une première théorie, pensée en amont, qui permettrait de saisir les mécanismes sous-jacents aux résultats expérimentaux. Il existe alors une tension importante entre validité interne et validité externe. Un arbitrage semble nécessaire. Cette tension fragilise les deux objectifs du J-PAL, les rendant contradictoires. Cette tension ou cette contradiction définit ce que j'appellerai une faille épistémologique. Cette faille épistémologique me semble expliquer la difficulté du J-PAL à produire des recommandations politiques claires, ainsi que sa difficulté à produire une théorie unificatrice.

Le récent tournant d'Esther Duflo vers un nouveau « paternalisme »²⁸, me semble illustrer parfaitement les conséquences de cette faille épistémologique. Duflo propose d'envisager un « paternalisme démocratique » qui, en imposant certains choix aux pauvres, augmenterait leur liberté. La notion de liberté à laquelle elle se réfère est celle qu'avait définie Amartya Sen dans le cadre de l'approche par les capacités²⁹. Mais, la proposition de Duflo s'éloigne sensiblement de la définition même des capacités pointant alors certaines contradictions au sein de sa proposition. Tout d'abord, le paternalisme envisagé n'est fondé sur aucune preuve, ce qui semble paradoxal compte tenu de la position initiale du J-PAL. Ensuite, la liberté, telle que la définit Sen, est à la fois instrumentale et substantielle. Dans cette perspective, le paternalisme ne peut être pensé comme un instrument vers plus de liberté, puisque la liberté

²⁸ Voir (Duflo, 2012a).

²⁹ Voir (Sen, 1982, 1992, 1999a, 1999b, 2009)

ne s'acquiert, chez Sen, que par la liberté. Enfin, retirer certains choix aux pauvres suppose d'avoir déterminé le processus de décisions des pauvres, ce qui reste absent de l'approche du J-PAL. Tous ces éléments pointent le fait qu'il est difficile, uniquement à partir d'expériences, de produire des recommandations politiques claires à destination des décideurs politiques. En d'autres termes, en se refusant à théoriser, à postuler des hypothèses, à mettre en évidence de mécanismes, l'approche du J-PAL fournit des résultats dont la validité interne est très forte, probablement beaucoup plus élevée que dans d'autres méthodes, mais elle ne parvient pas à guider les décisions politiques, forçant une telle approche à séparer ces deux objectifs. On peut alors comprendre pourquoi Duflo propose un nouveau paternalisme : la recommandation politique ne peut qu'être détachée de sa méthodologie.

La problématique de ce travail est donc la suivante : dans quelle mesure la faille épistémologique de l'approche du J-PAL explique-t-elle sa difficulté à fournir des recommandations politiques claires de lutte contre la pauvreté ? Ou, pour le formuler autrement, dans quelle mesure les deux objectifs que se fixe le J-PAL sont-ils contradictoires ? Mon travail vise à répondre à cette question. C'est à travers une analyse épistémologique que je définis la faille épistémologique de l'approche du J-PAL, ainsi que sa traduction politique. Ce qui me conduit à m'intéresser d'abord à une dimension méthodologique puis à une dimension théorique.

4. La méthode de la thèse

Mon travail se centre principalement sur les travaux d'Abhijit Banerjee et d'Esther Duflo. Leurs écrits constituent le corpus principal de cette thèse. Le J-PAL est désormais un laboratoire de recherche très influent, constitué de nombreux chercheurs - plus de quatre vingt -, à travers le monde. Le nombre d'évaluations ne cesse d'augmenter : le J-PAL compte, aujourd'hui, près de 300 expériences en cours ou réalisées. S'intéresser à l'ensemble des travaux du J-PAL aurait représenté un objet trop vaste. De plus, Banerjee et Duflo sont les principaux représentants et promoteurs d'une telle approche. D'une part, ils sont à l'initiative de la création du J-PAL et sont toujours les directeurs de ce laboratoire. D'autre part, ils ont écrit de nombreux articles cherchant à définir et justifier leur méthode. Ils sont donc pratiquement les seuls parmi les chercheurs du J-PAL à avoir produit des travaux réflexifs sur leur approche.

L'approche du J-PAL a fortement été critiquée, au cours de ces cinq dernières années. Deux principales vagues de critiques peuvent être distinguées. Une première vague de critiques provient directement des champs d'application du J-PAL, c'est à dire à la fois de l'économie du développement et de l'économie expérimentale. Une seconde vague de critiques émane de réflexions extérieures au domaine d'application du J-PAL et porte sur des questionnements de sociologie des sciences et de philosophie des sciences. L'approche du J-PAL étant récente, ces deux vagues de critiques constituent ma littérature secondaire sur le sujet, mon objectif étant de me positionner face à elle.

La randomisation utilisée par le J-PAL a surtout été questionnée au sein de sa sous-discipline : l'économie du développement. Ces travaux ont tous en commun d'insister sur le caractère dangereux de l'hégémonie croissante de la randomisation en économie du développement, ainsi que sur la faible validité externe de cette approche. Au premier rang de ces critiques se trouve celle d'Angus Deaton (2009, 2010a). Ce dernier montre que le principal problème de l'approche du J-PAL est son incapacité à rendre compte des mécanismes sous-jacents à ses résultats. Dani Rodrik (2007) insiste sur l'hégémonie d'une telle approche en économie du développement, en tentant de montrer que toute méthode possède ses forces et ses faiblesses, celle de la randomisation étant liée à sa fragile validité externe. Dans la même lignée, Martin Ravallion (2009a, 2009b) cherche à alerter sur la dominance du J-PAL en économie du développement, s'inquiétant de la transformation de l'agenda de recherche qui en découle. Michael Carter et Christopher Barrett (2010) s'inquiètent, eux aussi, de la faible validité externe de la randomisation et mettent l'accent sur le problème de l'hétérogénéité des effets de traitements. Ils produisent ainsi une étude poussée des problèmes éthiques qu'une telle approche soulève en économie du développement. Daron Acemoglu (2010) insiste, quant à lui, sur la nécessité d'intégrer la randomisation au sein d'analyses structurelles, en vue d'améliorer la validité externe des travaux du J-PAL. La randomisation du J-PAL a aussi été critiquée au sein, d'un autre domaine, l'économie expérimentale. Glenn Harrison (2011) donne, par exemple, une critique très sévère du J-PAL, montrant qu'une telle approche ne tient pas compte des résultats obtenus et développés au sein de l'économie expérimentale et comportementale. Enfin, les critiques les plus anciennes de la randomisation en sciences sociales, et sans doute les plus constructives sont celles de Heckman (1991, 2010), Heckman et Smith (1995), Heckman, Clement et Smith (1997) et Heckman, Ichimura, Smith et Todd (1998) qui portent à la fois sur la technique d'une telle méthode que sur sa portée.

Une seconde vague de critiques de l'approche du J-PAL provient de domaines extérieurs à cette dernière. En France, ces travaux émanent d'une réflexion de sociologie des sciences. Alain Desrosières (2013) s'intéresse au rôle des pays en développement dans l'inventivité et la création de nouvelles méthodologies. Agnès Labrousse (2010) s'appuie sur une perspective foucaldienne pour questionner la randomisation du J-PAL et donc la nouvelle forme de savoir qu'une telle approche peut offrir. Arthur Jatteau (2013) propose une synthèse des différents résultats obtenus par le J-PAL, il présente également la méthode et les limites de cette approche. Fabrice Bardet et Roser Cusso (2012) tentent d'analyser le contexte social et politique de la randomisation. En philosophie des sciences, Nancy Cartwright (2007a, 2009c, 2010b, 2011a) et Cartwright et Munro (2010) produisent une importante analyse de la randomisation, insistant sur sa faible validité externe et mettant en cause son statut de *gold standard* méthodologique. Cartwright ne se concentre pas sur la randomisation du J-PAL mais s'intéresse aux expériences randomisées en général. Vincent Guillin (2012) cherche, quant à lui, à confronter les travaux de Cartwright à la randomisation du J-PAL.

Contrairement, à la première vague de critiques, ce travail se situe en dehors de l'économie du développement et de l'économie expérimentale. Ma position est donc différente de celles des économistes du développement et des expérimentateurs. Mon analyse ne provient pas de l'« intérieur » des domaines d'application du J-PAL. Pour autant, les critiques de Deaton, Ravallion, Rodrik, Acemoglu, Carter et Barrett constituent une référence importante de ce travail. Je développerai précisément ces différentes critiques. Je tenterai de les unifier autour d'une unique question - celle de la généralisation d'un programme à une échelle plus large. Autrement dit, je tenterai d'unifier ces différentes critiques afin d'illustrer le second niveau de validité externe. Cela me permettra de penser simultanément les critiques faites en économie du développement, celles de Heckman et celle de Harrison. L'économie expérimentale occupe, également, une place particulière dans mon travail. Afin d'être en mesure de saisir pleinement les enjeux de la randomisation, je tente de situer la randomisation du J-PAL au sein de l'histoire de l'économie expérimentale. Et, ici les travaux de Harrison et List (2004), ainsi que ceux de Levitt et List (2008) offrent un cadre précieux.

Comparée à la seconde vague de critiques, cette thèse ne s'apparente pas à un travail de sociologie des sciences. Cela même si je retrace l'histoire du J-PAL, sa structure, son lien avec les différentes institutions internationales. Pour autant, je ne m'intéresse pas aux conditions sociales qui ont permis l'émergence et le développement de l'approche du J-PAL. A l'inverse, cette thèse se comprend comme un travail de philosophie des sciences. Je cherche

à questionner les fondements et les implications d'une approche spécifique en sciences sociales. En ce sens, c'est un travail d'épistémologie économique. Les travaux de Nancy Cartwright sont ici particulièrement éclairants. Son analyse offre à mon étude son cadre analytique. Cartwright propose sa propre philosophie des sciences, et c'est à l'aide de cette dernière qu'elle questionne la randomisation. L'analyse de Cartwright me permettra ainsi de traduire le premier niveau de la validité externe de la randomisation, c'est à dire la possible utilisation de ses résultats dans la sphère politique.

Ainsi, cette thèse devrait-elle permettre d'unifier les travaux des économistes du développement et ceux de Nancy Cartwright, afin de rendre compte et de questionner les deux niveaux de validité externe de l'approche du J-PAL que j'ai distingués. Mon travail cherche à aller plus loin, puisqu'il questionne aussi bien les apports théoriques du J-PAL que leurs implications politiques. Une analyse de la validité interne de l'approche du J-PAL me permet de déterminer et de définir son statut de *gold standard* méthodologique. Le questionnement méthodologique au travers de la dualité validité interne/validité externe me permet également de préciser la nature de la faille épistémologique d'une telle approche. Ces deux éléments apportent un éclairage nouveau sur les propositions politiques offertes par cette approche. En ce sens, ce travail s'inscrit dans le champ de l'histoire de la pensée économique³⁰ récente.

5. Structure et organisation de la thèse

La double dimension définissant mon analyse épistémologique structure l'argumentation de la thèse. Les deux premières parties de la thèse s'attachent à la dimension méthodologique. L'objectif de ces deux parties est de définir la faille épistémologique qui caractérise l'approche du J-PAL. La première partie s'attache à la validité interne des expériences randomisées et la deuxième partie à leur validité externe. La troisième partie traite de la dimension théorique, son objectif étant d'évaluer les apports théoriques du J-PAL à l'économie du développement.

5.1 La première partie : la force du J-PAL, son importante validité interne

La première partie de ce travail s'attache à présenter la force de la randomisation et son statut de *gold standard* méthodologique. Un premier chapitre se concentre sur l'histoire de la

³⁰ L'histoire de la pensée économique est ici entendue comme l'histoire d'un savoir économique (Cot et Lallement, 1995).

randomisation. J'y retrace l'histoire d'un tel procédé expérimental, depuis ses débuts en psychologie, jusqu'à sa définition précise et rigoureuse par Ronald Fisher, son essor en médecine au travers des essais cliniques médicaux et à son utilisation en sciences sociales. Ce retour historique me permet de montrer que cette méthode a toujours été utilisée et promue au nom de la fiabilité des évaluations qu'elle permet.

Le deuxième chapitre présente la méthode du J-PAL. Je présente, tout d'abord, l'histoire, la structure et le fonctionnement d'un tel laboratoire. Ensuite, je m'intéresse plus particulièrement à la volonté première de ce laboratoire qui consiste à : produire des preuves. Je montre aussi comment les expériences randomisées sont implantées ainsi que les différents enjeux rencontrés lors de cette implantation.

Le troisième chapitre présente le second objectif du J-PAL : guider la décision politique dans le cadre du mouvement d'*evidence-based policy*. Je reviens alors sur la naissance d'un tel mouvement, dans le but d'en saisir les enjeux, puis je montre comment le J-PAL se l'approprie. Cette première partie permet d'explicitier les deux objectifs du J-PAL, ainsi que leur articulation autour de la validité interne de la randomisation.

5.2 La deuxième partie : la faiblesse du J-PAL, son insuffisante validité externe

La deuxième partie passe d'un questionnement en termes de validité interne à un questionnement sur la validité externe de la méthodologie du J-PAL. L'objectif de cette partie est de montrer que la forte validité interne du J-PAL se fait au prix d'une très faible validité externe. De ce fait, cette partie définit la « faille épistémologique » d'une telle approche.

Le premier chapitre pose un cadre analytique à partir de l'épistémologie de Nancy Cartwright, qui permet de définir un premier niveau de validité externe. Cartwright questionne explicitement l'utilisation des preuves dans le domaine politique. Elle distingue le fait de « produire » les causes de leur utilisation.

Le deuxième chapitre interroge le second niveau de la validité externe de la randomisation. Le chapitre reprend les différentes critiques qui ont été adressées à la randomisation et tente d'en offrir une unité. Il permettra de montrer que pour produire des résultats utilisables l'approche du J-PAL doit perdre en validité interne au profit d'une plus grande validité externe. Ce à quoi une telle méthodologie semble se refuser, ce qui fait des résultats qu'elle produit sont comme une « boîte noire ». Le second niveau de la validité externe se trouve alors fragilisé.

Le troisième chapitre se centre sur une critique spécifique adressée à la randomisation du J-PAL, sa dimension a-théorique. Cette critique me semble englober toutes les autres. De plus, elle me permet d'envisager simultanément les deux niveaux de validité externe. Autrement dit, elle permet de penser simultanément l'analyse de Cartwright et les différentes critiques adressées à la randomisation. Je montre que la théorie dans l'approche du J-PAL ne se résume pas à une absence, mais qu'un double statut lui est accordé. Je définis ce double statut par (1) le refus des chercheurs du J-PAL de penser une théorie avant de conduire les expériences et (2) par la volonté de construire une théorie fondée et englobant les différents résultats. Cette absence de théorie pensée en amont fragilise la création d'une théorie *a posteriori* et de ce fait fragilise les deux niveaux de validité externe. Ce qui définit la faille épistémologique d'une telle approche qui me semble expliquer la tension existante entre validité interne et validité externe de l'approche du J-PAL.

5.3 La troisième partie : les apports théoriques du J-PAL, entre paternalisme et irrationalité

La troisième partie a pour objectif d'évaluer l'apport du J-PAL aux débats qui traversent l'économie du développement. Cette partie vise à montrer que l'approche du J-PAL peine à fournir les réponses politiques qu'elle aurait souhaité offrir. La position paradoxale de Duflo en faveur d'un paternalisme est expliquée par cette faiblesse. Cela tend à renverser la perspective initiale du J-PAL, et à offrir un éclairage sur la contradiction entre les deux objectifs que se fixe le J-PAL.

Le premier chapitre définit le contexte théorique auquel fait référence le J-PAL, contexte face auquel il souhaite se situer et apporter des réponses. J'y explicite l'effondrement du « consensus de Washington » qui a ravivé la question de l'aide au développement. Cette question de l'aide trouve dès lors une traduction politique clivée, soit en faveur d'une forte assistance des pays développés vers les pays en développement soit en faveur de la mise en avant d'*empowerments*.

Le deuxième chapitre cherche à évaluer l'apport concret du J-PAL à ces débats. Dans ce dessein, on retracera les expériences que Pascaline Dupas a menées sur le paludisme, qui constitue l'emblème du débat sur l'aide au développement. Ce chapitre permettra de montrer la force et la limite d'une telle approche. La première expérience de Dupas pointe une énigme (la faible utilisation des moustiquaires dans les zones étudiées), les expériences qui suivent cherchent à évaluer des dispositifs de *nudging* qui répondraient à cette énigme, mais aucun de ces dispositifs ne fonctionne. Ces différentes expériences posent donc très clairement la

question de la nécessité éventuelle de mesures paternaliste et laissent présager la proposition de Duflo pour un paternalisme fort.

Le dernier chapitre questionne le tournant récent d'Esther Duflo en faveur d'un paternalisme démocratique : expliquer ce tournant me permet d'analyser le cadre théorique et politique d'Esther Duflo. Elle envisage un paternalisme fondé sur l'approche par les capacités. L'objectif d'un tel paternalisme est d'imposer aux pauvres un certain nombre de biens en vue d'accroître leur liberté. Cette position philosophique me semble recouvrir deux principales confusions. Je montrerai alors que ces confusions ne font que traduire la faille épistémologique de l'approche dont Esther Duflo se fait la principale promotrice.

L'approche du J-PAL place au premier plan l'action mais en se refusant à théoriser en amont cette dernière est fragilisée. Pour reprendre les termes de Victor Hugo, la vive force du J-PAL peine alors à achever ce que l'idée aurait dû ébaucher.

Partie 1. La randomisation et le J-PAL : le rôle de la preuve

Introduction de la Partie 1.

« Je propose dans mes travaux la réponse suivante. Les sciences sociales peuvent accompagner la politique sociale dans un processus d'expérimentation créatrice. (...) La lutte contre la pauvreté est une réponse à une crise permanente. Elle a besoin d'expérimentation, dans les deux acceptions du terme : il faut essayer sans relâche de nouvelles approches ; mais il faut également se donner la chance de reconnaître ses erreurs et d'en apprendre quelque chose. Pour cela l'expérimentation doit être rigoureuse et scientifique. » (Duflo, 2009, 26-27)

La volonté des chercheurs du J-PAL est de proposer une lutte contre la pauvreté « rigoureuse et scientifique ». Les expériences randomisées apparaissent, pour ces derniers, être le meilleur outil pour réaliser un tel objectif. Ces expériences semblent permettre la rigueur et la scientificité voulues par les chercheurs du J-PAL. La randomisation est l'unique méthode permettant d'annuler un certain nombre de biais. C'est la dimension aléatoire caractérisant une telle méthode qui donne la possibilité d'annuler ces biais, offrant à la randomisation sa plus grande force : une importante validité interne. Pour autant, la volonté du J-PAL ne s'arrête pas là et entend guider la décision politique à l'aide des preuves que la randomisation lui permet de collecter. Les deux objectifs du J-PAL sont ici saillants (produire des preuves (1) afin de guider la décision politique (2)). Avec la randomisation du J-PAL, les preuves se retrouvent à un double niveau : elles sont d'abord collectées puis utilisées. L'objectif de cette partie est de développer le double rôle accordé à la preuve au sein de l'approche du J-PAL. Autrement dit, l'objectif est de présenter les deux objectifs du J-PAL ainsi que leur articulation. En ce sens, cette partie est centrée sur la notion de preuve, et s'intéresse donc essentiellement à la validité interne de la randomisation du J-PAL.

Dans ce but, le premier chapitre de cette partie vise à retracer une double histoire : celle de la randomisation et celle des expérimentations de terrain en économie. La dimension aléatoire qui caractérise la randomisation a toujours été utilisée, de ses débuts au XIX^{ème} siècle ou durant le XX^{ème} siècle, comme un gage de scientificité. Cette dimension aléatoire était considérée comme ce qui permet d'obtenir des résultats fiables. Charles Sanders Peirce intégra à son protocole expérimental une dimension aléatoire en vue d'obtenir des résultats solides. Ronald Fisher définira le cadre statistique et le protocole expérimental qui font de la randomisation une des méthodes les plus rigoureuses pour évaluer l'effet d'un traitement ou

d'un programme. Cela explique grandement que la médecine se soit emparée du protocole et du cadre statistique de Fisher en vue d'évaluer les effets de traitements médicaux. Retracer cette histoire me permet alors de mettre en avant une des dimensions fondatrices de l'utilisation de la randomisation par les chercheurs du J-PAL : le gage de scientificité et de rigueur. De plus, la randomisation du J-PAL se retrouve à la croisée d'une seconde histoire, celles des expériences de terrain en économie. Ces dernières, mises en places au XXème siècle visaient à évaluer des politiques publiques, intégrant souvent à leur protocole une dimension aléatoire. Retracer cette seconde histoire me permet de traduire le second aspect définissant l'approche du J-PAL : une volonté politique. Ces deux histoires me permettront alors d'explicitier les deux objectifs du J-PAL.

Le deuxième chapitre a pour objectif de définir le J-PAL, et donc de s'intéresser spécifiquement à la randomisation du J-PAL. Ce qui me semble définir le J-PAL est l'utilisation et la promotion de la randomisation. L'utilisation de la randomisation est justifiée, par les chercheurs du J-PAL, au travers de l'importante validité interne qu'offre une telle méthode. Ce deuxième chapitre traitera donc exclusivement du premier objectif du J-PAL : la production de preuves. Je m'attacherai à montrer comment le J-PAL utilise la randomisation, comment elle peut être mise en place sur le terrain, en vue de garantir la rigueur et la scientificité voulues par le J-PAL. De plus, ce chapitre sera l'occasion de présenter un tel laboratoire. Je ferai cela en insistant sur le fait que sa structure est pensée afin de promouvoir la randomisation en économie du développement. Ce chapitre me permettra alors de présenter et détailler le premier objectif du J-PAL.

Le troisième chapitre, s'intéressera lui, au second objectif du J-PAL. Le J-PAL s'insère explicitement au sein du mouvement récent d'*evidence-based policy*. Tout comme le souhaite le J-PAL, ce mouvement entend fonder la décision politique sur des preuves. Ces dernières doivent être rendues accessibles pour les décideurs politiques. C'est pourquoi le J-PAL propose de nombreuses synthèses de ses résultats en fonction de différents thèmes afin que les décideurs politiques puissent choisir compte tenu de leur budget et des effets qu'ils souhaitent. Ce mouvement d'*evidence-based policy* s'inspire d'un mouvement semblable, qui a d'abord eu lieu en médecine au travers de l'*evidence-based medicine*. Ce dernier cherche à fonder la pratique médicale sur des preuves. Ces dernières sont issues d'essais cliniques médicaux. Lors de ce chapitre, j'opérerai un double aller-retour avec la médecine. Le premier cherchera à démontrer que la randomisation du J-PAL est un essai clinique d'une nouvelle

forme, et qu'en ce sens elle produit le plus haut niveau de preuves de l'*evidence-based policy*. Le second aller-retour avec la médecine me permettra de définir l'*evidence-based medicine*, en vue d'obtenir une définition complète de l'*evidence-based policy*. Puis, je m'attacherai à montrer comment le J-PAL s'insère au sein de ce mouvement. Ce double aller-retour me permettra de montrer clairement le lien du J-PAL avec l'*evidence-based policy* et donc de montrer comment le J-PAL entend réaliser le second objectif qu'il se fixe.

Chapitre 1 : L'histoire de la « randomisation »

0. Introduction

« The tradition of experimental research using control groups and random allocation spans both social and natural science. However, the institutionalization of these methodological ideas and practices has been different in social and medical settings. There is no doubt that from the mid-19th century on, prospective experimental studies with control groups became more firmly ensconced and the method more clearly articulated in medicine. » (Oakley, 2000, 325)

La randomisation utilisée par le J-PAL est à la rencontre de deux histoires. Une première histoire relève directement de la méthode utilisée, c'est à dire un procédé expérimental intégrant une dimension aléatoire. Les premières expériences qui ont intégré une dimension aléatoire à leur protocole ont lieu en psychologie au début du XIX^{ème} siècle. Mais c'est en médecine, durant le XX^{ème} siècle, que ce type d'expériences connaît son plus grand succès. Cela explique que la randomisation soit souvent accolée à la médecine et donc aux essais cliniques. La dimension aléatoire introduite au procédé expérimental permet de garantir une importante fiabilité aux résultats obtenus par l'expérience. Cette dimension aléatoire est considérée comme une garantie de rigueur. Cette garantie de rigueur retrouve explicitement la première volonté du J-PAL – produire des preuves. La seconde histoire renvoie au domaine d'application de la randomisation du J-PAL, c'est à dire à l'économie expérimentale. Les expériences du J-PAL sont des expériences de terrain, en ce sens, elles se distinguent des expériences de laboratoire menées en économie. Les expériences de terrain menées en économie datent du milieu du XX^{ème} siècle. Certaines de ces expériences ont une visée directement sociale, on parle alors d'expériences sociales de terrain. Ces dernières ont pour objectif d'évaluer les effets d'interventions politiques. Cet objectif retrouve nettement la seconde volonté du J-PAL - évaluer des programmes de développement en vue de guider la décision politique.

Retracer ces deux histoires offre un double intérêt. Tout d'abord, cela me permet de traduire les deux objectifs du J-PAL. Ensuite, cela me permet de situer une telle approche, tant du point de vue de la méthode utilisée que de son domaine d'application. Je m'intéresserai, tout d'abord, à l'histoire d'une telle méthode ; puis, à l'histoire des expériences de terrain en économie. La première partie de ce chapitre s'intéresse aux origines de la randomisation en psychologie. Le philosophe pragmatiste, Charles Sanders Peirce est le

premier à intégrer une dimension aléatoire à son procédé expérimental. Il cherche à contrecarrer les résultats trouvés par Gustav Fechner. Ce dernier cherchait à montrer scientifiquement l'existence d'un seuil, à partir duquel on ne peut plus distinguer le degré d'un faible *stimulus*, c'est la fameuse loi dite de « Fechner-Weber ». Peirce, en souhaitant invalider ce résultat, évoquera pour la première fois le terme télépathie et intégrera, pour la première fois aussi, une dimension aléatoire au sein de ses expériences. De nombreuses expériences³¹ tenteront, après Peirce, de démontrer l'existence de la télépathie en se fondant sur des expériences comportant une dimension aléatoire. Les débuts de la randomisation sont donc caractérisés par cette dimension aléatoire et un objet commun : la télépathie.

La seconde partie de ce chapitre s'attache à l'histoire plus récente de l'utilisation de la randomisation. Elle s'intéresse alors aux travaux de Ronald Fisher. Ce dernier, au début du XX^{ème} siècle, offre à la randomisation un cadre expérimental et statistique rigoureux. Fisher définit alors clairement la structure de la randomisation : une expérience constituée de deux groupes dont l'attribution se fait aléatoirement. Fisher avait pensé un tel cadre pour l'agriculture. Cependant, ce cadre sera massivement utilisé non pas dans l'agriculture mais en médecine. C'est ce cadre qui définit aujourd'hui les essais cliniques médicaux et leur donne une dimension rigoureuse. Ces deux premières parties permettent donc de traduire le premier objectif du J-PAL – produire des preuves.

La troisième partie de ce chapitre relate l'histoire des expériences de terrain en économie. Pour cela, il est, tout d'abord, nécessaire de définir précisément ces dernières ; et en ce sens, de les distinguer des autres types d'expériences menées en économie, celles de laboratoire. Puis de s'intéresser particulièrement à un type d'expériences de terrain, les expériences sociales. Ces dernières sont caractérisées par la volonté d'évaluer des politiques sociales. Elles mêlent un intérêt politique à une volonté de rigueur scientifique. Cette dernière partie permet alors de traduire le second objectif du J-PAL – guider la décision politique.

1. Les origines de la randomisation : de la sensation à la télépathie

Les premières expériences randomisées sont menées par Charles Sanders Peirce (1839-1914), qu'il conclut en évoquant la télépathie³². Ce thème sera, ensuite, largement repris. Des

³¹ Comme on le verra, Charles Richet (1885) mènera plusieurs expériences dans ce but ainsi que John Edgar Coover.

³² Comme l'explique Hacking : « *the word telepathy³² was less than two years old when Peirce wrote these words, but the Eastern seaboard was awash with enthusiasm for the idea. The American Society for Psychical Research³² was about to be founded, and its first project was a survey to determine the frequency of the telepathic phenomena* ». (Hacking, 1988, 434)

milliers d'études randomisées auront lieu à l'université de Stanford, à la fin du XIX^{ème} siècle, afin de vérifier l'existence de la télépathie. En effet, pour Hacking, la télépathie constitue l'origine des expériences randomisées³³. La randomisation est ici justifiée par le caractère « empirique » des questions : « *the more « empirical » and nontheoretical a question, the more randomization makes sense* ». (Hacking, 1988, 434). Je présenterai, tout d'abord, le rôle de la psychologie dans l'histoire de la randomisation. Cela revient à expliciter les expériences de Peirce ainsi que le rôle de la télépathie au sein de ces dernières (1.1). Ensuite, je développerai les différentes expériences qui ont porté sur la télépathie à la fin du XIX^{ème} siècle (1.2). Ces deux points me permettront, en suivant l'histoire de Hacking (1988)³⁴, de définir les origines de la randomisation.

1.1 La psychologie comme discipline originelle de la randomisation

L'expérience menée par Charles Sanders Peirce et son étudiant Joseph Jastrow (1863-1944), de décembre 1883 à janvier 1884 (Stigler, 1992, 63), s'inscrit en « psychophysique » ou « physio-psychologie ». Ce terme est donné par Gustav Theodor Fechner³⁵ (Fechner, 1860) pour décrire les recherches qui visent à mettre en lien les *stimuli* physiques avec une forme de conscience telle que les sensations. Fechner, en créant la « psychophysique », cherche à donner à l'étude de l'esprit un caractère scientifique. Il s'est intéressé au rapport entre l'excitation et la sensation. L'excitation est le *stimulus* que l'on fait ressentir à un sujet et la sensation renvoie au moment où l'individu prend conscience du *stimulus* qu'on lui a appliqué. Fechner cherche à mesurer le rapport entre l'excitation et la sensation et tente d'établir un seuil entre les deux. Dans ce but, il a mené plusieurs expériences, mais ces dernières n'étaient pas randomisées (1.1.1). C'est Charles Sanders Peirce, qui en voulant contredire les résultats de Fechner, introduit la randomisation (1.1.2).

1.1.1 Sensations et seuil : les expériences de Fechner

³³ Le titre de l'article de Hacking (1988) est d'ailleurs explicite : « *Telepathy : Origins of Randomization in Experimental Design* ».

³⁴ Trudy Dehue (1997), Ann Oakley (2000) ou encore Stigler (1992) produisent une histoire similaire des origines de la randomisation. Pour autant, ils n'insistent pas sur la place de la télépathie au sein des premières expériences randomisées.

³⁵ Fechner (1801-1887) est l'instigateur de la « psychophysique », une telle discipline vise à mélanger la psychologie, la physique et la métaphysique. Les travaux de Fechner ont pour objectif de mesurer les liens de l'âme et du corps. C'est en cela qu'ils se retrouvent à la fois en psychologie, en physique ou encore en métaphysique.

Le seuil que cherche à définir Fechner (1860) correspond au niveau de stimulation qui marque l'entrée d'un événement dans la conscience de celui qui la reçoit. Autrement dit, ce seuil renvoie au moment où l'individu parvient à avoir conscience de la stimulation et donc au moment où il sent quelque chose. Ce seuil peut être mesuré à travers deux aspects : la sensibilité à un événement sensoriel, qui n'était pas présent avant ; et la sensibilité à un changement d'intensité dans un événement sensoriel qui est connu ou déjà présent. Fechner (1860) définit trois méthodes pour mesurer ces deux aspects : la méthode des limites, la méthode d'ajustement, la méthode du *stimulus* constant (autrement appelée la méthode des cas vrais et faux). Dans ses expériences, Fechner va privilégier les deux dernières méthodes. La méthode du *stimulus* constant lui permet d'appliquer une série de *stimuli* de façon répétée et dans un ordre aléatoire. Le sujet juge, pour chaque *stimulus*, s'il a ou non ressenti quelque chose. Par exemple, la sensibilité du toucher sur un avant-bras est testée en demandant aux individus, qui ont été piqués par une aiguille, s'ils ont été piqués une ou deux fois. La méthode des cas vrais et faux lui permet d'établir un tableau des réponses données par le sujet. Si ce dernier a bien évalué les *stimuli*, il s'agit d'un cas juste et s'il s'est trompé d'un cas faux. À l'aide de ce tableau, il est possible de déterminer le rapport entre les sensations et les *stimuli*.

Fechner a réalisé plusieurs fois ces expériences. Elles montrent que l'intensité de nos sensations croît en fonction de l'intensité de nos excitations. Par exemple, plus l'aiguille est enfoncée fortement dans l'avant-bras, plus le sujet aura une sensation de piquê intense. Il est, cependant, difficile de mesurer ce rapport, car une même excitation ne produit pas nécessairement la même sensation. Pour prendre un autre exemple, on entend le tic-tac d'une pendule distinctement durant la nuit, alors que nous n'entendons pas ce même tic-tac dans la journée, lorsque d'autres bruits s'y ajoutent. C'est pourquoi, pour Fechner (1860) l'intensité de la sensation ne croît pas de façon totalement similaire à celle de l'excitation, en fait, elle croît plus lentement. De ce constat découle la « loi de Fechner »³⁶ : la sensation varie de façon proportionnelle au logarithme de l'excitation³⁷. C'est-à-dire que pour une multiplication dans l'intensité du *stimulus*, l'intensité de la sensation ne fait qu'ajouter et ne multiplie pas.

1.1.2 Absence de seuil et randomisation

³⁶ Autrement appelée « loi de Weber-Fechner », Ernst Weber (1795-1878) était l'un des premiers à étudier quantitativement la corrélation entre sensation et excitation. Afin de lui rendre hommage, Fechner appellera sa loi, « loi de Weber ». Communément les deux termes sont accolés, donnant la « loi de Weber-Fechner ».

³⁷ La « loi de Fechner » s'écrit alors : $S = a \cdot \log I$; avec S pour sensation, et I pour intensité.

Comme cela a été soulevé, les expériences de Fechner ne sont pas des expériences randomisées ; cependant, elles sont centrales, car ce sont elles qui donnent à Peirce l'objet de ses expériences futures. C'est en s'opposant aux résultats de Fechner que Peirce envisage ses expériences. Avec son assistant et étudiant Joseph Jastrow durant l'année 1884, ils vont réaliser plusieurs expériences. La structure de ces expériences est beaucoup plus sophistiquée que celle des expériences qu'avait menées Fechner³⁸. Peirce et Jastrow s'intéressent, aussi, au lien entre les *stimuli* et les sensations, mais ils vont s'attacher à des variations de *stimuli* d'intensité très faible :

« Peirce wanted to measure extremely subtle sensations, the perception of very small incremental weights. And he had a wonderful idea: a blind randomized experiment. In order to eliminate the biases attendant on factors such as which weight was lifted first, or how the weights were arranged, or whether the subject knew which was which, Peirce worked with an assistant, Joseph Jastrow, who later had a distinguished career himself in psychology » (Stigler, 1992, 64).

Dans leurs expériences, Jastrow et Peirce font ressentir deux pressions différentes aux observateurs et leur demandent de les déterminer ; autrement dit, il est demandé aux observateurs de distinguer entre ces deux pressions. Pour calculer la probabilité d'erreurs de l'observateur, Jastrow et Peirce se réfèrent à la théorie mathématique des erreurs³⁹. Peirce explique que si les résultats montrent que le rapport entre les deux pressions est inférieur à un certain ratio, alors la théorie du seuil de Fechner est vérifiée. Par contre, si les résultats montrent que l'observateur ne peut pas différencier deux excitations lorsque l'une d'elle subit une variation très légère, alors toutes les raisons de croire à un seuil, comme l'envisage Fechner, sont détruites. La théorie mathématique a l'avantage d'offrir une précision plus forte que celle des physiologistes, et permet d'améliorer les méthodes d'observation. De plus, cela offre une méthode concrète pour mesurer la finesse d'une perception. Cela offre, aussi, la possibilité de comparer les résultats d'un observateur à un autre.

L'expérience de Peirce et Jastrow se déroule de la façon suivante : on applique au sujet deux poids différents sur le bout du doigt. Il doit, ensuite, déterminer si le premier poids est

³⁸ Une des explications est due au fait que, contrairement à Fechner, Peirce a travaillé avec son assistant. Cela lui a alors permis une sophistication expérimentale plus grande que celle de Fechner.

³⁹ La théorie des erreurs cherche à quantifier l'écart entre la mesure que l'on fait d'une variable et sa vraie valeur. Voir par exemple (Armatta, 2004).

plus lourd ou moins lourd que le second. Le sujet est obligé de donner une réponse, il ne peut pas dire qu'il ne constate aucune différence entre les deux poids. Ensuite, il doit établir le degré de confiance qu'il a envers sa propre réponse. Ce degré est de 0, s'il n'a aucune préférence entre sa réponse et une autre, de 1 s'il a un léger penchant pour l'alternative qu'il a choisie, de 2 s'il est confiant de sa réponse, et de 3 s'il est très confiant. Afin de savoir s'ils vont commencer par attribuer au sujet le poids le plus lourd des deux ou le poids le plus léger, Peirce et Jastrow introduisent une dimension aléatoire. C'est à partir d'un jeu de cartes, qu'ils vont décider si le poids présenté en premier sera le plus lourd ou le plus léger. Ils utilisent, pour cela, deux paquets de 25 cartes - l'un avec 12 cartes rouges et 13 cartes noires et l'autre avec 13 cartes rouges et 12 cartes noires. La couleur de la carte détermine l'ordre des poids. L'individu tire donc une carte et en fonction de celle-ci, l'ordre des pressions est établi, une fois qu'il a apprécié les différentes pressions il donne son jugement et son degré de confiance. La dimension aléatoire introduite par Peirce et Jastrow évite que les individus formulent leurs jugements en fonction des suppositions qu'ils appliquent aux expérimentateurs. Les résultats de l'expérience montrent que les sensations n'ont pas de seuil, comme l'affirmait Fechner. La conclusion principale de cette expérience est la suivante :

« The general fact has highly important practical bearings, since it gives new reason for believing that we gather what is passing in one another's minds in large measure from sensations so faint that we are not fairly aware of having them, and and give no account of how we reach our conclusions about such matters. The insight of females as well as certain "telepathic" phenomena may be explained in this way. Such faint sensations ought to be fully studied by the psychologist and assiduously cultivated by every man ». (Peirce et Jastrow, 1885, 83)

Les expériences de Peirce et Jastrow pointent deux des dimensions essentielles de la randomisation. Tout d'abord, le fait que la présentation des poids soit décidée aléatoirement, à l'aide d'un jeu de cartes ; puis, l'idée que cet aspect aléatoire garantit la fiabilité des résultats, en réduisant l'impact de facteurs extérieurs sur le jugement des individus⁴⁰. Les expériences de Peirce et Jastrow ont donc un double effet : l'introduction de la dimension aléatoire dans le processus expérimental ainsi que la suggestion d'un phénomène « télépathique », qui orientera bon nombre des expériences randomisées futures.

1.2 Randomisation et télépathie : de la psychologie à la parapsychologie

⁴⁰ « By means of these trifling devices the important object of rapidity was secured, and any possible psychological guessing of what change the operator was likely to select was avoided » (Peirce et Jastrow, 1885, 80).

Ian Hacking s'intéresse de plus près à la conclusion de Peirce et Jastrow en creusant le lien entre la randomisation et la télépathie. Il retrace de façon originale, ce qu'il appelle « l'histoire des origines de la randomisation » (Hacking, 1988), à travers le concept de télépathie. Il montre qu'après les expériences de Peirce et Jastrow, un médecin français, Charles Richet (1850-1935), s'est attaché à vérifier scientifiquement l'existence du phénomène de télépathie, ou ce qu'il appelle la suggestion mentale, qu'il définit de la façon suivante : « la suggestion mentale est l'influence que la pensée d'un individu exerce dans un sens déterminé, sans phénomène extérieur appréciable à nos sens, sur la pensée d'un individu voisin ». (Richet, 1885, 615). Cette suggestion mentale renvoie, pour Richet, aux vibrations de la pensée des individus (Richet, 1885, 617), et c'est cette vibration qu'il cherche à mesurer. Il mène plus de trois cents expériences en 1885, un an après celles de Peirce et Jastrow. Ces expériences ne sont pas randomisées, elles se fondent principalement sur les probabilités. Tous les résultats des expériences menées par Richet sont comparés à un calcul de probabilité. Cependant, ces expériences jouent un rôle central dans l'histoire des premières expériences randomisées. D'une part, puisqu'elles ont vocation à prouver l'existence de la télépathie et lancent donc le thème qui sera repris ensuite au sein de nombreuses expériences randomisées par John Edgar Coover (1872-1938). Et que Richet, d'autre part, souhaite donner une importante « scientificité » à ces expériences. Je commencerai par développer le protocole des expériences de Richet (1.2.1), puis j'explicitai les expériences randomisées de Coover (1.2.2).

1.2.1 Télépathie et probabilités : les expériences de Richet

Dans le but de donner un statut « scientifique » à sa démarche, Richet (1884) se fonde sur le calcul des probabilités. Il compare, tout d'abord, le résultat du calcul des probabilités à ceux de ces expériences. Il illustre cela de la façon suivante : un jeu de 52 cartes comprend quatre couleurs (carreau, trèfle, pique et cœur), la probabilité qu'un individu dise, au hasard, la couleur exacte de la carte qu'il pioche dans le jeu est de $\frac{1}{4}$. S'il y a 208 tirages, la probabilité que cet individu dise juste sur ces 208 tirages est donc de 52 ($208 \times \frac{1}{4}$). Richet fait cette expérience, seul. Il énonce au hasard la couleur de la carte avant de la piocher et note les résultats de ses 208 tirages. Il a eu, par le fruit du hasard, 50 succès. Il en conclut :

« Voici donc un premier point acquis, c'est qu'expérimentalement le calcul des probabilités est aussi vrai que théoriquement ; à savoir que, dès que l'expérience porte sur un suffisant nombre de chiffres, le calcul des probabilités se vérifie à peu de choses près ». (Richet, 1884, 619).

Il procède à quatre types d'expériences différentes. La première est de deviner la couleur de la carte qu'une autre personne a piochée - et qu'elle tente de faire deviner par télépathie. Il répète cette expérience plusieurs fois, le calcul de probabilité sur 1833 expériences est de 458 (1833/4), le résultat avec suggestion mentale (télépathie) est de 510. Il choisit, ensuite, de ne plus procéder avec un jeu de cartes mais avec des photographies : une personne pioche une image et une autre personne, près d'elle, doit deviner l'image. Ici, le calcul de probabilités sur 218 expériences offre un résultat de 42 (218/4), alors que l'expérience sur la suggestion mentale donne un résultat de 67. Dans la troisième expérience, des baguettes ont été cachées sous des boîtes, un des deux individus sait où se trouve la baguette cachée et doit guider, par la suggestion mentale, le deuxième individu qui lui ne sait rien. Le calcul de probabilité est ici de 18 sur 98 expériences (98/5, il y a cinq boîtes), et le résultat des expériences est lui de 44. La dernière expérience vise à présenter à deux individus une série de lettres en désordre, l'un va en déduire un mot (qui peut ne pas reprendre toute la série des lettres) et le second individu doit deviner le mot. Pour trouver une lettre du mot, le calcul de probabilité est 1/26 (puisque l'alphabet est constitué de 26 lettres). Pour un mot de n lettres, le calcul de probabilité est de 1/26 X n. A titre d'illustration, pour un mot de cinq lettres, comme « ESTHER » (Richet, 1885, 658), le calcul de probabilité est de 6/26 (1/26 X 6) ; et le résultat de l'expérience de Richet, pour ce mot, est d'une lettre. Autrement dit, la personne à qui le mot « ESTHER » a été suggéré a pu déterminer une lettre du mot, alors que la probabilité qu'elle y parvienne était de 6/26. Cela conduit Richet à conclure :

« Il est tout à fait invraisemblable que le hasard, sur 300 expériences environ ait pu me donner autant de fois ces séries remarquables (...) Autrement dit, en reprenant une des plus fameuses démonstrations de Pascal : *S'il fallait opter pour la réalité ou la non réalité de la suggestion mentale, je laisserais le hasard décider ; mais je donnerais deux chances à l'hypothèse que la suggestion mentale existe, et une chance seulement à l'hypothèse contraire* »⁴¹. (Richet, 1884, 670)

Richet (1884) fonde toutes ses expériences sur le calcul de probabilités, cela constitue son gage de « scientificité ». Cependant, sa conclusion souligne que l'on ne peut être certain de

⁴¹ Francis Ysidro Edgeworth (1885) note que la probabilité que le phénomène que met en évidence Richet ait été obtenu par chance est très faible ; elle est de 0,00004. (Hacking, 1988, 441)

l'existence de la télépathie. Il faut attendre plus d'une quinzaine d'années, pour que ce thème soit repris au sein de nombreuses expériences randomisées menées par John Edgar Coover (1872-1938).

1.2.2 Télépathie et randomisation : les expériences de Coover

La même année que celle des expériences de Peirce, Leland Stanford Junior, fils unique de monsieur et madame Stanford, décède. Ses parents décident de considérer tous les enfants de Californie comme leurs enfants ; c'est la raison pour laquelle, ils financent massivement l'université de Stanford. Ils offrent, par exemple, 10 000 livres à l'université pour la recherche psychique. L'université ne sait pas encore comment l'utiliser, mais en 1912, cet argent est placé dans le service de psychologie. Cette somme servira à continuer de financer la bourse de recherche du psychologue, John Edgar Coover. Ce dernier offre à la télépathie, dont Richet avait lancé la théorie, sa dimension expérimentale. De 1912 à 1917, Coover mènera 40 500 expériences en vue de tester l'existence du phénomène de télépathie. Ces expériences sont constituées de deux groupes : un groupe avec des étudiants et un autre avec des amateurs ou des professionnels de la parapsychologie. Dans les expériences constituées seulement d'étudiants, ceux-ci sont, là aussi, séparés en deux groupes : un groupe de 97 et un de 105. Le premier groupe pioche, au hasard, une carte et le second doit deviner la carte par télépathie. Dans les 3000 premières expériences, le groupe des 97 avait à changer de place tous les 20 essais. L'objectif était ici de déterminer si le fait de se situer près du groupe qui doit faire deviner la carte favorise la télépathie ou s'il vaut mieux se situer plus loin. Lors des 2500 essais suivants seule la durée du regard de la carte variait, allant de 20 à 60 secondes. L'apport majeur de ces expériences est, pour Hacking, l'utilisation de la randomisation au travers de la création d'un groupe de contrôle :

« the most notable innovation was the use of randomized control against regular experiments. In the regular experiment the agent drew a card at random and imagined it. In the control experiment the reagent guessed, but the agent did not look at the card at all. It was decided at random whether the trial would be regular or control ». (Hacking, 1988, 447)

Avec les expériences de Coover, on assiste à la naissance des expériences contrôlées⁴², c'est-à-dire une expérience fondée sur la distinction de deux groupes - un groupe de contrôle

⁴² Pour une histoire des expériences contrôlées, voir (Dehue, 2000).

et un groupe de traitement. La distinction de ces deux groupes, mise en avant par Coover, permet la comparaison⁴³. Mais les expériences de Coover ne sont pas seulement contrôlées, elles sont aussi randomisées :

« Thus in Coover's design we have both the use of the randomizer (playing cards) for choosing which object is to be guessed and the use of randomizer (the dice box) to assign one of the three possible types of treatment or a control experiment in which there is no treatment. Here we have the whole retenue of control, blindness, and randomization ». (Hacking, 1988, 448)

Les expériences de Coover reprennent le thème de Charles Richet (la télépathie), et elles ajoutent à la structure de Peirce la constitution de deux groupes et donc la dimension et le terme d'expériences contrôlées. Le choix aléatoire était déjà présent chez Peirce avec le rôle du jeu de cartes, qui comportait aussi un aspect aveugle. Les expériences de Coover offre, déjà, un socle conceptuel fort dans l'appréhension de la randomisation : la dimension aléatoire ainsi que le caractère contrôlé.

2. Naissance d'une structure expérimentale rigoureuse et son utilisation par la médecine

Ronald Fisher (1890-1962) et Austin Bradford Hill (1897-1991) sont les principales figures qui marquent le début de l'utilisation de la randomisation au sein d'une structure expérimentale définie précisément. Fisher construit les fondements statistiques d'une telle méthode, ainsi que la structure expérimentale que la randomisation doit prendre ; et Hill l'applique à un domaine particulier : la médecine. Les deux hommes ne travaillent pas ensemble et ont des visées très différentes. Fisher (1935 ; 1926) construit une structure statistique et expérimentale rigoureuse, en vue de l'appliquer à l'agriculture. Hill est médecin et mène des analyses épidémiologistes. Vivant à la même époque, le XX^{ème} siècle, ils échangent, cependant, sur leurs divers travaux⁴⁴. La naissance de la randomisation, dans son sens moderne, est donc due, d'une part à la structure que crée Fisher et d'autre part à l'utilisation qu'en fait Hill :

⁴³ Comme le souligne Trudy Dehue : « it clearly was Coover who first advocated group comparison as an imperative methodological condition ». (Dehue, 2000, 267).

⁴⁴ « It is clear, though, that during the development of Hill's ideas about clinical trials, they were on a good terms. During Fisher's Presidency of the Royal Society (1953-1954), when meetings were held at the London School of Hygiene, Fisher would frequently drop in to the Department of Medical Statistics before a meeting, to pass the time of day with Tony. » (Armitage, 2003, 927)

« Fisher was clearly the progenitor of randomization as an integral part of rigorous comparative experimentation (...) Hill, aided by colleagues such as Philip D'Arcy Hart and Marc Daniels in the streptomycin trial, had the ability and personality to persuade the medical profession that this was the way forward ». (Armitage, 2003, 927)

Hill, par sa position et ses relations, fait de la randomisation l'outil le plus utilisé dans la recherche clinique ; et Fisher offre à une telle méthode son statut de *gold standard*⁴⁵ méthodologique. Je commencerai par exposer la structure que propose Fisher (2.1), puis je détaillerai son application autour de ce qui est considéré comme le premier essai clinique en médecine - l'étude de Hill sur le cancer des poumons (2.2). Cela permettra d'offrir une définition précise du protocole expérimental de la randomisation ainsi que de la rigueur statistique qui lui est accolée.

2.1 Ronald Fisher ou la naissance de la structure expérimentale de la randomisation

Fisher offre la solution statistique et la structure qui définit, aujourd'hui encore, la randomisation⁴⁶. Mais avec Fisher, on quitte la psychologie - et la télépathie – pour un domaine tout autre : l'agriculture. L'objectif de Fisher était de pouvoir évaluer l'impact d'innovations agricoles sur les rendements. Fisher cherche à définir une structure expérimentale rigoureuse permettant de pallier le scepticisme que les résultats expérimentaux peuvent rencontrer. Fisher envisage des expérimentations fondées, à la fois sur un protocole expérimental rigoureux ; mais aussi sur des procédures statistiques, elles-mêmes rigoureuses. Ces deux aspects sont pour lui, en fait, une seule et même chose:

« If the design of an experiment is faulty, any method of interpretation which makes it out to be decisive must be faulty too. It is true that there are a great many experimental procedures which are well designed in that they may lead to decisive conclusions, but on other occasions may fail to do so; in such cases, if decisive conclusions are in fact drawn when they are unjustified, we may say that the fault is wholly in the interpretation, not in the design. But the fault of interpretation, even in these cases, lies in overlooking the characteristic features of the design which lead to the result being sometimes inconclusive, or conclusive on some questions but not at all. Statistical procedure and experimental design are only two different aspects of the same whole, and that whole comprises all the logical requirements of the complete process of adding to natural knowledge by experimentation. » (Fisher, 1935, 2-3)

⁴⁵ J'analyserai, avec plus de précision, ce caractère de *gold standard* et sa mise en avant dans le chapitre suivant.

⁴⁶ Pour une histoire de l'utilisation de la randomisation par Fisher, voir (Hall, 2007).

La structure expérimentale joue un rôle central pour Fisher. Si cette structure n'est pas solide et bien établie, les interprétations que l'on en peut tirer ne seront pas solides non plus. A l'inverse, si la structure expérimentale est solide, les interprétations qui en découlent auront tendance à l'être aussi, même si quelques exceptions peuvent exister. En ce sens, les procédures statistiques et le protocole expérimental sont deux faces d'une même médaille. Je commencerai par préciser ce qui définit un cadre statistique rigoureux pour Fisher (2.1.1), puis je développerai le protocole expérimental qu'il propose (2.1.2).

2.1.1 Un cadre statistique rigoureux, une illustration : « the lady tasting tea »

Afin de donner l'intuition de la randomisation, Fisher prend l'exemple de « *the lady tasting tea* »⁴⁷. Une jeune femme affirme qu'elle est capable de déterminer dans une tasse de thé si le lait a été versé avant ou après le thé⁴⁸. Afin d'infirmer ou de confirmer l'affirmation de la jeune femme, Fisher propose qu'on lui présente huit tasses de thé : pour certaines d'entre elles, le lait a été versé avant le thé et pour d'autres, il a été versé après. Les tasses lui sont présentées de façon aléatoire, il n'a pas été décidé à l'avance l'ordre dans lequel les différentes tasses seraient présentées. La jeune femme doit, une fois qu'elle a eu accès à toutes les tasses, les classer en deux groupes : un groupe où le thé a été versé en premier et un groupe où le thé a été versé en second. Selon Fisher, il est ensuite utile de prévoir les différents résultats possibles, selon la règle de « combinaisons et de permutations »⁴⁹, il existe 70 possibilités de regrouper les quatre tasses. C'est pourquoi Fisher choisit huit tasses car s'il n'y avait eu que six tasses, il n'y aurait que 20 possibilités⁵⁰ pour classer les tasses, ce qui aurait rendu le travail « trop facile » pour la jeune femme. L'objectif est ensuite de réduire au

⁴⁷ Pour une histoire du rôle de cet exemple dans les transformations des statistiques au XX^{ème} siècle, voir (Salzburg, 2002)

⁴⁸ Comme le souligne Ian Hacking (Hacking 1988), la capacité à déterminer si le lait a été versé après ou avant le thé peut paraître curieuse. Plus précisément, elle traduit l'appartenance à une certaine classe sociale, en Angleterre : la classe aisée sert en premier le thé puis le lait et la classe plus pauvre fait l'inverse.

⁴⁹ « *There are 70 ways of choosing a group of 4 objects out of 8. This may be demonstrated by an argument familiar to students of « permutations and combinations », namely, that if were to choose the 4 objects in succession we should have successively 8, 7, 6, 5 objects to choose from, and could make our succession of choices in 8x7x6x5, or 1680 ways. But in doing this we have not only chosen every possible set of 4, but every possible set in every order ; and since 4 objects can be arranged in order in 4x3x2x1, or 24 ways, we may find the number of possible choices by dividing 1680 by 24* ». (Fisher, 1935, 12)

⁵⁰ En reprenant la logique de « permutations » et de « combinaisons » mathématiques, il y aurait eu à choisir 3 objets et successivement on aurait eu 6x5x4 donc 120 possibilités, en arrangeant ces possibilités dans leur ordre : 3x2x1 (6), on trouve le nombre de choix possibles en divisant ces deux termes : 120/6, il y a donc 20 choix possibles.

maximum le fait que la jeune femme puisse placer les tasses dans le bon ordre par pure coïncidence :

« No such selection can eliminate the whole of the possible effects of chance coincidence, and if we accept this convenient convention, and agree that an event which would occur by chance only once in a 70 trials is decidedly « significant » in the statistical sense. » (Fisher, 1935, 13)

Il est impossible de supprimer complètement le rôle du hasard; pour autant, le fait qu'il y ait 70 essais permet d'obtenir un résultat significatif, au sens statistique. Fisher introduit ensuite le rôle de l'« hypothèse nulle ». Celle-ci postule une égalité entre deux données, et est toujours testée contre une autre hypothèse. Il s'agit ici d'une jeune femme qui a classé les tasses dans le bon ordre par pure coïncidence. Pouvoir tester cette hypothèse permet, lorsque le classement des tasses sera opéré, d'être sûr que l'ordre donné par la jeune fille est dû au fait qu'elle ait réussi à distinguer si le lait avait été versé en premier ou non. Une fois cette structure statistique définie, il reste à déterminer le cadre expérimental. Il apparaît évident qu'il faut que les tasses soient le plus identiques possibles, afin que la jeune fille ne bénéficie pas d'informations extérieures lui permettant de savoir si le lait a été versé avant ou après. Cependant, il semble difficile de rendre ces tasses parfaitement identiques⁵¹. La randomisation apparaît être une des principales solutions pour contrecarrer ce problème :

« The element in the experimental procedure which contains the essential safeguard is that the two modifications of the test beverage are to be prepared 'in random order.' » (Fisher, 1935, 19, souligné par moi)

La randomisation fait alors office de garantie. Elle permet d'éviter que d'autres facteurs puissent rentrer en jeu. Le fait que le lait soit versé avant ou après le thé est déterminé de façon aléatoire, afin que seul ce facteur puisse être évalué.

2.1.2 La randomisation : le cœur d'une structure expérimentale rigoureuse

⁵¹ « In practice it is probable that the cups will differ perceptibly in the thickness or smoothness of their material, that the quantities of milk added to the different cups will not be exactly equal, that the strength of the infusion of tea may change between pouring the first and the last cup, and that the temperature also at which the tea is tested will change during the course of the experiment. These are only examples of the difference probably present; it would be impossible to present an exhaustive list of such possible differences appropriate to any kind of experiment, because the uncontrolled causes which may influence the result are always strictly innumerable. » (Fisher, 1935, 18)

Fisher envisage une structure expérimentale au cœur de laquelle se trouve la randomisation. Cette structure s'articule autour de six principales idées : la comparaison (1), la randomisation (2), la construction de groupes (3), la réplication (4), l'existence d'un contraste orthogonal (5) et d'un plan factoriel (6). La comparaison (1) est l'un des enjeux centraux dans une expérience. L'objectif est, pour l'expérimentateur, d'être capable de pouvoir comparer deux situations : une situation dans laquelle un élément varie et une situation où ce même élément reste identique. Dans l'exemple de la jeune fille et du thé, l'objectif était de rendre parfaitement comparables les différentes tasses de thé, afin que la jeune fille puisse déterminer si le lait a été versé avant ou après le thé. La randomisation (2), comme cela a été vu, permet de rendre ces deux situations comparables, en allouant l'élément qui varie (le traitement ou le programme) de façon aléatoire. Cela permet de fixer les différences qui pourraient exister entre les deux situations. Afin que cette allocation aléatoire se matérialise et que la comparaison entre ces deux situations soit rendue possible, il est nécessaire de constituer deux groupes (3). Un premier qui recevra le traitement (l'élément variateur) et un second qui ne recevra rien et restera donc inchangé. Cette séparation en deux groupes est appelée le *randomized block*. La réplication (4) implique de refaire plusieurs fois l'expérience dans des conditions différentes afin d'être sûr de la validité des résultats et de leur indépendance à l'environnement dans lequel l'expérience a été réalisée. L'existence d'un contraste orthogonal (5) est une méthode de comparaison des moyennes, permise même lorsque les groupes ne sont pas tous égaux. Une expérience vise à répondre à plusieurs questions, chacune des questions représentant un contraste. Afin que les différents contrastes ne soient pas redondants, il faut qu'ils soient indépendants. Si tel est le cas, ils sont dits orthogonaux. L'existence d'un plan factoriel⁵² (6) permet à l'expérience de répondre à plusieurs questions distinctes, c'est-à-dire qu'il permet d'évaluer différents traitements différents au sein d'une même expérience.

Le principal intérêt de la randomisation est d'annuler toutes les autres sources de variations, autrement dit d'annuler tous les biais de sélection⁵³. Lorsque l'on réalise une expérience et que l'on fait varier un facteur dans le but de voir son impact, on ne peut jamais être sûr que cet impact soit imputable uniquement aux facteurs que l'on a fait varier, d'autres facteurs, ignorés, peuvent avoir une influence et être une partie explicative de cet impact. La

⁵² Un plan factoriel, renvoie à un plan d'expériences. Un plan d'expériences permet d'utiliser les mêmes groupes pour effectuer simultanément plusieurs comparaisons. Autrement dit, avec les mêmes groupes, il est possible d'évaluer simultanément plusieurs traitements.

⁵³ Je définirai et développerai plus largement ce terme dans le chapitre suivant.

randomisation, en distinguant deux groupes et par sa dimension aléatoire, permet d'annuler ce biais de sélection. La randomisation représente alors pour Fisher « *the essential safeguard* » (Fisher, 1935, 19).

2.2 Randomisation et application : les premiers essais cliniques

La structure définie par Fisher permet à la médecine de répondre à l'un de ses enjeux essentiels : comment savoir si la guérison est liée au traitement ou à un facteur autre ? Ou comment être sûr, par exemple, que la cigarette peut être une des causes du cancer des poumons ? Le premier essai clinique, considéré comme tel, est celui mené par Austin Bradford Hill et Richard Doll en 1950⁵⁴, cherchant à prouver le lien entre la cigarette et le cancer des poumons. Cependant, la première intuition de cette démarche remonte à l'essai réalisé par James Lind au XVIII^{ème} siècle⁵⁵. Je présenterai tout d'abord cette première intuition, à travers l'essai de Lind (2.2.1) ; puis je développerai l'essai clinique mené par Hill et Doll (2.2.2). Cela permettra de mettre en évidence l'application du cadre statistique et expérimental de Fisher à la médecine.

2.2.1 Genèse de l'essai clinique : l'essai de James Lind

Le premier essai clinique, sans qu'il soit défini comme tel, est mené par James Lind en 1747. James Lind tente de soigner le scorbut⁵⁶. Cette maladie était considérée comme « la peste des mers » et touchait à cette époque majoritairement des marins. Lind est un chirurgien de marine pour la *British Navy*. Il prend la mer à bord du HMS Salisbury, un bateau de cinquante canons, en 1747 à Portsmouth. Avant même de faire escale, plusieurs cas de scorbut se déclarent sur le bateau. Après avoir étudié la littérature sur cette maladie en vue d'établir un tableau clinique⁵⁷, il tente de déterminer le traitement permettant de lutter contre cette maladie. Pour cela, il sélectionne douze hommes sur le bateau qu'il sépare en deux groupes de six, et attribue à chaque groupe un régime différent lui permettant de comparer les

⁵⁴ Comme le note Marks (1997, 144) : « *For many physicians in the United-States, the writings of another British statistician, A. Bradford Hill, provided their initial introduction to the methods and purposes of the clinical trials.* »

⁵⁵ Pour une histoire des essais cliniques médicaux voir (Meldrum, 2000) ou (Marks, 1997), ou encore (Duncan, 1997).

⁵⁶ Le scorbut est une maladie due à une carence, il se traduit par un déchaussement des dents, une purulence des gencives, des hémorragies et dans le pire des cas la mort.

⁵⁷ Un tableau clinique est une description précise qui liste tous les symptômes et tous les états pathologiques observés cliniquement. Ce tableau constitue une des bases fondamentales du diagnostic.

effets des différents traitements et d'évaluer leur efficacité. Seul le premier groupe, qui a reçu une dose particulière de vitamine C, a survécu. James Lind a demandé aux douze marins d'ajouter à leur nourriture soit du cidre, soit du vinaigre, soit de l'eau de mer, soit du citron, ou encore de ne rien ajouter. C'est le citron qui a été le plus efficace et qui a permis de soigner le scorbut. James Lind est un des premiers à avoir expérimenté ce que l'on appellera plus tard un essai clinique.

2.2.2 Premier essai clinique et structure expérimentale

L'essai d'Austin Bradford Hill et de Richard Doll porte sur le lien entre le cancer des poumons et la cigarette. Hill et Doll partent du constat qu'au début des années 1950, en Angleterre – ils sont tous les deux britanniques –, on assiste à une importante hausse du nombre de morts dues au cancer des poumons⁵⁸. Ils cherchent alors à expliquer cette importante hausse. Selon eux, deux principales causes peuvent expliquer cette augmentation : la pollution et la cigarette. Il apparaissait difficile de tester les effets de la pollution sur l'augmentation de morts des cancers du poumon, c'est pourquoi Hill et Doll testent le lien entre la cigarette et le cancer des poumons. Leur étude procède de la façon suivante : « *Twenty London hospitals were asked to co-operate by notifying all patients admitted to them with carcinoma of the lung, stomach, colon or rectum.* » (Hill et Doll, 1950, 740) L'étude se fait sur une large partie de la population et ne s'arrête pas uniquement aux poumons puisque des patients atteints à l'estomac, au colon ou au rectum font partie de l'étude. Les patients de plus de soixante-quinze ans sont exclus de l'étude, et aucun des patients n'a refusé de participer, ce qui n'introduit, alors, aucun biais de sélection. Les cas de cancers sont divisés en deux groupes : un groupe A constitué des patients où le cancer a été diagnostiqué par autopsie ou biopsie, et un groupe B où le diagnostic a été fait par un autre type d'examen. Il y a alors 2475 patients, 1079 constituent le groupe A, 267 constituent le groupe B et 1129 patients ne sont dans aucun des deux groupes. Dans ces 1129 patients il y en a 709 qui souffrent d'une autre maladie que le cancer et ils constituent le groupe de contrôle. Hill et Doll soulignent que ces deux groupes sont parfaitement comparables en termes d'âge, de sexe, etc.

⁵⁸ « *The phenomenal increase in the number of deaths attributed to cancer of the lung provides one the most striking changes in the pattern of mortality recorded by the Registrar-General.* » (Hill et Doll, 1950, 739)

Hill et Doll font face à une première difficulté : qu'est-ce qu'un fumeur ? Doit-on considérer comme fumeur une personne qui ne fume plus mais qui a longtemps et beaucoup fumé ? Hill et Doll pointent cette difficulté en notant le fait que les habitudes de fumeurs varient dans le temps. Ils optent pour une définition large : « *A smoker was therefore defined in this inquiry as a person who had smoked as much as one cigarette a day for as long as one year, and any less consistent amount was ignored* ». (Hill et Doll, 1950, 741) La seule façon, pour Hill et Doll, de tester le lien entre la cigarette et le cancer des poumons est de distinguer, au sein des patients atteints de cancer du poumon, les fumeurs des non-fumeurs. Sur les 649 hommes atteints de cancer du poumon, 647 sont fumeurs et seulement 2 ne le sont pas. Pour le groupe de contrôle de 649 patients atteints d'une autre maladie, 622 sont fumeurs et 27 ne le sont pas. Pour ce qui concerne les femmes⁵⁹, sur 60 patientes atteintes du cancer des poumons, 41 sont fumeuses et 19 ne le sont pas. Dans le groupe de contrôle où les femmes sont atteintes d'une autre maladie 28, sont fumeuses et 32 ne le sont pas. Pour Hill et Doll, il existe une corrélation très forte entre le fait de fumer et le cancer du poumon : « *it must be concluded that there is a real association between carcinoma of the lung and smoking* ». (Hill et Doll, 1950, 746)

Cette étude, même si elle est considérée comme le premier essai clinique, ne s'apparente pas totalement à une expérience randomisée mais plutôt à une expérience contrôlée, c'est-à-dire à une expérience où la structure est définie par deux groupes distincts : un groupe de contrôle et un groupe de traitement. Il n'y a pas de sélection aléatoire au sein des patients, il y a uniquement l'existence d'un groupe de contrôle. On retrouve, ici, deux aspects qui serviront de socle à la randomisation : l'expérimentation et l'existence d'un groupe de traitement et d'un groupe de contrôle. De plus, l'expérience de Hill et de Doll ne retrouve pas la définition courante de l'essai clinique randomisé puisqu'il n'est pas intégré à un processus thérapeutique c'est la première fois que cela apparaît, il faut ou bien en discuter avant ; ou bien en discuter plus longuement ici. La randomisation, à proprement parler, était donc impossible ; puisqu'il fallait séparer les personnes atteintes du cancer des poumons puis celles fumeuses, dans le but d'établir potentiellement un lien. L'étude de Hill et Doll ouvre la voie à l'expérimentation en médecine, de façon moderne⁶⁰, et met en avant l'évaluation de facteurs causaux permettant

⁵⁹ Hill et Doll ont séparés les femmes des hommes, dans leurs études, afin d'éviter tous biais.

⁶⁰ Claude Bernard (1865) avait aussi initié l'introduction d'expériences en médecine, mais comme le souligne Foucault (1963), les expériences initiées par Claude Bernard se pratiquaient sur des cadavres, de plus aucune dimension aléatoire n'était introduite au sein de ces expériences ; c'est en ce sens qu'il y a une distinction

d'expliquer un phénomène ; une pathologie dans le cas de la médecine. C'est sans doute, au-delà de l'expérience, ce qui est repris par les chercheurs du J-PAL : mettre à jour les facteurs causaux dans le but de comprendre un phénomène. L'étude de Lind permet donc de donner l'intuition de ce que deviendront les essais cliniques médicaux, et celle de Hill et de Doll permet d'en offrir une illustration.

3. Expériences sociales et de terrain

L'Attribution, en 2002, du prix Nobel d'économie à Daniel Kahneman et Vernon Smith marque la reconnaissance et la légitimité de l'économie expérimentale comme une sous-discipline de l'économie. Pour Francesco Guala, ce prix Nobel reconnaît deux dimensions:

« First, it recognized the work of a scholar who according to the conventions to the contemporary academia should not be labeled as an « economist »⁶¹ (...) But secondly, and more importantly perhaps, the other half prize was devoted to recognizing a methodological innovation, rather than a contribution to the body of economic theory ». (Guala, 2005, 1-2)

L'attribution de ce prix Nobel reconnaît un nouveau domaine ; un nouveau domaine rendu possible, selon Guala, par des changements profonds au sein de la discipline⁶². L'économie expérimentale permet de tester des hypothèses, de mettre en évidence des liens causaux ou de prédire les comportements des individus. On distingue les expériences de terrain et les expériences de laboratoire⁶³. La distinction entre les deux n'est pas toujours claire, mais Harrison et List offrent une typologie permettant de les distinguer. Les expériences qui m'intéresseront particulièrement ici sont celles de terrain, car la randomisation utilisée par le J-PAL se fonde sur des expériences de terrain. C'est pourquoi, à partir de la typologie offerte par Harrison et List (2004), je tenterai de les définir (3.1). Je m'intéresserai, ensuite, à un type d'expériences de terrain particulier : les expériences sociales de terrain. C'est au sein de ces dernières que s'insère la randomisation du J-PAL (3.2). Je m'intéresserai donc ici à la seconde

majeure avec l'essai de Hill et de Doll, c'est pourquoi je parle d'expérimentation dans une appréciation « moderne ».

⁶¹ Daniel Kahneman est, au départ, psychologue et s'intéresse particulièrement aux biais cognitifs. C'est avec Amos Tversky, qu'il commence à étudier la prise de décision en économie et que débutera, ensuite, leurs travaux en économie expérimentale. Vernon Smith est, tout d'abord, ingénieur électricien, avant d'obtenir un doctorat en économie et d'enseigner l'économie du droit. Et c'est en enseignant les principes de la théorie microéconomique sur les équilibres de marchés concurrentiels, qu'ils décident de faire des expériences.

⁶² « *The success of experimental economics was made possible by several profound changes in the discipline of economics as a whole* ». (Guala, 2005, 3)

⁶³ Pour une histoire des expériences de laboratoire voir (Serra, 2012a ; 2012b) et pour une analyse épistémologique de ces dernières, voir, par exemple, (Guala, 2005) et (Bradsley et al., 2010).

histoire de la randomisation du J-PAL - l'histoire de la randomisation au sein de l'économie expérimentale. Cela permettra de situer précisément la méthode de Duflo et de Banerjee au sein des différentes approches expérimentales utilisées en économie.

3.1 Des expériences de laboratoire aux expériences de terrain en économie

Le principal intérêt d'une expérience de terrain est de permettre l'évaluation de l'effet d'une variation extérieure (un traitement ou un programme) en vue d'établir une causalité :

« Distinct from traditional empirical economics, field experiments provide an advantage by permitting the researcher to create exogenous variation in the variables of interest, allowing us to establish causality rather than mere correlation ». (Reiley et List, 2007, 2)

Les expériences de terrain offrent ainsi la possibilité de mettre en évidence une causalité. A première vue, une expérience de terrain est caractérisée par le fait qu'elle se réalise à l'extérieur ; alors qu'une expérience de laboratoire est rattachée à un espace précis et défini, le laboratoire. Par contraste, dans une expérience de terrain, celui-ci est mouvant : il ne renvoie pas à un espace clairement délimité. La seconde difficulté concerne le contenu de l'expérience : teste-on la même chose dans une expérience de terrain et dans une expérience de laboratoire ? L'espace est différent mais la forme d'une expérience de terrain est-elle similaire à une expérience de laboratoire ? Afin de déterminer précisément ce qu'est une expérience de terrain, je développerai tout d'abord la typologie qu'offrent Harrison et List (2004) (3.1.1). Ensuite, dans le but de saisir la place de la randomisation au sein de ces expériences, je m'intéresserai à l'évolution des expériences de terrain en économie à travers l'histoire qu'en proposent Levitt et List (2008) (3.1.2).

3.1.1 Qu'est-ce qu'une expérience de terrain en économie ?

Afin de distinguer les expériences de terrain de celles de laboratoire, Harrison et List (2004) retiennent six critères. Le premier s'attache aux sujets qui participent à l'expérience, qui peuvent être de deux types : « les sujets standards » et les « sujets non standards ». Les premiers sont les étudiants en économie, ce sont souvent eux qui participent le plus aux expériences de laboratoire ; les seconds sont des « extérieurs » (*outdoors*), c'est-à-dire des gens de dehors, extérieurs au monde académique. Ce sont souvent ces gens de dehors qui

caractérisent les expériences de terrain même si : « *the use of nonstandard subjects should not automatically qualify the experiment as a field experiment* ». (Harrison et List, 2004, 1012) Un deuxième critère s'attache à la nature de l'information que l'individu apporte : « *In the field, subjects bring certain information to their trading activities in addition to their knowledge of the trading institution.* » (Harrison et List, 2004, 1012) La nature des produits est un troisième critère : dans les expériences de terrain, le plus souvent, ce sont des produits « réels » qui sont mis en jeu et non des produits virtuels comme dans les expériences de laboratoire. Le quatrième critère est la nature de la tâche que l'individu doit réaliser au sein de l'expérience. Le cinquième critère concerne les paris que doit prendre l'individu lorsqu'il participe à une expérience. Le sixième critère est déterminé par l'environnement de l'expérience, par exemple les expériences de laboratoires se réalisent dans des salles de classes, ce qui n'est pas le cas des expériences de terrain.

Harrison et List proposent ensuite une terminologie des expériences : expérience de laboratoire ordinaire (1), expérience de terrain artefactuelle (2), expérience de terrain encadrée (4) et une expérience de terrain naturelle (5). Les six critères définis précédemment vont permettre de déterminer et de séparer ces différents types d'expériences. Une expérience de laboratoire ordinaire est une expérience dans laquelle les participants sont des étudiants (« *standard subject* »), l'encadrement est abstrait et les règles sont imposées. Une expérience de terrain artefactuelle est la même chose qu'une expérience de laboratoire ordinaire mais avec des participants non standards. Une expérience de terrain encadrée est identique à une expérience de terrain artefactuelle mais avec un contexte où les tâches, les produits et les informations peuvent être utilisés librement. Une expérience de terrain naturelle est similaire à une expérience de terrain encadrée mais l'environnement est l'environnement naturel de l'individu où rien n'est changé. La typologie d'Harrison et List (2004) permet donc de distinguer clairement les expériences de terrain des expériences de laboratoire, et offre une définition précise de ces dernières.

3.1.2 Évolution des expériences de terrain en économie

Steven Levitt et John List (2008) distinguent trois périodes dans les expériences de terrain. La première période est traversée par les travaux de Ronald Fisher et Jerzy Neyman (1920-1930). Ceux-ci sont rarement considérés comme des travaux économiques mais leurs expériences aident à répondre à des questions économiques liées avec la productivité agricole.

Les expériences de Fisher sont connues : « *to conceptualize randomization as a key element of the experimental method* » (Levitt et List, 2008, 3). J'ai développé cette période dans la deuxième partie de ce chapitre, lorsque je me suis intéressée au cadre de Fisher. La deuxième période (deuxième moitié du XX^{ème} siècle) est marquée par une série d'expériences sociales à grande échelle menées par des agences gouvernementales. Ces expériences ont eu une forte influence sur les politiques. Je m'intéresserai particulièrement à cette deuxième période à la fin de ce chapitre. La troisième période (ces dix dernières années) est caractérisée par trois types d'expériences de terrain : artefactuelles, encadrées et naturelles. Cette troisième période renvoie directement aux expériences randomisées en économie du développement conduites par le J-PAL. Je développerai précisément cette période dans le chapitre suivant⁶⁴. Levitt et List s'intéressent ensuite à la génération future d'expériences de terrain. Ils soulignent que la plupart des expériences de terrain ont été réalisées en partenariat avec l'Etat ou avec des organisations à but non lucratif, comme des organisations non gouvernementales. Selon eux, cela n'est pas surprenant car les expériences sociales qui ont eu lieu ces dix dernières années, par leur nature et les questions qu'elles soulevaient, rendaient l'intervention des gouvernements cruciale et parce que les ONG ont explicitement l'ambition d'améliorer le bien-être social. A l'inverse, peu d'expériences faites en lien avec le secteur privé ont eu lieu. Pour Levitt et List (2008), il existe beaucoup de questions économiques centrales qui pourraient être étudiées expérimentalement avec l'aide des firmes, comme l'impact d'une asymétrie d'informations, ou encore comment les marchés répondent à des chocs. C'est ici que réside l'avenir des expériences de terrain en économie pour Levitt et List⁶⁵. Selon Levitt et List, cette forme d'expériences de terrain a déjà émergé, et ce dans deux domaines : pour mesurer la réponse au prix du consommateur (1) et pour améliorer notre compréhension des firmes (2).

3.2 Des expériences de terrain aux expériences sociales

Ce qui semble, aujourd'hui, regrouper les expériences sociales de terrain, est le fait qu'elles cherchent toutes à évaluer une politique publique. L'expérimentation permet d'évaluer rigoureusement les effets d'une intervention et rend observables des dimensions qui

⁶⁴ Puisque le but est ici de comprendre ce qui a précédé l'utilisation massive de la randomisation en économie du développement.

⁶⁵ Ils soulignent, d'ailleurs, que cette forme d'expériences de terrain a déjà commencé à émerger, et ce dans deux domaines : pour mesurer la réponse au prix du consommateur et pour améliorer la compréhension des processus de production au sein des firmes.

ne l'étaient pas auparavant. C'est cet objectif que le J-PAL souhaite traduire au sein de ces expériences. C'est pourquoi, il est important de revenir à la définition des expériences sociales de terrain afin de saisir au mieux le second objectif du J-PAL. Je tenterai, tout d'abord, de définir les expériences sociales de terrain, afin d'en saisir les enjeux (3.2.1). Puis, je chercherai à illustrer ces dernières autour de deux exemples (3.2.2), cela me permettra de traiter explicitement de la deuxième période définie par Levitt et List (2004).

3.2.1 Les expériences sociales comme des expériences à vocation politique

Greenberg et Shroder (2004) définissent une expérience sociale selon quatre caractéristiques : une assignation aléatoire (randomisation) (1), une intervention politique (2), un suivi dans la collection de données (3), et une évaluation (4). L'objectif de ces expériences est de « parler aux décideurs politiques » (Greenberg et Shroder, 2004,4). Cette expérience randomisée (1) a pour but d'évaluer (4) une intervention politique (2) après avoir collecter des données (3). L'assignation aléatoire est donc caractéristique des expériences sociales. L'objectif étant d'évaluer les effets d'une intervention politique, l'évaluation et l'intervention publique sont au premier plan. Cette évaluation, réalisée à travers une expérience sociale, s'adresse directement au décideur politique et vise à lui fournir des éléments de décisions. Mais la définition des expériences sociales offerte par Greenberg et Shroder exclue les expériences de terrain faites en économie : « *Taken together, the second and third features of our definition exclude random-assignment experiments in medicine, psychology, economics, criminology, and education* ». (Greenberg et Shroder, 2004, 4) Greenberg et Shroder n'explicitent pas ce choix. Néanmoins, Ferber et Hirsch (1982) définissent les expériences sociales de façon similaire à celle de Shroder et Greenberg. La définition de Ferber et Hirsch (1982) est explicitement une définition des expériences sociales en économie. Ils les définissent de la façon suivante:

« *A publicly funded study that incorporates a rigorous statistical design and whose experimental aspects are applied over a period of time to one or more segments of a human population, with the aim of evaluating the aggregate economic and social effects of the experimental treatments.* » (Ferber et Hirsch, 1982, 7, souligné par moi)

Une expérience sociale menée en économie peut donc être définie autour de quatre caractéristiques :

1. C'est une étude financée par de l'argent public
2. C'est une étude fondée sur une structure statistique rigoureuse
3. Elle doit être appliquée ou faite sur des humains
4. Elle évalue les effets d'un traitement sur des dimensions économiques et sociales.

On retrouve ainsi certains des éléments de définition présents chez Shroder et Greenberg. La nouveauté apportée par Ferber et Hirsch réside dans le fait que ces expériences sont financées par de l'argent public. Cela signifie que ce sont des expérimentations coûteuses car de grande envergure⁶⁶. Les expériences sociales donnent aux économistes la possibilité de mesurer les effets d'une politique en rapport avec le comportement des individus. Ferber et Hirsch (1982) soulignent qu'il existait d'autres méthodes utilisées dans le passé pour mesurer de tels effets mais que tous les effets n'étaient pas observables :

« Economist have long faced the problem of how to measure the effect of changes in policy variables on the behavior of economic units and particularly of changes that have been merely contemplated (...) Social experimentation, however, does just that. It seeks to measure the effects of changes in policy variables by applying these changes to human population under conditions of controlled experimentation similar used in the physical and biological sciences. » (Ferber et Hirsch, 1982, 1-2)

L'expérimentation permet de rendre visibles des éléments que l'observation ne permet pas de voir. A travers l'existence d'un groupe de contrôle et d'un groupe de traitement, les expériences sociales rendent possible l'observation des effets de politiques publiques et offrent la possibilité de les mesurer. C'est ce qui justifie, selon Ferber et Hirsch (1982), l'utilisation de ces expériences en économie.

3.2.2 Les expériences sociales, des évaluations à grande échelle : une illustration

Une des expériences sociales phare en économie est l'expérience du *The New Jersey Income Maintenance*, menée aux États-Unis dans les années soixante. Cette expérience est due à l'initiative d'une doctorante du MIT, qui fut la première à utiliser une assignation aléatoire au sein de son expérience :

⁶⁶ « *large public funds are needed because the cost per experimental unit tends to be so high* ». (Ferber et Hirsch, 1982, 8). Je reviendrai largement sur ce point lorsque j'illustrerai ces expériences sociales en économie.

« Heather Ross, then a PhD student in MIT's economics department, was visiting the Brookings Institution and wrote a piece titled "A Proposal for Demonstration of New Techniques in Income Maintenance", in which she suggested a random assignment social experiment to lend insights into the debate. » (Levitt et Litt, 2008, 12)

L'expérience a été conduite en partenariat avec l'*Institute of Research on Poverty* à l'université du Wisconsin et le laboratoire *Mathematica* de l'université de Princeton. L'expérience commença en 1968 dans quatre endroits du New Jersey et de la Pennsylvanie et constitua la thèse de Heather Ross. L'idée de l'expérience était de tester l'effet d'un impôt négatif⁶⁷, thèse défendue par Milton Friedman. L'expérience dura de 1968 à 1972 et se fit sur un échantillon de 1357 ménages à bas revenus :

« The negative income tax rates works as follows. Assume that John is randomly inserted into the 100% guaranteed income (\$3300), 50% negative tax rate treatment. What this means is that when the policy binds, for each \$1 that John's family earns on its own, they receive \$0,50 less in federal benefits. Thus, if John's family earns \$2000 in year one, they would receive \$1000 less in program benefits, or \$2300, resulting in a total income of \$4300. In this case, if in any year John's family earns \$6600 or more, program benefits are zero. » (Levitt et Litt, 2008, 13)

Tous les trois mois, les participants avaient aussi à répondre à des questionnaires. Les résultats de l'expérience montrent, en partie, que le programme d'impôt négatif offre des incitations positives à travailler, ou du moins aucune incitation négative⁶⁸. Cette expérience est la première expérience sociale de grande échelle conduite aux États-Unis.

Une autre expérience sociale caractéristique en économie fut réalisée en Suède en 1975 et visait à améliorer les services qu'offraient les agences pour l'emploi (Greenber et Shroder, 2004). Une agence pour l'emploi, dans une petite ville, bénéficia d'un renforcement de son personnel pendant trois mois. 410 chômeurs furent séparés en deux groupes : un groupe de traitement et un groupe de contrôle. Le groupe de contrôle reçut un service normal, c'est-à-dire qu'il a pu bénéficier d'une moyenne d'une heure et demie des services de l'agence à l'emploi durant l'expérience. Quant au groupe de traitement, il a bénéficié des services de l'agence à l'emploi dans une moyenne de sept heures et demie durant l'expérience, ce qui leur a permis de travailler davantage sur leurs problèmes personnels. Les résultats de l'expérience

⁶⁷ Le concept d'impôt sur le revenu négatif, développé par Milton Friedman (1962), part de l'idée qu'au-dessus d'un certain seuil, un certain revenu, on paye des impôts et qu'en dessous de ce seuil on paye un impôt négatif. Autrement dit, on reçoit une aide sociale. C'est cette idée de seuil qui rend clair le terme négatif de l'impôt.

⁶⁸ Cette mesure n'a pour autant jamais été appliquée aux États-Unis.

ont été jugés positifs : les chômeurs qui ont obtenu un travail après l'expérience ont été influencés favorablement par les services proposés par l'agence à l'emploi

Depuis ces expériences, il y a eu 235 expériences sociales, évaluant des politiques publiques, s'intéressant à la santé et à la protection sociale (Greenberg et Shroder, 2004)⁶⁹. Ces expériences ont en commun l'utilisation de la randomisation pour évaluer une politique publique. C'est exactement ce que cherche à faire, aujourd'hui, le J-PAL en économie du développement.

4. Conclusion

Retracer l'histoire de la randomisation, depuis sa genèse jusqu'à la période précédant les expériences du J-PAL, permet de poser le cadre historique d'une telle méthode. C'est d'abord à travers la psychologie qu'une telle approche a été introduite. L'intégration d'une dimension aléatoire visait à rendre rigoureux le cadre expérimental. Ce cadre sera défini précisément par Fisher, puis popularisé en médecine par les protocoles d'essais cliniques médicaux. Ces derniers sont caractérisés par une importante rigueur statistique et par un protocole expérimental précis. C'est ce caractère « scientifique » que la randomisation du J-PAL souhaite importer.

L'histoire des expériences de terrain en économie nous permet de mettre en évidence le second aspect qui caractérise les expériences du J-PAL : l'évaluation des politiques de développement. De plus, cela nous permet de situer de telles expériences au sein d'une importante diversité. Pour cela, la typologie d'Harrison et List est particulièrement éclairante, ainsi que l'histoire que produisent List et Levitt. Les expériences du J-PAL possèdent donc une dimension sociale importante, cela est traduit par leur vocation d'évaluer des politiques – même si, on le verra, les expériences du J-Pal s'éloignent des premières expériences sociales qui avaient pour objectif d'évaluer des politiques à grande échelle. C'est donc la rencontre entre ces deux aspects de l'histoire : le caractère scientifique des expériences randomisées et la visée politique des expériences sociales qui font le socle de la randomisation du J-PAL.

⁶⁹ Ferber et Hirsch (1982) produisent aussi une revue de ces différentes expériences. Pour cela, ils reprennent la classification des expériences sociales d'Alice Rivlin (1974). Elle classe les expériences sociales selon quatre catégories : les expériences visant des marchés entiers (1), celles impliquant un changement dans la fonction de production des services publics (2), celles qui s'attachent aux effets d'une politique publique spécifique, et celles s'intéressant aux effets d'un changement microéconomique (4).

Chapitre 2 : Randomisation, économie du développement et le J-PAL : l'ambition de la preuve

0. Introduction

« Tout comme les recherches menées par le passé en économie du travail, de la santé et de l'éducation, la recherche expérimentale en économie du développement tente de résoudre la difficulté qu'il y a à identifier les effets d'un programme face à des chaînes causalité multiples et complexes. (...) 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, 692-693)

L'approche du J-PAL en économie du développement a pour objectif de produire des preuves, l'accent est donc porté sur la méthode utilisée. Comme cela a déjà été évoqué, la randomisation, par son protocole expérimental et son cadre statistique rigoureux, permet d'identifier les effets d'un programme de développement de façon fiable ; et en ce sens, permet de produire des preuves. A l'inverse, d'autres méthodes permettent d'évaluer les effets d'un programme de développement mais de façon moins fiable car certains biais persistent. Pour le J-PAL, la randomisation est donc considérée comme la méthode à utiliser. Le J-PAL tente alors d'offrir un rayonnement particulier à une telle méthode. Le cadre institutionnel et le fonctionnement d'un tel laboratoire tendent à participer très grandement à ce rayonnement. Le J-PAL en collaborant avec de nombreux partenaires de terrain, comme des entreprises ou des gouvernements, occupe une place importante sur le terrain. De plus, la présence de nombreuses antennes du J-PAL sur chaque continent participe à favoriser l'utilisation de la randomisation en économie du développement. C'est donc à la fois, la forte validité interne de la randomisation et le fonctionnement du J-PAL qui participent à faire de cette méthode un des outils les plus utilisés pour lutter contre la pauvreté.

L'objectif de ce chapitre est double. Il cherche à la fois à présenter le cadre institutionnel du J-PAL ainsi que la manière dont ce laboratoire utilise la randomisation. Ce sont donc ces deux dimensions, la méthode spécifique du J-PAL et son fonctionnement en réseau, qui permettent, selon moi, de définir un tel laboratoire, ce que je m'attacherai à faire dans ce chapitre. La première partie de ce chapitre cherche à définir le cadre institutionnel de ce laboratoire. Pour cela, je développerai son histoire, de sa création en 2003 à son essor en 2007. Il est aussi nécessaire de saisir la structure du J-PAL, afin de comprendre son fonctionnement. C'est pourquoi je définirai précisément l'organisation du J-PAL en

différentes antennes. Enfin, le J-PAL entretient un lien étroit avec le terrain, que je m'attacherai à définir. La deuxième partie de ce chapitre vise à présenter l'utilisation de la randomisation par le J-PAL. En ce sens, cette partie insiste sur le statut de *gold standard* méthodologique de la randomisation en la comparant à d'autres méthodes permettant de penser un contrefactuel. Enfin, la troisième partie de ce chapitre s'intéressera à la mise en place de la randomisation sur le terrain. Ces trois parties me permettront de définir précisément le J-PAL autour de son ambition première : produire des preuves.

1. Le cadre institutionnel du J-PAL

Le *Abdul Latif Jameel Poverty Action Lab* (J-PAL) domine, aujourd'hui, l'économie du développement et ses pratiques. Il s'apparente davantage à un réseau de chercheurs qu'à un laboratoire. Son organisation lui permet de jouer un rôle prépondérant en économie du développement. Le J-PAL est divisé en plusieurs antennes ; il dispose d'une antenne sur chaque continent. Ce laboratoire travaille également en lien étroit avec de nombreux centres de recherches ; ce qui permet aux chercheurs d'implanter leur méthode dans de nombreuses universités – et souvent dans les meilleures. La façon de procéder du J-PAL est de travailler « main dans la main » avec des partenaires de terrain. Tout cela explique le poids croissant de la randomisation en économie du développement. Je présenterai, tout d'abord, l'histoire du J-PAL (1.1), puis, j'explicitai sa structure organisationnelle (1.2) ; enfin, je m'intéresserai aux différents partenaires du J-PAL (1.3). Cela donnera quelques éléments d'explication à l'essor fulgurant du J-PAL et permettra de saisir au mieux le cadre institutionnel d'un tel laboratoire.

1.1 Histoire du J-PAL

Le J-PAL est créé en 2003, mais c'est à partir de 2007 que ce dernier connaît un important tournant, et qu'il devient l'un des laboratoires incontournables en économie du développement. Son rayonnement mondial est dû au fait qu'il bénéficie d'une antenne sur chaque continent. Il est alors possible de distinguer deux principales étapes au sein de l'histoire du J-PAL. La première est celle de sa création et de ses débuts (1.1.1) et la seconde celle de l'essor d'un tel laboratoire, celui-ci ayant pris une dimension internationale (1.1.2).

1.1.1 Les débuts (2003-2007)

Le Poverty Action Lab est créé en 2003 par trois professeurs du *Massachusetts Institute of Technology* (MIT) : Abhijit Banerjee, Esther Duflo et Sendhil Mullainathan. Abhijit Banerjee a suivi une formation en économie à l'université de New Dehli et fait sa thèse à Harvard sur le thème de l'information en économie. Il sera le directeur de thèse au MIT d'Esther Duflo, qui après des études d'histoire à l'École Normale Supérieure et sur les conseils de l'économiste français Thomas Piketty commence une maîtrise d'économie puis un DEA au DELTA⁷⁰. Sa thèse porte sur trois essais empiriques en économie du développement et vise à évaluer trois programmes de développement⁷¹. C'est ici que les évaluations randomisées en économie du développement trouvent leurs prémisses. Enfin, Sendhil Mullainathan⁷², après une thèse en économie à l'université d'Harvard, est engagé en 1998 comme *junior professeur* puis comme professeur au MIT qu'il quittera en 2008; abandonnant alors, à cette même période, la direction du J-PAL.

En 2004, Rachel Glennester entre à la direction du J-PAL. Après des études de philosophie, de politique et d'économie à l'université d'Oxford, Rachel Glennester fait une thèse en économie à l'université de Londres, puis devient *development associate* à Harvard et *senior economist* au FMI. Le *Poverty Action Lab* est renommé le *Abdul Latif Jameel Poverty Action Lab* en 2005, à la suite de trois dons conséquents faits par Mohammed Abdul Latif Jameel⁷³ : « grâce à son soutien, le J-PAL est capable de se développer de façon significative pour les années à venir »⁷⁴. Le J-PAL commence à former d'autres chercheurs dans le cadre de programmes d'entraînement (*training programs*) dont les deux premiers commencent cette même année à Cambridge, aux États-Unis et à Chennai en Inde. Soixante et une personnes sont formées aux évaluations randomisées et vingt évaluations ont lieu dans dix différents

⁷⁰ Le DELTA réunit le centre d'économie quantitative et comparative de l'École des hautes études en sciences sociales et le laboratoire d'économie politique de l'école normale supérieure. En 2005 le DELTA fusionne avec le CERAS afin de devenir une unité mixte de recherche (UMR). Cette section de recherche institue l'École d'Économie de Paris. Thomas Piketty dirige l'École d'Économie de Paris de sa création jusqu'en 2007. Depuis 2013, Pierre-Yves Geoffard en est le directeur.

⁷¹ « *Three Essays in Empirical Development Economics* », soutenue en 1999.

⁷² Sendhil Mullainathan est une figure importante de l'économie comportementale, surtout pour ses travaux en finance puis en économie du développement.

⁷³ Mohammed Abdul Latif Jameel est un homme d'affaire. Il est désormais président de l'entreprise Abdul Latif Jameel qui distribue les automobiles Toyota à travers le monde. Le laboratoire a été renommé par le nom de son père Abdul Latif Jameel. Les fonds offerts par Mohammed Abdul Latif Jameel ont permis d'instaurer une bourse d'étude finançant les études d'étudiants du moyen Orient ou d'Asie qui n'aurait pu étudier au MIT sans cette aide financière.

⁷⁴ *The History of J-Pal*, <http://www.povertyactionlab.org/>, (dernière consultation le 12 septembre 2013), avec ma traduction.

pays. En 2006, le J-PAL reçoit le soutien financier de la fondation William et Flora Hewlett Packard, de la fondation John et Catherine MacArthur, de la fondation Nike, et de la fondation Doug B. Marshall Junior et de Russ Siegelman. Il compte alors treize professeurs affiliés, trente évaluations sont en cours dans douze pays et quatre-vingt-deux personnes ont été formées à ses méthodes. En trois ans, le nombre de chercheurs associés au J-PAL a été multiplié par quatre et plusieurs évaluations commencent à être réalisées à travers le monde.

1.1.2 L'essor du J-PAL (2007-2011)

C'est en 2007 que le J-PAL connaît un tournant particulier, lorsqu'une nouvelle antenne du J-PAL est créée au sud de l'Asie, à Chennai avec le soutien de la fondation Mulago⁷⁵. Une autre antenne de formation est constituée en Afrique, à Abuja au Nigéria. Le J-Pal compte alors dix-huit professeurs affiliés, cinquante évaluations sont en cours dans quinze pays et deux cents personnes ont été formées à ses méthodes. En 2008, une autre antenne du J-PAL est établie en Europe, après « deux jours de conférences organisées par le ministère français du travail et inaugurée par Martin Hirsch, le Haut-commissaire aux solidarités actives contre la pauvreté »⁷⁶. Son siège est à Paris et le J-PAL Europe dépend de l'École d'Économie de Paris (EEP). Un centre de formations français est créé la même année à l'EEP. Le J-PAL est désormais constitué de trente professeurs affiliés, soixante-dix évaluations sont en cours dans vingt-deux pays et deux cents personnes ont été formées à ses méthodes. Le J-PAL constitue en 2009 un *policy group* dont l'objectif est de transmettre les résultats des évaluations disponibles aux ONG, aux instances internationales et aux décideurs politiques. Une autre antenne du J-PAL est établie au Chili grâce à l'aide financière de donateurs locaux. En 2010, une antenne du J-PAL est ouverte en Afrique du Sud. Il est désormais présent sur chaque continent. La même année, un comité de direction est créé. Il est composé de Bruno Crépon du J-PAL Europe, Benjamin Olken du J-PAL global, Sebastian Galiani du J-PAL Amérique latine, et des trois directeurs du J-PAL global, Abhijit Banerjee, Esther Duflo et Rachel Glennester et enfin de Iqbaq Dhaliwal pour la direction des politiques du J-PAL global.

Finalement, le J-PAL compte, en 2011, cinquante-neuf professeurs affiliés, vingt-neuf évaluations sont en cours dans quarante-trois pays et huit cent cinquante et une personnes ont été formées à ses méthodes. De nombreux projets et associations avec d'autres partenaires

⁷⁵ La fondation Mulago cherche à continuer le travail de Rainer Arnhold, pédiatre en Californie dévoué à l'humanitaire, pour plus d'informations voir : <http://www.mulagofoundation.org/>

⁷⁶ *The History of J-Pal*, <http://www.povertyactionlab.org/>, (dernière consultation le 12 septembre 2013), avec ma traduction.

sont mis en place au cours de cette même année. Avec le soutien de la Banque mondiale, le *Center for Learning in Evaluation and Results* (CLEAR) est fondé au J-PAL Asie du sud. La même année, le gouvernement chilien demande au J-PAL d'identifier les principaux défis en matière de politique sociale. Le J-PAL crée alors la *Compass Commission* qui se charge de déterminer ces défis et de proposer des programmes sociaux évaluables par randomisation. En matière de politique économique et de gouvernance, le J-PAL élabore avec le soutien de la fondation de William et de Flora Hewlett, la *Governance Initiative*. L'objectif d'une telle initiative est de financer des évaluations des programmes les plus prometteurs en matière de gouvernance. Le *Urban Service Associative* est créé au J-PAL Afrique avec le soutien de la fondation de Bill et de Melinda Gates. L'objectif de cette association est d'évaluer des solutions pour répondre aux problèmes d'eau potable, des sanitaires et des services d'hygiène dans les villes d'Afrique et d'Asie. Esther Duflo dirige cette initiative avec Sebastian Galiani et Mushfiq Mobarak.

Le J-PAL a donc mis en œuvre de nombreux projets et s'est associé à des partenaires très divers. Le nombre d'évaluations et de professeurs affiliés a fortement augmenté. En moins de dix ans, il est passé de trois chercheurs à plus de cinquante, et près de neuf cents personnes ont été formées à ses méthodes. L'ascension du J-PAL est fulgurante. Cette ascension est due à un soutien financier important provenant de nombreuses institutions, à un travail étroit avec les partenaires de terrain, et à la création de réseaux puissants avec d'autres centres de recherches utilisant la randomisation. Avant de développer en détail ces relations de réseaux, je vais m'attacher à définir la structure organisationnelle du J-PAL.

1.2 Structure du J-PAL

Le J-PAL Global est sans conteste l'antenne la plus structurante du J-PAL. Ses directeurs dans leur ensemble, sont tous rattachés à cette antenne. Son conseil d'administration est composé par les directeurs de chaque antenne - au total vingt chercheurs. Ce conseil définit la stratégie du J-PAL à travers le monde, il est « responsable de l'orientation stratégique générale et contribue au développement des activités de recherche, de formation et de diffusion des résultats »⁷⁷. Le J-PAL Global structure le fonctionnement de ce réseau de chercheurs, même si chaque antenne participe, au travers du conseil d'administration, à la

⁷⁷ <http://www.povertyactionlab.org/fr/j-pal/organigramme/conseil-dadministration>, dernière consultation le 9 juillet 2012.

ligne stratégique de ce réseau. Je présenterai, tout d'abord, les différentes antennes du J-PAL (1.2.1) ; puis, je porterai une attention particulière à l'antenne du J-PAL en France, car cette dernière soulève un certain nombre de spécificités (1.2.2).

1.2.1 Le J-PAL Global et ses différentes antennes

Chacune des antennes du J-PAL est structurée de la même manière. Elles sont séparées en trois catégories : les chercheurs affiliés, une équipe administrative et une équipe de recherche. Le J-PAL Global compte vingt-sept professeurs affiliés. Ces professeurs sont tous rattachés à des universités américaines ; souvent les plus prestigieuses, comme Harvard, Yale, Stanford, Berkeley, ou encore le MIT. Neuf⁷⁸ parmi ces vingt-sept professeurs affiliés sont en poste à Harvard. Cinq des professeurs affiliés sont rattachés au MIT (Michael Greenstone, Tavneet Suri, Abhijit Banerjee et Esther Duflo), université centrale du J-PAL. L'équipe de recherche du J-PAL Global est composée de tous ces professeurs affiliés ainsi que de sept managers de recherches qui aident les différents chercheurs à conduire leurs évaluations. L'administration du J-PAL est constituée de ces trois directeurs (Abhijit Banerjee, Esther Duflo et Rachel Glennester) ainsi que de Benjamin Olken. Les trois directeurs bénéficient, chacun, d'un assistant de recherche. L'administration comprend ensuite les chercheurs qui se chargent des politiques publiques, il y a dès lors trois possibilités : être directeur des politiques publiques, en être responsable ou encore en être analyste. Les directeurs des politiques publiques travaillent avec les décideurs politiques dans le but de les aider à identifier de nouveaux domaines d'évaluation⁷⁹.

Les autres antennes bénéficient d'une structure similaire mais plus petite. L'équipe de recherche du J-PAL Afrique est composée de deux chefs de projets (Raissa Fabregas et Bryan Plummer) s'occupant de coordonner les ressources sur les expériences de terrain, et de deux analystes de recherches (Megan Blair qui s'intéresse particulièrement à la micro-finance et Rebecca Metz qui se centre sur l'éducation). Il est rattaché à l'université de Cap Town en Afrique du Sud, et dirigé par deux des directeurs du J-PAL Global : Rachel Glennester et Abhijit Banerjee. Le J-PAL Amérique latine et Caraïbes est, lui, composé de huit chercheurs

⁷⁸ Ces neuf chercheurs sont : Nava Ashraf, Marianne Bertrand, Shawn Cole, Rema Hanna, Asim Khwaja, Michael Kremer, Dan Levy, Sendhil Mullainathan, et Rebecca Thornton

⁷⁹ L'objectif des responsables des politiques publiques est de mener des analyses coûts/bénéfices, de rédiger ensuite une synthèse afin de rendre compte aux décideurs politiques des leçons tirées des évaluations. Le travail des analystes des politiques publiques est similaire à celui des responsables. Leur objectif est d'analyser les résultats et d'écrire des résumés généraux à destination des ONG ou des décideurs politiques.

affiliés, tous provenant d'universités d'Amérique latine, à l'exception de deux chercheurs en provenance d'universités américaines - Sebastian Galiani (*Washington University*), Marcos Rangel (*University of Chicago*). Il convient aussi d'ajouter que cette antenne du J-PAL n'est rattachée à aucune université. Le J-PAL Asie du sud est étroitement lié à l'*Institute for Financial Management and Research*, regroupant un ensemble de centres de recherches qui promeuvent l'utilisation de la randomisation dans la valorisation des petites entreprises. Cette antenne est composée de deux chercheurs affiliés – Sharon Barnhardt (*Institute for Financial Management and Research*) et Raghabendra Chattopadhyay (*Indian Institute of Management*).

1.2.2 La Spécificité du J-PAL Europe

Le J-PAL Europe est rattaché à l'École d'Économie de Paris. Il entretient, d'ailleurs, avec cette institution une relation toute particulière. Sur vingt et un professeurs associés au J-PAL Europe, cinq chercheurs sont rattachés à l'École d'Économie de Paris. Un deuxième grand pôle de recherche français associé au J-PAL Europe est le CREST (Centre de Recherche en Économie et Statistique), avec deux professeurs associés au J-PAL Europe. Le dernier grand pôle de recherche associé au J-PAL Europe est la *London School of Economics*, avec trois professeurs associés. Six autres chercheurs associés sont rattachés à des universités européennes. La spécificité du J-PAL Europe est due aux liens étroits entre ce laboratoire et le haut-commissariat aux solidarités actives. Lorsque, Martin Hirsch devient commissaire aux solidarités actives en 2007, il envisage un ensemble de politiques en faveur de la jeunesse. L'objectif était de réduire d'un tiers la pauvreté en France. Il choisit de soumettre toutes ces politiques à une évaluation avant qu'elles ne soient mises en place. Le J-PAL est alors en charge d'une partie de ces évaluations. Le programme le plus connu évalué par le J-PAL lors de cette commission est la « mallette des parents »⁸⁰, financée par le ministère de la jeunesse. On retiendra, aussi, que le J-PAL n'a pas souhaité évaluer le projet central de Martin Hirsch, le « Revenu de Solidarité Active » (RSA). En effet, le RSA a été mis en place avant d'être évalué, ce qui ne permettait pas au J-PAL de mener une évaluation randomisée, comme les chercheurs souhaitaient le faire. Martin Hirsch et Esther Duflo apparaissent auprès du grand public comme étant les deux plus grandes figures de la lutte contre la pauvreté en France⁸¹. Le

⁸⁰ Voir (Francesco, Gurgand, Guyon et Maurin, 2010).

⁸¹ A titre d'illustration, *Télérama* titra en février 2012 : « Esther Duflo et Martin Hirsch réinventent le combat contre la pauvreté »

partage des tâches est claire : Martin Hirsch offre le budget et les politiques tandis que le J-PAL les évalue.

Le J-PAL bénéficie d'une structure qui lui permet de travailler avec un nombre de chercheurs important à travers le monde. Son fonctionnement en antenne lui donne la possibilité d'avoir des chercheurs qui connaissent les spécificités de leur pays, ou de leur continent. De plus, le fait que ces chercheurs soient rattachés à un pays ou un continent permet au J-PAL d'élargir son réseau avec des partenaires de terrain.

1.3 La randomisation et le terrain : une articulation autour de trois partenaires

Pour mener des évaluations, le J-PAL travaille en lien étroit avec trois partenaires :

« Pour un projet type, il faut toujours trois éléments. Il faut un partenaire de terrain qui a envie de mettre en œuvre le projet et qui désire une évaluation, en dépit du caractère contraignant de cette démarche. Ce partenaire de terrain peut être soit un gouvernement, soit une ONG, soit une entreprise privée. Ensuite, il faut un chercheur ou une équipe de chercheurs qui s'intéressent suffisamment au projet pour avoir envie de le mettre en œuvre. Et, enfin, il faut quelqu'un pour le financer ». (Duflo dans Mayneris, 2009, 3)

Le réseau du J-PAL n'est pas uniquement constitué de chercheurs rattachés aux différentes antennes du J-PAL, puisque ce réseau s'élargit à d'autres centres de recherches, à des partenaires de terrain et à des financiers. Au total, le J-PAL compte deux cent soixante-douze partenaires, tous difficiles à catégoriser⁸². Le J-PAL ne propose qu'une liste non exhaustive de ces différents partenaires, mais ne les associe à aucun rôle particulier – sans doute parce que ce rôle varie d'une évaluation à l'autre. J'ai tout de même cherché à regrouper les différents partenaires du J-PAL autour des trois rôles définis par Duflo. Je présenterai donc, tout d'abord, les différents partenaires de terrain du J-PAL (1.3.1) ; puis, les financiers (1.3.2) ; et enfin, les centres de recherches (1.3.3).

1.3.1 Les partenaires de terrain

Les partenaires de terrain les plus représentatifs sont les organisations non gouvernementales. Elles sont plus de soixante à avoir évalué des programmes en partenariat

⁸² Par exemple, un gouvernement peut tout aussi bien jouer le rôle d'un financier ou être un partenaire de terrain. Il en va de même pour les entreprises.

avec le J-PAL. Pour n'évoquer que deux d'entre elles, je retiens les ONG Seva Mandir et Pratham⁸³. Seva Mandir⁸⁴ est située à Udaipur en Inde et a été créée par Sinha Mehta en 1966. Cette ONG se concentre sur l'autonomisation des femmes, l'éducation primaire, la santé et la protection des enfants. Seva Mandir n'est pas une ONG locale, elle bénéficie d'une structure importante avec trois cents employés à plein temps et cent volontaires qui participent chaque année à différents projets. Seva Mandir a collaboré avec le J-PAL sur cinq évaluations⁸⁵ concernant la santé et l'éducation, à chaque fois en Inde. Pratham est une ONG dont l'objectif principal est l'accès à l'éducation primaire pour les enfants défavorisés. Cette organisation a été créée en 1994 par Madhav Chavan avec le soutien de l'UNICEF. Elle a depuis participé avec le J-PAL à huit évaluations. Le J-PAL a aussi bien travaillé avec des gouvernements de pays développés, comme la France ou les États-Unis, que des pays en développement comme le Kenya, le Bangladesh, l'Ouganda, l'Inde⁸⁶ ou encore la Zambie. Tout comme les gouvernements, les entreprises qui collaborent avec le J-PAL peuvent avoir un double rôle : elles peuvent financer l'évaluation ou être à l'initiative du projet. Une dizaine d'entreprises⁸⁷ ont collaboré avec le J-PAL. Sur deux cent soixante-douze partenaires, les entreprises ne représentent donc pas une part importante comparativement aux gouvernements ou aux ONG⁸⁸.

1.3.2 Les financiers

Au sein des partenaires du J-PAL le plus difficile, sans doute, est de déterminer qui finance les évaluations. Cependant, en recoupant la liste des partenaires avec la liste des expériences qui ont été menées par le J-PAL, il est tout de même possible d'en donner un aperçu. Au sein des financiers potentiels, je distingue : les banques, les institutions ou agences financières, puis les fondations. Un cas fait cependant exception : *Global Giving*, une

⁸³ Je vais m'attacher à présenter brièvement ces deux ONG car ce sont les deux ONG car elles ont collaboré avec le J-PAL sur plusieurs évaluations, cela depuis la création du J-PAL.

⁸⁴ Pour une revue détaillée des objectifs et du fonctionnement de Seva Mandir : <http://www.sevamandir.org/>

⁸⁵ Voir, (Banerjee, Duflo, Glennerster et Kohtari, 2010), (Banerjee, Duflo et Glennerster, 2008) et (Duflo, Hanna et Ryan, 2012).

⁸⁶ Par exemple, en Inde, la police du Rajasthan a participé à l'évaluation d'un programme dont le but était de renforcer la confiance des citoyens envers la police. Voir (Banerjee, Duflo et Chattopadhyay, 2012) et (Banerjee, Duflo, Keniston, et Singh, 2010).

⁸⁷ Les différentes entreprises qui travaillent avec le J-PAL sont : Alliance One International, Amendis, APA insurance, Carrefour, Cheetah Paprika Limited, Commercial Mexicana, EA Consultants, H&R Block, Mumias Sugar Company, SNV Peru, StickK, Sumito Chemical, Veolia Water.

⁸⁸ A titre d'exemple, Amendis (branche de Veolia Eau) a fourni l'accès à l'eau potable à Tanger au Maroc. Le programme visait à faciliter l'accès à l'eau potable à Tanger. Il a été financé par le programme de développement conçu par le roi Mohammed VI. Si l'évaluation du programme était jugée positive, Amendis serait le fournisseur d'eau à Tanger. Voir (Devoto, Duflo, Dupas, Pariente et Pons, 2011)

entreprise américaine non lucrative dont le but est de permettre aux donateurs de financer directement les projets qu'ils choisissent en ligne. Les banques, elles, sont originaires de tous les pays comme l'*Equity Bank* au Kenya et en Ouganda qui a participé à une évaluation sur l'accès au crédit et à l'assurance en 2011⁸⁹. Les institutions financières participent aussi à l'évaluation des programmes, comme l'*International Finance Corporation*, un groupe de la Banque mondiale qui fournit des services d'investissements afin de développer le secteur privé dans les pays en développement. De nombreuses fondations financent ces évaluations, comme la fondation Ford, ou la fondation de Bill Gates ou encore celle de Michael et Susan Dell.

1.3.3 Les centres de recherches

Le J-PAL collabore avec soixante-dix centres de recherches ou universités à travers le monde. Cependant six centres de recherche entretiennent un lien plus privilégié avec lui :

« les centres de recherches qui aident à gérer les évaluations randomisées du J-PAL et qui emploient l'équipe associée à ces évaluations sont : Innovation for Poverty Action (IPA), Centre for Microfinance, Center for international Development's Micro Development Initiative, Center of Evaluation for Global Action, Ideas 42, and Small Enterprise Finance Center »⁹⁰.

Ideas 42 et *Innovation for Poverty Action* (IPA) ont été créés par des membres du J-PAL et s'inscrivent exactement dans la même lignée. *Ideas 42* est fondé en 2008 à Harvard par deux grandes figures du J-PAL : Sendil Mullainathan (Harvard), Michael Kremer (Harvard). C'est un centre de « recherches en science sociale et un laboratoire en matière de développement à l'université de Harvard, qui ont pour objectif, en utilisant des idées scientifiques, de construire des produits et des politiques innovantes, que ce soit nationalement ou internationalement »⁹¹. *Innovation for Poverty Action* est créé en 2002 par Dean Karlan, un ancien étudiant d'Abhijit Banerjee, d'Esther Duflo et de Sendil Mullainathan. IPA est « une organisation non lucrative dédiée à découvrir ce qui marche et aide les pauvres du monde. Nous construisons et évaluons des programmes dans de réels contextes, avec des vraies personnes, et offrons une participation active pour élargir les

⁸⁹ Il n'y a pas, pour l'instant, de publication académique sur cette évaluation, mais le site du J-PAL en offre un aperçu : <http://www.povertyactionlab.org/node/5498>.

⁹⁰ « About J-PAL », <http://www.povertyactionlab.org/>, (dernière consultation le 24 octobre 2011).

⁹¹ « Origins », <http://ideas42.iq.harvard.edu/>, (dernière consultation le 24 octobre 2011), avec ma traduction.

programmes fructueux»⁹². A l'inverse du J-PAL, IPA n'est pas un laboratoire, mais une organisation non lucrative réunissant plus d'une centaine de chercheurs⁹³.

Par ailleurs, le *Centre for Microfinance* et *Enterprise Finance Center* sont tous les deux rattachés à une école indienne créée en 1970 : l'*Institute for Financial Management and Research* (IFMR). L'IFMR est directement relié à la branche du J-PAL en Asie du sud à Chennai en Inde, comme cela a été évoqué. L'objectif de l'IFMR est, tout comme celui du J-PAL, d'évaluer des programmes de développement à l'aide d'expériences randomisées. Mais l'IFMR est centré sur les programmes qui concernent plus spécifiquement le crédit, ainsi que le financement⁹⁴. Dans la même lignée, *Small Enterprise Finance Center* (SEFC) vise à développer les moyens de financements pour les pauvres, mais centrés cette fois-ci au niveau des entreprises. Le *Center for international Development's Micro Development Initiative* et le *Center of Evaluation for Global Action*, sont tous les deux rattachés à des universités américaines, le premier à l'université d'Harvard et le second à l'université de Berkeley. Le premier est constitué d'un groupe de chercheurs qui mènent des évaluations randomisées aussi bien en matière de santé, d'éducation, de gouvernance que de micro-finance. Il est aussi très étroitement relié au J-PAL, puisque tout comme lui, il offre des formations à cette méthode.

Le *Center of Evaluation for Global Action* est un centre de recherches sur le développement, et comme tous les autres, son objectif consiste à mener des évaluations randomisées. Mais il vise un niveau plus global, celui de l'élargissement des programmes déjà existants.

Pour résumer, la structure institutionnelle du J-PAL lui permet d'avoir une place de poids à travers le monde et contribue à sa position dominante en économie du développement, des chercheurs du monde entier pouvant être rattachés au sein de ses différentes antennes. Cette

⁹² « About IPA », <http://www.poverty-action.org/>, (dernière consultation le 24 octobre 2011), avec ma traduction.

⁹³ IPA est sans doute l'organisation avec laquelle le J-PAL entretient les liens les plus étroits. A titre d'illustration de cette proximité, on peut souligner que les personnalités les plus célèbres du J-PAL sont aussi rattachées à IPA en tant que chercheurs tels qu'Esther Duflo, Abhijit Banerjee, Sendil Mullainathan, Michael Kremer, Rachel Glennester. De plus, la sœur d'Esther Duflo, Annie Duflo, y dirige le réseau de chercheurs.

⁹⁴ « Les centres de recherches de l'IFMR construisent et testent des produits financiers appropriés pour les pauvres et les petits business, évaluent les stratégies anti-pauvreté des secteurs publics et privés, promeuvent la transparence des gouvernements à travers des meilleures mesures des effets des biens publics et des dépenses publiques, et avec une large gamme de partenaires du public, du privé et d'organisations non gouvernementales produise des recommandations claires pour les politiciens et les entrepreneurs sociaux ». (*IFMR Research*, http://www.ifmr.ac.in/academicprograms_pgdm_application.html, (dernière consultation le 24 octobre 2011), avec ma traduction).

séparation en antennes offre en outre la possibilité au réseau du J-PAL de se renforcer, chaque chercheur y apportant son propre réseau.

2. La randomisation : le *gold standard* ?

Cette place de poids s'explique aussi par l'avantage principal de la randomisation – sa forte validité interne. Je m'intéresserai ici à cette dimension. La randomisation est une des méthodes permettant d'obtenir le contrefactuel le plus solide. Un contrefactuel permet de répondre à la question suivante : que se serait-il passé si les individus n'avaient pas eu accès au programme ? Autrement dit, il permet de penser la situation où le programme n'a pas été mis en place. Donald Rubin (1974) offre la première modélisation de ce contrefactuel. Mais ce modèle doit ensuite être intégré au sein d'une méthode particulière. Plusieurs méthodes sont disponibles, la randomisation est l'une d'entre elles. La randomisation, par l'assignation aléatoire qui la caractérise, offre un contrefactuel permettant de limiter de nombreux biais, ce qui tend à faire d'elle le *gold standard* dans ce domaine. Afin de rendre compte de ce statut de *gold standard*, je présenterai, dans un premier temps le modèle développé par le statisticien Donald Rubin permettant de penser un contrefactuel (2.1). Puis, je m'intéresserai à sa transposition par les chercheurs du J-PAL en économie du développement (2.2), et enfin, je comparerai la randomisation aux autres méthodes qui permettent de penser un contrefactuel, afin de rendre saillant le statut de *gold standard* de la randomisation (2.3).

2.1 Le modèle de Rubin

Le modèle de Rubin s'inspire des travaux de Roy (1951). Roy s'intéresse aux choix des travailleurs lorsqu'ils doivent se décider entre pêcher ou chasser. Comment déterminer lequel de ces choix sera le plus profitable ? Les travailleurs doivent envisager deux options potentielles, mais ils ne peuvent savoir à l'avance celle qui leur rapportera le plus. En fonction de ces deux potentialités et de leurs compétences, ils ont à déterminer s'ils vont pêcher ou chasser. L'intuition la plus forte dans le travail de Roy est donc d'envisager deux potentialités, car les travailleurs ne peuvent pas en même temps pêcher et chasser, pour déterminer quelle activité leur est la plus profitable. Rubin va reprendre ces deux potentialités, mais cette fois-ci, non pas dans le choix d'une activité ou d'une autre, mais dans le but de déterminer rigoureusement les effets d'un traitement. Je commencerai par donner les intuitions du modèle de Rubin (2.1.1), puis, je développerai son application statistique (2.1.2).

2.1.1 Intuitions

Le modèle causal de Rubin (MCR) pose le problème de l'inférence causale et cherche à mettre en évidence un cadre qui pourra la justifier. Il cherche à montrer une causalité particulière, par exemple, que le traitement A a causé B. En d'autres termes, il s'intéresse à l'effet de A. Pour Rubin il n'y a pas de causalité sans manipulations, c'est-à-dire qu'il faut que quelque chose soit changé, manipulé, pour pouvoir ensuite questionner le lien de causes à effets. C'est dans le domaine médical que le modèle de Rubin a connu l'écho le plus important. Le modèle de Rubin⁹⁵ a trois caractéristiques principales. Tout d'abord, il se sert d'un résultat potentiel (*potential outcome*) dans le but de déterminer l'inférence causale (1). Il utilise, ensuite, un mécanisme d'assignation probabiliste (2), « *which mathematically describes how treatment are given to units, with possible dependence on background variables and the potential outcomes themselves* » (Imbens et Rubin, 2008, 1). Et enfin, il est caractérisé par une distribution de probabilités optionnelles sur toutes les variables (3). L'élément central du modèle de Rubin reste, cependant, le concept de « résultats potentiels ». Ce dernier offre une solution à un problème intuitif bien connu (*i.e.*, comment être sûr qu'une causalité n'est pas une corrélation). Par exemple, comment savoir si l'aspirine qu'un individu a prise est responsable de la disparition de son mal de tête ou si cela est dû à un autre facteur, comme le fait de s'être allongé dans le noir ? L'individu ne peut à la fois prendre l'aspirine et ne pas la prendre, s'il fait le choix de prendre une aspirine, il n'y a aucun moyen de savoir ce qui se serait passé s'il ne l'avait pas prise. C'est à cette dernière question que le modèle de Rubin tente d'offrir une réponse. L'objectif de Rubin est de parvenir à trouver un moyen pour penser les deux états en même temps : envisager à la fois chez un même individu la prise et la non prise d'aspirine. Les « résultats potentiels » lui permettent de penser ces deux états, car ils sont tous deux réduits à une potentialité. Pour analyser l'effet d'un traitement, il y a alors deux résultats possibles à envisager : le résultat lorsque l'individu prend l'aspirine et un résultat lorsqu'il ne la prend pas.

2.1.2 Application statistique

Il y a donc deux potentialités : prendre une aspirine (que nous appellerons la potentialité T) et, par exemple, boire un verre d'eau (que nous appellerons la potentialité C). Ces

⁹⁵ J'utiliserai indifféremment les termes modèles de Rubin et modèle causal de Rubin.

potentialités impliquent deux résultats, eux-mêmes éventuels ; en d'autres termes ils vont potentiellement créer deux effets particuliers. Nous appellerons $Y(T)$ le résultat observé lorsque l'individu prend une aspirine et $Y(C)$ le résultat lorsque l'individu boit un verre d'eau. $Y(T)$ représente alors l'effet causal de la prise d'aspirine, autrement dit l'effet causal de T ; et $Y(C)$ représente l'effet causal de la prise du verre d'eau. Y est le résultat potentiel de chacune de ces possibilités. Pour obtenir l'effet de l'aspirine comparativement à celui de la prise d'un verre d'eau sur la migraine – ou ce que Rubin appelle « l'effet causal de T versus C sur Y », il suffit alors de différencier les deux termes : $Y(T) - Y(C)$. Néanmoins, les deux potentialités ne seront pas observées chez un même individu : il ne peut pas à la fois prendre l'aspirine et ne pas en prendre. Afin d'affiner l'intuition de Rubin, élargissons cet exemple à deux personnes. Chacune de ces deux personnes a, donc, deux résultats potentiels, comme on l'a vu plus haut. Pour la personne 1, ces deux possibilités seront $Y_1(T)$ et $Y_1(C)$; pour la personne 2 ces deux possibilités seront $Y_2(T)$ et $Y_2(C)$. L'effet de l'aspirine est alors trouvé par un double contrefactuel puisque deux personnes rentrent en jeu, désormais. L'effet comparatif est alors obtenu par : $\frac{1}{2} [Y_1(T) - Y_1(C) + Y_2(T) - Y_2(C)]$. Dans la réalité, seules deux possibilités sont observables :

1. $Y_1(T) - Y_2(C)$, dans ce cas, on différencie la situation où la personne 1 prend une aspirine et celle où la personne 2 boit un verre d'eau.
2. $Y_2(T) - Y_1(C)$, à l'inverse, dans ce cas, on compare la situation où la personne 2 prend une aspirine à celle où la personne 2 boit un verre d'eau.

Comme un individu ne peut à la fois prendre une aspirine et ne pas en prendre, pour obtenir une différence entre ces deux effets, il est nécessaire de prendre en considération un autre individu, qui permettra la comparaison. Pour que la détermination de l'inférence causale de la prise d'aspirine soit scientifiquement solide, il faut que ces deux individus soient le plus comparable possible. Autrement dit, il faut que $Y_1(T) - Y_2(C) = Y_1(T) - Y_1(C)$, ou que $Y_2(T) - Y_1(C) = Y_2(T) - Y_2(C)$. Cette modélisation permet de passer de l'idée abstraite de contrefactuel à un cadre analytique plus concret.

2.2 Économie du développement et transposition du modèle de Rubin

Dans le but de déterminer les effets d'une politique de lutte contre la pauvreté, les chercheurs du J-PAL, vont appliquer le modèle de Rubin à l'économie du développement.

Leur objectif central est de trouver un contrefactuel parfait, afin de déterminer une inférence causale solide. Dans cette perspective, la randomisation apparaît être le moyen le plus fiable. Afin de mettre cela en évidence, et de présenter la justification du J-PAL en faveur de la randomisation, je développerai leur utilisation du modèle de Rubin, ainsi que le principal avantage de la randomisation (2.2.1), l'annulation des biais de sélection (2.2.2).

2.2.1 Contrefactuel et économie du développement⁹⁶

Afin de fournir une évaluation des programmes de développement rigoureuse, les chercheurs du J-PAL tentent donc de créer un contrefactuel. Cette fois-ci, on se sépare de l'aspirine pour s'intéresser à un programme de développement ; la question devient : que se serait-il passé si le programme n'avait pas été mis en place ? L'évolution de la situation est-elle due au programme ou à d'autres facteurs ? Sans le programme aurait-on observé une évolution similaire ? Les chercheurs du J-PAL s'inspirent, ici, du modèle de Rubin et utilisent son cadre analytique. On l'a vu avec le modèle de Rubin, les deux résultats potentiels d'une personne ne sont pas observables en même temps, il n'est donc possible de déterminer que les effets du traitement qu'en terme de moyenne. Par exemple, si l'on s'intéresse à l'effet des manuels scolaires sur l'absentéisme des élèves, on se centrera sur la moyenne des effets des élèves qui ont eu accès à ces manuels et à la moyenne de ceux qui n'en ont pas eu accès. Dans le but de déterminer l'effet des manuels scolaires, on différenciera, ensuite, ces deux effets. Dans ce cas on obtient l'équation suivante : $E[Y_i(T) - Y_i(C)]$, où E est l'espérance mathématique et i représente les différentes écoles sélectionnées. Supposons, comme (Duflo, Glennerster et Kremer, 2007), que l'accès aux données dépende d'un nombre important d'écoles (i) dans une région donnée et que certaines écoles bénéficient de manuels scolaires et d'autres n'en bénéficient pas alors : « *one approach is to take the average of both groups and examine the difference between average tests scores in schools with textbooks and in those without* » (Duflo, Glennerster et Kremer, 2007, 3900). Cette approche consiste à regarder la « moyenne des effets du traitement » – ici les manuels scolaires. Dans ce cas, on s'intéresse à la fois aux personnes traitées – bénéficiant de manuels scolaires – et aux personnes non traitées. Une autre possibilité aurait été de regarder les effets du traitement uniquement dans le groupe traité (ici, celui qui a accès aux manuels scolaires). Selon Duflo, Glennerster et

⁹⁶ Cette sous-partie reprend les étapes définies par Esther Duflo, Rachel Glennester et Michael Kremer dans : « *The Use of Randomization in Development Economics : a Toolkit* » (2007).

Kremer (2007), lorsque l'on observe cette moyenne sur un large échantillon, cela converge vers l'équation D :

$$D = E[Y_{iT} / \text{écoles qui bénéficient des manuels scolaires}] - E[Y_{iC} / \text{écoles qui ne bénéficient pas des manuels scolaires}] = E[Y_{iT}/T] - E[Y_{iC}/C].$$

Si l'on reprend le modèle de Rubin, exposé plus haut, on retrouve la même équation mais écrite différemment. On imagine deux groupes : un groupe qui reçoit les manuels scolaires ($Y_i/\text{écoles qui bénéficient des manuels scolaires} = Y_{iT}$) et un groupe qui ne les reçoit pas ($Y_i/\text{écoles qui ne bénéficient pas des manuels scolaires} = Y_{iC}$). Chacun de ces deux groupes a deux résultats potentiels, comme dans le modèle de Rubin. Cela s'exprime de la façon suivante :

- Pour le groupe qui bénéficie des manuels scolaires (Y_{iT}) :
 1. (Y_{iT}/T), le résultat potentiel du groupe des écoles i bénéficiant des manuels scolaires lorsqu'il les reçoit effectivement.
 2. (Y_{iC}/T), le résultat potentiel du groupe des écoles i bénéficiant des manuels scolaires s'il n'en avait pas bénéficié, qu'il n'avait pas eu de manuels scolaires (c'est ici que le contrefactuel agit car on intègre une situation hypothétique).
- Et pour le groupe qui ne bénéficie pas des manuels scolaires (Y_{iC}) on a :
 1. (Y_{iC}/C), le résultat potentiel du groupe des écoles i qui ne bénéficie pas des manuels scolaires lorsqu'ils n'en bénéficient réellement pas.
 2. (Y_{iT}/C), le résultat potentiel du groupe des écoles i qui ne bénéficie pas du programme si elles avaient reçues le programme.

On comprend donc mieux l'équation « D ». Elle correspond à la différence entre la moyenne des effets du traitement dans le groupe traité, diminué des effets de l'absence de traitement dans le groupe non traité. Ces deux variables sont donc observables, comme on l'a vu chez Rubin. En ajoutant et en retirant $E[Y_{iC}/T]$ ⁹⁷ à l'équation (D), ce qui mathématiquement ne change rien, cela permet de mettre en évidence un concept central : le biais de sélection. L'équation (D) devient la suivante :

⁹⁷ C'est à dire le résultat potentiel du groupe non traité (qui ne bénéficie pas des manuels scolaires) s'il avait reçu les manuels scolaires.

$$D = E[Y_{iT}/T] - E[Y_{iC}/T] - E[Y_{iC}/C] + E[Y_{iC}/T]$$

Cette équation peut ensuite être réécrite de la façon suivante :

$$D = E[Y_{iT} - Y_{iC}/T] + E[Y_{iC}/T] - E[Y_{iC}/C]$$

Le premier terme ($E[Y_{iT} - Y_{iC}/T]$) traduit les effets du traitement puisqu'il exprime la différence entre les deux résultats potentiels au sein du groupe traité – qui bénéficie des manuels scolaires, ce terme permet de répondre à la question suivante : quel est exactement l'effet du traitement ? Le second terme ($E[Y_{iC}/T] - E[Y_{iC}/C]$) représente la différence entre la moyenne des effets sur le groupe traité lorsqu'il ne bénéficie pas du traitement et la moyenne des effets sur le groupe non traité lorsqu'il ne reçoit pas le traitement. Ce second terme représente ce que l'on appelle le *biais de sélection*.

2.2.2 La notion de biais de sélection

Le biais de sélection représente le fait que la différence entre les deux groupes peut s'expliquer par d'autres facteurs que ceux du programme évalué. Il peut exister des différences au sein des deux groupes expliquant la différence des effets observés. Cela rend les résultats inutilisables car, dans ce cas, ils ne traduisent pas les effets du traitement mais la différence inhérente aux deux groupes. Pour reprendre un exemple de Duflo, Glennerster et Kremer (2007), imaginons que le groupe d'écoles qui reçoit les manuels scolaires (Y_i/T) soit composé d'écoles dans lesquelles les parents sont particulièrement investis dans l'éducation de leurs enfants et font de cette dernière une priorité. Dans cette situation, les parents encouragent leurs enfants à travailler et à faire leurs devoirs. Si ce groupe bénéficie effectivement du programme, la différence entre le résultat potentiel de ce groupe lorsqu'il reçoit les manuels scolaires (Y_{iT}/T) et celui du groupe non traité qui ne bénéficie pas des manuels scolaires (Y_{iC}/C) sera beaucoup plus importante ; autrement dit $E[Y_{iT}/T] - E[Y_{iC}/C]$ sera plus importante. Cette plus grande différence est due au fait que $E[Y_{iC}/T]$ et $E[Y_{iC}/C]$ ne sont plus comparables car ils ne sont plus identiques. $E[Y_{iC}/T]$, les écoles bénéficiant des manuels scolaires ont, même lorsqu'ils ne reçoivent pas le programme, des résultats aux examens supérieurs grâce à l'investissement des parents ; ce qui n'est pas le cas

pour les écoles qui ne reçoivent pas les manuels $E[Y_{iC}/C]$. Cette différence plus importante entre les deux groupes n'est pas due aux effets importants du programme. Elle vient du fait que ces deux groupes d'écoles partagent une différence intrinsèque : le premier groupe est constitué d'enfants dont les parents s'investissent considérablement dans l'éducation de leurs enfants alors que dans le second groupe l'investissement des parents n'est pas présent. L'effet du programme ne sera donc pas évalué à sa juste valeur. L'évaluation souffre d'un biais de sélection. Les populations sélectionnées sont trop différentes et ne sont donc pas parfaitement comparables, rendant donc impossible une évaluation correcte des effets du programme.

Il est très difficile de mesurer le biais de sélection, puisqu'il est impossible d'observer le résultat potentiel du groupe d'école qui bénéficie des manuels scolaires lorsqu'elle n'en bénéficie pas. Ainsi : « *An essential objective of much empirical work is to identify situations where we can assume that the selection bias does not exist or find a way to correct it.* » (Duflo, Glennerster et Kremer, 2007, 3901) La randomisation permet d'annuler ce biais de sélection grâce à l'assignation aléatoire des deux groupes⁹⁸.

« *The treatment group is then exposed to the « treatment » (their treatment status) while the comparison group (treatment status C) is not. Then the outcomes Y is observed and compared for both treatment and comparison groups. For example, out of 100 schools, 50 are randomly chosen to receive textbooks, and 50 do not receive textbooks. The average treatment effect can then be estimated as the difference in empirical means of Y between the two groups* ». (Duflo, Glennerster et Kremer, 2007, 3901)

Il est alors possible d'évaluer les effets du programme. A l'aide de la randomisation l'équation D est la suivante :

$$D = E[Y_{iT}/T] - E[Y_{iC}/C]$$

Comme les individus sont assignés au groupe de traitement et au groupe de contrôle de façon aléatoire, la seule différence qui existe entre les deux groupes serait la possibilité d'être exposé au traitement. Si le groupe de traitement n'avait pas reçu le traitement, leur résultats potentiels auraient été les mêmes. Cela implique que le biais de sélection est nul, autrement

⁹⁸ James Heckman, qui conceptualise cette notion de biais de sélection, nuance - avec Jeffrey Smith - le fait que la randomisation permette d'annuler tous les biais de sélection et montre que la randomisation résout ce biais en le balançant entre les deux groupes (Heckman et Smith, 1995) : « *Randomized social experiments solve the problem of selection bias (...) Finally note that random assignment does not remove selection bias, but instead balances the bias between the participant and nonparticipant samples.* » (Heckman et Smith, 1995, 88-89).

dit $E[Y_{iC}/T] - E[Y_{iC}/C] = 0$. Il n'y a donc pas de différence entre le résultat potentiel du groupe d'école qui ne reçoit pas le traitement et celui du groupe d'école qui le reçoit s'il ne l'avait effectivement pas reçu. Dans ce cas, pour estimer les effets du programme, il suffit juste de traduire cette équation dans une régression économétrique. Toutes les données définissant D ne sont pas toutes disponibles ou observables, il n'est possible que d'estimer D . La régression qui permet d'estimer D est : $Y = \alpha + \beta T + \epsilon_i$, où T est un « *dummy* » pour traduire l'assignation au groupe de traitement. Cette équation est estimée par la méthode économétrique des moindres carrés ordinaires.

2.3 Les autres méthodes permettant de réduire les biais de sélection

Il existe plusieurs méthodes pour minimiser les biais de sélection : des méthodes expérimentales et donc *a priori* ainsi que des méthodes non expérimentales – donc *a posteriori*. La randomisation est l'une des méthodes expérimentales, tout comme les expériences contrôlées⁹⁹. Les méthodes non expérimentales tentent de résoudre le problème du biais de sélection à l'aide d'une série d'hypothèses qui leur permet de s'approcher au plus près de la démarche expérimentale. Les trois méthodes les plus utilisées de ce type sont : la méthode d'appariement (2.3.1), la double différence (2.3.2) et la discontinuité (2.3.3). Je développerai chacune d'entre elles afin de mettre en évidence le principal attrait de la randomisation, annuler les biais de sélection, alors que les autres méthodes ne permettent que de les réduire.

2.3.1 La méthode d'appariement et d'appariement par score de propension

La méthode d'appariement tente de reconstituer à partir de variables observables un groupe de contrôle et un groupe de traitement. L'objectif est donc de trouver pour chaque individu qui a reçu le traitement un individu comparable qui, lui, ne l'a pas reçu. Il s'agit de constituer, à partir de caractéristiques similaires observables, deux groupes – un de traitement et un de contrôle – afin d'être en mesure de les comparer et de déterminer l'effet du traitement. Le biais de sélection est ici réduit à partir des variables observables.

⁹⁹ Les expériences contrôlées, tout comme les expériences randomisées, sont constitués d'un groupe de contrôle et d'un groupe de traitement. La seule différence est que la constitution de ces deux groupes ne s'est pas faite aléatoirement.

Dans le même but, on peut aussi utiliser un score de propension. Un score de propension est la probabilité pour un individu de recevoir une intervention (traitement) conditionnellement à ses caractéristiques. Ce score permet de déterminer le groupe de contrôle - c'est pourquoi cette méthode est aussi appelée appariement par score de propension. Avec un score de propension, il est possible de prendre en compte un plus grand nombre de variables. Le principal avantage de cette méthode par rapport à la randomisation est qu'en n'étant pas expérimentale, elle permet d'évaluer des programmes qui ont déjà été mis en place. De plus, elle coûte beaucoup moins cher. Cependant, elle sous-entend un travail de collecte de données conséquent pour être en mesure de constituer les deux groupes.

Une de ses limites majeure est la possible persistance des biais de sélection. La différence intrinsèque entre les deux groupes peut résider dans des variables non observables, par exemple si les individus des deux groupes se distinguent par une motivation différente, un accès à l'information, lui aussi, différent. Dans ce cas, cette méthode ne permet pas d'éliminer les biais de sélection. En d'autre terme, l'appariement permet d'annuler uniquement les biais de sélection causés par des variables observables, mais ne peut pas rendre compte des variables non observables.

2.3.2 La méthode de double différence

Une autre possibilité, si le programme a déjà été mis en place, est d'utiliser la méthode de double différence. L'idée centrale est de comparer des données avant et après la mise en place du programme, en constituant, là aussi, un groupe de contrôle et un groupe de traitement, à partir de ces données. La première étape est alors de collecter les données sur le groupe traité et sur le groupe non-traité. Il s'agit d'une double différence car on soustrait la différence « avant-après » des moyennes observées dans le groupe des non traités à la même différence observée dans le groupe de traitement. Il y a donc, au sein de chaque groupe, une première différence, qui cherche à comparer un avant et un après programme. Et c'est cet « avant-après » qui est ensuite comparé entre les deux groupes, par la deuxième différence. Cette méthode permet donc de réduire les biais de sélection, à partir d'une collecte de données observables opérée à deux moments particuliers – avant et après la mise en place du programme. Le fait de comparer deux périodes dans le temps, permet d'éviter un biais que le temps aurait pu causer, comme l'existence d'un cycle conjoncturel ; cet effet permet d'être évité s'il est le même pour les deux groupes.

Cet avantage a pour corollaire la plus importante limite de cette méthode. Si l'effet du temps est différent pour les deux groupes, il existera alors un biais de sélection, que cette méthode exacerbera. Une seconde limite¹⁰⁰, tout comme pour la méthode des appariements, est due à la collecte de données des variables uniquement observables, dans ce cas : « si les deux groupes devaient évoluer de manière très différente (par exemple si les personnes qui participent au programme avaient une propension à augmenter leurs revenus plus rapidement que ceux qui ne participent pas), le biais de sélectivité resterait important car cet écart dû aux inobservables est attribué au programme » (Parienté, 2008, 7).

2.3.3 La méthode de discontinuité

La méthode de la discontinuité permet d'assigner les individus à un traitement en fonction d'une variable extérieure qui définit un seuil. En fonction de ce seuil, la séparation entre les deux groupes est opérée. Il faut donc trouver une variable au sein de laquelle on peut observer un seuil, comme par exemple l'âge ou le résultat aux examens. C'est pourquoi on parle de discontinuité car on choisit un moment de rupture, à l'aide du seuil, pour rendre compte de l'effet d'un traitement. Les initiateurs de cette méthode, Donald Thistlewaite et Donald Campbell (1960), cherchent à évaluer l'effet des certificats aux mérites sur les résultats futurs des étudiants qui en ont bénéficié. Le problème étant que les étudiants les plus performants sont ceux qui vont bénéficier de ces certificats. Afin de pallier cela, l'objectif est de comparer les notes des étudiants qui sont juste en dessous et au-dessus du seuil d'attribution des résultats. C'est la note aux examens permettant de bénéficier d'un certificat qui va faire l'objet du seuil. L'idée étant qu'autour de ce seuil, les étudiants, avec une note légèrement inférieure ou légèrement plus élevée, sont parfaitement comparables. Une des principales limites de cette méthode est qu'il n'est pas toujours possible de définir une variable observable permettant de définir un seuil. De plus, il faut que les règles soient strictes et précises, et il se peut que cela ne soit pas le cas. Duflo souligne que dans les pays en développement :

« Il est souvent possible que les règles ne soient pas appliquées de manière suffisamment stricte pour engendrer des discontinuités qui puissent être utilisées pour des motifs d'identification. Par exemple, les chercheurs ont tenté d'utiliser comme source d'identification la discontinuité introduite dans la politique de la Grameen (la plus importante organisation de microcrédit au Bangladesh) par le fait de ne prêter qu'aux

¹⁰⁰ Bertrand, Duflo et Mullainathan (2004) produisent une analyse des avantages et des limites d'une telle méthode.

personnes propriétaires de moins d'une acre de terrain (Land et Khandker, 1998). Il se trouve qu'en pratique, la banque Grameen prête à beaucoup de personnes possédant plus d'une acre de terrain et qu'il n'y a pas de discontinuité dans la probabilité d'emprunt au seuil (Morduch, 1998). » (Duflo, 2005, 192)

Dans ce cas, si les règles ne sont pas respectées en pratique, la discontinuité n'existe plus et ne peut donc pas être utilisée pour évaluer un effet causal.

3. La randomisation en pratique

Etant la seule méthode qui permet d'annuler tout biais de sélections, la randomisation bénéficie d'une solide fiabilité scientifique. Cependant, comme tout processus expérimental, sa mise en pratique est complexe et peut impliquer certaines limites. Tout d'abord, il est nécessaire de réfléchir à certains enjeux avant la mise en place de l'expérience. Ensuite, il est important de déterminer comment la randomisation sera introduite, au sein du processus expérimental. Malgré son statut de *gold standard* méthodologique, la randomisation rencontre certaines limites intrinsèques à sa mise en place. Ici, je m'intéresserai donc à la mise en place de la randomisation sur le terrain ainsi qu'à certaines de ces limites internes. Cela me permettra de présenter les différentes façons possibles, pour le J-PAL de procéder sur le terrain. Je présenterai tout d'abord les enjeux rencontrés avant la mise en place de l'expérience (3.1) ; puis comment la dimension aléatoire peut être introduite au sein d'un procédé expérimental (3.2) ; et enfin je développerai certaines des limites inhérentes à la mise en place de la randomisation (3.3).

3.1 Enjeux rencontrés avant la mise en place de la randomisation

Avant la mise en place de l'expérience dans une région particulière, les chercheurs font face à trois enjeux particuliers. Le premier est de savoir s'ils ont besoin de collecter des données en amont afin d'obtenir des informations supplémentaires sur la population qui fera partie de l'expérience (3.1.1). Le deuxième enjeu est de déterminer le niveau de l'assignation aléatoire, autrement dit si elle doit se réaliser au niveau des individus ou de groupes d'individus (3.1.2). Enfin, le troisième enjeu est de définir la structure de la randomisation, à savoir si l'expérience va évaluer plusieurs effets ou un unique effet (3.1.3). Je vais désormais m'attacher à développer chacun de ces trois enjeux.

3.1.1 Collecte de données

Il n'est pas nécessaire, lorsqu'on réalise une expérience randomisée de collecter des données dans le but de créer un groupe de contrôle. Cependant, il peut être utile de bénéficier de données supplémentaires afin d'avoir une compréhension plus précise des résultats. C'est la raison pour laquelle, avant de mettre en place l'expérience, une des questions principales que se posent les chercheurs est de savoir s'ils ont besoin de collecter des données qui fournissent une information sur la situation, s'il est nécessaire de mener une *baseline survey*. La randomisation rend, en principe, la *baseline survey* inutile¹⁰¹. Cependant, cette étude permet de générer des variables de contrôle qui vont réduire la variabilité du résultat final et donc réduire la nécessité d'un échantillon plus large. Le coût financier joue un rôle important dans la décision de mener ou non une *baseline survey* : si l'intervention est chère et que la collecte de données est relativement peu coûteuse, mener une *baseline survey* permettra d'économiser des fonds. Dans le cas inverse (lorsque la collecte des données est chère) il n'est pas profitable de mener une *baseline survey*.

La *baseline survey* permet d'examiner les interactions entre les conditions initiales et l'impact du programme, elle offre aussi la possibilité de vérifier que la randomisation a été faite correctement. Si la collecte de données est trop contraignante, il est en règle générale possible de bénéficier facilement des données administratives. Les partenaires de terrain fournissent alors ces données aux chercheurs du J-PAL avant l'expérience. Ces données peuvent se substituer à la *baseline survey* :

« Using administrative data (data collected by the implementing organization as part of their normal functioning) linked to information on treatment can greatly reduce the cost of data collection and reduce attrition. » (Duflo, Glennerster et Kremer, 2007, 3933)

Malgré tout, il est important de s'assurer que les données sont comparables entre les groupes de traitement et de contrôle ; il pourrait être tentant de ne collecter des données que dans la zone où a eu lieu le programme pour réduire les coûts de collecte de données :

« However this could introduce biases as a difference in measured outcomes between treatment and comparison areas could reflect different data collection methodologies ». (Duflo, Glennerster et Kremer, 2007, 3933-3934)

¹⁰¹ *« In principle, randomization renders baseline surveys unnecessary, since it ensures the treatment and comparison groups are similar in expectation. »* (Duflo, Glennerster, Kremer, 2007, 3933)

Même si les données au sein du groupe de traitement sont collectées, cela ne permettra pas d'appréhender les résultats de façon plus précise. Au contraire, cela introduira un biais, car la collecte de données aura été différente d'un groupe à l'autre.

3.1.2 Le niveau de randomisation

La sélection aléatoire pour l'assignation des groupes peut avoir lieu sur des individus ou sur une population. Pour déterminer le niveau de randomisation, plusieurs facteurs rentrent en compte. Tout d'abord, les externalités peuvent biaiser l'évaluation du traitement. Pour y pallier, il peut être utile de randomiser à un niveau qui permet de capturer les effets du traitement sans être biaisé par les externalités. Par exemple, cela a été évoqué, Kremer et Miguel (2004), trouvent un effet plus important des médicaments pour les vers intestinaux que ceux des randomisations qui avaient été faites auparavant au niveau individuel ; cela s'explique par le fait que ces médicaments ont une externalité, c'est-à-dire qu'ils ont un effet sur les personnes ne prenant pas le traitement¹⁰². La comparaison individuelle rend donc la différence entre le groupe de traitement et le groupe de contrôle plus faible. Miguel et Kremer font alors la randomisation sur un échantillon plus large (les écoles), les externalités sont toujours présentes mais leur impact est plus faible. De plus, sélectionner la population aléatoirement sur un échantillon important peut rendre la mise en place de l'expérience plus simple. Par exemple, pour une intervention qui a des coûts fixes élevés, il est plus avantageux de sélectionner un échantillon très large. Lorsque Banerjee, Duflo et Glennerster (2011) évaluent l'impact de plus de fer dans la farine sur le bien être des individus, une grande partie de la population est sélectionnée. Une autre raison, ici, de privilégier l'assignation aléatoire au niveau du groupe plutôt qu'au niveau individuel est que si le programme est perçu comme désirable dans le village cela peut créer des tensions avec l'organisation qui implante le programme. En somme, le choix du niveau de la randomisation est contextuel, il dépend de la nature de l'intervention et de l'interaction avec les individus traités et les individus non traités.

3.1.3 Structure de la randomisation

Indépendamment du niveau choisi de randomisation, il est possible d'évaluer plusieurs éléments en même temps. Lorsque plusieurs éléments sont évalués simultanément, on parle alors de structure factorielle ou *cross-cutting design* : « *In cross-cutting designs several different treatments are tested simultaneously with randomization being conducted so the treatment are orthogonal to each other* » (Duflo, Glennerster et Kremer, 2007, 3930). On

¹⁰² Les personnes désormais traitées ne peuvent plus transmettre de vers intestinaux aux personnes qui ne sont pas traités. Il y a alors une diminution de la transmission des vers intestinaux.

retrouve ici une des dimensions mises en avant par Fisher¹⁰³. Il y a deux façons d'envisager une étude factorielle : elle peut être utilisée pour tester une variété d'interventions ou elle peut être mise en place pour voir si les traitements créent une interaction d'effets. Par exemple, le programme PROGRESA¹⁰⁴ a une structure factorielle parce qu'il est une combinaison de plusieurs programmes (un transfert d'argent, une redistribution des ressources tournées vers les femmes, et une dimension incitative). Pour un décideur politique évaluer l'ensemble peut être suffisant pour décider si oui ou non le programme va se continuer, mais on peut aussi vouloir comprendre quelle composante du programme PROGRESA doit être étendue à d'autres programmes. On peut vouloir savoir si la dimension incitative est essentielle ou pas, si la distribution d'argent aux femmes plutôt qu'aux hommes compte, etc. Une randomisation avec une structure factorielle permet de séparer les différentes composantes d'un programme et de les étudier simultanément. C'est pourquoi Duflo, Glennerster et Kremer (2007, 3931) affirment : « *In principle, a cross-cutting design could have been used in order to disentangle the various component of PROGRESA* ». Evaluer l'ensemble des composantes du programme possède un inconvénient : on ne sait pas quels sont les éléments permettant d'expliquer la réussite ou l'échec du programme ; on sait uniquement que ce programme fonctionne dans ce contexte¹⁰⁵. Une analyse factorielle peut aussi être utile pour tester plusieurs hypothèses en même temps, avec une faible augmentation du coût de l'expérimentation (car le principal coût est de conduire les études qui établiront les conditions de départ et la mesure du résultat des variables¹⁰⁶). Duflo souligne que comme leur coût est plus faible qu'une expérience non factorielle, cela peut alors représenter une opportunité pour les étudiants ou autres personnes qui ont des ressources limitées d'implanter une randomisation avec des traitements additionnels dans un projet plus large.

3.2 L'introduction de la randomisation

Une fois que les trois points qui viennent d'être décrits sont déterminés il reste à introduire la dimension aléatoire au sein de l'expérience. Il existe quatre façons d'introduire cette dimension, chacune se rattachant à un ensemble différent de justifications. La randomisation peut être introduite au sein d'une population large, parce qu'il existe trop de

¹⁰³ La structure expérimentale définie par Fisher suppose l'existence d'un plan factoriel.

¹⁰⁴ Le programme PROGRESA est un programme mis en place au Mexique et avec le partenariat de la banque mondiale.

¹⁰⁵ Je reviendrai très largement sur ce problème dans la deuxième partie de cette thèse.

¹⁰⁶ Pour un exemple d'expérience factorielle, voir (Miguel et Kremer, 2004).

personnes éligibles pour bénéficier du programme (*oversubscription*) (3.2.1). A l'inverse, elle peut intervenir lorsque toute la population pourrait participer au programme et vise dans ce cas à évaluer des mécanismes incitatifs (*encouragement design*) (3.2.2). La randomisation peut aussi être introduite non pas au sein d'un groupe mais au sein de sous-groupes (*within-group randomization*) (3.2.3). Et enfin, elle peut être introduite progressivement dans le temps (*Phase-In*) (3.2.4). Je vais désormais m'attacher à développer chacune de ces approches.

3.2.1 « Oversubscription »

La randomisation, on l'a vu, tire son principal intérêt de sa solidité scientifique. Mais en matière d'économie du développement, elle offre un second intérêt. Dans les pays pauvres, les ressources sont souvent rares. Parfois, il se peut donc que toute la population ne bénéficie pas d'un programme. La randomisation, en sélectionnant aléatoirement les personnes qui auront accès au programme, permet de résoudre la question éthique de l'allocation des ressources. La randomisation peut donc avoir lieu dans l'unique but de sélectionner une population qui aura accès au programme, car sans l'assignation aléatoire, il faudrait déterminer un autre critère d'allocation pertinent. Dans ce cas, la randomisation est introduite car il existe un nombre trop important de personnes par rapport aux ressources disponibles. On parle alors d'*oversubscription*.

3.2.2 « Encouragement design »

À l'inverse de l'*oversubscription* il est possible que tout le monde ait accès au traitement. Dans ce cas, la randomisation peut tout de même intervenir. Elle est utilisée pour augmenter la participation des individus au programme. Tout le monde aura accès au programme, mais une partie de la population qui a été sélectionnée aléatoirement bénéficiera d'incitations supplémentaires. C'est ce que l'on appelle une structure incitative ou un *encouragement design*. Par exemple, Duflo, Kremer et Robinson (2011) étudient l'impact d'une démonstration d'engrais sur l'utilisation de l'engrais par les fermiers. Tous les fermiers bénéficient de l'engrais, mais une partie seulement de ces fermiers est sélectionnée aléatoirement pour assister à une démonstration de l'utilisation de cet engrais. Comme l'invitation est assignée de façon aléatoire, cela fournit une variable instrumentale avec laquelle on peut évaluer l'impact du traitement.

3.2.3 « Phase-In »

La randomisation peut être introduite graduellement dans le temps. Les expériences randomisées durent souvent plus d'une année. Il est possible de diviser le temps, et d'introduire la randomisation en fonction de cette séparation. Par exemple, l'évaluation fameuse de Kremer et Miguel (2004) discutée plus haut s'est étalée sur quatre ans (de 1998 à 2002), et s'est intéressée à trois groupes d'école. Ils ont sélectionné soixante-quinze écoles, chaque groupe étant constitué de vingt-cinq écoles. Le premier groupe a bénéficié du programme en 1998, pendant que les groupes 2 et 3 ne recevaient rien. En 1999, le groupe 2 a lui aussi bénéficié du traitement, pendant que seul le groupe 3 n'en bénéficiait pas. Enfin, en 2000, le groupe 3 a reçu le traitement. L'impact du programme a été évalué en comparant les résultats du groupe 1 à ceux du groupe 2 et 3 en 1998 et en comparant les résultats du groupe 1 et 2 en 1999 à ceux du groupe 3. L'attribution du traitement s'est donc faite graduellement dans le temps, ce qui soulève cependant un certain nombre de questions, par exemple : est-il possible que les individus changent leurs anticipations en vue du traitement futur ? Si tel est le cas, cela introduirait un biais important, puisque les individus transformeraient leurs comportements. Ou encore, est-ce que les effets du traitement sont suffisamment rapides pour qu'ils se produisent avant que le dernier groupe ne soit traité ? Si tel n'était pas le cas, la comparaison entre les différents groupes n'aurait aucun sens puisque les effets du traitement ne sont observés nulle part.

3.2.4 *Randomisation au sein des groupes*

La randomisation, on l'a vu, permet de sélectionner aléatoirement deux groupes. Mais il est possible d'introduire un niveau de randomisation supplémentaire, en l'appliquant au sein de sous-groupes. Par exemple, lorsqu'une école entière ne veut pas participer à une expérience, le programme est donné aléatoirement à des sous-groupes au sein de ces écoles. Un exemple est ici l'évaluation de Banerjee, Duflo, Cole et Linden (2007) en Inde qui s'intéresse à l'effet d'un enseignant supplémentaire sur le résultat des élèves¹⁰⁷. Les écoles ne souhaitent pas y participer entièrement, c'est pourquoi au sein des écoles qui ont participé,

¹⁰⁷ Je détaillerai longuement cette évaluation dans le chapitre suivant.

le programme était donné, dans certaines écoles à des élèves de degré 3 et dans d'autres écoles de degré 4¹⁰⁸.

3.3 Limites (internes) de la randomisation

Malgré son caractère de *gold standard* méthodologique, la mise en place de la randomisation rencontre, tout de même, certaines limites. Ces limites tiennent à la fois de la validité interne et de la validité externe de la randomisation. Je m'attacherai, ici, uniquement aux problèmes relevant de la validité interne de la randomisation¹⁰⁹. La mise en place de la randomisation peut transformer les comportements des individus. Les individus faisant partie du groupe de traitement peuvent agir plus favorablement car ils savent qu'ils font partie du groupe traité ; alors que les individus du groupe de contrôle peuvent agir défavorablement compte tenu du fait qu'ils ne bénéficient pas du programme. On parle pour le premier cas d'effets Hawthorne et pour le second d'effet John Henry (3.3.1). Une seconde limite relève du fait que les individus ne souhaitent pas toujours participer à l'expérience, entraînant alors un biais de sélection (3.3.2). Je développerai, ici, ces deux limites.

3.3.1 Les effets Hawthorne et John Henry

L'une des premières limites vient du fait que les individus peuvent changer, transformer leurs comportements, lorsqu'ils savent qu'ils font partie d'une expérience, soit pour faire plaisir à l'expérimentateur ou le contraire. Ces deux réactions sont associées à ce que l'on appelle l'effet Hawthorne et l'effet John Henry, respectivement. L'effet Hawthorne¹¹⁰ concerne le groupe traité : lorsqu'il est reconnaissant d'être traité, il agit comme il pense que les expérimentateurs voudraient qu'il agisse ; ce qui biaise les résultats de l'expérience puisque ce n'est pas cette motivation comportementale que les expérimentateurs souhaitent étudier.

¹⁰⁸ Le degré trois correspond à la troisième année d'éducation élémentaire. Le système scolaire Indien est construit sur la base de huit années d'études élémentaires et de deux années d'études secondaires. Les cinq premières années sont des années d'éducation élémentaire de base, elles sont divisées en cinq degrés, le premier degré correspondant à la première année. Il y a ensuite trois années d'éducation élémentaires supérieures, toujours séparées en degré, plus celui-ci est élevé plus le niveau d'étude l'est aussi.

¹⁰⁹ La partie 2 de cette thèse portera sur la question de la validité externe de la randomisation.

¹¹⁰ L'effet Hawthorne tire son nom d'expériences réalisées de 1924 à 1932 à Chicago au sein d'une usine électrique, la « *Hawthorne Works* ». Ces expériences s'intéressaient à la productivité des ouvriers de l'entreprise.

L'effet John Henry¹¹¹ renvoie au groupe de contrôle, dont les membres en réaction hostile au fait de ne pas faire partie du groupe de traitement, peuvent changer leurs comportements¹¹².

Duflo (2005) envisage deux possibilités pour contrecarrer ces effets. D'une part, il s'agit de mener des études d'impact à long terme, comme continuer d'observer les effets du programme, lorsque l'expérience est terminée mais que le programme est toujours en place¹¹³. D'autre part, les évaluations peuvent être conçues de telles façons qu'elles empêchent ces effets. Un des exemples que cite Duflo et Banerjee (2009) est l'expérience de Ashraf, Karlan et Yin¹¹⁴, qui évalue le programme SEED (qui concerne l'engagement à épargner). Dans cette étude, les auteurs sont soucieux du fait que certains individus peuvent avoir épargné plus car une équipe leur avait suggéré d'épargner plus et non à cause du programme SEED. Pour contrecarrer cela, ils ont créé un groupe de traitement supplémentaire à qui ils adressaient un programme d'épargne normal (c'est-à-dire sans engagements comme le SEED).

3.3.2 Tout le monde ne veut pas participer

Comme le souligne Heckman (1991), il peut arriver que la population ne souhaite pas participer à l'expérience. Ceci introduit un biais important car si certaines personnes, au sein de la population choisie, ne souhaitent pas participer, elles sont exclus d'emblée, ce qui introduit un biais de sélection – que la randomisation est sensé annuler. La sélection totalement aléatoire au sein de la population ne peut avoir lieu. Et comme le note Duflo et Banerjee : « il est vrai que toutes les ONG ou tous les gouvernements ne sont pas volontaires pour participer à des évaluations aléatoires » (Banerjee et Duflo, 2009, 707). Mais selon Banerjee et Duflo, ce problème tend à être, désormais, moins important dans les pays en développement. Le travail avec des partenaires de terrain en lesquels la population a confiance, permet de diminuer ce biais. À une échelle plus grande, pour les ONG ou les gouvernements qui ne souhaitaient pas participer, le travail des grandes institutions, comme la Banque mondiale, avec ces gouvernements ou ONG permet, elle aussi, de réduire ce biais. Duflo et Banerjee nuancent, alors, la limite lancée par Heckman :

¹¹¹ Le terme d'effet John Henry a été introduit par Gary Saretsky. John Henry est un héros américain, il représente de façon mythique la classe ouvrière américaine du XIX^{ème} siècle. La légende raconte que pour sauvegarder son emploi John Henry défie le propriétaire de chemin de fer pour lequel il travaille. John Henry remporte le défi mais en meurt d'une crise cardiaque.

¹¹² Par exemple un enseignant du groupe de contrôle peut choisir de paresser car il ne fait pas partie du groupe de traitement.

¹¹³ L'expérience de Duflo, Hanna et Ryan (2012) est un bon exemple de cette pratique.

¹¹⁴ Voir : (Ashraf, Karlan, et Yin, 2006).

« Notre expérience suggère que, dans le contexte des pays en développement, ce problème tend à disparaître car les évaluations aléatoires sont de mieux en mieux acceptées, et les partenaires se multiplient. La situation va continuer de s'améliorer si les évaluations aléatoires sont recommandées par les donateurs dans la mesure où les partenaires se diversifieront encore davantage. C'est déjà ce qui est en train de se passer. De nombreux chercheurs et acteurs de terrain de la Banque mondiale travaillent notamment avec les gouvernements de pays en développement en vue de lancer un ambitieux programme d'évaluation ». (Banerjee et Duflo, 2009, 708)

Banerjee et Duflo sont ouverts au problème soulevé par Heckman (1991). Néanmoins, selon eux, ce problème est moins important dans les pays en développement grâce à la collaboration que leur laboratoire entretient avec le terrain.

4. Conclusion

Le J-PAL est désormais un laboratoire central en économie du développement. Sa structure, tout comme son organisation lui permettent d'être extrêmement présent sur le terrain, ainsi que dans de nombreuses universités. Etudier le cadre institutionnel du J-PAL permet d'offrir une explication à son essor. De plus, cela offre un éclairage particulier sur le fonctionnement d'un tel laboratoire, ce qui donne alors une définition précise de ce dernier. Le J-PAL se définit par l'utilisation et la promotion de la randomisation. La structure organisationnelle du J-PAL montre comment le J-PAL promeut sa méthode, en étant très présent sur le terrain et dans de nombreuses universités ou laboratoires de recherche. L'utilisation de la randomisation par le J-PAL s'inspire directement du modèle de Rubin (1974). En ce sens, elle permet d'offrir un contrefactuel comportant très peu de biais que d'autres méthodes ne parviennent pas à supprimer. Cette caractéristique lui offre le statut de *gold standard* méthodologique en matière d'évaluation d'impacts. De plus, l'utilisation de la randomisation par le J-PAL est aussi caractérisée par sa mise en pratique sur le terrain ; la dimension aléatoire peut être introduite de quatre façons différentes. Cette rencontre avec le terrain permet d'unir les deux éléments de définition du J-PAL : l'utilisation et la promotion de la randomisation. L'utilisation de la randomisation du J-PAL, par sa mise en place sur le terrain, implique une étroite collaboration avec les partenaires qui s'y trouvent. Et, cette étroite collaboration tend à promouvoir la randomisation du J-PAL et donc à offrir un élément d'explication à l'essor d'un tel laboratoire ces dix dernières années.

Chapitre 3 : L'approche du J-PAL et la médecine : de l'essai clinique à l'*evidence-based policy*

0. Introduction

« Il y a une tradition en économie qui refuse de se prononcer sur la politique et qui prétend écrire le monde tel qu'il est ; mais il y a aussi une tradition normative, dans laquelle notre travail s'inscrit, qui considère que l'économiste dans son domaine particulier a des choses à dire sur la meilleure façon de combler par exemple une défaillance du marché, ou de mieux organiser le fonctionnement d'un service public. Donc on ne s'interdit pas d'avoir des idées. Soit on a des idées nous-mêmes directement, auquel cas on fait des suggestions directement. Soit on utilise la connaissance de l'ensemble du domaine pour évaluer les idées des autres. » (Duflo in Mayneris, 2009, 14)

Les preuves collectées par le J-PAL sont dédiées aux décideurs politiques des pays en développement. C'est la deuxième partie de la maxime du J-PAL « transformer la recherche en action ». En ce sens, comme le souligne Duflo, le J-PAL porte en lui une vocation normative et ne souhaite donc pas uniquement décrire le réel. La rencontre avec le terrain et les différents partenaires du J-PAL offre à une telle approche une dimension politique. Le J-PAL s'inscrit dans le mouvement de la « politique fondée sur les preuves ». Ce mouvement fait suite à un courant, développé en médecine, cherchant à fonder les décisions médicales sur des preuves provenant d'essais cliniques. L'objectif de ce chapitre est de définir la dimension normative du J-PAL ; autrement dit, de définir son second objectif - guider la décision politique à l'aide de preuves.

Dans le but de définir cette dimension normative, il semble nécessaire d'opérer un double aller-retour avec la médecine. D'une part, il apparaît fructueux de procéder à un premier aller-retour en vue de saisir le lien qui unit les essais cliniques médicaux à la randomisation du J-PAL. Les essais cliniques médicaux fournissent les preuves nécessaires à l'*evidence-based medicine*, et la randomisation du J-PAL entend faire de même au sein de l'*evidence-based policy*. Faire ici le parallèle avec la médecine semble alors particulièrement éclairant. D'autre part, il est utile de mener un second aller-retour avec la médecine afin de déterminer précisément ce qui constitue l'*evidence-based policy*. L'*evidence-based medicine* étant à l'origine de l'*evidence-based policy*, s'intéresser à la naissance de ce courant en médecine permet d'en saisir les principaux enjeux qui tendent à tramer ceux de l'*evidence-based policy*. Ces deux allers-retours avec la médecine offriront un cadre précis afin de définir précisément

le rôle du J-PAL au sein de l'*evidence-based policy*, et donc la dimension normative d'un tel laboratoire.

Dans la première partie de ce chapitre, je tâcherai de montrer que la randomisation du J-PAL est analogue à celle des essais cliniques médicaux. Le lien entre les deux a déjà été souligné lors du premier chapitre, mais ici je m'attacherai à définir précisément, non pas le premier essai clinique, mais la forme standard des essais cliniques médicaux. Je confronterai ensuite cette présentation à celle des expériences du J-PAL ; afin que les similarités entre les deux méthodes apparaissent saillantes. Dans la deuxième partie de ce chapitre, j'opérerai un second aller-retour avec la médecine, en développant le mouvement de l'*evidence-based medicine*. La présentation de ce mouvement me permettra de définir les fondements de l'*evidence-based policy*. Finalement, dans la troisième partie de ce chapitre, je montrerai comment le J-PAL s'insère au sein de l'*evidence-based policy*.

1. Essais cliniques versus randomisation en économie du développement

Peter Piot, épidémiologiste et ancien directeur de l'ONUSIDA¹¹⁵, a succédé à Esther Duflo à la chaire annuelle du Collège de France « Savoir contre pauvreté »¹¹⁶. Il peut paraître étrange qu'un médecin reprenne une chaire sur la pauvreté. Cela l'est moins quand on sait qu'Esther Duflo l'a précédé. La relation entre santé et pauvreté est aujourd'hui explicite. La santé constitue l'un des éléments majeurs du développement, comme l'illustre l'« Indice de Développement Humain » (IDH), dans lequel elle constitue l'un des trois piliers. Cela étant, la relation unissant les deux disciplines - la médecine et l'économie du développement - qui traitent de deux objets distincts - la santé et la pauvreté - reste à démontrer. Il a déjà été noté que la randomisation trouve son essor au sein des essais cliniques en médecine ; le premier essai clinique mené par Hill et Doll en 1950 a été présenté. Ici, nous nous intéresserons, plus particulièrement, à l'essai clinique dans son sens moderne, c'est-à-dire rattaché à la thérapeutique. Mon objectif sera de montrer que la randomisation utilisée par les chercheurs du J-PAL a pour vocation de s'approcher au plus près de celle utilisée en médecine ; et donc de s'éloigner de celle utilisée dans les expériences sociales de la moitié du XXe siècle. Pour cela, je développerai, dans un premier temps, la structure des essais cliniques modernes (1.1) ;

¹¹⁵ L'ONUSIDA est un programme de l'ONU cherchant à coordonner les différentes actions de l'ONU dans la lutte contre le sida. Le directeur de l'ONUSIDA est désormais Michel Sibidé.

¹¹⁶ Cette chaire est créée en 2009 en partenariat avec l'agence française de développement (AFD), Esther Duflo inaugure cette chaire la même année.

puis, je montrerai en quoi la randomisation du J-PAL peut être comprise comme un essai clinique (1.2).

1.1 Essai clinique en médecine

L'objectif d'un essai clinique, cela a été soulevé, est d'évaluer l'efficacité d'un traitement. Bertran Auvert (2005) cherche, par exemple, à mesurer l'impact de la circoncision dans la transmission du VIH¹¹⁷. A l'inverse de l'essai de Hill et de Doll, la randomisation est au cœur de cet essai clinique. La structure est la même que celle définie par Fisher (1935) précédemment. Cependant, un essai clinique n'est qu'une partie d'un essai plus large, l'essai thérapeutique. Afin de définir au mieux les essais cliniques standards en médecine, je montrerai, tout d'abord, comment l'essai clinique s'insère au sein de l'essai thérapeutique (1.1.1). Puis, j'ajouterai à la structure expérimentale de Fisher (1935) quelques dimensions méthodologiques qui offrent aux essais cliniques une très forte validité interne (1.1.2). Enfin, j'illustrerai tout cela à l'aide de l'essai de Bertran Auvert (1.1.3).

1.1.1 L'essai clinique : une partie d'un essai plus large

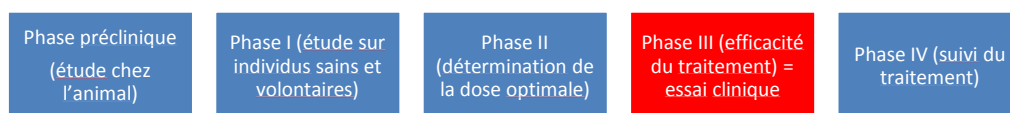
Un essai clinique ne constitue, en fait, qu'une partie d'un essai plus large : l'essai thérapeutique¹¹⁸. L'essai clinique est la dimension quantitative de l'essai thérapeutique. Un essai thérapeutique est constitué de cinq phases : la phase préclinique (1), la phase I (2), la phase II (3), la phase III (4) et la phase IV (5). La phase préclinique précède le passage du traitement de l'animal à l'homme, elle vise à définir « l'effet principal du produit ou effet thérapeutique (dose efficace, type de durée de l'effet sur l'organe isolé ou dans différentes espèces animales), son mécanisme (mise en jeu de médiateurs, de récepteurs, etc.), ses effets secondaires (dits indésirables lorsqu'ils sont nocifs) et les doses relatives entraînant effet principal et effets secondaires. » (Bouvenot, 2006, 2) Cette première phase s'adresse donc particulièrement à l'évaluation d'un médicament. La phase I renvoie à l'étude de la première administration chez l'homme : « elle vise à déterminer les conditions de tolérance humaine » (Bouvenot, 2006, 6). Elle a pour objectif de mettre en évidence les effets secondaires et l'effet du traitement souhaité. Elle est souvent menée sur un groupe d'individus volontaires et sains,

¹¹⁷ Cet essai clinique est le premier essai médical sur le rôle de la circoncision dans la transmission du VIH. Il a depuis, été largement repris.

¹¹⁸ La thérapeutique renvoie à la partie de la médecine qui s'intéresse aux moyens de guérir ou soulager les maladies. Un essai thérapeutique vise à tester un de ces moyens particulier, par exemple un médicament ou une pratique médicale. Pour une analyse du « raisonnement médical », voir (Masquelet, 2006).

si la toxicité du médicament n'est pas trop élevée. Cette phase concerne toujours l'évaluation d'un médicament ou d'une pratique médicale qui pourrait engendrer une certaine toxicité. La phase II étudie spécifiquement l'efficacité du traitement, elle s'adresse alors aussi bien à une pratique médicale qu'à un médicament : « elle vise à déterminer les conditions de l'efficacité et à définir les modalités thérapeutiques optimales » (*Ibid.*). Elle se fait sur un petit groupe d'individus, malades ou sains, et elle a pour principal objectif de déterminer la dose optimale qui sera utilisée dans la phase suivante. La phase III étudie l'efficacité du traitement avec la dose optimale : elle se fait sur un groupe homogène d'individus malades et suit un protocole précis (dépendant de la structure de Fisher). C'est cette phase qui caractérise l'essai clinique. Le but de cette phase est de pouvoir comparer les effets du traitement. C'est ici que la randomisation a lieu. Parmi la population, deux groupes sont constitués aléatoirement : un qui recevra le traitement et un qui ne recevra rien. La dernière phase, la phase IV, est conduite une fois le traitement terminé et s'intéresse à la mise sur le marché du médicament. Elle a pour principal objectif de déceler des effets secondaires rares et de suivre les effets du traitement à long terme.

Seule la phase III est un « essai clinique » en tant que tel. Le tableau suivant illustre les différentes phases de l'essai thérapeutique. Et cet essai thérapeutique fait suite à un processus clinique qui a permis de déterminer, à l'aide d'un diagnostic fait en amont, un remède particulier. L'essai thérapeutique ne vise qu'à évaluer ce remède.



1.1.2 La méthode d'un essai clinique

En plus de la structure définie par Fisher, un essai clinique peut être composé de plusieurs études - on en dénombre six. Pour que l'essai clinique soit parfaitement fiable, il faut qu'il englobe la totalité de ces études, même si peu d'essais cliniques parviennent à mener l'ensemble de ces études simultanément. L'étude doit être prospective (1) : elle vise à constituer le scénario des possibles et construit, de ce fait, un contrefactuel. L'étude doit aussi être contrôlée (2)¹¹⁹, c'est-à-dire qu'elle doit être constituée par un groupe de contrôle et un

¹¹⁹ Le fait que l'essai soit contrôlé permet de réduire le biais de confusion. Ce biais traduit la confusion entre les effets du traitement et les conséquences de la maladie.

groupe témoin. Ou encore, l'étude peut être randomisée (3). On peut aussi mener un essai clinique en aveugle (4)¹²⁰. Le patient ne sait pas s'il fait partie du groupe témoin ou du groupe de contrôle, aucun facteur psychologique ne peut donc rentrer en compte. Lorsque l'expérimentateur n'est pas non plus informé des traitements qu'il donne, on parle d'étude en double aveugle. Si tous les acteurs connaissent les différentes appartenances au groupe témoin ou de contrôle, il s'agit d'une étude ouverte. L'étude est dite multicentrique (5) si elle se déroule dans plusieurs lieux au même moment. Enfin l'étude peut être menée « en intention de traiter » (6) : les participants continuent à être étudiés même s'ils n'ont pas été jusqu'à la fin du protocole, ou d'autres groupes sont ajoutés, pour réduire le biais d'attrition¹²¹. Si tel n'est pas le cas, on parle d'étude *per protocole*. Pour que l'essai clinique soit le plus « solide » possible, il faut qu'il soit constitué de ces six études. Mais il peut aussi être réalisé par des combinaisons diverses entre ces études. Ronald Fisher a donc défini la structure de ces expériences, qui ont ensuite été couplées à différents types d'études.

1.1.3 Sida et circoncision : illustration d'un essai clinique médical

Afin de rendre saillantes les similitudes entre la méthode utilisée par le J-PAL et celle des essais cliniques médicaux, je vais présenter, brièvement, un essai clinique médical récent. Cet essai est réalisé par Bertran Auvert *et al.* (2005)¹²² en Afrique du Sud de juillet 2002 à février 2003 et vise à évaluer les effets de la circoncision dans la transmission du VIH¹²³. L'essai porte sur 3274 hommes qui ont entre 18 et 24 ans et qui ne sont pas circoncis. Ces 3274 hommes ont été attribués aléatoirement au groupe de contrôle et au groupe de traitement. Au sein de ce dernier groupe, les participants devaient se faire circoncire dans la semaine. Les frais d'intervention chirurgicale étaient financés par l'expérience. Ces frais étaient aussi pris en charge pour les participants du groupe de contrôle à la fin de l'expérience (en février 2003). Chaque participant des deux groupes a eu trois visites médicales pendant la période,

¹²⁰ Mener un essai clinique en double aveugle permet de réduire les biais de suivis. Ce biais traduit le fait qu'il pourrait exister une différence de prise en charge entre le groupe de traitement et le groupe de contrôle, compte tenu du fait que le personnel médical connaît le groupe traité.

¹²¹ Un biais d'attribution représente la différence entre les groupes initiaux (lors de la mise en place de l'essai) et les groupes finaux. Cette différence est importante lorsque des individus abandonnent l'essai clinique.

¹²² Cet essai est, désormais, largement cité par les chercheurs du J-PAL. Une expérience randomisée a d'ailleurs été menée par les chercheurs du J-PAL dans ce domaine. Cette expérience évaluait l'impact d'une campagne de prévention (qui expliquait les effets positifs de la circoncision dans la transmission du VIH) sur les comportements sexuels des individus. Voir (Godlonton, *et al.*, 2012)

¹²³ L'essai clinique d'Auvert est le premier essai clinique qui évalue l'impact de la circoncision, d'autres essais ont depuis été réalisés ; un en Ouganda voir, (Gray, R.H, *et al.*, 2007), et un autre au Kenya, voir (Bailey, *et al.*, 2007).

une au troisième mois, une autre au douzième mois et une dernière à la fin de l'expérience, c'est à dire au 21^{ème} mois. Durant ces visites, les participants répondaient à un questionnaire, faisaient une prise de sang, avaient un examen génital puis bénéficiaient d'une consultation individuelle. Durant l'essai, 49 personnes du groupe de contrôle ont contracté le virus du sida alors que dans le groupe de traitement, 20 personnes ont été infectées. Cela revient à dire, selon les auteurs, que la circoncision a empêché six des dix possibles infections. L'essai clinique conclut que « *MC [Man Circumcision] provides a degree of protection of acquiring HIV infection equivalent to what a vaccine of high efficacy would have achieved* ». (Auvert *et al.*, 2005, 1120) L'essai d'Auvert *et al.*, afin de réduire au maximum tous les biais et d'offrir un résultat fiable, est donc à la fois randomisé (1), contrôlé (2) et prospectif (3) ; permettant d'annuler les biais de sélection et de confusion. Par contre, il n'a pas été multicentrique (4) car il ne s'est réalisé simultanément que dans une région du monde, ici, en Afrique du sud. L'essai a été mené en intention de traité (5) car certains hommes sont sortis de l'essai et l'étude a été ouverte (6).

1.2 Une nouvelle forme d'essai clinique : l'évaluation des programmes d'aide au développement

La méthode des chercheurs du J-PAL, cela a été soulevé, consiste à placer la randomisation au centre de leur démarche. Mais leur ambition est plus grande, ils souhaitent se rapprocher au plus près de la méthode des essais cliniques :

« Le principe général c'est de s'approcher au mieux de la méthode de l'essai clinique. » (Duflo dans Mayneris, 2009, 2)

« Il faut adosser le développement de la santé et de l'éducation dans le monde à une technologie de l'évaluation et poser la question du choix : comment déterminer la meilleure politique, celle qui sera la plus efficace pour parvenir au but qu'on s'est fixé ? Cette question a des répercussions très concrètes. L'école doit-elle être gratuite ou payante ? Quelle est la taille optimale des classes ? Faut-il construire des centres de santé tout près des villages ou acheminer les malades des zones rurales vers des hôpitaux urbains ? Pour choisir entre différentes options, l'intuition et le raisonnement *in abstracto* sont des guides bien incertains. La seule solution consiste à tester rigoureusement chacune de ces politiques et à en comparer le prix ainsi que les effets. ». (Duflo, 2010a, 16)

Ici, je chercherai à creuser cette analogie entre les essais cliniques médicaux et la démarche du J-PAL. Cette analogie me permettra de faire un lien explicite entre le

mouvement de l'*evidence based-medicine* et celui de l'*evidence based policy*. Dans ce but, je comparerai l'essai d'Auvert à une expérience du J-PAL, afin de rendre saillant les similarités entre les deux (1.2.1). Ensuite, je m'attacherai à montrer que les expériences du J-PAL ne s'apparentent qu'à la phase III de l'essai thérapeutique tout comme l'essai clinique, ce qui insiste, encore une fois, sur la similarité des deux méthodes (1.2.2).

1.2.1 La méthode des essais cliniques en économie du développement

Afin d'illustrer au mieux les liens entre la randomisation utilisée par les chercheurs du J-PAL et les essais cliniques médicaux, nous partirons d'une étude randomisée qu'Esther Duflo a menée avec Abhijit Banerjee, Shawn Cole et Leigh Linden (2007), en Inde, de 2001 à 2004, avec l'ONG Pratham¹²⁴.

L'un des deux programmes envisagés par l'ONG consiste à engager des jeunes filles qui ont terminé leurs études secondaires, afin d'aider les enfants en difficulté scolaire - ceux qui n'ont pas acquis les bases de lecture et d'écriture¹²⁵. Les jeunes filles travaillent avec quinze à vingt élèves en difficulté pendant deux heures chaque semaine. Ce programme avait déjà été mis en place par Pratham en 1998 à Bombay. L'expérience l'évalue, dans un premier temps, de 2001 à 2002, dans une autre ville indienne, à Vadodara. Cinquante-huit écoles parmi les cent vingt-deux que compte la ville participent à cette évaluation. Ces cinquante-huit écoles sont divisées en deux groupes : le groupe A, dont font partie les élèves de degré trois¹²⁶ et le groupe B dont font partie les élèves de degré quatre. A l'intérieur du groupe A, deux groupes sont sélectionnés aléatoirement : l'un recevra le programme (groupe de contrôle) et l'autre ne le recevra pas (groupe témoin). La même sélection aléatoire est faite pour le groupe B.

Ce programme a été évalué l'année suivante, de 2002 à 2003. Cette fois-ci, L'attribution du programme est inversée : les enfants du groupe A, qui ne recevaient pas le programme, en bénéficient désormais, et ceux qui en bénéficiaient ne le reçoivent plus. Le groupe B subit la même inversion. L'année suivante, de 2003 à 2004, aucun des groupes ne recevait les

¹²⁴ Cette évaluation a déjà été évoquée lors du chapitre précédent.

¹²⁵ Le second programme vise à fournir des ordinateurs aux écoles en vue d'impliquer de, façon ludique, les enfants dans l'apprentissage. Nous reviendrons sur ce dernier dans la deuxième partie de notre travail.

¹²⁶ Pour rappel : le degré trois correspond à la troisième année d'éducation élémentaire. Le système scolaire Indien est construit sur la base de huit années d'études élémentaires et de deux années d'études secondaires. Les cinq premières années sont des années d'éducation élémentaire de base, elles sont divisées en cinq degrés, le premier degré correspondant à la première année. Il y a ensuite trois années d'éducation élémentaires supérieures, toujours séparées en degré, plus celui-ci est élevé plus le niveau d'étude l'est aussi.

programmes. Ce protocole a mesuré l'impact du programme. Il se fonde, tout comme en médecine, sur une étude contrôlée : la population participant à l'évaluation est séparée en deux groupes - l'un de contrôle et l'autre de traitement. C'est aussi une évaluation prospective : elle vise l'avenir, et envisage le « scénario des possibles ». Enfin, elle est faite par randomisation : le programme est assigné à un groupe de façon aléatoire.

Afin de voir si ce programme était généralisable, il a été évalué, durant la même période et selon le même schéma, à Bombay - la première ville où il avait été mis en place par l'ONG en 1998. Soixante-dix-sept écoles ont été ajoutées à l'évaluation. Les élèves de degré trois constituaient le groupe C. La première moitié, choisie aléatoirement, bénéficiait du programme et l'autre moitié ne le recevait pas. Les élèves de degré quatre constituaient le groupe C, seule la moitié du groupe disposait du programme. L'année suivante, comme à Vadodara, ceux qui bénéficiaient du programme n'en bénéficient plus et ceux qui ne le recevaient pas le reçoivent, et enfin l'année d'après, plus aucun groupe ne le reçoit. Les résultats de l'évaluation ont été considérés comme satisfaisants:

« The Balsakhi program appears to be succesful: in all years, for both subjects, and in both cities, and for all subgroups, the difference in post-test scores between treatment and comparison groups is positive and, in most instance significant. » (Banerjee, Cole, Duflo et Linden, 2007, 10)

Mais cette évaluation a rencontré une difficulté majeure : beaucoup d'enfants assignés au groupe qui recevait le programme ne l'ont, en réalité, pas reçu. Les chercheurs ont alors ajouté un troisième groupe, afin de réduire au maximum le biais d'attrition - causé par le fait que certains enfants, en ne recevant pas le programme, étaient en dehors de l'expérience. Ils ont alors mené une étude « en intention de traiter ». On retrouve, avec cette évaluation, cinq des six dimensions qui peuvent constituer un essai clinique : l'expérience de Banerjee, Cole, Duflo et Linden (2007) est prospective, contrôlée, randomisée, multicentrique (l'évaluation a eu lieu simultanément à Bombay et à Vadodara), et elle est faite « en intention de traiter ». Mais elle n'a pu être réalisée en aveugle ou en double aveugle, tous les participants étant informés sur la nature du traitement qu'ils reçoivent, ce dernier est visible¹²⁷. L'approche du J-PAL, par la nature des évaluations qu'elle mène en économie du développement, s'apparente, dans le cas que l'on vient de discuter, parfaitement à la méthode d'un essai clinique.

¹²⁷ Pour une analyse de l'importance du rôle des études en aveugle ou en double aveugle en médecine et en sciences sociales, voir (Teira, 2013a ; 2013b).

1.2.2 Nouvelle forme d'essai clinique et pratique

A l'instar des essais cliniques médicaux, la randomisation du J-PAL ne s'apparente qu'à la phase III d'un essai thérapeutique. En effet, les évaluations par assignation aléatoire du J-PAL peuvent évaluer un projet pilote ou un programme déjà existant dans le but de l'étendre. Le projet pilote est la première évaluation d'un programme avant qu'il soit mis en place : « avant qu'un programme ne soit lancé à grande échelle, un projet, nécessairement d'envergure limitée, est souvent mis en place » (Duflo, 2005, 193). Si les résultats sont jugés satisfaisants, le programme peut être réévalué afin d'être étendu.

Un des programmes pilotes souvent pris pour exemple par les chercheurs du J-PAL¹²⁸ est une étude de Banerjee, Jacon et Kremer (2001), menée en Inde avec l'ONG Seva Mandir. Celle-ci désirait évaluer l'impact d'un enseignant supplémentaire dans les centres d'éducation qu'elle a créés. Le niveau d'absentéisme des élèves et des enseignants étant extrêmement élevé dans les pays en développement, et l'objectif du programme était de le réduire afin d'augmenter les performances des élèves. Ce programme a eu un effet sur l'absentéisme mais pas sur les performances :

« Le programme a réduit les nombres de jours de fermeture des écoles : les écoles avec un enseignant sont fermées 39% du temps alors que les écoles avec deux enseignants sont fermées 24% du temps. L'assiduité des filles a augmenté de 50%. Cependant, il n'y a eu aucune différence dans les résultats de tests. » (Duflo, 2005, 194)

Ce programme n'a pas été étendu car les résultats ont été jugés trop insuffisants. Si les résultats du programme avaient été jugés satisfaisants, il aurait été réévalué et étendu. Une des limites majeures de cette méthode est de réfléchir uniquement en équilibre partiel, la généralisation est donc souvent problématique¹²⁹. Lorsque l'on déplace un programme d'une région à une autre, les caractéristiques des individus changent. Un programme qui a été « efficace » dans une région particulière peut s'avérer « inefficace » dans une autre région. La population n'est pas la même, notamment parce que l'environnement social est différent. C'est pourquoi les programmes sont souvent réévalués avant d'être étendus. Il est alors possible, comme en médecine, de mener des études multicentriques - dans plusieurs régions différentes en même temps - afin de renforcer la solidité de l'évaluation et de ses résultats.

¹²⁸ Par exemple, voir (Duflo, 2005).

¹²⁹ Je reviendrai plus amplement sur cette limite dans la partie suivante de ce travail.

Sinon, il est possible de mener, dans un premier temps, une évaluation pilote, puis de réévaluer le programme dans d'autres régions : on parle alors d'expansion de projet pilote.

Duflo (2005) prend comme exemple d'expansion de projet pilote un programme mené par le gouvernement mexicain, le PROGRESA¹³⁰, dont l'idée était de distribuer des bourses d'études aux jeunes filles en fonction de leur niveau d'assiduité scolaire et de leur capacité à suivre des programmes de mesures préventives (en matière de santé surtout). Le programme, dans un premier temps, a été évalué dans une seule région, sous forme de projet pilote (le budget du gouvernement ne permettait pas de faire plusieurs évaluations en même temps et donc de mener une évaluation multicentrique). Les résultats de ce projet pilote ont été jugés très satisfaisants en matière d'éducation et de santé :

« Les enfants connaissent une réduction de 23% dans l'incidence de la maladie, une augmentation de 1-4% de leur taille et une réduction de 18% de l'anémie. Les adultes connaissent une réduction de 19% du nombre de jours perdus à cause de la maladie. Schultz (2001) trouve une augmentation moyenne de 3,4% de la scolarisation pour tous les élèves des niveaux 1 à 8 ; l'augmentation était la plus forte parmi les filles ayant obtenu le niveau 6, avec 14,8%. » (Duflo, 2005, 195)

Ce programme a été élargi au Mexique, même lorsque le gouvernement a changé en 2000. Les évaluations du J-PAL se font, généralement, en deux étapes principales : le projet pilote et l'expansion du programme. Ces deux étapes s'articulent toujours autour d'une évaluation randomisée. En ce sens, cette approche n'utilise de l'essai thérapeutique que la randomisation, autrement dit la phase III : celle de l'essai clinique. C'est pourquoi les évaluations du J-PAL peuvent être comprises comme des essais cliniques. Il n'y a pas de tests en amont, en laboratoire, par exemple (phase pré-clinique), il n'y a pas de détermination de la tolérance du traitement et de ses modalités thérapeutiques (phase I et II) et il n'y a pas de suivi du programme à long terme (phase IV), qui pourrait mettre en évidence des effets non visibles à court terme. La nouvelle forme d'essai clinique que développe le J-PAL et les autres chercheurs du J-PAL emprunte à la médecine la totalité de ce qui caractérise un essai clinique - les différents types d'études possibles et une seule des phases de l'essai thérapeutique. L'approche du J-PAL n'est donc pas seulement analogue aux essais cliniques médicaux : elle est en elle-même un essai clinique. Elle s'en distingue par l'objet qu'elle étudie - la pauvreté. De plus, elle s'en distingue par le fait que la randomisation n'est pas

¹³⁰ D'abord évalué en 2001 par Paul Gertler et Simone Boyce, voir (Gertler et Boyce, 2001).

encastrée dans un processus plus large comme l'est l'essai clinique médical au sein de l'essai thérapeutique. C'est pourquoi je parle de nouvelle forme d'essais clinique.

2. De l'evidence based medicine à l'evidence based policy

Les essais cliniques sont devenus un élément central du mouvement de l'*evidence-based médecine*, au début des années 1990. Ce terme renvoie à deux impératifs. Le premier est de s'intéresser à des preuves scientifiques (c'est ici que les essais cliniques jouent un rôle central). Le second impératif est clinique, avoir conscience de la spécificité de chaque patient et donc posséder une expertise clinique. Cet aspect permet, pour Sackett (un des fondateurs de l'*evidence-based médecine*) d'éviter de donner aux preuves la première position et donc d'oublier le rôle du patient¹³¹. Ce mouvement de la médecine fondée sur les preuves a ensuite donné lieu à un mouvement similaire ; mais, cette fois-ci, dans le domaine politique : l'*evidence-based policy*. L'idée est la même qu'en médecine, il s'agit de fonder les décisions politiques sur des preuves. C'est donc le lien entre *evidence-based medicine* et *evidence-based policy* qui m'intéressera ici. D'une part, car il permet de saisir les fondements de l'*evidence-based policy* ; et, d'autre part en ce qu'il permet d'éclairer une des volontés du J-PAL - guider la décision politique. J'opérerai alors, encore une fois, un aller-retour avec la médecine. Il apparaît nécessaire de saisir ce qu'est exactement l'*evidence-based medicine* et le rôle des essais cliniques au sein de ce mouvement, afin de définir précisément l'*evidence-based policy*, en vue d'éclairer une des volontés du J-PAL. Dans ce but, je présenterai tout d'abord le mouvement de l'*evidence-based médecine* (2.1), puis je développerai celui de l'*evidence-based policy* (2.2).

2.1 Le paradigme de la médecine fondée sur les preuves

Le terme *evidence based-medicine* a été utilisé pour la première fois par Gordon Guyatt (1995) pour caractériser l'utilisation de preuves scientifiques dans la pratique de la médecine quotidienne. Cependant, c'est un texte écrit par une trentaine de chercheurs en 1992¹³² qui institue la médecine fondée sur les preuves. Ce texte est tiré des résultats d'un groupe de travail, l'*evidence-based Medicine Working Group*, dont l'une des figures principales est

¹³¹ « Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. » (Sackett et al., 1996, 72)

¹³² Voir (*Evidence-based Medicine Working Group*, 1992)

Sackett. Ce groupe de travail définit l'*evidence-based medicine* comme l'utilisation de preuves afin de prendre les décisions les plus pertinentes pour le patient :

« *Evidence based medicine is the conscientious, explicit, and judicious use of the current best evidence in making decisions about the care of individual patients.* » (Sackett et al, 1996, 71)

Ce mouvement lance alors un nouveau paradigme¹³³. Afin de rendre compte de ce nouveau paradigme, je développerai, tout d'abord, ce qui le fonde, c'est à dire la volonté de contrecarrer la faible utilisation des avancées médicale par les praticiens en vue de réunir la médecine de ville et la recherche médicale (2.1.1). Ensuite je m'attacherai aux différentes étapes qui définissent l'*evidence-based medicine* (2.1.2) ainsi qu'au niveau de preuves défini par un tel mouvement (2.1.3). C'est ici que l'essai clinique joue un rôle central, car ce dernier offre les preuves considérées comme les plus « solides ». Enfin, je développerai brièvement quelques limites adressées à l'*evidence-based medicine* (2.1.4).

2.1.1 Contrecarrer une faible utilisation des connaissances scientifiques en unissant la médecine de ville et la médecine scientifique

De nombreuses enquêtes réalisées auprès de médecins généralistes ont montré que la plupart d'entre eux n'avaient pas connaissance des avancées scientifiques médicales. Leurs pratiques quotidiennes de la médecine se rapprochent alors plus de la routine et se fondent sur leurs anciennes connaissances universitaires, mais malheureusement ces dernières n'ont pas été réactualisées. Cette situation est alors dommageable pour les patients, qui ne peuvent pas bénéficier des avancées médicales :

« Les patients n'avaient donc pas la garantie de profiter du traitement le plus adéquat, ni de bénéficier d'une information complète sur les solutions thérapeutiques disponibles et sur leur efficacité. Mais les médecins interrogés ont fait valoir qu'il leur était impossible de suivre une littérature scientifique toujours plus foisonnante, d'en faire des synthèses, d'en actualiser les critiques et d'en évaluer l'efficacité pour la pratique clinique. » (Laurent et al., 2009, 855)

C'est paradoxalement l'abondance d'avancées scientifiques, dans des domaines divers, et sa littérature toujours croissante qui tendent à rendre la pratique quotidienne de la médecine plus routinière ; car il semble trop ardu d'obtenir une information précise au sein d'une abondance

¹³³ « *A new paradigm for medical practice is emerging* ». (Evidence-based Medicine Working group 1992, 2420)

d'informations. Les initiateurs du mouvement de la médecine fondée sur les preuves, dans le but de contrecarrer cette pratique quotidienne et de la rendre plus « solide », mettent en avant plusieurs types d'actions à réaliser. Tout d'abord, ils proposent de transformer la formation universitaire en sensibilisant les futurs médecins à la recherche scientifique (1). Afin que les médecins ne soient plus noyés dans une littérature trop importante, ils envisagent de créer des analyses synthétiques (2) sur l'état des connaissances pour des questions précises. Pour que la qualité des « preuves » soit commune et donc unifiée, les initiateurs du mouvement de la médecine fondée sur les preuves, expliquent que les critères de qualité des preuves doivent être rendus explicites (3). Enfin, pour faciliter l'accès à ces analyses synthétiques, ils proposent de mettre en place une organisation collective (4).

La naissance de la médecine fondée sur les preuves est due à un manque d'accès à la connaissance scientifique, rendue trop opaque par sa précision et son abondance. L'objectif premier de la médecine scientifique est donc de rendre possible cet accès dans le but d'une pratique médicale quotidienne plus rigoureuse, chaque décision médicale doit être appuyée par une preuve scientifique. Il est alors nécessaire de trouver un moyen d'unir la médecine de ville à la médecine scientifique. Afin d'unir ces deux dimensions, la médecine fondée sur les preuves se conjugue autour de trois points : l'expertise du clinicien (1), le patient (2) et les données cliniques scientifiques (3). L'expertise du clinicien (1) relève de la pratique propre de chaque clinicien et se définit par son interaction avec le patient (2) :

« By individual expertise we mean the proficiency and judgement that individual clinicians acquire through clinical experience and clinical practice. Increased expertise is reflected in many ways, but especially in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients' predicaments, rights, and preferences in making clinical decisions about their care. » (Sackett et al., 1995, 71)

Les preuves scientifiques (3) sont offertes par la recherche médicale et sont classées selon une hiérarchie précise (que je développerai plus loin). Mais elles se fondent aussi sur les examens cliniques réalisés par le clinicien¹³⁴. Ici, le lien entre la recherche médicale et sa pratique apparaît saillant.

2.1.2 Les quatre étapes constitutives de la médecine fondée sur les preuves

¹³⁴ « By best available external clinical evidence we mean clinically relevant research, often from the basic sciences of medicine, but especially from patient centred clinical research into the accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic makers, and the efficacy and safety therapeutic, rehabilitative, and preventive regiments. » (Sackett et al., 1996, 71-72)

Dans le but de rendre compte de ces trois points, la médecine fondée sur les preuves se déroule en quatre étapes¹³⁵. Tout d'abord, il faut formuler une question clinique claire à partir d'un problème bien défini (1). Afin de répondre au mieux à cette question, la deuxième étape suggère de procéder à une analyse de la littérature scientifique sur le sujet (2) et, ensuite, d'évaluer la validité des résultats, autrement dit de déterminer le niveau et la solidité de la preuve (3). La dernière étape est d'utiliser les résultats cliniques en vue d'une prise en charge personnalisée du patient (4).

La formulation d'une question clinique précise (1) se fonde sur le modèle PICO, élaboré par la collaboration Cochrane¹³⁶. Le modèle PICO permet de traduire quatre questions distinctes. La première question relève du patient et de ses caractéristiques (P) ; la deuxième de l'intervention (I), demandant alors ce qui va être fait ou envisagé ; la troisième est la comparaison (C) d'efficacité par rapport à une autre intervention : et enfin, la quatrième concerne les résultats des études cliniques (O pour *outcomes*). Pour procéder à l'analyse des résultats scientifiques (2) beaucoup de ressources sont disponibles. De nombreuses bases de données ont été créées afin de centraliser les informations scientifiques disponibles et donc offrir différentes bases de données¹³⁷. L'évaluation de la preuve (3) se fait selon une hiérarchie très précise, définie par des hautes autorités, sur lesquelles nous reviendrons plus en détail. Avant cela, il convient de discuter la prise en charge personnalisée du patient (4), qui renvoie à son choix. Autrement dit, la décision médicale est prise par le patient ; mais, ce dernier ne possède pas toute l'information, généralement, ses connaissances médicales sont nettement moins importantes que celles du médecin. Cette décision est donc, bien souvent, orientée par le point de vue du médecin. La participation du patient varie, d'ailleurs, beaucoup selon les patients, le stade de la maladie etc.

¹³⁵ Pour un développement non exhaustif, voir (McGovern, Valori, Summerskill et Levi, 2001).

¹³⁶ La collaboration Cochrane est une association à but non lucratif dont le but est de promouvoir *l'evidence-based medicine*.

¹³⁷ Par exemple, le *British Medical Journal* a mis en place une base de données, *BMJ Clinical Evidence*, qui propose des synthèses de résultats des interventions cliniques. La collaboration Cochrane a, elle aussi, créé une base de données, qui fait office de référence : la *Cochrane Library*. Elsevier a constitué une base de données, « EMBASE », où cinq mille périodiques sont indexés. Une version publique de la base de donnée *Medline* a été créée, « pubmed ». Elle indexe les références scientifiques de 1950 à aujourd'hui. L'« infobanque AMC » est une base de données canadienne offrant des guides de pratique clinique fondée sur des preuves. Et enfin, le *national guidelines clearinghouse* est un site internet américain, subventionné par des agences médicales américaines, qui donne des recommandations sur les bonnes pratiques cliniques.

2.1.3 La hiérarchisation des preuves au sein de l'evidence-based medicine

Il existe plusieurs niveaux de preuves¹³⁸. Comme cela a été vu, l'essai clinique est une partie d'un essai plus large, l'essai thérapeutique. Il y a alors une graduation du niveau de la preuve pour l'essai clinique et une graduation pour l'essai thérapeutique. Pour l'essai clinique, les preuves les plus « solides » proviennent de données publiées dans des revues scientifiques et les résultats sont tirés d'au moins une étude randomisée ou de méta-analyses d'essais randomisés. Un niveau plus faible de preuves, mais toujours fortement valide, est fourni par des données publiées dans des revues scientifiques mais les résultats, cette fois-ci, ne proviennent pas d'études randomisées mais d'études de cohorte ou d'études de cas-contrôle¹³⁹. Enfin, le dernier niveau de preuves pour un essai clinique se fait sur la base d'analyses d'experts, se fondant sur d'autres données disponibles, qui ont elles aussi été publiées dans un journal scientifique à comité de lecture. On voit alors nettement apparaître la valeur de scientificité contemporaine qui est associée à la randomisation en médecine. Concernant les essais thérapeutiques et les recommandations qui en découlent, le niveau de preuve est défini par la haute autorité de santé. Il existe alors quatre niveaux de preuves. Le niveau 1 (NP1) s'appuie sur des essais randomisés de « forte puissance », c'est-à-dire réalisés sur un échantillon large ; ce qui signifie que le résultat est « prouvé ». Le niveau 2 (NP2) s'appuie sur des essais randomisés où l'échantillon est trop petit ou sur des études de cohorte, ce qui donne un caractère « probable » à la preuve. Le niveau 3 (NP3) se fonde sur des cas témoins et la preuve y est considérée comme « acceptée ». Enfin, le niveau 4 (NP4) est caractérisé par des études comparatives comprenant beaucoup de biais, comme les études rétrospectives ou les séries de cas. Les niveaux de preuves, pour l'essai thérapeutique et clinique, sont, en fait, pratiquement semblables.

2.1.4 Critiques de l'evidence-based medicine

Malgré son poids de plus en plus dominant en médecine, l'*evidence-based medicine* connaît aussi des détracteurs. Trois principales critiques lui sont adressées. Il arrive que les preuves n'existent pas sur un thème particulier (1), dès lors cette pratique devient impossible.

¹³⁸ Pour une discussion de ces différents niveaux de preuves, voir (Worall, 2002).

¹³⁹ En épidémiologie, une étude de cohorte revient à comporter deux cohortes, une qui a accès au traitement et une qui n'y a pas accès ; mais l'attribution au traitement ne se fait pas de façon aléatoire. L'étude cas-témoin est une étude rétrospective qui vise à comparer les caractéristiques des malades (les cas) à celles d'individus non malades (les témoins).

Il apparaît parfois difficile d'appliquer les conclusions provenant d'études rigoureuses à un individu particulier, c'est-à-dire à un patient précis¹⁴⁰ (2). Les preuves tirées d'essais cliniques, par exemple, se fondent sur des modèles statistiques au sein desquels la pathologie n'est pas envisagée du point de vue de l'individu, mais du point de vue d'un groupe. Ces études s'appliquent donc à des « patients standards », mais ces derniers peuvent s'écarter très largement du patient précis qui doit être traité ; les études randomisées ne prennent pas en compte les dimensions contextuelles propres à chaque individu. À ces deux premières objections, s'en ajoute une troisième : il n'existe pas de preuves concernant tous les actes cliniques (3). C'est pourquoi, selon Naylor (1995), il existe alors des « zones grises » :

« The craft of caring for patients can flourish not merely in the grey zones where scientific evidence is incomplete or conflicting but also in the recognition that what is black and white in the abstract may rapidly become grey in practice as clinicians seek to meet their individual patients' needs. » (Naylor, 1995, 841)

Un des enjeux est de parvenir à envisager l'*evidence based-medicine* dans une perspective à la fois institutionnelle et individuelle, c'est à dire adaptée à un patient particulier.

2.2 L'*evidence based Policy* ou la politique de la preuve

L'*evidence based policy* s'inspire directement de la médecine fondée sur les preuves. L'objectif est, cette fois-ci, d'éclairer la décision politique en la fondant sur des preuves. Au cœur de l'*evidence based policy* on retrouve aussi l'utilisation de la randomisation. Afin de définir au mieux ce mouvement et d'en saisir les enjeux en vue de questionner un des deux objectifs du J-PAL (guider la décision politique), je présenterai, tout d'abord, l'histoire de l'*evidence-based policy* (2.2.1). Cette dernière est constituée de plusieurs éléments qui tendent à les définir, je développerai donc ensuite ces éléments (2.2.2). Ensuite je m'attacherai à montrer qu'un tel mouvement peut se fonder sur une pluralité de méthodes, mais ces méthodes partagent toutes un socle commun que je tenterai de définir (2.2.3). Enfin, je détaillerai la hiérarchisation des preuves qu'offre l'*evidence-based policy*, même si cette hiérarchie est moins précise qu'en médecine (2.2.4).

¹⁴⁰ On retrouve, ici, une des limites de la randomisation que nous développerons plus largement dans la suite de ce travail (partie 2 – chapitre 2), la question de l'hétérogénéité des effets de traitement.

2.2.1 Histoire de l'evidence-based policy

L'*evidence based policy* a été popularisée¹⁴¹ par le gouvernement de Tony Blair, en Angleterre, au début des années 2000: « *the term 'evidence-based policy making' has been most recently popularised by the Blair Government, which was elected on a platform of 'what matters is what works'. Blair spoke of ending ideologically-based decision making and 'questioning inherited ways of doing things' (Blair et Cunningham 1999) »*. (Banks, 2009, 3) Souhaitant rompre avec le gouvernement de Margaret Thatcher, Tony Blair lance un vaste programme d'évaluation de politiques publiques. Ce programme cherche à évaluer un ensemble de politiques comme : le *new deal for communities*, the *employment new deals*, *on-track et sure start* and the *children's fund*. Ces programmes ont en commun de viser une évaluation de grande envergure. Les États-Unis représentent aussi un pôle important dans l'evidence-based policy : « *the United-States has long been the major global location for policy analysis and evaluation professionals, both within government and in other policy-relevant sector* » (Head, 2009, 15). Un autre pays où l'*evidence based policy* s'est fortement développée est l'Australie. Le Premier ministre y a annoncé en avril 2008 qu'il existait pour lui un lien très fort entre l'*evidence-based policy* et une bonne gouvernance¹⁴². Les implications pratiques restent encore à être mises en place mais le rôle de l'*evidence-based policy* est ici clairement lancé. L'*evidence-based policy* connaît un fort rayonnement dans les pays anglophones mais peu dans les pays non anglophones :

« Dans nombre de pays non anglophones, comme la France, les discussions sur ces approches sont ignorées ou occultées. En France, les recherches sur les relations entre science et décision publique sont en grande partie des travaux relevant de la sociologie des sciences. » (Laurent *et al.*, 2009, 858)

Même si l'*evidence-based policy* ne connaît pas le même rayonnement en France que dans les pays anglophones, le Haut-commissariat aux solidarités actives de Martin Hirsch a tenté d'offrir à cette dernière un écho similaire¹⁴³. Ce mouvement montre que la recherche et l'évaluation sont appréciées par les décideurs politiques, poussant chacune des disciplines

¹⁴¹ On considère, d'ailleurs, que la première utilisation de l'*evidence-based policy* a eu lieu en Angleterre : « *Probably the oldest example, or longest-standing one, would be tariff making, which for many years was required under legislation to be informed by a public report produced by the Tariff Board and its successor organizations* ». (Banks, 2009, 3)

¹⁴² Le premier ministre a affirmé : « *policy innovation and evidence-based policy making is at the heart of being a reformist government* ». (Rudd, 2008)

¹⁴³ Ce que j'ai développé dans le chapitre précédent.

engagées à faire un effort : les chercheurs doivent adopter une perspective plus compréhensive et les décideurs politiques, s'engager en matière d'évaluation.

2.2.2 Les différents « ingrédients »¹⁴⁴ qui fondent l'evidence-based policy

Il est difficile de définir précisément ce qu'est l'*evidence-based policy*, car l'idée de fonder les décisions politiques sur des preuves scientifiques peut s'apparenter à différentes façons de procéder¹⁴⁵. Pour Gary Banks (2009), il y a cinq « ingrédients » essentiels à l'*evidence-based policy*. Ces cinq ingrédients permettent alors de donner une définition précise de l'*evidence-based policy*. Le premier ingrédient est la méthode, l'*evidence-based policy* se fonde sur une méthode rigoureuse. Un deuxième ingrédient est caractérisé par l'existence de données. Il faut pour mesurer l'impact d'une politique qu'on puisse utiliser des données disponibles. Les interventions politiques ont toujours des effets sur plusieurs dimensions, l'identification peut alors être difficile. Les données permettent, déjà, d'avoir une vue un peu plus précise. Obtenir ces données peut être compliqué et on peut rencontrer certains problèmes qui rendent, de ce fait, la force de la preuve plus faible : « *data deficiencies inhibit evidence-based analysis for obvious reasons.* » (Banks, 2009, 11) Le troisième ingrédient est la transparence. Pour Banks (2009), de nombreuses analyses politiques sont menées de façon extrêmement opaque¹⁴⁶. La transparence implique de regarder de près les méthodes utilisées, les hypothèses sous-jacentes et les données pour que l'évaluation puisse être répliquée. Il est aussi intéressant de prendre en compte le retour des personnes concernées par le programme, de ne pas s'en tenir qu'aux points de vue des experts. Le problème est que cette transparence ne peut pas toujours avoir lieu car elle implique du temps et de l'énergie qui ralentissent le processus de prise de décisions, ce qui traduit le troisième « ingrédient » de l'*evidence-based policy* : cette dernière prend du temps. On voit ici une contradiction existante entre le point de vue du chercheur sur le besoin de « bonnes preuves » et le point de vue de l'homme politique qui lui est pressé par le temps. Le quatrième « ingrédient » est l'importance de l'expertise : « *you can't have good evidence, you can't have good research, without good people* ». (Banks, 2009, 15) Les personnes qui fournissent les preuves peuvent être variées, cela peut être des consultants, des personnes travaillant pour le gouvernement ou des chercheurs. Les experts peuvent, alors, être très

¹⁴⁴ Je reprends, ici, le terme utilisé par Gary Banks (2009).

¹⁴⁵ (Davies, Nutley, et Smith, 2000), discutent des différents résultats obtenus en sciences sociales et de leurs mises en pratique.

¹⁴⁶ « *Much policy analysis (...) actually occurs behind closed doors.* » (Banks, 2009, 14)

divers. Le dernier point est considéré par Banks, comme le plus important, le processus de prise de décision politique doit être « réceptif à la preuve ». Il doit s'agir d'un processus qui part d'une question et non pas d'une réponse.

2.2.3 Pluralité des méthodes, communauté des caractéristiques

L'*evidence based-policy* ou plus largement l'idée d'*evidence-based decision* s'applique à des disciplines très diverses comme l'éducation, la santé, les transports, ou encore la criminologie. Tout cela explique qu'il existe plusieurs méthodes fondées sur des faits pour éclairer la décision politique. Les études peuvent être qualitatives ou quantitatives. Les méthodes quantitatives relèvent alors d'approches plurielles comme l'anthropologie ou l'ethnographie et se fondent majoritairement sur des entretiens. Les méthodes quantitatives sont variées, elles peuvent s'apparenter à des expériences dont l'illustration est la randomisation, ou être quasi expérimentales. Pour autant, ces différentes méthodes partagent toutes les caractéristiques suivantes. Elles visent toutes à tester une théorie ou une proposition (1). Elles utilisent toutes un contrefactuel (2). Elles impliquent une quantification de l'impact évalué (3). Elles s'intéressent à la fois aux effets directs du programme et à ces effets indirects (4). Elles construisent une structure permettant de contrecarrer les autres facteurs qui peuvent rentrer en compte dans les variables observées (bien souvent la création d'un groupe de contrôle). Cette structure leur permet ainsi d'éviter certains biais (5). Enfin, elles peuvent être répliquées par un tiers (6). La représentation la plus courante que prend l'*evidence-based policy* est celle d'analyses coûts/bénéfices :

« *Most evidence-based methodologies fit broadly within a cost benefit (or at least cost effectiveness) framework, designed to determine an estimated (net) payoff to society. It is a robust framework that provides for explicit recognition of costs and benefits, and requires the policy maker to consider the full range of potential considerations and community norms that are best made by accountable political representatives* ». (Banks, 2009, 10-11)

L'*evidence-based policy* peut donc s'apparenter à une pluralité de méthodes. Pour autant, ces différentes méthodes se retrouvent sur de nombreuses caractéristiques ; et bien souvent l'*evidence based policy* est matérialisée par une analyse coût/bénéfices.

2.2.4 Types et niveaux de preuves

Laurent *et al.* (2009) définissent plusieurs types de preuves : des « preuves d’existence », des « preuves de causalité », des « preuves d’efficacité », des « preuves d’innocuité ». Les premières, les « preuves d’existence », prouvent que quelque chose existe. Elles peuvent être : « [des] observations naturalistes pour inscrire des espèces dans un inventaire utilisable pour la gestion de la biodiversité, [des] recensements pour dénombrer une population. » (Laurent *et al.*, 2009, 865) Les « preuves de causalité » visent à rendre compte d’une relation causale entre deux éléments, « par exemple, toute choses égales par ailleurs, deux pratiques de pâturages distinctes ont des effets différents observables de façon répétée au moyen d’indicateurs de biodiversité des prairies concernées ». (*Ibid.*) Les « preuves d’efficacité »¹⁴⁷ prouvent que le programme évalué a les effets souhaités, par exemple une mesure budgétaire qui parvient effectivement à réduire le déficit. Enfin, les « preuves d’innocuité » montrent que l’intervention n’a aucun effet négatif, par exemple une politique de taxation qui n’a pas d’effets négatifs sur la consommation des ménages. Ces preuves renvoient donc à des actions particulières, elles ne cherchent pas à prouver la même chose ; les séparer peut être important car cela permet de distinguer des phénomènes qui parfois s’entremêlent.

En *evidence-based medicine*, comme cela a été développé, les preuves sont hiérarchisées et impliquent des recommandations particulières. Les niveaux de preuves en *evidence-based policy* sont moins clairs. Laurent *et al.* (2009) ont tenté de transposer le système de hiérarchie des preuves de l’*evidence-based medicine* à celui de l’*evidence-based policy*. L’essai contrôlé randomisé retrouve alors la même place et garde, une fois de plus, le statut de *gold standard*. Nancy Cartwright, en reprenant les règles du Maryland, qui hiérarchise les preuves en criminologie, les classe en quatre groupes¹⁴⁸. On ne s’attache ici qu’aux preuves d’efficacité, plus précisément aux preuves d’*effectiveness*. La première catégorie, « *works* » (1), doit relever d’au moins deux études faites à l’aide d’une méthode rigoureuse. La deuxième catégorie, « *does not work* » (2), doit, comme la première, relever d’au moins deux études faites à l’aide d’une méthode rigoureuse. La troisième, « *promising* » (3), doit contenir au moins une étude s’appuyant sur une méthode rigoureuse. Et enfin, la quatrième, « *don’t know* » (4), contient les preuves empiriques qui ne rentrent dans aucune des trois autres catégories. L’objectif de l’*evidence-based policy* est donc de parvenir à utiliser la preuve

¹⁴⁷ Comme le souligne Laurent *et al.* (2009), il est important de noter que le terme d’efficacité peut renvoyer à trois dimensions précises. L’efficacité renvoie d’une part à l’« efficacy » : est-ce que le programme peut bien fonctionner ? Elle renvoie aussi à l’« efficiency » : est-ce une bonne solution compte tenu des ressources dont on dispose ? Et enfin, elle peut renvoyer à l’« effectiveness » : est-ce que le programme fonctionne dans la pratique ? On reviendra plus largement sur la distinction de ces termes dans la partie suivante de ce travail.

¹⁴⁸ Cartwright mène une importante analyse des enjeux épistémologiques que rencontrent le mouvement de l’*evidence-based policy* ; voir, par exemple, Cartwright (2008 ; 2009a ; 2009b ; 2011b). Je reviendrai largement dessus dans la prochaine partie de ce travail (partie 2 – chapitre 1).

après hiérarchisation. Elle implique un lien étroit entre le producteur de preuves (les chercheurs) et ses utilisateurs (les décideurs politiques).

3. Le J-PAL et l'*evidence-based policy*

Le J-PAL s'insère dans le mouvement d'*evidence-based policy*. Il a un rôle de producteur de preuves et il cherche à rendre ces preuves disponibles pour la décision politique. Dans ce but, il procède à des synthèses de résultats à destination des décideurs politiques. La pierre angulaire de l'*evidence-based policy* est ici saillante : la production et l'utilisation de preuves. Afin de rendre clairement compte de ces deux objectifs, le J-PAL sépare ses publications en deux niveaux : les publications académiques et les « mallettes politiques ». Les premières renvoient directement à la production de preuves et détaillent les expériences ainsi que leurs résultats. Elles sont à destinations, principalement, des chercheurs, car publiées dans des revues académiques. À partir de ces publications, le J-PAL choisit de réaliser des analyses coûts/efficacité qui permettent de résumer les effets des différents programmes et de les mettre en lien avec leurs coûts ; cela afin que les décideurs politiques aient la possibilité de choisir entre différents programmes en fonction de leur budget et des effets souhaités. Ces analyses coûts/efficacité sont ensuite traduites sous forme de « mallettes politiques » à destination des décideurs politiques. Elles représentent une synthèse des analyses coûts/efficacité et ont un format court et simplifié. Il y a donc ici trois dimensions : les publications académiques, les analyses coûts/efficacité et les publications politiques. Afin de montrer comment le J-PAL s'inscrit au sein de l'*evidence-based policy*, je développerai, tout d'abord, le rôle des publications académiques (3.1), puis celui des analyses coût/efficacité (3.2), et enfin celui des synthèses politiques (3.3).

3.1 Publications académiques et biais de publication : le rôle des agences politiques

Les différentes expériences menées par le J-PAL, une fois terminées, sont publiées et font office de publications académiques. Ces expériences sont, majoritairement, publiées dans les meilleures revues économiques, comme l'*American Economic Journal* ¹⁴⁹ et tendent, désormais, à définir ce qui se fait de « mieux » en économie du développement. Le rôle de *gold standard* méthodologique attribué à la randomisation y est sans doute pour beaucoup.

¹⁴⁹ D'ailleurs la plus grande figure du J-PAL, Esther Duflo, est éditorialiste de ce journal.

Mais une des principales explications me semble être l'organisation que choisit le J-PAL pour rendre ses résultats accessibles autant aux chercheurs qu'aux politiciens. En publiant dans les meilleures revues académiques, le J-PAL connaît, alors, une reconnaissance scientifique forte. Pour autant, ces publications relèvent d'évaluations diverses et provenant de domaines particuliers. Les résultats des évaluations apparaissent souvent trop éclatés les uns des autres et sont toujours rattachés à une publication particulière. Il apparaît difficile pour les décideurs politiques comme pour les chercheurs de s'y retrouver. Afin de contrecarrer cela, le J-PAL classe ses publications académiques autour de sept domaines. Je commencerai par définir le classement des publications académiques autour de ces sept domaines (3.1.1). Puis, j'insisterai sur le biais de publication, biais qui montre que seuls les résultats originaux et positifs tendent à être publiés (3.1.2). Les chercheurs du J-PAL en appellent alors au rôle des agences internationales afin de contre carrer ce biais. Finalement, je définirai le rôle que les chercheurs du J-PAL attribuent aux agences internationales en vue de réduire le biais de publication (3.1.3).

3.1.1 Un classement des publications autour de sept domaines

Dans un souci de plus grande visibilité et de clarté, les chercheurs du J-PAL classent leurs résultats en fonction de sept domaines : l'agriculture, la santé, l'éducation, le marché du travail, l'économie politique et la gouvernance, l'environnement, et la micro-finance. Il est intéressant de noter que chacune des personnes ayant un rôle dans la direction du J-PAL est directeur d'un de ces domaines. Les publications du J-PAL relevant de l'agriculture renvoient à des évaluations de programmes qui ont pour objectif de faciliter l'accès à de nouvelles technologies, afin que l'agriculture devienne un domaine plus rentable pour les pauvres. Ils évaluent donc des programmes qui cherchent à promouvoir ce type de stratégies. Rachel Glennester et Christopher Udry dirigent les évaluations concernant ce domaine. Les programmes portant sur l'éducation sont, quant à eux, encadrés par Abhijit Banerjee et Paul Glewwe. Les programmes envisagés partent de résultats d'expériences précédentes, montrant que lorsque les enfants vont à l'école, ils apprennent en fait peu de choses. Les programmes évalués visent à améliorer la qualité de l'éducation. En matière d'environnement et d'énergie, l'objectif des programmes est de montrer les bénéfices que l'on peut tirer de la qualité environnementale en termes de santé et de bien-être. Michael Grennstone les supervise. Les programmes concernant la finance tentent de favoriser l'accès au financement pour les pauvres. En matière de santé, les différents programmes visent à promouvoir l'amélioration

de la santé dans la réduction de la pauvreté, c'est Esther Duflo qui encadre ce domaine. Concernant le marché du travail, le J-PAL ne définit pas ce qu'il entend par ce terme, et ce qu'il veut y évaluer. Une seule évaluation est disponible sur le sujet, elle concerne l'impact de la formation professionnelle sur le marché du travail, la migration, la santé et la fertilité au Kenya¹⁵⁰. Marianne Bertrand et Bruno Crépon dirigent les programmes s'y rattachant. Les programmes relevant de l'économie politique et de la gouvernance sont regroupés autour d'un seul et même programme le *Political Economy and Governance Program* (PEGP). Il regroupe des initiatives tentant de comprendre les déterminants de la participation politique citoyenne ainsi que l'impact de la corruption. Benjamin Olken et Rohini Pande dirigent le *Political Economy and Governance Program*. Chaque publication est donc rapportée à un domaine spécifique, et ce domaine cherche à évaluer des programmes dans un but précis.

3.1.2 Le biais de publications

Cependant, tout comme en médecine, les revues académiques ne tirent que peu d'intérêts à publier des résultats d'expériences randomisées négatifs. De plus, ces revues cherchent aussi à publier des résultats originaux et dont la validité statistique est importante. Cette originalité peut être due aux programmes qui sont évalués, ou aux résultats qui apparaissent contre-intuitifs. Ces problèmes définissent le biais de publication. Les chercheurs du J-PAL sont sensibles à ces problèmes :

« the selection by journals of papers with significant results introduces another level of publications bias. Moreover citation of papers with extreme results by advocates on one side or another of policy debates is likely to compound publication bias with citation bias. The cumulative result of this process is that even when a program has no effect, strongly positive and/or strongly negative estimates are likely to be published and widely cited. » (Duflo, Glennerster et Kremer, 2007, 3909)

Le biais de publication est un biais très important car il tend à orienter la recherche vers ce qui apparaît être « publiable ». Il peut laisser de côté des évaluations dont le thème semble trop simpliste, mais les résultats de ces évaluations peuvent être essentiels pour guider la décision politique. Les résultats non positifs sont presque aussi intéressants que les résultats positifs, car, pour les chercheurs du J-PAL, le but n'est pas uniquement de déterminer ce qui « fonctionne » mais aussi ce qui « ne fonctionne pas » pour lutter contre la pauvreté. Un

¹⁵⁰ Cette évaluation n'est pas encore publiée mais on peut trouver de nombreux éléments sur cette dernière sur le site du J-Pal : <http://www.povertyactionlab.org/labor-markets>.

résultat négatif dit presque autant de choses qu'un résultat positif en matière de lutte contre la pauvreté. De plus, comme Duflo le souligne, au biais de publication s'ajoute un biais de citation. Les expériences dont les résultats sont solides statistiquement, tendent à être plus cités, offrant alors un rayonnement plus important. Ces biais apparaissent donc être un problème pour les études randomisées mais il est nuancé, par Duflo car, selon elle, il est encore plus important au sein des évaluations d'impacts rétrospectives¹⁵¹. Selon Duflo, il y a plus de chances que les études rétrospectives soient manipulées en vue d'une publication :

« Le biais de publication a des chances d'être un problème particulier concernant les études rétrospectives. *Ex post*, les chercheurs ou les évaluateurs définissent leur propre groupe de comparaison et devraient alors être capables de choisir une variété de groupes de comparaison plausibles ; en particulier, les chercheurs obtenant des résultats négatifs avec des techniques rétrospectives vont probablement tenter des approches différentes, ou ne pas publier (...) À l'opposé, les évaluations par assignation aléatoire s'engagent à l'avance avec un groupe de comparaison particulier : une fois que le travail de mener une évaluation par assignation aléatoire prospective est effectué, les résultats sont d'ordinaire documentés et publiés même si les résultats suggèrent des effets plutôt modestes voire même aucun effet du tout ». (Duflo, 2005, 210-211, souligné par moi)

Le biais de publication est donc moins présent dans les études randomisées que dans les études rétrospectives, mais il se peut que les résultats d'évaluations randomisées ne soient pas publiés, ce que le « d'ordinaire » suggère. C'est pourquoi Duflo accorde un rôle important aux agences internationales dans la réduction de ce biais.

3.1.3 Le rôle des agences internationales

Le rôle des agences internationales, dans la réduction de ce biais de publication, passe, pour Duflo, par la création d'une base de données qui regrouperait toutes les évaluations menées :

« Dans l'idéal la base de données inclurait toute information qui pourrait être utile à l'interprétation des résultats (estimation, taille de l'échantillon, région et époque, type de projet, coût, analyse coûts/bénéfice, avertissements, etc.) ainsi que la référence aux études en rapport. La base de données pourrait inclure à la fois les évaluations d'impact avec et sans assignation aléatoire satisfaisant certains critères à condition que les différents types d'évaluation soient clairement libellés. Les évaluations devraient satisfaire des exigences

¹⁵¹ Les évaluations d'impact rétrospectives renvoient à des évaluations qui sont réalisées après la mise en place du programme et non en même temps que sa mise place, comme le permet la randomisation.

de rapport minimum pour être incluses dans la base de données et tous les projets supportés par l'unité devraient être inclus dans la base de données quels que soient leurs résultats ». (Duflo, 2005, 222)

Les agences internationales en ayant un rôle centralisateur permettraient de réduire ce biais de publication car elles rendraient compte de tous les résultats, à travers l'existence d'une base de données. Cela est dans la droite lignée des bases de données déjà existantes en médecine, qui cherchent à centraliser les différents résultats provenant d'essais cliniques.

3.2 Les analyses cout/efficacité

Beaucoup de preuves existent sur des thèmes différents, il est alors difficile pour les décideurs politiques de s'y retrouver. Un des moyens de pallier à cela consiste à mener des analyses synthétiques ; par exemple, des analyses coûts/efficacité, c'est-à-dire une analyse qui compare les effets et les coûts de programmes différents qui visent un même objectif. Selon Duflo, cette façon de procéder se fait rare en économie du développement, il faut alors la développer. C'est ce que les chercheurs du J-PAL proposent afin de créer une collaboration étroite entre les chercheurs et les décideurs politiques. Il existe une autre façon de mener des analyses synthétiques, pour que les décideurs politiques puissent choisir : les analyses coûts/bénéfices. Dans ce type d'analyse, l'objectif est de mettre ensemble les différents bénéfices d'un programme en les rapportant à une dimension commune : la dimension monétaire. L'un des principaux avantages de cette approche est de pouvoir rendre compte des effets d'un programme qui a plusieurs objectifs, comme la construction d'infrastructure, par exemple. Comme le souligne Duflo, l'avantage des analyses coûts/efficacité par rapport aux analyses coûts/bénéfices, est qu'il laisse le décideur politique libre de choisir entre plusieurs programmes en fonction de leurs coûts et de leurs différents apports. C'est pourquoi les chercheurs du J-PAL privilégient ces dernières et que leurs synthèses de résultats sont issues uniquement d'analyses coûts/efficacité. Je définirai, tout d'abord, ces analyses coûts/efficacité (3.2.1). Puis je développerai deux de leurs enjeux, le premier étant de parvenir à quantifier l'impact des programmes évalués (3.2.2), et le second étant de réussir à quantifier les différents coûts du programme évalué (3.2.3).

3.2.1 Les analyses coût/efficacité : quelques éléments de définition

Le but de ce type d'analyses consiste à comparer l'impact de différents programmes en les ramenant à leurs coûts. Ces analyses permettent de résumer clairement les effets d'un programme et offrent une mesure commune pour comparer différents programmes ; ce qui fait d'elles un outil important de la prise de décision politique¹⁵². Comme cela a déjà été souligné, les évaluations du J-PAL visent des programmes locaux, c'est pourquoi :

« The methodology that is adopted in the examples given here (and by J-PAL as an organization) is intended for an audience of policymakers in governments, foundations, international development agencies, and NGOs that have particular policy objective in mind and are trying to decide between a range of different options for achieving that policy objective. We are not trying to help the prime minister of a country, or the chairman of a foundation, to decide whether to put their money into education versus health (...) instead we are taking the perspective of, for example, the minister of education of a state in India, or an education program officer in a foundation, who aims to maximize their impact on a particular objective like student attendance within a budget constraint ». (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 9-10)

L'idée du J-PAL est donc d'aider des décideurs politiques sur des objectifs précis et particuliers, non pas de réfléchir avec les décideurs politiques à des « solutions » sur des problèmes plus généraux. Les analyses coûts/efficacité impliquent cependant plusieurs défis et une méthode particulière. Le premier enjeu est donc de fournir un élément de comparabilité entre les différents programmes qui visent un objectif similaire. Cet élément est le coût. Et dans un deuxième temps, il faut parvenir à quantifier l'impact des programmes évalués.

3.2.2 Quantifier l'impact

L'un des premiers problème est de quantifier l'impact des programmes. Par exemple, comment faire lorsque les programmes ont plusieurs effets ? Une analyse coût/efficacité se centre sur un impact particulier, il est alors difficile de rendre compte des impacts qui ont plusieurs effets¹⁵³. Une des solutions envisagées par Duflo est de comprendre le programme comme un « package », autrement dit un ensemble qui a différents effets. On ramène alors tous ces différents effets à une même unité. Un autre problème est de savoir à partir de quelle estimation on intègre un programme dans l'analyse coût/efficacité. Autrement dit, à partir de quel niveau une estimation est jugée positive ou négative, à partir de quel critère une

¹⁵² « When done correctly, such analysis can be a useful tool for decision makers in organization that fund or implement education or other social programs in developing countries, allowing them to compare results of alternative programs when deciding how to allocate resources. » (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 5)

¹⁵³ C'est ici qu'il est plus intéressant de mener une analyse coût/bénéfices.

estimation est jugée significative. Le J-PAL choisit de la façon suivante : « *programs whose impact is significant at 10% or better are included in J-PAL analysis* » (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 14). Une autre source de problèmes pour l'estimation de l'impact est liée à l'existence d'externalités. Un programme peut générer des effets positifs et négatifs au sein de populations qui ne bénéficient pas, ou pas encore du programme. Ces externalités doivent-elles être prises en compte ? Et si oui, comment les prendre en compte et les intégrer à l'analyse ? Pour cette question, les chercheurs du J-PAL décident ou non d'intégrer les externalités existantes lorsque le programme sera élargi à une population plus large : « *in deciding whether or not to include the effect of spillovers in a cost-effectiveness analysis, we must assess whether spillovers would take place even when a program was scaled up* ». (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 16) Une fois les impacts pris en compte déterminés, il reste alors à les agréger pour obtenir l'impact total du programme. Pour ce faire, les chercheurs du J-PAL multiplient l'impact du programme par unité à la taille de l'échantillon puis à la durée du programme (Impact total du programme = impact (par unité) * taille de l'échantillon * durée du programme).

3.2.3 Quantifier les coûts

Une fois l'impact rendu quantifiable, reste la deuxième partie de l'analyse : la quantification des coûts. Elle consiste, tout d'abord, à collecter des données sur les coûts afin de pouvoir les intégrer à l'analyse. En pratique, cela est difficile à obtenir, notamment parce que les informations sur les budgets sont parfois incomplètes. Les chercheurs du J-PAL sont actuellement en train de créer une fiche technique à l'usage des autres chercheurs pour qu'ils puissent se repérer dans la collecte de données sur les coûts :

« *We are currently developing a general worksheet for researchers to use as a template for collecting cost data as they run evaluations. This will not only make it much easier to perform future cost-effectiveness analyses, but will also help improve comparisons of the different program as part of cost effectiveness analysis.* » (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 19)

Pour modéliser les coûts, les chercheurs du J-PAL se réfèrent aux coûts marginaux. Par exemple le rapport coût/efficacité d'un programme en matière d'éducation est calculé comme un changement marginal dans le résultat aux examens des élèves, qui est compris comme le résultat du programme – puisque son objectif est d'améliorer les résultats aux examens des

élèves –, ensuite divisé par le changement marginal, cette fois-ci en termes de coût. On obtient alors :

- Ratio coût/efficacité = (résultats aux examens avec programme) – (résultats aux examens sans programme) / (coûts avec programme) – (coûts sans programme).

Plusieurs coûts sont difficiles à établir. Par exemple, comment rendre compte des biens et services fournis gratuitement ? Doit-on s'intéresser à ceux qui fournissent ces biens et donc en supportent les coûts ? Ou doit-on considérer la société dans son ensemble et donc prendre en compte les bénéfices que la société peut tirer de ces biens ? L'approche du J-PAL choisit de tenir compte de la société dans son ensemble. Bénéficier d'un programme peut aussi représenter un coût. Par exemple, si un programme en faveur de l'éducation nécessite la participation des parents, cela leur demandera du temps, comme se rendre à des réunions ou participer aux devoirs des enfants. Lorsque le programme nécessite un tel engagement, les chercheurs du J-PAL choisissent d'inclure ces coûts dans leur analyse :

« In many cases, programs also requires beneficiaries to spend time contributing to the program – for instance, when parents must attend meetings to get information about the returns to education or give consent for the administration of deworming drugs. (...) because this time is a requirement of the program and represents a real cost to the user, we have chosen to include such costs wherever programs required users to commit their time. » (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 22)

Le programme peut aussi impliquer des frais de gestion, comme une durée de travail supplémentaire pour la bonne poursuite du programme. Ces frais sont très difficiles à intégrer car ils sont difficilement quantifiables. C'est pourquoi, les chercheurs du J-PAL choisissent de ne pas en tenir compte, leur évaluation serait trop imprécise¹⁵⁴. Il peut aussi exister une différence entre le coût d'un programme lorsqu'il est évalué à petite échelle et lorsqu'il est pensé à plus grande échelle. Les chercheurs du J-PAL ne tiennent pas non plus compte de ces coûts¹⁵⁵. Le dernier aspect important de la quantification des coûts est la détermination d'une

¹⁵⁴ « One of the most difficult cost items to incorporate in cost-effectiveness analysis is the incremental cost of indirect overhead. (...) These additional costs are almost never reported, especially at the pilot or research stage when they represent a small amount in a (likely) much larger organizational budget. As such, they are almost impossible to observe and any estimations by us would be extremely imprecise. » (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 27-28)

¹⁵⁵ « The cost of a program evaluated in its pilot phase may be different from the actual costs if one were to massively scale up the program. (...) Without good reason, it is often better to stick with the actual costs and actual benefits of the program as it was tested. » (Dhaliwal, Duflo, Glennerster et Tulloch, 2011, 28-29)

unité commune. Le temps et la monnaie varient en fonction du programme. Un programme est évalué à un certain moment. Lorsque l'analyse coût/efficacité est réalisée, les prix peuvent être différents de ceux que l'on avait durant l'évaluation du programme. Il faut, dans ce cas, harmoniser les prix pour que l'analyse soit viable. Il est aussi nécessaire de tenir compte de l'inflation et des différents taux de change. Il faut, ensuite, déterminer une monnaie de référence ; celle utilisée par le J-PAL est le dollar.

3.3 Les « mallettes » politiques

Afin de présenter ces analyses et de les rendre abordables pour les décideurs politiques, ces dernières sont présentées sous forme de « mallettes politique »¹⁵⁶ dans lesquelles le problème et les enjeux des différents programmes sont, tout d'abord, présentés ; ensuite, les résultats des différentes stratégies sont exposés autour d'une analyse coûts/efficacité ; puis, suivent des « leçons politiques » à destination des décideurs politiques. Afin d'illustrer précisément ce que peuvent être ces « mallettes » je présenterai l'une d'elles. Dans un souci d'exhaustivité, j'ai choisi de n'en présenter qu'une, le format et la façon de procéder étant identiques pour chaque « mallette ». Elles présentent toutes les différents résultats sur un thème particulier puis tentent d'en tirer les principales leçons politiques. Je commencerai par présenter la « mallette » choisie (3.3.1). Puis, je développerai les différents résultats mis en avant au sein de cette « mallette » (3.3.2). Enfin, je détaillerai les leçons politiques qui en découlent (3.3.3).

3.3.1 Les comportements sexuels à risques : illustration d'une « mallette politique »

La mallette que je choisis d'étudier¹⁵⁷ s'intéresse aux comportements sexuels des adolescents en vue de diminuer les risques de transmission du VIH. La « mallette » expose tout d'abord le problème puis reprend des enjeux étudiés dans des évaluations randomisées menées par les chercheurs du J-PAL. Dans cette « mallette », il est expliqué que pratiquement trois millions de personnes sont atteintes du VIH en Afrique sub-saharienne, que très peu des personnes qui devaient bénéficier du traitement en bénéficient réellement (une personne sur six), et que la prévention contre le sida reste encore très fragile. Cette « mallette » se

¹⁵⁶ Le terme exact utilisé par le J-PAL est : « *policy briefcase* », mais il est traduit par les chercheurs du J-PAL par « mallette politique ».

¹⁵⁷ Policy Briefcase No. 3 : « *Cheap and Effective Ways to Change Adolescents' Sexual Behavior* ».

concentre sur une évaluation particulière : évaluer l'impact d'un programme qui fournit aux élèves des connaissances sur le VIH dans le but d'observer un comportement sexuel plus prudent. Il renvoie à une évaluation menée par Esther Duflo, Pascaline Dupas, Michael Kremer et Samuel Sinei¹⁵⁸ en 2006 au Kenya sur le rôle que peut jouer l'introduction d'uniforme. Elle reprend aussi les résultats du programme « *sugar daddies* »¹⁵⁹. Cette mallette est donc constituée de deux évaluations randomisées publiées concernant le lien entre la prévention et les comportements sexuels des adolescents. Il est souligné qu'elles sont toutes deux des solutions peu chères et souvent considérées comme efficaces : « *not only are these programs effective, they are also cheap* ». (Policy Briefcase No.3, 1)

3.3.2 Résultats des différents programmes

L'ensemble des résultats des différents programmes est ensuite résumé sous forme de tableau. La « mallette » se réfère à quatre programmes : offrir une formation aux enseignants en matière de prévention sur le VIH (1), encourager des débats au sein des écoles (2), le programme « *sugar daddy* » (3), et réduire les coûts de l'éducation en fournissant gratuitement des uniformes aux élèves afin qu'ils restent plus longtemps à l'école¹⁶⁰ (4).

La formation des enseignants visait à leur apprendre à rendre compte du programme voulu par le gouvernement, ce qui incluait de donner des connaissances sur la maladie, d'apprendre aux jeunes filles à refuser certaines relations sexuelles et mettait en avant l'abstinence avant le mariage. Cela a eu plusieurs impacts : le taux de mariage parmi les jeunes filles qui ont eu des enfants a augmenté. Cependant, il n'y a pas eu d'impact sur le niveau de grossesse des jeunes filles et sur leur niveau de connaissances en matière de VIH¹⁶¹. Le coût de ce programme est estimé à un dollar par élève. Le deuxième programme (encourager les débats au sein des écoles), a augmenté les connaissances des élèves, et la probabilité qu'ils utilisent des préservatifs ; mais il n'a eu aucun impact sur l'activité sexuelle des étudiants. Son coût

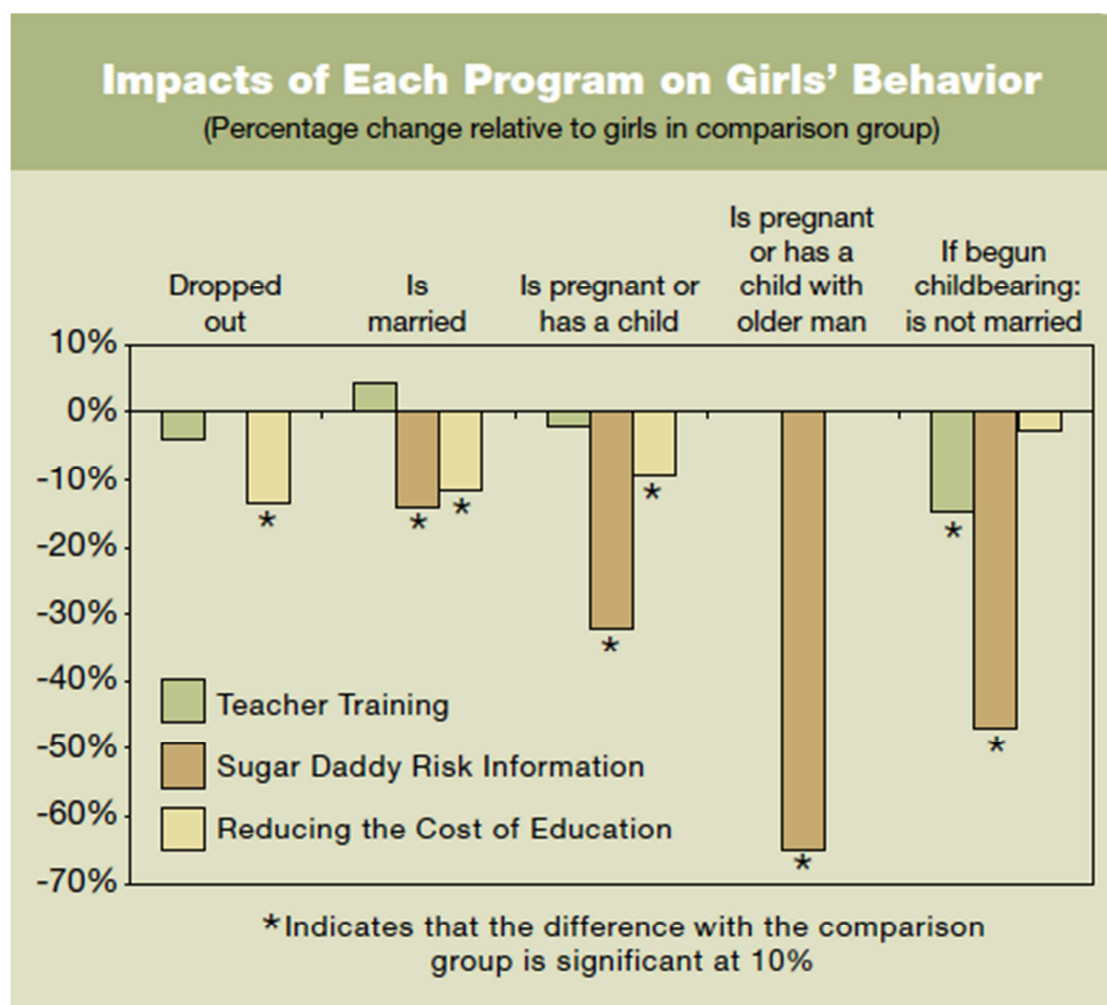
¹⁵⁸ Voir (Duflo, Dupas, Kremer et Sinei, 2006)

¹⁵⁹ Voir (Duflo, Dupas et Kremer, 2011)

¹⁶⁰ Les uniformes sont payants au Kenya et représentent donc un coût qui empêche certains adolescents de rester à l'école. Rendre ces uniformes gratuits permet alors aux adolescents de rester plus longtemps à l'école et cela tend à diminuer le nombre de grossesses et de mariages : « *While there are no school fees in Kenya, purchasing a school uniform is expensive and stops many from attending. In some schools, students were provided free uniforms. This reduced dropout by 17 percent for boys and 14 percent for girls, and reduced teenage childbearing and marriage among girls by 9 percent and 12 percent respectively.* » (Policy Briefcase No.3, 3)

¹⁶¹ Arcand et Djimeu Wouabe (2010), évaluent aussi, au Cameroun, l'impact d'une formation similaire sur les grossesses des adolescentes, ils montrent que ce type de formation réduit de 7 à 10% les grossesses chez les jeunes filles de 15 à 17 ans.

est, lui, de deux dollars par élève. Le troisième programme (*sugar daddy*) offrait, aux écoles sélectionnées, des sessions de quarante minutes où une ONG venait exposer les risques qu'implique une relation sexuelle avec un homme âgé. Ce programme réduit les grossesses au sein des adolescentes ayant une relation avec des hommes âgés, car il réduit les relations sexuelles non protégées avec des hommes plus âgés. Le coût par étudiant est d'un dollar et le coût pour chaque grossesse évitée est de quatre-vingt-onze dollars. Le dernier programme (fournir des uniformes gratuitement) réduit le taux d'abandon à l'école ainsi que les grossesses et les mariages. Il coûte onze dollars par garçon et douze dollars par fille, le coût des grossesses évitées est de sept cent cinquante dollars. Les différents impacts et coûts sont ensuite résumés sous forme de graphique :



Policy Briefcase No.3 (page 3)

Le changement des comportements des adolescentes, suite aux différents programmes évalués, a été mesuré par la réduction des grossesses et le mariage. Ces réductions sont

envisagées comme traduisant la réduction d'un comportement sexuel à risque¹⁶². Le programme « *sugar daddy* » est celui qui engendre le plus de changements dans le comportement des adolescentes. Pour autant, le programme de formation des enseignants reste le moins coûteux.

3.3.3 Les « leçons » politiques

Une fois les résultats et les coûts des programmes mis en lumière, la « mallette » conclut par des leçons politiques à destination des décideurs politiques. Ces leçons politiques ont pour but de rompre avec l'idéologie et d'encourager une démarche politique fondée sur des preuves :

« The debate about how to help prevent HIV/AIDS infection has too often degenerated into ideological battles based on little hard evidence. This research finds that there are cheap and effective ways to change the behavior of adolescents that should be acceptable to all those seeking to address the suffering caused by HIV/AIDS. » (Policy Briefcase, No.3, 3, souligné par moi)

La conclusion reprend ensuite brièvement l'ensemble des résultats. Avertir les adolescentes que les hommes plus âgés ont plus de risques d'avoir le sida, diminue le niveau de grossesses des adolescentes avec un homme plus âgé qu'elles. Le niveau de grossesse est aussi diminué en fournissant gratuitement des uniformes. Les conclusions politiques ne donnent aucune prévalence à un programme mais cherche à laisser les décideurs politiques choisir en fonction de leur budget. C'est pourquoi, sans doute, dans ces conclusions, l'accent est mis sur le rôle des preuves et le danger de l'idéologie, selon les auteurs.

4. Conclusion

Le J-PAL ne souhaite pas uniquement produire des preuves, il a aussi vocation à les rendre utilisables. Ces preuves visent à guider la décision politique. La maxime du J-PAL « transformer la recherche en action » prend donc tout son sens. Cela signifie deux choses : collecter des preuves et permettre aux décideurs politiques de les utiliser afin que les politiques établies soient fondées sur des preuves. Cette double dimension traduit clairement

¹⁶² De plus, la « mallette politique » suggère que même si la réduction des grossesses n'est qu'un équivalent à la réduction des comportements sexuels à risque, elle est en elle-même un objectif : « *Even in the absence of HIV, reducing pregnancy among primary school students is arguably a goal in its own right.* » (Policy Briefcase, No.3, 3)

le lien souhaité par le J-PAL entre sa méthode et celle des essais cliniques médicaux. Tout d'abord, la randomisation du J-PAL vise à se rapprocher au plus près des essais cliniques dans le but de posséder la même rigueur scientifique. Ensuite, l'approche du J-PAL s'inscrit dans un mouvement similaire à celui opéré en médecine – la médecine fondée sur les preuves – dont le cœur est l'essai clinique. Opérer un double aller-retour avec la médecine m'a permis d'éclairer ces deux aspects.

La volonté normative du J-PAL se traduit par le souhait de rendre ses résultats accessibles aux décideurs politiques. Cela implique de mener des analyses coûts-bénéfices afin de pouvoir comparer sur un même objet différentes possibilités ainsi que leurs coûts. De plus, afin de rendre ces résultats encore plus accessibles pour les décideurs politiques, le J-PAL constitue des « mallettes politiques ». Ces dernières sont dans la droite continuité des analyses coûts/bénéfices, mais visent à les illustrer en les rendant plus faciles d'accès. C'est pourquoi, pour les chercheurs du J-PAL, les agences internationales ont tout intérêt à favoriser de telles initiatives. Ce soutien apparaît nécessaire, compte tenu de l'important biais de publication dont souffre une telle approche. C'est donc, en partie, avec l'appui des agences internationales que le J-PAL entend étendre son rôle politique dans la lutte contre la pauvreté.

Conclusion de la Partie 1.

Retracer l'histoire de la randomisation m'a permis de mettre en évidence une des dimensions centrales de l'approche du J-PAL : la recherche de preuves. Les débuts de la randomisation en psychologie à travers les expériences de Peirce introduisent la dimension aléatoire comme un gage de validité pour les résultats obtenus. L'utilisation de cette dimension aléatoire dans les expériences en psychologie qui suivirent celles de Peirce autour de la télépathie joue un rôle identique. C'est pourtant Fisher qui, en 1935, offrira le cadre statistique et expérimental de cette approche. Fisher, à partir de techniques statistiques sophistiquées et d'un cadre expérimental rigoureux, met au cœur de son analyse la randomisation. Bien que ce cadre fût pensé pour l'agriculture, c'est en médecine qu'il connaîtra un écho important : c'est la médecine qui institutionnalisera une telle méthode, à travers les essais cliniques. Retracer la seconde histoire qui traverse la randomisation du J-PAL, l'histoire des expériences de terrain en économie m'a permis de traduire le second objectif du J-PAL – guider la décision politique. De plus, cela m'a permis de situer les expériences du J-PAL au sein d'une importante diversité d'expériences en économie. Ce premier chapitre, à l'aide des deux histoires qui traversent la randomisation du J-PAL, a permis de pointer les deux objectifs du J-PAL.

Le deuxième chapitre s'est particulièrement intéressé au premier objectif du J-PAL, en cherchant à définir un tel laboratoire. Ce laboratoire est défini par l'utilisation et la promotion de la randomisation. L'utilisation d'une telle méthode unit tous les chercheurs de ce laboratoire. Cette méthode, contrairement à d'autres méthodes non expérimentales, permet de créer un contrefactuel fiable, directement inspiré du modèle de Rubin. L'attribution aléatoire aux deux groupes constituant l'expérience permet d'annuler les biais de sélection, faisant d'une telle méthode le *gold standard* méthodologique. Cela permet donc au J-PAL de produire des preuves. De plus, la rencontre avec le terrain offerte par cette méthode permet au J-PAL de collaborer avec de nombreux partenaires. Cela tend à promouvoir une telle méthode. Cette promotion est aussi principalement garantie par la structure institutionnelle du J-PAL. Ce dernier est désormais présent sur chaque continent et travaille avec plus de 500 partenaires. Ce deuxième chapitre, à travers la définition du J-PAL, a permis de développer le premier objectif du J-PAL.

Finalement, le J-PAL ne cherche pas uniquement à produire des preuves, il cherche à les rendre utilisables. L'objectif est de fonder la lutte contre la pauvreté sur des preuves, retrouvant le mouvement de l'*evidence-based policy*. Le dernier chapitre s'est alors intéressé à montrer comment le J-PAL s'insère au sein de ce mouvement. Le lien qui unit la médecine à l'approche du J-PAL est donc, ici, particulièrement éclairant. C'est pourquoi, je me suis tout d'abord attachée à démontrer les similarités entre les essais cliniques médicaux et la randomisation, puisque ces derniers permettent de collecter les preuves qui seront utilisées au sein de l'*evidence-based medicine* pour la médecine et de l'*evidence-based policy* pour l'économie du développement. Puis, ce chapitre a opéré un second aller-retour avec la médecine afin de définir précisément l'*evidence-based policy*, puisque ce mouvement n'est que la traduction en politique de celui initié en médecine au début des années 1990. Ces deux allers-retours m'ont alors permis de définir le second objectif du J-PAL ainsi que la manière dont les chercheurs du J-PAL entendaient le réaliser. Les deux objectifs du J-PAL sont désormais explicites et se retrouvent autour d'une dimension principale : la preuve. C'est pourquoi cette partie a traité explicitement de la forte validité interne de la randomisation.

Partie 2. De la validité interne à la validité externe de la randomisation du J-PAL : la fin du *gold standard* méthodologique ?

Introduction de la Partie 2.

« Il y aurait donc un choix à faire entre les études par assignation aléatoire dont la validité interne est forte, et les études non expérimentales dont la validité externe est plus grande. Pourtant, cela n'est pas nécessairement vrai. Une partie du problème est de savoir ce que l'on veut dire par effet généralisable : cela signifie que si l'on mène la même action dans un lieu différent, on obtiendra le même résultat. Mais de quelle action et de quel résultat parlons nous ? » (Banerjee et Duflo, 2009, 706)

La méthode du J-PAL bénéficie d'une importante validité interne, faisant de cette dernière le *gold standard* méthodologique en matière d'évaluation d'impact. Cependant, les expériences du J-PAL sont fortement critiquées sur leur faible validité externe. De telles expériences offrent des résultats solides, mais ce dans le cadre très précis de l'expérience. Or, si l'effet positif d'un programme est évalué dans une région particulière du monde, on ignore si les mêmes effets seront obtenus dans une autre région. Les résultats obtenus par le J-PAL dépendent donc, d'une certaine manière, de leur contexte. De plus, si les expériences menées par le J-PAL se réalisent à petite échelle, il n'est pas sûr qu'étendre un programme à une plus grande échelle garantisse les mêmes effets. La question de la validité externe se pose alors à deux niveaux. Le premier niveau traduit l'utilisation politique des résultats offerts par le J-PAL, tandis que la généralisation d'un programme à une échelle plus large renvoie au second niveau de la validité externe d'une telle approche. Il semble que l'approche du J-PAL se concentre sur la validité interne de ses expériences laissant de côté les deux niveaux de la validité externe. Cela tend à fragiliser le second objectif que se fixe le J-PAL : guider la décision politique. Afin de réaliser ces deux objectifs, le J-PAL aurait donc un arbitrage à faire entre sa validité interne et sa validité externe.

L'objectif de cette partie sera de s'intéresser à l'arbitrage entre la validité interne et la validité externe de la randomisation du J-PAL. L'analyse de Nancy Cartwright me permettra de poser un cadre épistémologique précis, tout en rendant compte du premier niveau de la validité externe de l'approche du J-PAL. Le second niveau, quant à lui, sera développé au travers des différentes critiques adressées à l'approche du J-PAL, ce qui me permettra d'analyser ces critiques. Enfin, selon moi ces deux niveaux de la validité externe se retrouvent autour du rôle accordé à la théorie dans la randomisation du J-PAL.

Le premier chapitre de cette partie développera le cadre analytique de Cartwright afin de penser le premier niveau de la validité externe du J-PAL. Cartwright distingue la recherche de preuve, de son utilisation ; offrant alors un cadre épistémologique extrêmement précis sur la randomisation. Aussi, si l'on considère que la justification de la preuve renvoie à la validité interne, et son utilisation renvoie à la validité externe; la randomisation, selon Cartwright, permet d'offrir des preuves très solides, mais dont la portée demeure limitée. Même si l'analyse de Cartwright permet de penser le premier niveau de la validité externe de la randomisation, le second niveau n'y est pas particulièrement explicité.

Le deuxième chapitre de cette partie s'intéressera au second niveau de la validité externe de la randomisation du J-PAL. Pour cela, il se centrera sur les critiques adressées ces cinq dernières années par de nombreux économistes du développement. L'ensemble de ces critiques s'accorde sur le fait que la randomisation du J-PAL, en se centrant sur la question de savoir « ce qui fonctionne », et non « pourquoi un programme fonctionne ? », produit des résultats identiques à une boîte noire. En ne s'intéressant pas à « pourquoi un programme fonctionne ? », il apparaît difficile de déterminer les effets qu'un tel programme puisse avoir à une échelle plus large. Cette vague de critiques met alors l'accent sur le second niveau de la validité externe, questionnant la généralisation des résultats obtenus par le J-PAL, et presse ce dernier de s'intéresser à « pourquoi un programme fonctionne? ». Cela sous-entend de formuler des hypothèses, et de ne pas se limiter à l'étude d'une dimension, en abordant une perspective plus structurelle. Ces suggestions tendent à fragiliser la validité interne de la randomisation du J-PAL, tout en favorisant sa validité externe. Cela pointe donc très clairement la tension existante au sein de telles expériences : un arbitrage entre sa validité interne et sa validité externe. En ce qui concerne le J-PAL, celui-ci privilégie explicitement la validité interne de son approche, au détriment de sa validité externe.

Le troisième chapitre portera sur le statut de la théorie au sein de la randomisation du J-PAL. Les principales critiques à l'encontre de la randomisation pointent donc le fait que celle-ci ne se concentre pas suffisamment sur l'explication et la mise en évidence des mécanismes sous-jacents aux résultats qu'elle produit. C'est la dimension a-théorique d'une telle approche qui me semble unifier l'ensemble des critiques adressées au J-PAL. Lors de ce chapitre, je montrerai que le statut de la théorie au sein de la randomisation du J-PAL ne se limite pas à sa simple absence. Tout d'abord, afin de garantir la validité interne de leur approche, les chercheurs du J-PAL se refusent à toute théorie *a priori*. Néanmoins, leur objectif étant de

parvenir - à partir des résultats obtenus - à la construction d'une théorie *a posteriori*, il apparaît difficile d'envisager cette théorie *a posteriori* sans évoquer une théorie préalable, qui guiderait les différentes expériences. Par ailleurs, ce défi théorique possède sa traduction politique : comment passer de la recommandation politique locale à des recommandations politiques globales ? Cette question traduisant explicitement le second niveau de la validité externe de la randomisation du J-PAL, ce passage reste opaque dans l'approche de Duflo et Banerjee. Questionner le statut de la théorie au sein de la randomisation du J-PAL me permettra donc de penser simultanément des deux niveaux de la validité externe d'une telle approche. Je montrerai alors que ces deux niveaux sont extrêmement fragiles ne permettant pas au J-PAL de réaliser pleinement le second objectif qu'il se fixe. L'accent porté sur la preuve empêche une telle approche de guider la décision politique, rendant ces deux objectifs contradictoires.

Chapitre 1 : La randomisation à la lumière de l'épistémologie de Nancy Cartwright

0. Introduction

« There is no gold standard; no universally best method. Gold methods are whatever methods will provide a) the information you need, b) reliably, c) from what you can do and from what you can know on the occasion. Often Randomized Controlled Trials (RCTs) are very bad at this and other methods very good. What method best provides the information you want reliably will differ from case to case, depending primarily on what you already know or can come to know. » (Cartwright, 2007a, 11)

Il paraît difficile d'utiliser le résultat d'une expérience dans un autre contexte que celui dans lequel elle a été menée. Il peut être efficace de donner des médicaments contre les vers intestinaux dans une région particulière du Kenya pour réduire l'absentéisme des élèves, afin de favoriser l'éducation. Mais cela veut-il dire que ce traitement sera efficace dans une autre région du monde pour favoriser l'éducation ? Il semble impossible de répondre à cette question. Pour autant, l'approche du J-PAL souhaite s'adresser directement aux décideurs politiques. Ainsi, les résultats obtenus au Kenya doivent pouvoir être utilisés, d'une manière ou d'une autre, par un décideur politique zambien qui souhaiterait lutter contre la pauvreté de son pays. Mais s'il est impossible de savoir si le traitement qui est efficace au Kenya peut être efficace dans une autre région du monde, comment les différents décideurs politiques pourront-ils utiliser les résultats de ces expériences ?

L'approche en termes de capacités causales, développée par Nancy Cartwright, peut permettre de répondre à ces interrogations. Tout d'abord, Cartwright (2009c) distingue l'efficacité réelle de l'efficacité potentielle¹⁶³. L'efficacité réelle renvoie au résultat d'une expérience particulière, dans une situation donnée. Alors que l'efficacité potentielle renvoie à l'efficacité des résultats en dehors de l'expérience et rend compte de leur efficacité en pratique, de leur utilisation. Pour reprendre l'exemple cité plus haut, le fait de considérer que les médicaments contre les vers intestinaux sont efficaces pour améliorer l'éducation au Kenya renvoie à une efficacité réelle, car il s'agit d'une expérience donnée menée dans un cadre particulier. Déterminer si ce traitement peut être efficace ailleurs, en Zambie par exemple, relève de l'efficacité potentielle du résultat obtenu au Kenya. La distinction entre efficacité réelle et efficacité potentielle permet de penser le passage entre des résultats

¹⁶³ Nancy Cartwright parle « d'*efficacy* » et « d'*effectiveness* », je traduis le premier terme par efficacité réelle et le second par efficacité potentielle.

obtenus dans une situation particulière à une autre situation. En ce sens, cette distinction permet de penser l'utilisation possible que les décideurs politiques peuvent avoir des résultats des expériences randomisées du J-PAL. Cette distinction est donc particulièrement éclairante pour questionner la validité externe des expériences de Duflo et de Banerjee. La distinction opérée par Cartwright permet de traduire le premier niveau de la validité externe de telles expériences : l'utilisation des résultats des expériences.

La question devient alors : comment passer de l'efficacité réelle à l'efficacité potentielle ? Pour Nancy Cartwright, il est possible de passer de l'efficacité réelle à l'efficacité potentielle d'un résultat si l'on a *une bonne raison de penser* que le résultat obtenu peut avoir les mêmes effets ailleurs. Cette *bonne raison de penser* pointe l'existence de ce que Cartwright appelle les capacités causales. Un exemple simple de capacités causales serait : l'aspirine a la capacité de réduire les migraines. Sachant cela, il devient possible de dire à quelqu'un qui souffre de migraines de prendre une aspirine afin que sa migraine disparaisse. On peut donc ici, passer de l'efficacité réelle de l'aspirine à son efficacité potentielle. Pour Cartwright ce passage se fait grâce à l'explicitation des capacités causales en jeu. Cartwright définit la validité externe de telles expériences à travers ce passage. L'objectif de ce chapitre est donc de définir l'épistémologie de Cartwright afin de rendre compte du premier niveau de validité externe des expériences du J-PAL, et donc de questionner la portée de telles expériences.

En distinguant l'efficacité réelle de l'efficacité potentielle, Cartwright sépare deux niveaux d'analyses : la garantie de l'affirmation causale (comment la preuve permet-elle de garantir une efficacité réelle ?) et son utilisation (l'affirmation causale d'efficacité réelle peut-elle être utilisée ailleurs, peut-on supposer l'efficacité potentielle d'une telle affirmation causale ?) Pour plus de clarté, je distinguerai ces deux niveaux d'analyses. La première partie de ce chapitre traitera de la garantie des affirmations causales. Cela me permettra de définir exactement quel type de garantie causale offre la randomisation pour Cartwright. Dans la deuxième partie de ce chapitre, je m'intéresserai à l'utilisation de ces garanties causales. Cela me conduira à définir la théorie des capacités causale de Cartwright. Cette théorie permet de penser le passage de l'efficacité réelle à l'efficacité potentielle. Enfin, dans la troisième partie de ce chapitre, j'appliquerai ce cadre d'analyse à la randomisation de Duflo et de Banerjee, afin de déterminer la portée de leur approche en termes d'utilisation politique. Cela me permettra de montrer que le premier niveau de la validité externe de la randomisation du J-PAL est faible.

1. La randomisation « idéale » de Nancy Cartwright

L'épistémologie de Cartwright se fonde sur celle de John Stuart Mill. Cette dernière cherche à penser une pluralité de causes. Dans cette perspective, un phénomène est le fruit de plusieurs causes. À l'instar de Mill, Cartwright considère que ces causes ne sont pas déterminées, qu'elles ne sont pas stables. C'est pourquoi, elle ne propose pas une approche déterministe de la causalité mais une approche probabiliste. Cartwright reprend l'une des méthodes inductives, fondée sur les probabilités, définie par Mill (1843) : la « méthode de différence »¹⁶⁴. Selon elle, cette méthode s'apparente à la randomisation. En ce sens, Cartwright offre un premier niveau de définition de la randomisation : c'est une méthode qui garantit une causalité probabiliste, fondée sur la méthode de différence de Mill. De plus, la randomisation offre des résultats dont la validité interne est forte, mais dont la validité externe reste fragile. En ce sens, la randomisation est ce que Cartwright appelle un « vérificateur ». Ces deux dimensions me permettent de définir ce que la randomisation permet d'offrir comme garantie causale. Celles-ci définissent ce que Cartwright nomme « randomisation idéale ». Cette dernière est donc un « vérificateur » qui se fonde sur une théorie probabiliste de la causalité. Je définirai tout d'abord, ce qu'est un « vérificateur » chez Cartwright, cela me permettra de montrer comment les preuves peuvent être justifiées au sein de l'épistémologie de Cartwright (1.1). Puis, je m'attacherai à développer le deuxième niveau de définition de la randomisation pour Cartwright, c'est à dire sa dimension probabiliste (1.2). Cela me permettra de définir le type de preuves que la randomisation permet de produire.

1.1 La justification de la preuve

Les méthodes qui permettent de justifier des affirmations causales se retrouvent, pour Cartwright en deux catégories : il y a les méthodes qui « vérifient » et celles qui « garantissent »¹⁶⁵ :

¹⁶⁴ Mill (1843) définit cinq méthodes inductives : la méthode d'accord direct (1), la méthode de différence (2), la méthode de différence et d'accord (3), la méthode de résidus (4) et la méthode de variations concomitantes (5). Je ne développerai dans ce chapitre que la méthode de différence, puisque c'est la méthode dont s'inspire Cartwright.

¹⁶⁵ Cartwright parle de « clincher » et de « voucher », je choisis la traduction suivante : vérificateur pour « clincher » et garantisseur pour « voucher ».

« Methods for warranting causal claims falls into two broad categories. There are those that clinch the conclusion but are narrow in their range of application; and those that merely vouch for the conclusion but are broad in their range of application. » (Cartwright, 2007b, 25)

Les premières, les méthodes qui « vérifient » , permettent d’obtenir des conclusions solides et fiables dans un cadre précis et étroit, mais ces conclusions sont difficilement généralisables. A l’inverse, les « garantisseurs » proposent des conclusions moins solides, mais leur généralisation est plus facile. On a, avec cette opposition, un balancement clair entre d’un côté une forte validité interne et une faible validité externe (ie. « vérificateur »), et de l’autre une faible validité interne et une forte validité externe (ie. « garantisseur »). Comme intermédiaire entre ces deux catégories de justification causale, Cartwright cite la méthode hypothético-déductive. Cette dernière offre des justifications moins solides que les « vérificateurs » mais plus solides que les « garantisseurs ». Cependant, leur champ d’application est plus large que celui des « vérificateurs », mais moins grand que les « garantisseurs ». Je présenterai, tout d’abord, les méthodes qui « garantissent » (1.1.1), puis la méthode hypothético-déductive (1.1.2), et finalement les méthodes qui « vérifient ». Cela me permettra de montrer la justification causale que permet la randomisation, tout en la situant au sein du cadre analytique de Cartwright.

1.1.1 Les méthodes qui offrent des « garantisseurs »

Les méthodes qui « garantissent » sont les suivantes : les méthodes qualitatives comparatives (1), les raisonnements à partir de modèles ou les systèmes-modèles (2), les méthodes ethnographiques (3), et des supports indirects mixés (4). La méthode qualitative comparative débute avec les travaux de John Leslie Mackie (1988) sur la condition INUS (insufficient but non redundant part of an unnecessary but sufficient condition) dans la causalité¹⁶⁶. Pour Mackie (1988), il existe une pluralité de causes et selon lui ces causes sont des conditions INUS. La condition INUS s’oppose à l’idée d’existence de conditions nécessaires et suffisantes, puisqu’ici la causalité est pensée à l’aide d’un ensemble de facteurs. Angus Deaton considère, d’ailleurs, que la randomisation ne produit au mieux que des

¹⁶⁶Un exemple de condition INUS : si un court-circuit est la cause du feu d’une maison, cela sera une condition INUS. Le court-circuit est à lui seul insuffisant pour causer le feu, il faut y ajouter d’autres facteurs comme la présence de matériaux inflammables, la présence d’oxygène etc. Le court-circuit est aussi non-redondant car sans lui les autres facteurs ne sont pas suffisants pour mettre le feu. Le court-circuit n’est qu’une partie d’un ensemble d’autres conditions suffisantes (oxygène, matériaux inflammables, etc.), mais cet ensemble de conditions suffisant n’est pas nécessaire car un ensemble d’autres conditions peuvent brûler une maison comme un pyromane muni d’essence et d’un briquet peut mettre le feu à une maison.

affirmations causales qui sont des conditions INUS. La deuxième méthode (les raisonnements à partir de modèles ou les systèmes-modèles) permet de tirer des conclusions causales à partir d'un modèle. Quant à la troisième (les méthodes ethnographiques¹⁶⁷) elle est tournée vers l'étude descriptive et analytique de différentes sociétés en vue d'en sortir des conclusions causales. Enfin, la dernière méthode (les supports indirects mixés) est illustrée par Cartwright autour du travail de Michael Marmot¹⁶⁸. La question de Marmot est la suivante : les faibles revenus causent-ils une mauvaise santé, ou est-ce l'inverse? Marmot montre que le stress causé par de faibles revenus engendre une mauvaise santé. Pour cela, il s'appuie sur de nombreux supports : des études statistiques, des études médicales sur la relation physiologique entre le stress et des problèmes de santé, et des études longitudinales. Ces méthodes permettent toutes de penser une pluralité de causes, cela à l'aide de divers supports méthodologiques, ce qui leur offre un champ d'application large. En revanche, la fiabilité de leur résultats n'est pas toujours solide : « *the evidence is symptomatic of the conclusion but not sufficient for it.* » (Cartwright, 2007b, 25)

1.1.2 La démarche hypothético-déductive : un « intermédiaire » chez Cartwright

Cartwright (2007a, 2007b) illustre la méthode hypothético-déductive à l'aide d'une des recherches d'Angus Deaton (2001) sur le lien entre la précarité économique et une santé fragile. Je reprends cette illustration, car elle permet de situer la méthodologie de Deaton¹⁶⁹. A l'inverse de Marmot, pour Deaton les individus en mauvaise santé peuvent difficilement travailler, ce qui tend à diminuer leur revenus, et finalement à les placer dans une précarité économique. Deaton fonde son argumentation sur une étude de la *National Longitudinal Mortality Study*. Dans cette étude, il existe une corrélation entre de faibles revenus et un faible niveau d'éducation, puis une corrélation entre de faibles revenus et la mortalité. C'est alors que Deaton établit l'ordre du lien entre mauvaise santé et faibles revenus : une mauvaise santé entraîne de faibles de revenus. Néanmoins, Cartwright souligne que l'affirmation de Deaton se limite à l'étude sur laquelle elle se fonde. Les hypothèses reçoivent confirmation, mais « *it is very hard to say how much confirmation to award it nor how far beyond the National Longitudinal Mortality Study data set to suppose it will hold.* » (2007b, 54)

¹⁶⁷ Pour plus de précisions sur les méthodes ethnographiques voir (Cartwright et Efstathiou, 2008).

¹⁶⁸ Voir (Marmot, 2004).

¹⁶⁹ Je reviendrai sur la méthodologie de Deaton, plus en détail, et questionnerai son approche dans le chapitre suivant.

La méthode hypothético-déductive permet d'infirmer ou de confirmer une hypothèse, mais elle le fait de façon moins solide que les méthodes qui « vérifient » ; pour autant, son champ d'application est plus large.

1.1.3 Les méthodes qui offrent des « vérificateurs »

Les méthodes qui « vérifient » sont les suivantes : les dérivées de la théorie (« *derivation from theory* ») (1), les mécanismes (2), le réseau Bayésien (*Bayes-net*) (3), les méthodes économétriques (4), les expériences contrôlées ou naturelles (5), causalité probabiliste (6) et la randomisation (7). La première méthode cherche à mettre en lumière des inférences causales à partir de la théorie. La deuxième cherche à retracer le procédé causal qui relie une cause à un effet. La troisième est un modèle graphique probabiliste, comme la relation probabiliste entre un symptôme et différentes maladies. Cartwright définit la quatrième méthode de la façon suivante :

« Econometrics has well-developed 'structural' methods that allow the deduction of the strength of causal connections between factors in a preselected variable set, provided stringent are met. These methods begin by assuming that a particular set of functional forms correctly represents the causal structure generating the observed data. What is to be discovered are the parameters that turn these functional forms into real functions. » (Cartwright, 2007b, 62)

Les méthodes économétriques permettent de rendre compte de relations causales entre des facteurs et des variables sélectionnées. Les expériences contrôlées ou naturelles tentent de mettre en évidence des inférences causales à partir d'expériences de terrain. Les deux dernières méthodes sont, pour Cartwright, intimement liées. Cela a été déjà évoqué, la randomisation est, pour Cartwright, fondée sur une causalité probabiliste. La randomisation offre des affirmations causales qui « vérifient », c'est à dire qui possède une solide validité interne mais dont le champ d'application est restreint.

1.2 « Randomisation idéale » et théorie probabiliste de la causalité

Ce que Cartwright appelle « une randomisation idéale » se caractérise par l'aspect « vérificateur » des affirmations causales que la randomisation offre. Mais de quoi sont constituées ces affirmations causales ? C'est ici que la causalité probabiliste entre en jeu. Elle va permettre de définir le contenu des affirmations causales produites par la randomisation :

« *I have claimed that ideal RCTs¹⁷⁰ are clinchers. That of course depends on how they are defined. But they are perfectly natural definitions from which it can be proved that RCTs, as thus defined, allow causal claims about the population in the study to be deduced from probability differences between treatment and control group.* » (Cartwright, 2007a, 15)

La randomisation est donc un « vérificateur » permettant de définir une causalité à l'aide de la différence de probabilités entre le groupe de traitement et le groupe de contrôle. Je définirai, tout d'abord, ce que Cartwright entend par théorie probabiliste de la causalité (1.2.1), puis je montrerai comment cette théorie de la causalité se lie à la randomisation (1.2.2). Cela permettra de définir précisément la nature des affirmations causales offertes par la randomisation.

1.2.1 Une théorie probabiliste de la causalité

L'idée centrale de cette théorie est que les dépendances probabilistes doivent avoir une explication causale. Son objectif est de déterminer, à l'aide de la théorie des probabilités, la relation qui unit une cause à un effet. L'idée sous-jacente est que les causes entraînent les probabilités de leurs effets. Le fait de se référer aux probabilités s'oppose à une vision déterministe de la causalité. Cette vision déterministe suppose que si C cause E, alors C doit toujours être suivi par E. L'idée de causalité probabiliste est moins stable, dans une perspective probabiliste C cause E, si l'occurrence de C augmente la probabilité de E. Cartwright traduit la théorie probabiliste de la façon suivante :

« *C causes E in K_i iff $P(E/C \& K_i) > P(E/\neg C \& K_i)$ and for any population A, C causes E in A iff C causes E in some K_i that is a subset of A.* » (Cartwright, 2010b, 62)

En résumé, C cause E dans une sous-population K_i si la probabilité que C cause E dans cette sous-population est supérieure à la probabilité que E ne cause pas C dans la sous-population K_i . Ici, c'est donc à l'aide d'une différence de probabilité, qu'une causalité peut être établie.

¹⁷⁰ RCTs renvoie à *randomized controlled trials* (essais contrôlés randomisés).

1.2.2 Ce que garantit la randomisation à l'aide des probabilités

Pour mettre en lien la théorie probabiliste de la causalité et la randomisation, Cartwright reprend la méthode par différence de Mill. Cette méthode compare les différences de probabilités entre le résultat (O) avec ou sans intervention (T) au sein de deux groupes. Cette approche suppose une « fixation causale »¹⁷¹. La randomisation n'est rien d'autre, pour Cartwright, que la méthode par différence de Mill. Compte tenu de la théorie probabiliste de la causalité et de ce fixage causal, la randomisation permet de déterminer que T cause O dans une sous-population donnée :

« Given CF¹⁷², if a positive probabilistic difference obtains and if the two groups have identical distributions of other causal factors, it follows T causes O in some subpopulation ϕ , of the population, X, of the individuals in the study that is causally homogeneous with respect to O. » (Ibid)

La randomisation permet de déterminer un lien causal dans une population restreinte, ϕ . On retrouve (ici illustré) ce que Cartwright expose lorsqu'elle définit la randomisation comme un « vérificateur » : une méthode qui fournit des liens causaux solides (fondée sur une théorie probabiliste et comprenant un fixage causal) mais dont la portée est réduite à une sous-population (ϕ). La « randomisation idéale » offre des preuves certaines d'efficacité réelle¹⁷³. Une fois cela posé, on peut, désormais, se demander si ces preuves d'efficacité réelle peuvent devenir des preuves d'efficacité potentielle.

2. Rendre les preuves utiles ou la nécessité d'un modèle causal

Cartwright distingue la « randomisation idéale » de la « randomisation en pratique ». Cette dernière renvoie à l'utilisation des affirmations causales de la randomisation. Autrement dit, elle traite de leur efficacité potentielle. Selon Cartwright, pour passer de la « randomisation idéale » à la « randomisation en pratique » il est nécessaire de disposer d'une théorie de la preuve. Cette théorie nécessite de penser un modèle causal qui rendrait compte des différentes causes et de leurs interaction, afin d'être en mesure de penser si l'on peut transposer les résultats d'une expérience à une autre situation. L'explicitation de ce modèle causal passe, pour Cartwright, par la mise à jour des capacités causales à l'œuvre. Ce modèle

¹⁷¹ « The probability of an effect is fixed by the values taken by a full set of its causes. » (Cartwright et Munro, 2010, 261)

¹⁷² CF renvoie à *Causal Fixing* que je traduis par fixage causal.

¹⁷³ Je reviendrai sur cette notion dans la suite de ce chapitre.

causal permettrait donc de penser l'efficacité potentielle à partir de l'efficacité réelle. Le problème est que, pour Cartwright, cette théorie de la preuve est manquante. Je commencerai par expliciter pourquoi cette théorie de la preuve est manquante (2.1), puis je développerai la notion de modèle causal chez Cartwright (2.2), et enfin je définirai le rôle des capacités au sein de ce modèle causal (2.3). Cela montrera comment la notion de capacités causales permet de penser le passage de l'efficacité réelle à l'efficacité potentielle.

2.1 Une théorie manquante de la preuve ?

Selon Cartwright, afin d'évaluer l'efficacité potentielle d'une politique proposée, on a besoin des preuves de l'efficacité de cette politique et l'on a besoin de savoir comment utiliser ces preuves. :

« In evaluating the effectiveness of a proposed policy we need credible evidence that speaks for or against the policy and we need to know what to do with the evidence when we have it. » (Cartwright, 2008, 1)

Cela génère trois questions : (1) qu'est ce qui compte comme une preuve crédible ? (2) Quelles preuves sont pertinentes ? (3) Comment les preuves doivent-elles être intégrées ? Ces trois questions traduisent trois dimensions de la preuve : sa crédibilité, sa pertinence, et son intégration. C'est en répondant à ces trois questions qu'il devient possible d'utiliser les preuves dans le domaine politique. La première question, à savoir la question de la crédibilité de la preuve, renvoie à sa probabilité d'être vraie : quelle preuve est potentiellement vraie ? La réponse à cette question dépend de comment la preuve a été « faite », de sa solidité interne, donc de sa justification (ce que j'ai développé lors de la partie précédente - c'est-à-dire s'il s'agit d'un « vérificateur » ou d'un « garantisseur »). Les deux dernières questions renvoient à l'utilisation de la preuve, à sa dimension externe de la preuve. Ce sont ces deux dernières questions que je vais m'attacher à développer, puisque j'ai déjà traité de la première. Je vais donc, tout d'abord m'intéresser à la question de la pertinence de la preuve (2.1.1), puis à celle de son intégration (2.1.2).

2.1.1 La pertinence de la preuve

En matière d'*evidence-based decisions*¹⁷⁴, les preuves sont hiérarchisées¹⁷⁵. Pour Cartwright, cette hiérarchie souffre de nombreuses déficiences. Tout d'abord, seules les méthodes qui offrent des « vérificateurs » sont intégrées à cette hiérarchie. Ce classement se centre sur l'efficacité réelle, et omet de nombreux aspects comme les effets indésirables, la question de l'implantation, les effets moraux et culturels. Cette hiérarchie ne rend pas compte de la pertinence de la preuve. Afin de questionner cette pertinence, Cartwright prend un exemple particulier. Elle imagine que l'on souhaite évaluer l'effet d'un antibiotique dans le but de réduire la mortalité infantile des enfants atteints du sida. Le traitement n'aurait pas pour objectif de guérir le sida, mais de renforcer les défenses immunitaires, afin qu'ils ne meurent pas à cause de maladies opportunistes que le système immunitaire trop faible n'aurait pu combattre. Une expérience randomisée pourrait montrer que ces antibiotiques réduisent de 40% le risque de décès des enfants vivant avec le sida en Zambie. Que faire de ce résultat ? Comment l'interpréter, mais surtout comment l'utiliser ? Selon Cartwright, ce résultat n'a de sens que si l'on suppose certaines hypothèses :

1. Il existe une population d'enfants infectés par le VIH, pour qui une expérience randomisée (en Zambie, par exemple) établit que l'antibiotique réduit la mortalité de 40%.
2. La population cible (la population qui intéresse le décideur politique) est semblable à la population Zambienne de l'expérience.
3. Alors, si le décideur politique administre l'antibiotique à la population cible, la mortalité infantile va diminuer.

Pour Cartwright, avec la randomisation tout se passe comme si ces trois hypothèses étaient vérifiées. Autrement dit, tout se passe comme s'il était possible de passer de la preuve (la randomisation en Zambie) à une conclusion politique dans un autre contexte. Cela est problématique, pour Cartwright, car le résultat obtenu en Zambie peut très bien n'avoir aucun impact ailleurs :

¹⁷⁴ L'*evidence based decision* renvoie au mouvement politique récent qui souhaite que les décisions politiques soient fondées sur des preuves.

¹⁷⁵ Cartwright se réfère, principalement, à la hiérarchisation des preuves offertes en criminologie par l'université du Maryland, que j'ai développé lors du chapitre 3 de la partie 1. Il avait été souligné que cette hiérarchisation était identique à celle proposée en médecine autour des essais cliniques.

« *What's wrong here is what generally wrong in supposing you can read off conclusions about single cases from scientifically established principles : almost all principles are defeasible and those that are not (like 'All men are mortal') do not provided very detailed advice.* » (Cartwright, Goldfinch et Howick, 2008, 13)

Les conclusions obtenues en Zambie peuvent être invalides dans un autre contexte, et comme le souligne Cartwright, même si la conclusion obtenue avait été plus générale, elle aurait sans doute été moins informative. On est alors face à deux problèmes majeurs :

- 1 On sait seulement que la population de l'expérience est zambienne, mais on ne connaît pas ses caractéristiques. Cela signifie que le résultat obtenu en Zambie peut être extrêmement intéressant mais inutile pour une autre population qui se distingue très largement de la Zambie¹⁷⁶.
- 2 Et même si la population cible s'apparentait à la population zambienne, l'implantation de la politique peut transformer la distribution et la structure causale sous-jacente, et de ce fait modifier les caractéristiques de la population cible.

Pour réaliser le passage de l'efficacité réelle à l'efficacité potentielle, trop d'informations sont manquantes. On ne sait pas explicitement ce qui permet de produire l'effet (O), dans la population (1). On ignore également si la pertinence des facteurs perdurera dans une population apparemment semblable à celle de la population cible (2).

2.1.2 Intégration de la preuve et théorie de la preuve

Pour ce qui concerne la troisième question, celle de savoir comment combiner les preuves ensembles, il semble impossible d'y répondre si on ne sait pas ce qu'offrent exactement les preuves de la randomisation. Si l'on ne peut juger de la pertinence des preuves fournies par la randomisation, il est impossible, alors, de les combiner. L'intégration de la preuve reste une question sans réponse. Au final, pour Cartwright, la question de l'intégration de la preuve, comme celle de la pertinence sont délaissées au profit de celle de la crédibilité de la preuve :

¹⁷⁶ « *The guidelines may be able to provide sound advice but it is not practicable advice: we don't know how to tell whether we are following it or not.* » (Cartwright, Goldfinch et Howick, 2008, 13)

« We put a huge amount of effort and expense into question 1, which ensures that there really is a causal connection between programme and outcome somewhere. We do so in aid of evidence-based policy. But question 1 is not a policy question; it is a question of pure science. » (Cartwright, 2008, 4)

Cartwright explique alors que différents récits plus ou moins contruits sont développés afin de contrecarrer l'absence de réponse claire aux questions 2 et 3¹⁷⁷. Il manque pour Cartwright une théorie de la preuve qui permettrait d'échapper à ces différents récits et qui offrirait un éclairage aux deux dernières questions, et donc de penser l'efficacité potentielle de la randomisation. Pour Cartwright la tâche de la philosophie est, ici, très importante. Cette théorie doit avoir une dimension pratique qui doit rendre audible les preuves aux décideurs politiques. Cette théorie de la preuve passe, pour Cartwright, par la définition d'un modèle causal.

2.2 Un modèle causal : éléments de définition

Avec la randomisation, les résultats arrivent de façon séparée, parfois éclatée. Il faut alors réunir ces résultats, pour cela un modèle causal est nécessaire :

« So we need more than a piecemeal knowledge of what causes what. We need a causal model. Again, our methodologies for how to construct causal models for new target situations from even highly stable well-warranted causal claims are very poor. » (Cartwright, 2007b, 77)

La seule possibilité pour envisager la validité externe de la randomisation est de définir un modèle causal. Ce modèle permettra de rendre compte de la pertinence des preuves, et donc de les combiner. Je commencerai par illustrer ce qu'est un modèle causal pour Cartwright (2.2.1), puis je montrerai qu'il est ensuite nécessaire de rendre ce modèle causal explicite (2.2.2).

2.2.1 Le modèle causal : une illustration

Afin d'expliciter ce qu'est un modèle causal, Cartwright choisit d'imaginer une situation de la vie quotidienne. Elle imagine qu'elle a accès à un aimant de bureau et à un aimant industriel (beaucoup plus imposant), et qu'elle souhaite soulever une masse qui bloque son

¹⁷⁷ *« We begin to construct a variety of different narratives, some more plausible or more probable than others. » (Cartwright, 2009c, 17)*

allée. Elle connaît parfaitement la force de ces aimants, elle connaît parfaitement l'efficacité réelle de chacun des aimants. Elle sait qu'ils sont capables de soulever des masses, mais qu'ils ont aussi besoin d'être liés à d'autres facteurs pour pouvoir soulever une masse. Par exemple, un aimant ne lui sera d'aucune utilité si ce que Cartwright souhaite dégager de son allée est une allumette. Pour que l'aimant puisse soulever une masse, il faut que cette masse soit métallique, non en bois. Il y a donc des facteurs auxiliaires nécessaires à connaître pour déterminer l'efficacité potentielle d'un aimant pour dégager un objet de l'allée. Deuxièmement, il faut déterminer quel aimant choisir, par exemple l'aimant industriel sera utile pour soulever une voiture mais ne sera d'aucune utilité pour soulever un trombone. Et finalement, il est nécessaire de comprendre comment tous ces facteurs interagissent ensemble pour produire un résultat, par exemple dégager notre allée d'une voiture. Ensuite, il est important de déterminer si l'aimant choisi sera efficace pour dégager notre allée :

« the point is whether the magnet will be effective at all in the target situation and to what extent depends on the causal structure of the situation. So the most direct way of predicting effects is to construct a causal model of the situation and estimate them. » (Cartwright, 2008, 5)

Pour Cartwright, chaque individu réalise ce processus dans la vie quotidienne. Cartwright souligne que si les individus sont capables de faire cela dans la vie de tous les jours, pourquoi ne serait-il pas possible de le faire en sciences sociales ?

2.2.2 Rendre le modèle causal explicite ou produire une histoire causale

Construire un modèle causal peut s'avérer compliqué. Cependant, lorsqu'on pense une politique on suppose, tout de même un modèle causal :

« that do not alter the fact that when we buy a policy we are betting on a causal model, willy-nilly, whether we wish to think about it or not. Generally then it is better to think about it than not, and to do so in a systematic and deliberate way. » (Cartwright, 2008, 6)

Le modèle causal doit être explicite. Et lorsqu'on l'explique, on raconte, en fait, une histoire causale. Un modèle causal n'est rien d'autre qu'une histoire causale. Pour reprendre l'exemple de Cartwright, si l'on souhaite dégager une voiture de son allée, on choisira un aimant industriel, car la voiture est un objet métallique, donc l'aimant peut la soulever, un aimant de bureau ne pourrait pas soulever une masse aussi importante, c'est pourquoi on

choisit un aimant industriel. En faisant cela, on ne fait rien d'autre que raconter une histoire permettant de choisir entre les deux types d'aimants.

Ce modèle causal offre un éclairage aux deux questions restées sans réponse, celle de la pertinence et celle de l'intégration de la preuve. Il rend compte de la pertinence de la preuve. Ainsi, dans notre exemple, l'affirmation causale « l'aimant de bureau peut soulever des importantes masses métalliques » est pertinente si l'on souhaite soulever une voiture pour dégager son allée. Les preuves sont, ici, combinées. On a séparé deux niveaux de preuves : les aimants peuvent soulever des masses métalliques (1), les aimants industriels des masses métalliques importantes et les aimants de bureau des masses métalliques plus légère (2). Il a fallu combiner ces deux preuves pour aboutir à l'idée qu'il fallait utiliser un aimant industriel pour soulever la voiture de l'allée.

2.3 Les capacités comme fondement d'un modèle causal

Dans le modèle causal envisagé par Cartwright, différentes chaînes causales sont reliées, reste à savoir comment elles sont déterminées. Ces chaînes causales renvoient à ce que Cartwright nomme la capacité causale. Certains phénomènes possèdent la capacité de causer des effets¹⁷⁸. Et c'est en mettant à jour ces capacités que l'on construit un modèle causal. Ces capacités ne sont donc pas fixes, elles peuvent se transformer, et en ce sens elles s'opposent à ce que Cartwright appelle les lois causales. Pour Cartwright, les sciences sociales - particulièrement l'économie - produisent des lois causales mais font comme si c'étaient des capacités. Cartwright l'illustre à travers l'économétrie. Afin de comprendre comment ces capacités causales se définissent et s'intègrent à un modèle causal, je développerai tout d'abord la distinction entre capacités causales et lois causales chez Cartwright (2.3.1). Ensuite, je montrerai comment, pour Cartwright, la méthodologie économique produit, sans s'en rendre compte, des capacités causales (2.3.2). Et enfin, j'explicitai comment ces capacités causales s'imbriquent au modèle causal (2.3.3).

¹⁷⁸ Cartwright s'inspire ici explicitement de la notion de tendances définies par John Stuart Mill : « *I suggest that the reader take my 'capacity' and Mill's 'tendency' to be synonymous.* » (Cartwright, 1989, 170). La notion de tendances chez Mill renvoie au fait que les phénomènes possèdent une façon d'agir d'une façon plutôt qu'une autre. De plus ces tendances évoluent et peuvent se transformer dans le temps.

2.3.1 Distinction loi causale et capacités

Les lois causales sont des affirmations causales très localisées¹⁷⁹. Les lois causales sont des « vérités locales »¹⁸⁰, en ce sens elles s'apparentent aux « vérificateurs ». Cartwright cherche, de fait, à trouver un concept de causalité plus général. Un concept qui ne dirait pas seulement C cause E au sein de la population T mais qui permettrait de dire : C cause E. Pour Cartwright, si C cause E, c'est que C porte en elle la capacité de produire E. Cartwright définit les capacités causales de la façon suivante :

« The capacity is something they can be expected to carry with them from situation to situation. So if the probability goes up in one test situation, thus witnessing to the capacity, it will do the same in all the others. Hence the universal quantifier¹⁸¹. » (Cartwright, 1989, 145)

En définitive, les lois causales sont un fait causal inféré par les probabilités et elles sont valables que dans une situation spécifique ; alors que les capacités causales infèrent une affirmation plus forte, et elles supposent que le lien causal établi peut continuer à exister dans une autre situation. Un fait causal général serait : « l'aspirine soulage les maux de têtes ». Pour Cartwright l'aspirine, par le fait d'être une aspirine, soulage des maux de têtes. Cartwright souligne tout de même le trouble que peut susciter l'expression « par le fait d'être une aspirine » :

« The troublesome phrase 'because of being an aspirins' is put there to indicate that the claim is meant to express a fact about properties and not about individuals : the property of being an aspirin carries with it the capacity to cure headaches. » (Cartwright, 1989, 141)

Malgré le trouble de cette phrase ce qui intéresse Cartwright est l'idée qu'il existe dans l'aspirine une propriété qui permet de soulager les migraines. Autrement dit, l'aspirine a la capacité causale d'apaiser les migraines. C'est pourquoi, pour Cartwright, la notion de

¹⁷⁹ « If in a given test population we see the increase in probability that we are looking for, that guarantees that Cs cause Es there in that population. But it does not tell us any more. Since it is probabilities and not mere frequencies that are involved, it is assured that the causing of Es by Cs happen regularly – but regularly in that kind of population. Who knows what happen elsewhere? » (Cartwright, 1989, 144)

¹⁸⁰ « local truths », (Cartwright, 1989, 106)

¹⁸¹ En logique, un quantificateur est un opérateur permettant de relier plusieurs variables à une quantité. Par exemple, en mathématiques le symbole $\forall$ est un opérateur, il est utilisé pour indiquer qu'un prédicat est vrai pour l'ensemble des composantes de $\forall$.

capacité causale permet un niveau de généralité beaucoup plus grand que ne le permet la notion celui de loi causale. Et c'est là que réside l'intérêt majeur de cette notion :

« My arguments for a distinction between causal laws ascriptions of capacity rest in part on claims about how general causal facts are established. Causal laws, which are relativized to certain given kinds of situation, can be inferred from the probabilities that obtain in those situations. To infer the stronger claim – what I call capacity claim – one must suppose that the causal possibilities that are established in that situation continue to obtain in various kind of situation. » (Cartwright, 1989, 147)

Il est donc nécessaire pour Cartwright de supposer l'existence de capacités causales, pour qu'un résultat obtenu dans un contexte particulier puisse être le même dans une autre situation.

2.3.2 La notion de capacité

L'objectif de Cartwright est de montrer que dans les sciences sociales – en particulier – on a recours au concept de capacité, et non à celui de lois causales lorsque l'on utilise les probabilités comme socle de référence. Cartwright le démontre en analysant une simple régression économétrique linéaire. Elle reprend l'équation de la courbe de demande (D) :

- $q = \alpha p + u$

« q » représente donc les quantités demandées, « p » les prix et « u » un choc aléatoire ; ce dernier terme fait donc passer l'équation d'une forme déterminée à une forme probabiliste en introduisant une dimension aléatoire et sur laquelle il n'y a aucune prise. Il est important de noter que cette équation traduit une relation causale ; les prix influencent les quantités qui vont être demandées. Les effets sont donc placés du côté gauche et les causes du côté droit de l'équation. Mais cette équation ne suppose pas seulement qu'il y ait une causalité, elle suppose aussi que le prix a une tendance stable, déterminée et mesurable à influencer la demande, et c'est ce que traduit le terme « α ». En étant accolé de cette façon au prix, « p », il exprime le fait qu'il existe dans « p » la possibilité « α » que « p » agisse sur les quantités demandées :

« The methods presuppose that causes have stable tendencies of fixed strengths that they carry about with them from situation to situation. What p contributes to q – its total influence, αp – depends on p alone, and is the same no matter what goes on elsewhere. This is, in effect, the consequence of the commitment to

the existence of a single linear theory that is supposed to hold across all different contexts and across all different statistical regimes. » (Cartwright, 1989, 153)

En partant de l'une des équations les plus simples et les plus connues en économie, Cartwright montre que les méthodes économétriques présupposent que les phénomènes économiques sont régis par des capacités. Elle va plus loin, en affirmant que cela est valable pour toutes les méthodes fondées sur les statistiques¹⁸², telles que la médecine, la physique, l'économie, ou encore la sociologie.

2.3.3 Modèle causal et capacités

Les capacités désignent donc ce qui est à l'œuvre au sein d'un modèle causal, même si Cartwright n'explicite pas clairement leur lien. C'est parce qu'il existe des capacités causales que l'on peut construire un modèle causal. C'est en connaissant les capacités causales de l'aimant, leurs interactions et les facteurs dont elles ont besoin pour se réaliser, qu'il est possible de construire un modèle causal et donc établir une causalité plus générale. Pour reprendre l'exemple développé ci-dessus, c'est parce que l'aimant industriel a la capacité de soulever des objets métalliques (1) et des objets métalliques lourds (2) que l'on peut prédire que l'aimant industriel permettra de dégager notre allée d'une voiture (3). C'est donc en ayant défini les capacités (1) et (2) qu'il devient possible de prédire l'effet de l'aimant dans cette situation (3). Les capacités (1) et (2) constituent en partie un modèle causal, puisqu'ajoutées l'une à l'autre elles permettent de comprendre leurs interactions et leur potentialité de succès (3). Aussi, pour être en mesure de dire que T a une capacité stable d'impliquer E plusieurs conditions doivent être réunies :

- un mode opératoire doit clairement être identifié répondant à la question de savoir comment T opère pour impliquer E. (dans l'exemple de Cartwright : comment l'aimant industriel opère pour soulever une voiture ?)
- les auxiliaires nécessaires qui permettent à T de réaliser E doivent être définis répondant à la question : qu'est ce qui doit être mis en place pour que T opère et

¹⁸² « What I claim for econometrics will be equally true in any field that uses conventional statistical methods to make inferences about causes. For it has been my thesis throughout that it takes something like the structural equations of econometrics (or path analysis) to secure a reliable connection between causes and probabilities; and it is just the feature of stability that I have pointed to in econometric structures that must be presupposed in any discipline if we are to infer from statistics we observe in one set of conditions to the effects that might be achieved by varying these conditions. » (Cartwright, 1989, 158)

implique E ? (qu'est qui doit être mis en place pour que l'aimant industriel opère et soulève une voiture ?)

- Les destructeurs doivent aussi être déterminés : qu'est ce qui pourrait détruire l'opération de T ? (qu'est ce qui pourrait détruire le fait que l'aimant industriel soulève une voiture?)
- Les autres capacités qui peuvent intervenir dans E doivent aussi être déterminées : quelles autres capacités jouent un rôle pour E ? (quelles autres capacités permettent de soulever une voiture ?)
- Il est aussi nécessaire de connaître les règles de combinaison c'est-à-dire : qu'est-ce qu'il advient quand plusieurs capacités sont à l'œuvre ? (qu'est-ce qu'il advient quand d'autres capacités que celles de l'aimant sont à l'œuvre ?)

C'est pourquoi les capacités ne peuvent avoir de sens que dans un ensemble de connaissances ; seule elles n'ont pas de sens, à part d'exister : « *So capacity claims make sense only relative to a far larger body of knowledge. That is, ultimately we need what we would call a theory.* » (Cartwright et Munro, 2010, 263) L'importance des capacités et leur rôle au sein d'un modèle causal est désormais explicite. Ce sont ces deux éléments qui définissent la théorie de la preuve que souhaite développer Cartwright.

3. La randomisation du J-PAL à l'épreuve de l'épistémologie de Cartwright

La position de Cartwright est empiriste, bien qu'elle s'éloigne sensiblement de l'empirisme de David Hume¹⁸³. Le cœur de la perspective empirique de Cartwright est la notion de capacités causales. Pour Cartwright, les capacités existent en elles-mêmes, elles ne sont pas le fruit de l'esprit humain : « *in this way capacities are much like essences.* » (Cartwright, 1989, 146) Cartwright va même plus loin en affirmant que leurs dimensions ne sont plus seulement existentielle mais réelle :

¹⁸³ L'empirisme de Hume suppose que toute connaissance est le fruit de l'expérience sensible. Pour autant, chez Hume, les liens de causalité qui unissent les différents événements observés lors d'expériences ne peuvent être le fruit d'une empiricité, ils sont le fruit des habitudes. Un individu observe plusieurs fois que A entraîne B, par habitude il anticipe donc que lorsque A se produit B sera entraîné. Cartwright résume l'empirisme de Hume autour de deux points : « *(1) for Hume, singular causal facts are true in virtue of generic causal facts (...) (2) generic causal facts are reducible to regularities.* » (Cartwright, 1989, 2) Cartwright s'oppose alors à l'empirisme de Hume sur ces deux points : « *My claims, then, are doubly anti-Humean. I take singular causes to be primary, and I endorse capacities.* » (Cartwright, 1989, 3)

« it is also an important feature of the philosophical story, for it fits the capacity into nature in a realistic way. They do indeed endure; on the other hand, their characteristics may evolve naturally through time, and they may be changed in systematic, even predictable, ways as a consequence of other factors with which they interact. All this speaks in favor of their reality. » (Cartwright, 1989, 157)

Ces capacités existent et le but de la science est de les révéler. Et c'est par la mesure qu'il sera possible de les révéler¹⁸⁴. Cela permet à Cartwright d'affirmer : *« science is measurement »*. (Cartwright, 1989, 1) Les capacités causales peuvent être mesurées alors que les lois causales peuvent être uniquement testées. L'empirisme de Cartwright est donc fondé sur la mesure. Ici, je chercherai à questionner l'empirisme de Cartwright. Je souhaite déterminer si la perspective empiriste – et la notion de capacité qui lui donne sens – permet d'envisager la validité externe de la randomisation menée par Duflo et Banerjee. Je développerai, tout d'abord, les « ingrédients » nécessaires à la validité externe de la randomisation pour Cartwright (3.1). Puis, à l'aide du travail de Vincent Guillin, je regarderai si les « ingrédients » nécessaires définis par Cartwright sont présents dans les expériences de Duflo et Banerjee. Cela me permettra de déterminer si les résultats du J-PAL peuvent passer de l'efficacité réelle à l'efficacité potentielle (3.2).

3.1 Le passage de l'efficacité réelle à l'efficacité potentielle de la randomisation : les trois ingrédients essentiels

La perspective empiriste de Cartwright se veut double : à la fois ontologique et pragmatique. Ces deux dimensions se retrouvent au sein des capacités. Ces dernières sont ontologiques puisqu'elles existent, qu'elles sont réelles et que l'objectif de la science est de les définir. Les capacités, en ayant un caractère stable et réel, permettent de transposer des résultats d'une situation à une autre, autrement dit de passer de l'efficacité réelle à l'efficacité potentielle. Elles permettent de prédire et en définitive de proposer des recommandations politiques claires. C'est en ce sens qu'elles ont une dimension pragmatique, car elles offrent la possibilité d'agir sur le monde. C'est cette dimension que je questionnerai ici. Pour cela, je définirai, précisément, la distinction entre efficacité réelle et efficacité potentielle (3.1.1), puis je développerai les différents « ingrédients » définis par Cartwright permettant de passer de l'efficacité réelle à l'efficacité potentielle (3.1.2).

¹⁸⁴ *« Measuring instruments have this kind of ability to read nature. »* (Cartwright, 1989, 6)

3.1.1 De la confusion entre l'efficacité réelle et l'efficacité potentielle dans la randomisation

La façon standard de comprendre l'efficacité réelle de la randomisation est de se référer au modèle de Rubin – que j'ai développé dans la partie précédente de ce travail (ie. Partie 1 – chapitre 2). L'efficacité réelle, dans ce cadre, renvoie à la différence des moyennes entre deux groupes – l'un de traitement et l'autre de contrôle. Cependant, pour Cartwright, ces moyennes sont conditionnées par plusieurs facteurs : « *observed values of treatment outcomes, of covariates, a missing-indicator variable and, (...) a 'treatment assignment variable' W.* » (Cartwright, 2009c, 5) Cette dernière dimension tient compte de deux types d'informations : la façon dont l'assignation aléatoire a été introduite et la population sélectionnée. Il apparaît de ce fait impossible, pour Cartwright, de parler de la comparaison des moyennes des deux groupes sans spécifier qu'il s'agit d'une population spécifique. C'est pourquoi, selon elle, l'effet causal est, ici, toujours défini de façon relative, dépendant d'une structure expérimentale particulière. Et dans ce cas :

« *It does not make sense to talk about the cause being efficacious in another setting. The concept only applies to those situations in which the experimental design obtains, and in no others.* » (Cartwright, 2009c, 6)

La randomisation ne produit alors que des preuves d'efficacité réelle mais pensent produire des preuves d'efficacité potentielle. Des preuves d'efficacité potentielle ne dépendraient pas d'une population spécifique, puisqu'il serait possible de les penser d'un contexte à un autre. Cette possibilité renvoie exactement à ce que Cartwright entend par capacités. Pour passer de l'efficacité réelle à l'efficacité potentielle, il faut donc mettre à jour, mesurer des capacités causales.

3.1.2 Les trois « ingrédients » nécessaires à l'efficacité potentielle

Afin de mettre à jour ces capacités causales et donc de penser l'efficacité potentielle de la randomisation, il est nécessaires de répondre à trois demandes :

1. On a besoin d'une bonne raison de penser que l'effet produit dans l'expérience est un effet durable.
2. On doit avoir de bonnes raisons de penser que l'effet propre a été identifié.

3. On a besoin de savoir comment est composé l'élément que l'on a identifié et de comprendre ce à quoi il contribue.

Vincent Guillin (2013) nomme ces trois dimensions : robustesse (1), irréductibilité (2) et compositionnalité (3). En reprenant, les trois dimensions de Cartwright avec la terminologie de Guillin la première demande renvoie à la robustesse du résultat et cette robustesse dépend de l'identification d'une capacité. L'irréductibilité renvoie au fait qu'on a bien identifié la capacité et qu'elle explique les phénomènes sans être expliquée par eux. La compositionnalité est le fait de connaître la composition de cette capacité. Cela renseigne sur la façon dont les différentes forces interagissent entre elles, et ou avec d'autres facteurs. Pour Cartwright, lorsque ces trois demandes sont réalisées, il est possible de penser l'efficacité potentielle de la randomisation :

« Meeting these three demands provides the three ingredients necessary to turn efficacy measured in an RCT into evidence for effectiveness in a new setting. It's like making pancakes. RCTs are baking powder. But the baking powder is useless without the flour, milk and eggs. » (Cartwright, 2009c, 17)

Afin de penser la validité externe de la randomisation, il est nécessaire que cette dernière ajoute à sa méthodologie les trois ingrédients nécessaires. Cartwright, dans chacun de ses travaux, parle aussi bien de la randomisation en médecine, qu'en sciences sociales, elle ne s'intéresse pas spécifiquement à celle développée par Duflo et Banerjee, alors que la randomisation en médecine et en économie du développement se retrouve parfois mais s'éloigne souvent.

3.2 De la « randomisation idéale » de Cartwright à la randomisation de Duflo et de Banerjee

Vincent Guillin souligne, d'ailleurs, la spécificité des expériences randomisées de Duflo et de Banerjee en définissant leur approche comme des « expérimentations créatives »¹⁸⁵ (Guillin, 2013, 86). Il définit alors ces expérimentations de la façon suivante : « L'« économie expérimentale créative » propose de se rapprocher du terrain pour y tester directement, en s'appuyant sur la méthodologie des essais contrôlés randomisés, des interventions ou des programmes concrets et ciblés visant à réduire sélectivement certains des

¹⁸⁵ Ce que faisait déjà Duflo dans sa leçon inaugurale au Collège de France, Voir (Duflo, 2009, 54).

« pièges à pauvreté » dans lesquels sont prises certaines populations. » (*Ibid*) De cette définition ressort deux principales caractéristiques d'une telle méthodologie : le fait de se rapprocher du terrain (1) et la volonté de réduire les « trappes à pauvreté » (2). Guillin s'attache à déterminer si les expériences du J-PAL permettent de mettre en évidence des capacités causales. Selon lui, il existe un raisonnement théorique patent au sein de ces expériences¹⁸⁶. Je développerai, tout d'abord, ce point (3.2.1), puis je montrerai comment Guillin questionne l'existence des trois ingrédients définis par Cartwright au sein des expériences du J-PAL (3.2.2).

3.2.1 Un raisonnement théorique patent dans les expériences du J-PAL

C'est autour d'une expérience faite sur l'utilisation des engrais en économie du développement que Guillin va tenter de mettre en évidence l'aspect patent de la dimension théorique de ces expérimentations créatives. L'expérience évoquée par Guillin est conduite au Kenya par Duflo, Kremer et Robinson¹⁸⁷, et montre que l'engrais chimique permet une augmentation importante de la production. Le gouvernement proposait de donner des engrais aux différents agriculteurs. Dans un premier temps, l'expérience évalue l'efficacité de ces engrais, et conclut positivement. Duflo Kremer et Robinson évaluent, dans un second temps, l'utilisation de ces engrais par les fermiers. Le programme offrait les engrais pendant la première saison mais les fermiers devaient ensuite les acheter à un moindre coût (pendant la deuxième saison) : ils devaient donc disposer d'une petite épargne pour pouvoir se les procurer. L'expérience étudie six saisons. Au début, 60% des fermiers utilisent ces engrais, mais après trois saisons seuls 29% des fermiers les utilisent. Le résultat n'est donc pas concluant, et l'action des fermiers, en n'utilisant pas ces engrais, est inefficace. Une des principales raisons invoquée est qu'il était ressenti comme difficile pour les fermiers de constituer cette petite épargne, même si elle leur permettait d'espérer de meilleurs rendements. Ces individus seraient alors incohérents temporellement. Guillin montre que pour exprimer ce résultat paradoxal les chercheurs du J-Pal font appel à des heuristiques comportementales ou ce qu'il nomme des « dispositions à agir ».

¹⁸⁶ Lors du chapitre 3 de cette partie 2, je m'attacherai longuement à définir le statut de la théorie au sein des expériences du J-PAL.

¹⁸⁷ Voir (Duflo, Kremer et Robinson, 2008). Cette expérience a déjà été évoquée lors de la partie 1 de cette thèse.

Pour Guillin, une fois que les mécanismes à l'œuvre dans ce genre de comportement auront été identifiés, il sera alors possible de déterminer sur quel facteur jouer afin d'aboutir à un comportement qui mène à des résultats efficaces. Cette façon de procéder est pour : « indissolublement théorique (pour avoir une explication causale complète du phénomène de la sous-utilisation des engrais, il faut identifier les circonstances précises de déclenchement des heuristiques psychologiques associés à la procrastination et au biais de projection) et pratique (parce que ces mécanismes sont robustes, c'est sur les facteurs déclencheurs qu'il va falloir intervenir pour les neutraliser) ». (Guillin, 2013, 93) Pour Guillin, l'expérimentation créative dispose alors d'« un appareil conceptuel » (Guillin, 2013, 94) afin d'être à même d'identifier des liens causaux se caractérisant par « l'invocation de disposition à agir ou de procédures de prises de décisions » (*Ibid*). Même si Guillin souligne que cet « appareil conceptuel » n'est pas toujours mis en avant, il est pour lui sous-jacent à l'analyse. Mais des questions demeurent : les mécanismes sous-jacents sont-ils explicites ? Les « expérimentations créatives » de Duflo et Banerjee permettent-elles de mesurer des capacités causales ?

3.2.2 La randomisation du J-PAL : des capacités à l'œuvre ?

Guillin analyse, ensuite, si les trois « ingrédients » définis par Cartwright sont présents dans la randomisation de Duflo et Banerjee, à savoir : la robustesse (1), l'irréductibilité (2) et la compositionnalité (3). En ce qui concerne la robustesse (1), pour Guillin l'approche de Duflo permet d'identifier des dispositions, mais ces dernières semblent être le fruit de la structure expérimentale, un artefact en somme, et donc ne pas renvoyer pleinement à des capacités causales. Bon nombre d'effets disparaissent lorsque l'expérience a lieu dans un autre cadre et on retombe, sans que Guillin l'évoque, sur les problèmes de validité interne que sont les effets John Henry et Hawthorne¹⁸⁸. Il apparaît aussi difficile de déterminer si l'effet observé est irréductible (2) : « dans la mesure où ce champ de recherches est encore majoritairement dans une phase d'identification ou d'inventaire de ces dispositions ». (Guillin, 2013, 100) La compositionnalité (3) est elle aussi délicate, ce qui suit logiquement le problème précédent, sans que soient identifiées clairement des dispositions.

¹⁸⁸ L'effet John Henry et Hawthorne renvoie au fait que les individus peuvent changer leurs comportements, sachant qu'ils font partie d'une expérience, j'ai précédemment développé (partie 1 – chapitre 2) ces effets.

A l'instar de Cartwright, Guillin souligne que sans une « théorie unifiée » établissant une hiérarchie de ces dispositions et de ces mécanismes, il ne sera pas possible de répondre aux trois critères énoncés par Cartwright. Guillin souligne que Cartwright est pessimiste quant à la possibilité que cela soit possible en sciences sociales :

« Cartwright se montre tout à fait pessimiste sur ce point : « dans bien des cas, nos difficultés ne sont pas dues au fait que nous n'avons pas encore trouvé la théorie [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 ». » (Guillin, 2013, 100)

Guillin est moins pessimiste que Cartwright. Il préfère laisser du temps à ce programme de recherche pour que les chercheurs du J-PAL « s'intéressent aux ressources cognitives et émotionnelles qui sont mobilisées par l'action humaine. » (*Ibid*), et puissent donc développer une théorie unifiée. D'un côté Cartwright est pessimiste et de l'autre Guillin préfère laisser du temps au programme de recherche du J-PAL en espérant que ce dernier progresse et apprenne des comportements individuels en vue de guider la décision politique. Faut-il être pessimiste ou optimiste ? Les résultats des RCTs du J-PAL seront-ils transposables dans l'action politique ? C'est par la transposition des résultats des expériences randomisées dans le discours politique que Cartwright pense leur validité externe, ou ce qu'elle préfère appeler le passage de l'efficacité réelle à l'efficacité potentielle. Néanmoins, la validité externe des expériences de Banerjee et de Duflo relève de deux niveaux : la transposition des résultats de leurs expériences à d'autres situations mais aussi la généralisation de leurs résultats en vue de proposer des politiques globales de lutte contre la pauvreté. L'analyse de Cartwright ne permet pas de penser ce second niveau.

4. Conclusion

A l'aide de la théorie des capacités de Cartwright, j'ai pu montrer que la randomisation du J-PAL permettait d'offrir des affirmations causales, mais que la portée de ces affirmations était limitée, se cantonnant au cadre de l'expérience. Les affirmations causales offertes par la randomisation sont, dans les termes de Cartwright, des « vérificateurs » dont la causalité est fondée sur une théorie probabiliste. En ce sens elles n'offrent que des preuves d'efficacité réelle. Pour que ces preuves puissent devenir des preuves d'efficacité potentielle, il faudrait

que la méthode du J-PAL soit tournée vers la recherche de capacités causales permettant de définir un modèle causal. Guillin questionne, lui, l'existence des capacités causales au sein de la randomisation du J-PAL. Il montre que ce programme de recherche est sans doute trop récent pour rendre compte des capacités causales à l'œuvre. Pour l'instant, les « expérimentations créatives » ne permettent pas de produire des résultats utilisables dans d'autres situations. Cela tend à fragiliser très fortement la validité externe de telles expériences. En se centrant sur le caractère utilisable des résultats produits par la randomisation, Cartwright et Guillin laissent de côté un des niveaux de la validité externe de telles expériences : penser les politiques à une échelle élargie. C'est donc le premier niveau de la validité externe qui est ici fragilisé.

Chapitre 2 : Les « randomistas » ou le difficile arbitrage entre la validité interne et la validité externe de la randomisation du J-PAL

0. Introduction

« As for me, I prefer to carry parameters on my journey so I know where I am and where I am going, not travel stoned on the latest euphoria drug. This is the story of Tantalus, grasping for knowledge that remains always beyond reach. (...) If we stopped grasping for the apple of the Asymptotia, we could discover that our pool of Tantalus is full of small but enjoyable insights and wisdom. Can we economists agree that it is extremely hard work to squeeze truths from our data sets and what we genuinely understand will remain uncomfortably limited? We need words in our methodology vocabulary to express the limits. » (Leamer, 2010, 32)

Les chercheurs du J-PAL, cela a été vu, se sont faits les principaux promoteurs de la randomisation en économie du développement, cherchant à l'imposer face aux autres méthodes permettant d'évaluer des programmes de développement. Angus Deaton (2009) traduit ce mouvement par ce qu'il appelle les randomistas. Les principaux randomistas sont, bien sûr, Banerjee et Duflo. Leur principale caractéristique est de considérer que seule la randomisation peut produire des résultats fiables ; et en ce sens, la randomisation devient la seule méthodologie à utiliser, si l'on veut lutter efficacement contre la pauvreté. Transformant alors l'agenda de recherche en économie du développement, seules les questions pouvant être traitées à travers une expérience randomisée sont à mettre en avant tandis que les autres sont reléguées au second plan, voire abandonnées. Par exemple, les « grandes questions économiques » ne peuvent plus être adressées puisque la randomisation ne peut pas y répondre. Il est impossible d'évaluer l'impact d'une institution, par exemple. Ce qui explique, sans doute très grandement, le fait que Banerjee et Duflo refusent toutes ces grandes questions. Et qu'à l'inverse, ils proposent une approche fondée sur des changements à la marge en espérant que ces derniers aient *in fine* un impact global. De ce fait, tout ce qui n'est pas *randomisable* ne rentre pas en compte dans l'agenda de recherche. Ce que souligne très fortement Martin Ravallion :

« From the point of view of development policy-making, the main problem in the randomistas agenda is that they have put their preferred method ahead of the questions that emerge from our knowledge gaps. Indeed, in some respects (such as sectoral allocation research) the randomistas' success may have made things worse. The risk is that we end up with lots of social experiments that provide evidence on just one or two parameters for a rather narrow set of assigned interventions and settings. The knowledge gaps persist and even widen. » (Ravallion, 2009b, 3)

Le point de départ de l'importante vague de critiques adressées à la randomisation ces trois dernières années est, sans aucun doute, la dimension exclusive et dominante adoptée par les chercheurs du J-PAL, faisant de la randomisation le seul outil efficace pour lutter contre la pauvreté. Transformant alors l'agenda de recherche et balayant, de ce fait, cinquante ans de recherches dans le domaine. Ce qui a donné lieu aux questions suivantes : la randomisation est-elle réellement la meilleure méthode pour lutter contre la pauvreté ? S'auto-suffit-elle ? Nombreux économistes du développement ont posé ces questions et tenté d'y répondre¹⁸⁹, tout comme des économètres¹⁹⁰, ou encore des expérimentalistes¹⁹¹. Ce qui m'intéresse, tout particulièrement ici, c'est la critique adressée par les économistes du développement car elle provient du même champ disciplinaire que celui du J-PAL. Ces critiques permettent de questionner précisément l'apport de la randomisation en économie du développement. Pour autant, ce premier volet de critique ne peut être pleinement compris sans être lié aux deux autres vagues de critiques qui lui sont attachées. L'effort que j'essaierai de fournir ici est de produire une unité au sein de ces différentes critiques. Pour ce faire mon analyse se fonde, principalement, sur la critique adressée par Angus Deaton. D'une part car c'est, d'une certaine façon, la plus complète ; et d'autre part, car c'est celle qui a connu l'écho le plus important au sein des randomistas et du champ. Cependant, cette critique reprend de nombreux points déjà développés par James Heckman presque deux décennies auparavant. En ce sens, elles ne peuvent être pleinement séparées, c'est pourquoi je tenterai de les relier. De plus, ces différentes critiques, pensées ensemble, me permettent de traduire le second niveau de la validité externe de la randomisation du J-PAL - généraliser un programme à une échelle plus large-, ce que l'analyse de Cartwright ne m'avait pas permis de faire.

En cherchant à relier ces différentes critiques, il m'est apparu que la question de la distribution des effets de traitement - rendue invisible par la randomisation - était la clef de voûte de toutes ces critiques. C'est donc cette question qui sera le point de départ de mon analyse. La première partie de ce chapitre s'attachera à définir ce point de départ. C'est à partir de la question de la distribution des effets de traitement que la forte validité interne de la randomisation va se retourner et se transformer en une faible validité externe. Je considère

¹⁸⁹ Les principaux économistes à avoir critiqués la randomisation sont : Angus Deaton (2009 ; 2010a ; 2010b), Martin Ravallion (2009a et 2009b), Michael Carter et Christopher Barrett (2010 et 2012), Dani Rodrik (2007), et Daron Acemoglu (2010).

¹⁹⁰ Les économètres qui se sont questionné sur la randomisation sont principalement Leamer (2010) et James Heckman (1991, 2010), Heckman et Smith (1995), Heckman, Clement et Smith (1997).

¹⁹¹ La randomisation est critiquée partiellement par certains expérimentalistes, seul Harrison (2010) produit une réelle critique de la randomisation en économie du développement.

cette difficulté comme un premier symptôme qui laisse, ensuite, place à un problème plus important : une hétérogénéité indifférenciée. La deuxième partie de ce chapitre définira cette hétérogénéité indifférenciée. C'est à travers cette dernière que je questionnerai explicitement, lors de la troisième partie de ce chapitre, le second niveau de la validité externe de la randomisation du J-PAL. Enfin, dans une quatrième partie je montrerai qu'il est impossible pour la randomisation du J-PAL de viser une certaine validité externe sans perdre en validité interne.

1. L'absence de la distribution des effets de traitement : un premier symptôme de la faible validité externe de la randomisation du J-PAL

La randomisation permet de comparer les effets d'un programme entre le groupe de traitement et le groupe de contrôle ; cependant, les résultats qui sont comparés sont des moyennes¹⁹². C'est à dire que l'information disponible, à partir de la randomisation, est uniquement la moyenne des effets du programme dans le groupe traité et celle du groupe non traité. Ces moyennes ne permettent pas de rendre compte de la distribution des effets du programme pour les individus constituant les deux groupes. Pour autant, les effets d'un programme peuvent être sensiblement différents d'un individu à un autre ; pour des raisons physiologiques, environnementales, institutionnelles, etc. La randomisation, à elle seule, ne rend pas visible l'hétérogénéité potentielle des effets de traitement. L'ensemble des critiques qui se sont attachées à la randomisation soulève ce problème¹⁹³, comme Deaton, par exemple :

« A fact that is often quoted by critics of RCTs, but is often ignored by practitioners, at least in economics: RCTs are informative about the mean of the treatment effects, $Y_i1 - Y_i0$, but do not identify other features of the distribution. For example, the median of the difference is not the difference in medians, so an RCT is not, by itself, informative about the median treatment effect, something that could be of as much interest to policy makers as the mean treatment effect. It might also be useful to know the fraction of the population for which the treatment effect is positive, which once again is not identified from a trial. » (Deaton, 2009, 27)

En ne s'intéressant pas à la distribution des effets du traitement au sein des groupes, la randomisation tend alors à considérer un individu moyen. Cela n'est pas sans conséquences,

¹⁹² Dans la première partie de ce travail, j'ai, à partir du modèle de Rubin (1974), explicité la comparaison de ces deux moyennes et j'ai développé comment elles étaient obtenues. J'ai aussi montré en quoi la comparaison de ces moyennes permettait à Rubin de penser un contrefactuel à travers la construction artificielle de *potential outcomes*, qui est au cœur de la randomisation utilisée en économie du développement.

¹⁹³ Voir : Angus Deaton (2009 ; 2010a ; 2010b), Martin Ravallion (2009a et 2009b), Michael Carter et Christopher Barrett (2010 et 2012), Dani Rodrik (2007), et Heckam (1997, 2010).

lorsque l'on souhaite lutter contre la pauvreté - où la distribution des effets de traitement peut être d'un intérêt sensible. Ici, je chercherai à montrer l'importance de l'hétérogénéité des effets de traitement en matière de lutte contre la pauvreté. L'objectif est donc d'explicitier le point de départ de mon analyse. Pour cela, je poserai tout d'abord le problème à travers la médecine, car c'est dans ce domaine, qu'il a d'abord été identifié et défini (1.1). Puis, je relierai ce problème à l'économie (1.2). Et enfin, je traiterai de son importance (1.3). Cette partie se centre majoritairement sur les travaux de James Heckman, car c'est celui qui en a traité de la façon la plus systématique, et qui a tenté d'en révéler les hypothèses sous-jacentes ainsi que leurs implications.

1.1 L'hétérogénéité des effets de traitement : l'illustration de la médecine

Comme cela a été montré dans la première partie de ce travail, la randomisation a connu son essor le plus frappant durant la seconde moitié du 20^{ème} siècle à travers les essais cliniques médicaux. Sans grande surprise, le problème de l'hétérogénéité des effets de traitement a d'abord été soulevé au sein de cette discipline. C'est à travers l'explicitation de ce problème en médecine que je poserai les enjeux de cette première limite adressée à la randomisation. En médecine, le problème de l'hétérogénéité des effets de traitement soulève l'idée que deux individus peuvent réagir de façon différente à un traitement identique. Kravitz, Duan et Braslow (2004) donnent quatre raisons à la différence des effets de traitements. La première est le fait que le patient peut connaître des risques sans être traité : « *All else being equal, the absolute benefits of a treatment increase along with the increasing baseline risk (i.e., heightened susceptibility or worse prognosis).* » (Baslow, Duan et Kravitz, 2004, 669) La deuxième explication est la réceptivité, tous les êtres humains ne sont pas réceptifs de la même façon au traitement, « *for example, 7 percent of Caucasians cannot metabolize codeine into morphine and so do not get any pain relief from codeine.* » (Ibid.) À l'inverse certaines personnes répondent trop fortement au traitement, ce qui est une troisième justification à la différence des effets d'un traitement (la susceptibilité). Enfin, les patients ont des préférences dans la façon d'être soigné, c'est la dimension la plus arbitraire pouvant expliquer des effets de traitement non identiques au sein d'un groupe. Il est donc possible qu'en moyenne le traitement soit jugé efficace, alors que plusieurs individus au sein du groupe de traitement ne réagissent pas positivement à ce dernier. L'hétérogénéité des effets de traitement est alors définie de la façon suivante en médecine :

« Heterogeneity of treatment effects is the magnitude of the variation of individual treatment effects across population. In statistical terms, HTE [heterogeneity of treatment effects] is equivalent of the interaction between treatment effect and individual patient effect. » (Braslow, Duan et Kravitz, 2004, 664)

C'est donc l'interférence entre les effets du traitement et les effets produits sur le patient qui définit l'hétérogénéité des effets du traitement. Considérer l'homogénéité de ces effets, c'est donc ne plus s'intéresser aux effets produits sur un patient. Et en ce sens, c'est ne plus considérer un patient avec des caractéristiques propres ; c'est considérer un individu moyen, un individu extrait d'une moyenne. Cet individu est donc un artefact, le fruit des résultats exprimés par la randomisation. Il est, de plus, important de noter que les essais randomisés sont caractérisés par de faibles critères d'inclusion. En d'autres termes, les individus choisis pour participer aux expériences ne subissent pas de très grandes sélections. Cela tend à augmenter, potentiellement, l'hétérogénéité des effets de traitement, et rend donc la prise en compte d'un individu moyen encore plus discutable.

1.2 L'hétérogénéité des effets de traitement en économie

L'hétérogénéité des effets de traitement n'est pas présente uniquement en médecine. Cette hétérogénéité peut, d'ailleurs, être encore plus importante dans les sciences sociales où les individus sont difficilement détachables de leurs environnements, de leurs histoires personnelles, et des institutions auxquelles ils font face. De plus, dans le domaine sociétal les individus ont tendance à interagir entre eux, compliquant un peu plus l'interprétation des effets de traitements. C'est pourquoi, pour Barrett et Carter (2010), le problème de l'hétérogénéité des effets de traitement est sans doute encore plus criant dans les sciences sociales qu'en médecine :

« RCTs¹⁹⁴ are invaluable tools for biophysical scientist, where the mechanisms involved are more mechanical than is the case in behavioral and social sciences, and where virtually all conditions can be controlled in the research design. Human agency complicates matters enormously (...). As a result, impacts and behaviors elicited experimentally are commonly endogenous to environmental and structural conditions that vary in unknown ways within a necessarily highly-stylized experimental design. » (Barrett et Carter, 2010, 524)

¹⁹⁴ RCT renvoie à *randomized controlled trials*, autrement dit à un essai contrôlé randomisé ; qui tient donc de synonyme pour randomisation.

Avec l'hétérogénéité des effets de traitement on retombe dans la problématique, en partie relevée par Rubin, du contrefactuel. Celui que crée la randomisation est artificiel. Il est impossible d'observer simultanément un même individu qui bénéficierait du programme et qui n'en bénéficierait pas. En ce sens, il est impossible de déterminer l'impact du programme pour chaque personne faisant partie de l'évaluation. Cependant, se cantonner à la moyenne et non à la distribution peut cacher un nombre d'éléments importants. La moyenne observée peut se traduire par des distributions totalement différentes, comme le souligne Harrison (2011) en montrant une même moyenne dont deux distributions opposées sont possibles :

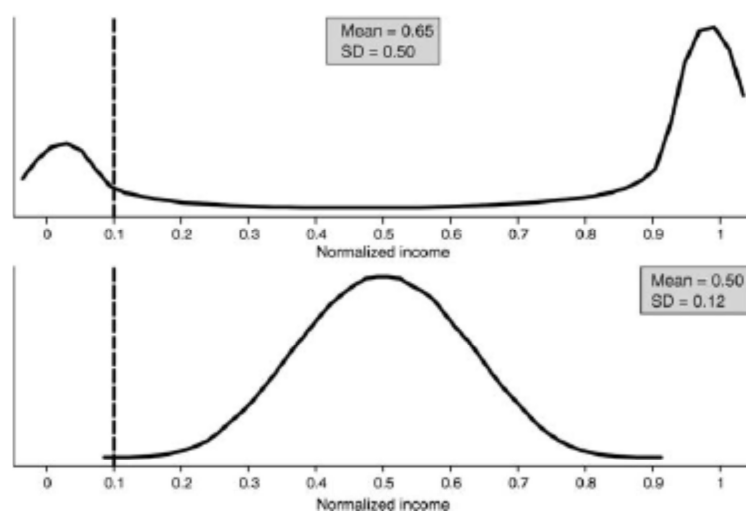


Figure 1: Why Average Effects Are Not Everything.

Source : (Harrison, 2011, 637)

De plus, Heckman souligne qu'il n'est absolument pas rare que les effets du programme varient d'un individu à un autre, et que cela constitue d'ailleurs un phénomène empirique important. Afin de rendre compte de l'hétérogénéité des effets de traitements, Heckman analyse les effets d'un programme qu'il a beaucoup discuté dans d'autres de ces travaux : le *Job Training Partnership Act Title II-A*¹⁹⁵. Il montre qu'à partir des moyennes des effets de traitement deux histoires sont possibles :

¹⁹⁵ Le *Job Training Partnership Act Title II-A* est une évaluation demandée par le gouvernement américain en 1986. Ce programme est un sous-programme d'un programme plus large le *Job Training Partnership Act* mis en place en 1982 sous le gouvernement de Reagan. Le *Job Training Partnership Act Title II-A* visait à fournir à des adultes et à des jeunes adultes souhaitant entrer sur le marché du travail, mais rencontrant d'importantes barrières (car peu qualifiés) une aide financière et une formation afin de déterminer si cela avait un effet positif sur l'embauche. L'évaluation débute en 1986 et se traduit par une expérience randomisée. Heckman, dans sa critique de la randomisation, prend cette expérience comme un emblème.

« The first story is that the JTPA programme benefits many people by facilitating their employment but also harms many people who would have worked if they had not participated in the program. The second story is that the program benefits and harms few people. » (Heckman, Smith et Clements, 1997, 504)

Ne pas s'intéresser à la distribution des effets du traitement laisse alors l'interprétation des résultats suspendus à deux histoires possibles. Ces deux histoires sont sensiblement différentes et ont donc des implications distinctes. Dans l'une, le programme est efficace pour bon nombre des participants, mais lèse aussi de nombreux participants. Dans l'autre histoire, le programme est efficace pour une petite partie des participants et nuit à une petite partie des participants. La même moyenne peut traduire ces deux effets distincts, alors que les implications politiques qui sous-tendent ces deux distributions doivent être différentes.

1.3 De l'importance de la distribution des effets de traitement en matière de lutte contre la pauvreté

La différence des effets de traitement a une importance toute particulière en matière de pauvreté, puisque la distribution de ces effets génère des implications politiques différentes. Lorsque l'on cherche à évaluer l'impact d'un programme de développement, l'unique moyenne de ces effets laisse de côté des informations importantes, particulièrement pour les décideurs politiques. De plus, nombreux sont les programmes de développement qui visent une population spécifique, comme les femmes, les enfants, ou le décile le plus pauvre. Connaître les effets d'un programme sur ces différentes populations peut être d'une grande importance en matière de décisions politiques. Ne s'intéresser qu'à la moyenne des effets de traitement, en matière de lutte contre la pauvreté, repose alors, selon Heckman, sur deux hypothèses principales :

« (a) that increases in total output increase welfare; and

(b) that undesirable distributional aspects of programmes are either unimportant or are offset by transfers governed by a social welfare function, or its counterpart for families or groups. » (Heckman, Smith et Clements, 1997, 488)

Pour Heckman, lorsque l'on se centre sur la moyenne des effets de traitement, cela suppose qu'« un dollar est un dollar » peu importe qui le reçoit. Or, en matière de lutte contre la pauvreté, déterminer « qui reçoit le dollar » semble crucial. Se contenter de la moyenne des effets du traitement, c'est supposer que la moyenne est représentative. C'est donc aussi

supposer que l'impact est le même pour tous les individus, ou encore que les effets d'un programme sont transférables d'un individu à un autre. Cependant, Heckman explique que les effets de certains programmes ne sont pas transférables. Par exemple, les effets positifs de la vaccination sur un individu ne peuvent pas être transférés à un autre individu ; alors qu'un dollar peut être transféré d'un individu à un autre. Le problème est que tous les programmes ne sont pas exprimables sous forme monétaire - ce qui permettrait de penser cette transférabilité. L'absence de visibilité de cette distribution limite énormément les résultats de la randomisation.

2. D'une hétérogénéité des effets de traitement à une hétérogénéité indifférenciée : la question du pourquoi un programme fonctionne

Selon Deaton, le problème de l'hétérogénéité des effets de traitement n'est pas un problème technique que les économètres doivent tenter de résoudre¹⁹⁶ « *but [it] is a reflection of the fact that we have not started on our proper business, which is trying to understand what is going on.* » (Deaton, 2009, 11) Pour Deaton, il est important d'opérer un changement de perspective : ne plus se demander uniquement « si un programme marche » mais se demander « pourquoi un programme marche ». En pointant du doigt que les randomistas ne se demandent pas pourquoi les programmes fonctionnent, Deaton reprend une critique déjà soulevée par (Heckman, 1991), (Heckman et Smith, 1995), et (Pawson et Tiley, 1997) ; puis largement reprise, plus récemment, en économie du développement par (Ravallion, 2009a et 2009b), (Carter et Barrett, 2010), et (Rodrik, 2007). Sans aucun questionnement sur le « pourquoi un programme fonctionne » les résultats produits par la randomisation ne sont qu'une « *black-box test of « what works »* ». (Deaton, 2010a, 451) On est face à une hétérogénéité indifférenciée qui ne dépend plus uniquement de la distribution des effets du traitement, mais qui dépend du fait qu'il devient difficile de rendre compte des différents effets du programme car on souffre d'un manque d'explications. On passe alors d'une hétérogénéité d'effets de traitement à une hétérogénéité indifférenciée. L'absence de réflexion sur le « pourquoi un programme fonctionne » place les randomistas dans une position difficile où les résultats ne sont pas expliqués mais sont souvent suspendus à des suppositions fragiles, comme le souligne Deaton :

¹⁹⁶ En cherchant, par exemple, des méthodes économétriques qui permettraient de rendre compte de la distribution des effets du traitement. Voir, par exemple : (Crump, Hotz, Imbens et Mitnik, 2008).

« For example, when two RCTs in two cities in India find that children's scores improved less in Mumbai than in Vadodara, the authors state « this is likely related to the fact that over 80% of the children in Mumbai had already mastered the basic language skills the program was covering », Duflo, Glennerster, and Kremer (2008). It is not clear how *“likely”* is established here, and there is certainly no evidence that conforms to the “gold standard” that is seen as one of the central justifications for the RCTs. » (Deaton, 2009, 30, souligné par moi)

Afin de traduire explicitement à quoi renvoie cette hétérogénéité indifférenciée, je vais reprendre l'expérience que Deaton évoque¹⁹⁷. Cela me permettra ; d'une part, de rentrer dans la boîte noire, et d'autre part d'étayer la critique de Deaton. Pour cela, je présenterai l'expérience (2.1). Puis, je la questionnerai afin de déterminer précisément à quelles questions cette expérience permet de répondre (2.2).

2.1 De Bombay à Vadodara : rentrer dans la boîte noire

La citation de Deaton est tirée du chapitre sur la randomisation, écrit par Duflo, Glennerster et Kremer¹⁹⁸, dans le manuel de référence en économie du développement¹⁹⁹. Ce chapitre vise à offrir une « boîte à outils²⁰⁰ » afin d'utiliser la randomisation. Les auteurs y discutent le résultat de certaines expériences. Celle qu'évoque la citation de Deaton est une étude randomisée menée par Abhijit Banerjee, Esther Duflo, Shawn Cole et Leigh Linden (2007), en Inde, de 2001 à 2004, avec l'ONG Pratham²⁰¹. L'ONG Pratham propose deux programmes cherchant à améliorer l'apprentissage des enfants à l'école dans les économies en développement. Ces deux programmes cherchent à répondre à une tension importante entre ; d'une part, le souhait des Nations Unies de rendre l'école primaire universelle²⁰², et d'autre part le constat que les enfants n'apprennent pas à l'école. Ces programmes avaient donc pour

¹⁹⁷ Deaton évoque aussi une étude de Duflo et Pande (2007), qui n'est pas une expérience randomisée, cherchant à évaluer l'impact de plans d'irrigation sur la production agricole en Inde. De façon similaire à l'expérience de Banerjee, Cole, Duflo et Linden (2007), les résultats de l'étude de Duflo et Pande (2007) peuvent difficilement être généralisés : « *Similarly, if we find that dams in India do not reduce poverty on average, as in Duflo and Pande's (non experimental) 2007 study, there is no implication about any specific dam, even one of the dams included in the study, yet it is always a specific dam that a policy maker has to decide about.* » (Deaton, 2009, 21)

¹⁹⁸ Voir (Duflo, Glennerster et Kremer, 2007)

¹⁹⁹ Voir Schultz et Strauss (2007).

²⁰⁰ C'est d'ailleurs le titre du chapitre : “Using Randomization in Development Economics Research: A Toolkit”

²⁰¹ J'ai déjà développé une partie de cette expérience dans la première partie de ce travail (i.e. Partie 1 – chapitre 2), lorsque je me suis intéressée au lien entre les essais cliniques médicaux et la randomisation du J-PAL.

²⁰² Améliorer l'éducation est l'un des huit objectifs du millénaire définis par les Nations Unies en 2005. Les sept autres objectifs du millénaire sont : réduire l'extrême pauvreté et la faim (1), promouvoir l'égalité des genres et l'autonomisation des femmes (2), réduire la mortalité infantile (3), améliorer la santé maternelle (4), combattre le sida, le paludisme et les autres maladies (5), assurer un environnement humain durable (6) et construire un partenariat mondial pour le développement (7). Je reviendrai sur ces différents objectifs dans la suite de la thèse (partie 3 – chapitre 1 et chapitre 2).

objectif d'améliorer la qualité de l'éducation en Inde. L'expérience consiste à évaluer simultanément deux programmes. Le premier programme²⁰³ vise à fournir aux écoles des ordinateurs, ces derniers étant envisagés comme une manière ludique d'impliquer les enfants dans l'apprentissage. Pour ce premier programme, l'évaluation prend place à Vadodara de 2002 à 2004, et se divise en deux périodes : 2002-2003 et 2003-2004. 111 écoles sont sélectionnées aléatoirement, 55 écoles ont bénéficié de cet enseignement assisté par ordinateur, les 56 autres écoles constituent le groupe témoin durant la première période (2002-2003). Lors de la seconde période (2003-2004) les groupes de traitement et les groupes de contrôle sont inversés ; ceux qui n'avaient pas bénéficié du programme lors de la première période le reçoivent désormais, et ceux qui en avaient bénéficié n'en bénéficient plus. Afin d'évaluer si le programme a permis d'améliorer l'apprentissage des élèves (Banerjee, Cole, Duflo et Linden, 2007), ils comparent les résultats aux examens avant l'implantation du programme et après l'implantation, pour chacun des deux groupes. Les résultats sont décrits de la façon suivante :

« The CAL Program has a strong effect on math scores (0.35 standard deviations in the first year (year 2) and 0.47 standard deviations in the second year (year 3)). It has no discernible impact on language scores (the point estimates are always very close to zero). This is not surprising, since the software targeted exclusively math skills, although some spillover effects on language skills could have occurred (for example because the program increase attendance, or because the children got practice in reading instructions, or because the teachers had reallocated time away from math to reading) ». (Banerjee, Cole, Duflo et Linden, 2007, 1251)

Le second programme (*Balshaki*), -cela a été développé dans la première partie de ce travail-, consiste à embaucher des jeunes filles qui ont terminé leurs études secondaires afin qu'elles aident les élèves en difficulté scolaire. Cela a aussi été vu, ce programme est d'abord évalué à Bombay puis à Vadodara. Les résultats de l'évaluation ont été considérés satisfaisants:

« Comparing Mumbai and Vadodara, the effects are very similar for math in both years (0.19 in Vadodara vs. 0.16 in Mumbai in year 1 and 0.37 vs. 0.32 in year 2), but in Mumbai, the effects for language are weaker and insignificant in both year (0.09 and 0.07 in year 1 and year 2), while they are significant in both years in Vadodara. The lower impact of language in Mumbai is consistent with the fact observed above, that most children (83.7 percent) in Mumbai already had some basic reading skills and are therefore

²⁰³ Dans la première partie de la thèse je me suis attaché à développer le second programme envisagé par l'ONG, je n'y avais donc pas développé ni défini ce premier programme.

less in need of a remedial program that targets the most basic competencies in language. In math, where more lag behind, the program was as effective as it was in Vadodara.” (Banerjee, Cole, Duflo et Linden, 2007, 1251)

2.2 L’absence de réponses au « pourquoi un programme fonctionne »

Après être entré dans la boîte noire, l’objectif est désormais de déterminer à quelles questions exactement l’expérience a permis de répondre et quelles questions elle a laissé de côté. J’ai distingué trois principales conclusions auxquelles arrivent les auteurs :

1. Le programme *Balshaki* est efficace à Vadodara en matière de langue, alors qu’il est moins efficace concernant les mathématiques.
2. Ce programme est aussi efficace à Bombay, en matière de langue, mais de façon moindre qu’à Vadodara ; et il est aussi moins efficace en mathématiques qu’il ne l’est pour les élèves de Vadodara.
3. Le programme d’enseignement assisté par ordinateur est efficace mais uniquement en ce qui concerne les mathématiques.

À partir de ces expériences, il est possible de dire, comme le souligne Deaton, *si* le programme fonctionne. Le programme d’enseignement assisté par ordinateur fonctionne pour améliorer l’éducation des élèves en ce qui concerne les mathématiques mais pas en langue. Le programme *Balshaki* fonctionne pour améliorer l’apprentissage des élèves en matière de langue, cependant il fonctionne mieux à Vadodara qu’à Bombay, mais n’a que peu d’effets sur l’amélioration de l’éducation en mathématiques. À partir de ces expériences, la seule information dont on dispose est de savoir *si* le programme fonctionne. Il n’est jamais question, explicitement, du « pourquoi le programme fonctionne ». Les seules questions adressées par les auteurs sont : pourquoi le programme d’enseignement assisté par ordinateur favorise les mathématiques aux langues ? (1) Et pourquoi les résultats en matières de langues sont meilleurs à Vadodara qu’à Bombay? (2) Comme l’avait souligné Deaton les réponses à ces deux questions restent fragiles, se référant à des suggestions, des probables, ou encore des explications possibles. La conclusion des auteurs explicite clairement cela :

« Some may be puzzled by the effectiveness of these two programs and the lack of spillovers of the Balsakhi Program to the other children given that the Balsakhis have less training than the formal teachers and that Computer-Assisted Learning Programs have not been shown effective in developed country

settings. We see two plausible explanations. First, teachers teach to the prescribed curriculum and may not take time to help students who are behind catch up, ending up being completely ineffective for them [Banerji, 2000]. Second, students share a common background with the balsakhis but not with the teachers”. (Banerjee, Cole, Duflo et Linden, 2007, 1262, souligné par moi)

La réponse à la première question posée par les chercheurs - savoir pourquoi le programme d'enseignement assisté par ordinateur favorise plus les mathématiques que les langues - a déjà été évoquée, lorsque j'ai présenté les résultats de l'expérience. Cela est dû, selon eux, au fait que le logiciel installé dans l'ordinateur est un logiciel d'entraînement aux mathématiques²⁰⁴. La réponse au second *pourquoi* est celle qu'évoque Deaton, en soulignant que les auteurs justifient l'efficacité moindre du programme *Balshaki* à Bombay, par le fait que cela est probablement lié au fait que les élèves de Bombay possédaient déjà les connaissances nécessaires en matière de langue. Comme cela apparaît désormais clairement, la question du « pourquoi le programme fonctionne » est absente de la discussion par les auteurs. Ne pas savoir « pourquoi le programme fonctionne » mais juste *si* il fonctionne, retrouve clairement le problème de l'hétérogénéité des effets de traitements. Il retrouve l'idée qu'on est face à une boîte noire. Ne pas savoir « pourquoi le programme fonctionne » laisse, comme Deaton le souligne, face à une hétérogénéité indifférenciée :

« The general lesson is once again the ultimate futility of trying to avoid thinking about how and why things work – if we do not do so, we are left with undifferentiated heterogeneity that is likely to prevent consistent of any parameter of interest.” (Deaton, 2010a, 432, souligné par moi)

Cela a des implications politiques importantes : comprendre pourquoi un programme n'a pas fonctionné permet de le retravailler et de cibler au mieux les populations. Heckman pointe ce problème en montrant, que ces informations sont d'une grande importance pour les décideurs politiques :

« Furthermore, policymakers often do not care solely about whether or not a particular program « works » in the sense of having benefits that exceed its costs. When program fail, it is important why they do not work. Without this information, which is not available from typical black-box experimental analyses, the only alternative open to politicians is to eliminate one program completely and start fresh with another. With the additional information available through nonexperimental methods, or through experiments designed to uncover the parameters of invariant structural models, it can be determined which services

²⁰⁴ *“This is not surprising, since the software targeted exclusively math skills”* (Banerjee, Cole, Duflo et Linden, 2007, 1251)

offered by a program work and for whom, thus allowing politicians to retarget and redesign existing programs, whose effectiveness will increase along with the store of social science knowledge about them.”
(Heckman et Smith, 1995, 94, souligné par moi)

Ne pas savoir « pourquoi un programme fonctionne » ne permet pas d’apprendre sur ce dernier, et ne permet donc pas de fournir les informations dont ont besoin les décideurs politiques. De plus, lorsqu’on ne sait pas « pourquoi un programme fonctionne » il est difficile d’envisager sa généralisation. La question du *pourquoi* est donc importante à un triple niveau : ne pas être laissé face à une hétérogénéité indifférenciée (1), afin d’être en mesure de fournir les informations nécessaires aux décideurs politiques (2) et donc envisager une possible généralisation des résultats (3).

3. Le problème de la généralisation des résultats ou la faible validité externe de la randomisation

La difficile généralisation des résultats constitue une des principales limites de la randomisation. C’est d’ailleurs une critique unanime, que même les chercheurs du J-PAL soulèvent²⁰⁵. Cette difficile généralisation semble être directement liée au problème développé plus haut. En ne comprenant pas « pourquoi un programme fonctionne » dans une certaine région du monde, il est difficile d’envisager qu’il puisse fonctionner dans une autre région. De plus, il est difficile de savoir s’il continuera à être efficace s’il étendu à une population plus grande. Les deux niveaux de la validité externe sont alors rendus difficiles. La forte validité interne de la randomisation semble être au dépend de sa validité externe. L’objectif est ici de développer les différentes critiques adressées à la validité externe du J-PAL par les économistes du développement. Ces critiques portent sur les deux niveaux de la validité externe mais insistent principalement sur le second niveau – la difficulté à étendre un programme à une échelle plus large. En ce sens, ces critiques s’ajoutent à l’analyse de Cartwright. L’objectif est ici de rendre compte de ces critiques et de montrer comment l’hétérogénéité indifférenciée, définie plus haut, donne lieu à une faible validité externe. Je commencerai par présenter les différentes critiques des économistes du développement portant sur la validité externe de la randomisation du J-PAL (3.1). Cela me permettra de situer ce volet de critiques par rapport à l’analyse de Cartwright ainsi que d’y ajouter de nouvelles

²⁰⁵ Voir par exemple (Banerjee et Duflo, 2009), (Banerjee, 2005b), (Duflo, Glennerster et Kremer, 2007), (Duflo, 2005), et (Duflo, 2004).

dimensions. Ensuite, je m'intéresserai particulièrement aux critiques traitant du second niveau de la validité externe (3.2).

3.1 La question de la généralisation des résultats.

Le problème de la validité externe de la randomisation est d'abord exprimé en médecine. Les économistes du développement reprennent ce problème de façon similaire à la médecine. C'est pourquoi, comme précédemment, je poserai, tout d'abord le problème en médecine (3.1.1). Celui-ci est exprimé, par la distinction entre efficacité potentielle et réelle, déjà présente chez Cartwright. Les économistes du développement qui visent à limiter la validité externe de la randomisation opèrent un lien direct avec cette distinction en médecine²⁰⁶. C'est pourquoi, je montrerai, ensuite, comment les économistes du développement reprennent et adressent ce problème à la randomisation du J-PAL (3.1.2).

3.1.1 La question de la validité externe de la randomisation en médecine : efficacité réelle vs efficacité potentielle

La validité externe de la randomisation est la principale critique faite aux essais cliniques en médecine, comme en souligne cette affirmation de Rothwell : « *Lack of consideration of external validity is the most frequent criticism by clinicians of RCTs* ». (Rothwell, 2005, 82) De plus, il existe en médecine de nombreux travaux prouvant cette faible validité externe²⁰⁷. Comme le souligne Rothwell, cette faible validité externe n'a été que très peu prise en compte par la médecine. Le problème de la validité externe des essais randomisés est d'une grande importance, car il permet de savoir si un traitement sera applicable dans des circonstances réelles. La distinction entre efficacité réelle et potentielle de Cartwright (2009c) provient, en fait, de la médecine. El-Serag (2007) définit l'efficacité réelle et l'efficacité potentielle de la façon suivante :

« *Efficacy*²⁰⁸ is largely determined by the biological effects of a therapy, but effectiveness takes into account external factors, such as individual patient characteristics, health system features, or societal influences. These are not evaluated easily in clinical trial settings. In addition to pharmacologic and

²⁰⁶ Voir par exemple (Deaton, 2009 ; 2010a) ou (Rodrik, 2007).

²⁰⁷ Voir par exemple (Jacobson, *et al.*, 1997) ou (Smith et Egger, 1998).

²⁰⁸ Tout comme je l'ai fait au chapitre précédent, je traduis les termes « *efficacy* » par efficacité réelle et « *effectiveness* » par efficacité potentielle, ce qui est la traduction commune en médecine.

physiologic effects, other issues must be considered for an intervention to be effective. These include (1) availability of the intervention to patient who can obtain maximum benefit, (2) identification of patients who are appropriate for the intervention, (3) recommendation of the intervention by providers, (4) acceptance of intervention by patients, and (5) adherence to treatment at the recommended dosing for therapeutic coverage to achieve fully the benefits of therapy. Effectiveness can be viewed from a patient or consumer perspective: If I receive the treatment as prescribed, will I benefit? Effectiveness is also viewed from a societal perspective: Are the benefits of this therapy reaching all those with the targeted condition?” (El-Serag, 2007, 8, souligné par moi)

La distinction entre efficacité réelle et efficacité potentielle faite par El-Serag (2007) comprend les deux niveaux de la validité externe alors que la distinction de Cartwright (2009c) ne pensait que le premier niveau. L'efficacité potentielle, telle que définie en médecine, est directement liée à la question de l'hétérogénéité indifférenciée des effets du traitement. D'une part, car elle se place du point de vue du patient et non de l'individu moyen. En ce sens, elle cherche à sortir d'une hétérogénéité indifférenciée. D'autre part, elle intègre aussi une dimension sociétale, en déterminant si le traitement sera efficace pour une autre population cible et s'il peut être étendu à une population plus large. L'efficacité potentielle cherche, ici, à appréhender si la généralisation d'un traitement n'entraîne pas d'influences sociétales, transformant les résultats obtenus au sein d'une expérience donnée. Malheureusement, la randomisation en médecine est principalement tournée, comme le soulignait déjà Cartwright, vers des preuves d'efficacité réelle et non d'efficacité potentielle. Dans cette perspective, la randomisation en médecine se centre sur sa validité interne, à défaut de sa validité externe. Il en va de même en économie du développement, ce à quoi je vais désormais m'intéresser.

3.1.2 La question de la validité externe en économie du développement

Tout comme en médecine, l'efficacité potentielle des résultats est d'un important intérêt en économie du développement puisqu'elle permet de penser la validité externe de la randomisation. Tout comme Cartwright, les différentes critiques adressées à la randomisation du J-PAL en économie du développement pointent cette faible validité externe²⁰⁹. Ils posent donc des questions similaires à celles posées par Cartwright : comment penser la généralisation des résultats des expériences ? Les résultats des expériences ne sont-ils pas

²⁰⁹ Angus Deaton (2009 ; 2010a), Martin Ravallion (2009a et 2009b), Michael Carter et Christopher Barrett (2010 et 2012), Dani Rodrik (2007), et Daron Acemoglu (2010).

dépendants de leurs contextes, les empêchant d'être envisagés dans un autre contexte ? Le contexte joue, souvent, un rôle central dans les programmes de développement ; s'interroger sur les effets de ce dernier permet de comprendre, on l'a vu, pourquoi le programme a fonctionné ou pourquoi il n'a pas fonctionné, et donc d'envisager qu'il puisse fonctionner ailleurs. Comme le souligne explicitement Martin Ravallion :

« One must recognize the importance of context since this can be key to draw valid lessons for other settings. Relevant contextual factors may include the circumstances of participants, the economic, the cultural and political environment, and the administrative context. Unless we understand how much such factors influence the outcomes of an intervention, the evaluation will have weak external validity. »
(Ravallion, 2009b, 39)

Tous les résultats provenant d'expériences randomisées s'appliquent à des régions, des domaines particuliers. Un programme qui a fonctionné dans une région particulière du monde peut ne pas fonctionner dans une autre. Ce que, par exemple, Rodrik regrette explicitement :

« the typical study based on a randomized field experiment says very little about external validity. If there are some speculations about the background conditions which may have influenced the outcomes and which do or do not exist elsewhere, they are offered in passing and are not central to the flow of the argument. Most importantly, the typical field experiments make no claims about the generalizability of the results ». (Rodrik, 2007, 21-22)

Les expériences randomisées du J-PAL en économie du développement souffrent du même problème que celui dont souffrent les essais randomisés en médecine et peu semble être fait pour contrecarrer cela. Tout cela tend à affaiblir très fortement l'utilisation possible des résultats des expériences randomisées. On en revient alors à la question du *pourquoi*, savoir pourquoi un programme fonctionne permet de rendre compte du contexte, des différentes interactions et donc d'être en mesure de savoir si ce programme pourrait fonctionner ailleurs. Cette question me semble nouer les deux principaux problèmes que j'ai tenté de développer durant ce chapitre : l'hétérogénéité des effets de traitement et la validité externe des expériences randomisées. Répondre à la question « pourquoi un programme fonctionne » permet à la fois de rentrer dans la « boîte noire » des résultats fournis par l'expérience et de sortir d'une hétérogénéité indifférenciée. Cela aidera à favoriser la validité externe de la randomisation du J-PAL.

3.2 Les effets négligés de la randomisation du J-PAL

Même si les différentes critiques adressées aux randomistas par d'autres économistes du développement portent sur les deux niveaux de la validité externe de la randomisation, ils insistent nettement plus sur le second niveau. Toujours dans le but d'offrir une vision globale de ces critiques, je m'attacherai ici à détailler leur critique portant sur la difficulté pour la randomisation du J-PAL de penser la généralisation d'un programme à une échelle plus large. En effet, il semble que la randomisation du J-PAL oublie de tenir compte de deux effets principaux. Ces derniers transformeraient les résultats obtenus à une échelle locale lorsque ceux-ci seraient étendus à une échelle plus large. Acemoglu (2010) insiste très largement sur ces deux effets, c'est pourquoi, c'est principalement sa critique que je reprendrai ici. Je présenterai tout d'abord le premier effet : l'effet d'équilibre (3.2.1). Puis je développerai le second effet : l'effet des facteurs politiques (3.2.2). Cela me permettra d'illustrer le second niveau de la validité externe du J-PAL ainsi que sa fragilité.

3.2.1 Effets d'équilibre : le passage d'un équilibre partiel à un équilibre général

Lorsqu'un programme est étendu à une population plus large, de nombreux changements peuvent opérer transformant complètement les résultats obtenus en équilibre partiel. Les évaluations du J-PAL sont faites à petite échelle et considèrent les autres facteurs fixés. Autrement dit, la randomisation fonctionne « toutes choses égales par ailleurs ». En ce sens, les résultats obtenus par la randomisation en économie du développement sont en équilibre partiel. Lorsque l'on étend le programme on envisage, cette fois-ci, un équilibre plus général. Le passage de l'équilibre partiel à un équilibre général peut entraîner de nombreux effets. Acemoglu cherche à montrer que ces effets d'équilibre ont une importance cruciale en économie du développement, ne pas en tenir compte affaiblirait donc sensiblement le résultat des expériences. Deaton et Ravallion pointent aussi très clairement ce problème, mais Acemoglu (2010) est le seul à avoir consacré un article presque entier à ce problème. C'est pourquoi, c'est son analyse que je développerai plus précisément ici.

Acemoglu (2010) définit trois effets d'équilibre qui ne sont pas pris en compte par les expériences randomisées. Le premier est lié au fait, qu'en réponse à une intervention de plus grande envergure les facteurs de productivités et de prix peuvent changer, entraînant alors un bouleversement profond. Un autre effet d'équilibre oublié par les expériences randomisées est

le fait que la généralisation à plus grande échelle peut créer des réponses technologiques endogènes. Et enfin, le troisième effet d'équilibre oublié est l'apparition d'effets de compositions résultant d'autres facteurs. Pour illustrer ce troisième effet, Acemoglu reprend un argument de Duflo et Banerjee portant sur les marchés du crédit. Banerjee et Duflo (2005c) expliquent que les petites et moyennes entreprises sont contraintes par le crédit dans les pays en développement. Un meilleur accès au crédit leur permettrait d'augmenter la production de leurs entreprises. Cependant, Acemoglu explique que si l'accès au crédit est facilité à une plus grande échelle, la conclusion de Banerjee et de Duflo (2005) ne tient plus. Pour Acemoglu, si toutes les entreprises ont accès au crédit, le profit que pouvaient tirer les petites et moyennes entreprises de cette facilité d'accès au crédit en augmentant leur production ne tient plus. La recommandation de Duflo et Banerjee (2005) avait un sens car cela se faisait au détriment d'autres entreprises, alors que si toutes les entreprises ont accès au crédit, c'est comme si rien n'étaient changé. Ici, Banerjee et Duflo ne tiennent clairement pas compte de ces effets.

La notion d'équilibre général est donc étroitement liée à la question de la validité externe de ce type d'expériences puisque ces effets d'équilibres sont le fruit d'une faible validité externe. Si une réflexion avait été amorcée pour déterminer ces effets, ils pourraient alors être évités et la généralisation serait sans doute possible. Acemoglu définit l'effet d'équilibre comme ce qui se : « *refers to factors that become important when we consider counterfactuals in which large change are contemplated.* » (Acemoglu, 2010, 17) Il est donc important de se questionner afin de déterminer, d'anticiper quels facteurs peuvent devenir importants et donc éviter ces effets d'équilibre. L'idée est, pour Acemoglu, de déterminer la « grande image » et non pas de se centrer sur une « petite image » qui peine à sortir du cadre.

3.2.2 Économie politique : la « grande image »

Acemoglu (2010) ajoute une autre dimension importante, en vue de ne pas perdre de vue la « grande image » des programmes de développement. Cette autre dimension est ce qu'il appelle l'économie politique et qu'il définit de la façon suivante : « *political economy refers to the fact that the feasible set of interventions is often determined by political factors and that large counterfactuals will induce political responses from various actors and interest groups.* » (Acemoglu, 2010, 17) Acemoglu retrace l'expérience du Ghana en 1971, lorsque le pays signe avec le FMI un accord de dévaluation de la monnaie en vue d'améliorer la

situation économique du pays. Après l'annonce par le gouvernement de l'application de la dévaluation de la monnaie le gouvernement de Kofi Busia fut renversé par le colonel Acheampong qui, une fois au pouvoir annula la dévaluation de la monnaie. Acemoglu conclut que la dévaluation était une bonne mesure économique mais une mauvaise mesure politique. C'était une mauvaise mesure politique, car elle laissait de côté le fait que le contrôle de prix était une partie importante du patronage au Ghana et que chaque politicien qui perdait ce patronage était susceptible de subir un coup d'état, ou de ne pas être réélu. Cet exemple pointe l'importance, en matière de programme de développement, de la prise en compte des facteurs politiques, pouvant annuler totalement les effets escomptés de la politique économique mise en place. Ce qui semble être aussi oublié par la randomisation. Afin d'anticiper ces effets, il est nécessaire de penser plus précisément aux structures sous-jacentes. Sans cela, la randomisation utilisée par Duflo et Banerjee souffre d'une faible validité externe, fragilisant ses résultats.

4. L'impossible arbitrage entre validité interne et validité externe ou la fin du gold standard

L'estimation de l'impact d'un programme que la randomisation offre est une estimation en forme réduite, comme l'explique Duflo : « *Results from randomized evaluations (and from internally valid program evaluations) provide reduced form estimates of the impacts of the treatment* » (Duflo, Glennerster, et Kremer, 2007, 3903) Cela signifie que l'estimation des effets des programmes porte sur un modèle simplifié. Cette estimation en forme réduite s'oppose à une estimation structurelle. L'approche structurelle vise à modéliser l'ensemble des relations économiques pertinentes, alors que la forme réduite ne s'attache qu'à une relation particulière. Ce faisant, l'estimation structurelle rend compte de tous les paramètres du modèle alors que l'estimation en forme réduite se centre sur un paramètre d'intérêt. Cette dimension de structure est directement liée aux deux effets développés ci-dessus. Modéliser l'ensemble des relations économiques pertinentes suppose de s'être questionné en amont sur le « pourquoi un programme peut fonctionner » et permet alors d'y offrir une réponse. La théorie joue ici un rôle important ; car c'est elle qui permet, en amont, de définir la relation structurelle. Modéliser des relations structurelles permettrait de penser les effets d'équilibres et donc de ne pas perdre de vue la « grande image ». Une dimension plus structurelle permettrait de pallier à la faible externe de la randomisation. Les principales critiques adressées à la randomisation préconisent la prise en compte de relations structurelles comme

(Ravallion, 2009a), (Deaton, 2009, 2010a), (Harrison, 2010), (Heckman, 2010) et (Acemoglu, 2010) afin de contrecarrer la faible validité externe de la randomisation.

Cependant, seuls Heckman (2010) et Acemoglu (2010) explicitent ce qui devrait être fait pour rendre compte de cette dimension structurelle. Tous deux offrent deux approches différentes. Acemoglu, propose d'intégrer des paramètres structurels dans les estimations. Ce qui suppose d'ajouter un certains nombres d'hypothèses fragilisant la validité interne de la randomisation pour une plus grande validité externe. Alors qu'Heckman propose d'estimer une combinaison de paramètres, ce qui suppose d'ajouter là aussi un certain nombre d'hypothèses, mais moins que celles que suppose l'approche d'Acemoglu. Ce qui rend l'arbitrage entre validité interne et validité externe moins fort. Heckman vise à proposer un véritable pont entre les deux approches. Je commencerai par développer l'approche proposée par Acemoglu (2010) (4.1), puis je développerai celle d'Heckman (2010) (4.2).

4.1 l'estimation de paramètres structurels au sein de la randomisation : une plus grande validité externe pour une plus faible validité interne

Acemoglu (2010) propose d'utiliser la randomisation pour évaluer des paramètres structurels, afin d'être à même de rendre compte des effets d'équilibre et d'économie politique. L'objectif est donc : « *not [to] lose sight of the bigger picture of the problem of economic development.* » (Acemoglu, 2010, 30) Pour Acemoglu, la prise en compte de ces effets d'équilibre et la prise en compte de la dimension politique passe par la compréhension des facteurs politiques à l'œuvre. L'objectif est, pour Acemoglu, de définir les facteurs à l'œuvre, et de les intégrer ensuite dans l'estimation que produit la randomisation. Il propose d'estimer un paramètre structurel qui tiendrait compte de ces facteurs, et non plus un unique paramètre d'intérêt qui n'attrape qu'un élément comme le fait la randomisation de Duflo. Afin de rendre compte de la proposition d'Acemoglu et de la tension, qu'elle me semble pointer entre validité externe et interne, j'expliquerai, tout d'abord, comment il montre qu'il est possible de traduire une estimation en forme réduite en forme structurelle (4.1.1). Puis, je questionnerai l'impact de cette proposition en termes de validité externe et interne (4.1.2).

4.1.1 La possibilité de traduire la forme réduite en forme structurelle

Le rôle de la théorie est d'autant plus important chez Acemoglu que c'est grâce à elle qu'on est à même de définir les paramètres structurels du modèle. C'est pourquoi il affirme :

« I therefore define « structural parameters » as those that provide external validity and would thus be useful in testing theories or policy analysis beyond the specific environment and sample from which they are derived. External validity becomes particularly challenging task in the presence of general equilibrium and political economy considerations, and a major role of economic theory is in helping us overcome these problems or at the very least alerting us to their importance. » (Acemoglu, 2010, 18)

Afin d'expliciter le rôle que peut jouer la théorie dans la détermination de paramètres structurels, Acemoglu commence par distinguer précisément, autour d'un exemple sur l'éducation, ce qu'offre l'estimation d'un modèle en forme réduite. Il s'intéresse alors à la relation qui peut exister entre les coûts de scolarité et les décisions en matière d'éducation (qui se résume ici à aller à l'école ou non). Dans un modèle de forme réduite, la relation entre les deux pourrait être explicitée de la façon suivante :

- Année de scolarité = caractéristiques des individus que l'on veut contrôler – les années d'études suivies.
- De façon économétrique le modèle pourrait se traduire de la façon suivante :
- $\text{Log}(S_i) = X_i' \beta - \alpha \log(c_i) + \varepsilon_i$

Où i correspond aux individus de l'échantillon, S_i renvoie aux années d'étude suivies par les individus de l'échantillon, c_i correspond aux coûts de scolarité que doivent payer les individus de l'échantillon, X_i est un vecteur capturant les caractéristiques des individus que l'on veut contrôler afin que l'estimation ne soit pas biaisée, β est un vecteur du paramètre X_i , ε correspond au terme d'erreur et enfin, α est le paramètre d'intérêt que l'on souhaite évaluer. Ce que l'on cherche donc à évaluer ici est l'impact du coût de scolarité sur les années d'études suivies par les individus de l'échantillon. En forme réduite, -ce que fait la randomisation de Duflo-, on évalue donc ce paramètre en contrôlant pour les autres variables qui pourraient elles aussi jouer un rôle sur le nombre d'année d'études suivies. Afin d'estimer ce paramètre, la méthode des moindres carrés (celle privilégiée par Duflo) sera utilisée.

Pour Acemoglu, il est possible d'arriver à la même équation économétrique tout en proposant cette fois-ci un modèle plus sophistiqué fondé sur une théorie spécifique. Le modèle intégrerait le capital humain comme fonction du niveau d'étude des individus. On aurait alors un niveau de capital humain donné pour un individu, et de ce niveau de capital humain dépendrait un certain niveau de salaire que l'individu pourrait espérer. Les années d'éducation seraient traduites par un coût à la fois monétaire et non monétaire. On suppose ensuite que les individus vont chercher à maximiser leur revenu. Ce modèle implique la même équation que le précédent modèle, sauf que le paramètre d'intérêt α est désormais plus sophistiqué et correspond à l'arbitrage des individus entre le coût de l'éducation et le salaire qu'ils peuvent obtenir en fonction de leur capital humain. La théorie qui permet de construire ce modèle s'inspire directement de la théorie du capital humain de Gary Becker (1964).

4.1.2 Un gage de validité externe pour une plus faible validité interne ?

La question que soulève, ensuite, Acemoglu est de savoir : « *in what sense can we think of it as a « structural relationship » ?* » (Acemoglu, 2010, 19) Pour Acemoglu, cela relève directement de la question de la validité externe :

« Suppose we now ask the question : what would be the effects of subsidies to reduce the cost of schooling, c_i , for a set of individuals ? This counterfactual experiment could be motivated by a potential policy that is being by a potential policy that is being contemplated or it may be used for undersatnding and testing the implications of our theory. The question might be for the same sample on which the initial estimation was performed or it could be for an entirely different sample or population. In either case, one answer to the above question readily follows from using the estimate α to compute the increase in the years of schooling for individuals whose cost of schooling has declined. » (Ibid.)

Mais comment être convaincu de la validité externe du résultat obtenu ? Pour Acemoglu, la question est vite réglée: puisque que c'est une relation structurelle on doit croire le résultat. Cependant, et c'est ici que la théorie va jouer un rôle important pour lui, la théorie va permettre de justifier le « paramètre clé ». Elle va permettre de justifier en quoi ce dernier joue un rôle majeur en toutes circonstances. Le premier rôle joué par la théorie est donc de justifier explicitement l'équation qui va être estimée et donc, de ce fait, le paramètre d'intérêt. La théorie de Becker peut permettre d'explicitier cela. De plus, pour Acemoglu, la théorie offre le moyen le plus pertinent pour comprendre les résultats obtenus, et donc les interpréter. Cette interprétation permet de répondre à la question du « pourquoi un programme

fonctionne » soulevée plus haut, et d'appréhender la généralisation de tels résultats. Néanmoins, Acemoglu souligne que l'approche structurelle fondée sur la théorie peut rencontrer certaines difficultés. Afin de permettre une validité externe, il est nécessaire de faire plusieurs hypothèses, et les théories permettant de faire ces hypothèses sont des abstractions. De plus, le paramètre structurel d'une théorie peut être endogène dans une autre théorie et donc ne pas être pertinent en toutes circonstances. Malgré tout, pour Acemoglu la définition d'une relation structurelle à travers la théorie reste le meilleur moyen pour aboutir à une importante validité externe :

« These challenges notwithstanding, it is clear that economists often have to take position about the parameter being estimated corresponding to structural parameters (at least for well-defined though perhaps limited set of variations in environment and policy). Otherwise, we will have no way performing counterfactual exercises and making predictions about policy changes (Imbens, 2009). But this necessitates a claim to external validity (even if it is only implicit), and economic theory is our best guide for formulating the appropriate models and justifying such claim to external validity. These issues become only more central in the presence of general equilibrium effects and policy factors ». (Acemoglu, 2010, 22)

Une fois la question de la validité externe réglée de cette façon, c'est désormais la question de la validité interne qui se pose. L'intérêt de la randomisation est de pouvoir évaluer l'impact d'un programme en faisant très peu d'hypothèses. En utilisant explicitement une théorie et en déterminant des relations structurelles, la randomisation perd en validité interne, et de ce fait perd son statut de *gold standard*, mais gagne potentiellement en validité externe. Ce que, par ailleurs, Banerjee et Duflo soulignent eux même, non sans réticence :

« Très clairement, un raisonnement a priori nous aide peu – le raisonnement économique de base pose que la distribution d'uniformes n'aura d'effet que pour les populations dont le revenu moyen n'est pas très élevé par rapport au prix des uniformes – mais où se situe « pas très élevé » ? Si nos théories sont suffisamment précises pour nous permettre de le savoir, ou si nous posons le principe qu'elles le sont, alors nous n'avons sans doute plus besoin d'expérimentations : la théorie suffit à nous donner une idée de qui aura tendance à obtenir un uniforme ou non, et nous pourrons utiliser cette restriction pour évaluer de manière convaincante les modèles structurels portant sur l'impact de la distribution d'uniformes scolaires. En d'autres termes, sans hypothèses, les résultats d'une expérimentation ne peuvent être généralisés au-delà de leur contexte, mais avec suffisamment d'hypothèses, les données non expérimentales peuvent suffire. Pour justifier l'expérimentation, nous devons nous placer au milieu du gué. » (Banerjee et Duflo, 2009, 703)

La faible validité externe de la randomisation semble pouvoir être contrecarrée mais au prix d'une plus faible validité interne, impliquant des hypothèses supplémentaires, permettant

de quitter une hétérogénéité indifférenciée pour une hétérogénéité supposée et explicitée. La préconisation d'Acemoglu met l'accent sur la validité externe de la randomisation affaiblissant très clairement sa validité interne et faisant perdre, d'une certaine façon, l'attrait que la randomisation possède. Est-il alors possible de déterminer le milieu qu'évoquent Duflo et Banerjee, renforçant de ce fait la validité interne et externe de la randomisation ?

4.2 Un pont possible entre l'approche structurelle et la randomisation ?

Heckman montre que la randomisation suppose une définition de la causalité particulière. La notion de causalité qui sous-tend la randomisation se fonde sur la notion d'équilibre partiel développée par Alfred Marshall²¹⁰. De ce fait, comme on l'a vu plus haut, la randomisation passe à côté des effets d'équilibre et oublie un certain nombre d'autres chaînes causales. Alors que l'approche structurelle tient compte de ces différentes chaînes causales, mais aux dépens de la validité interne du résultat. Heckman propose un pont entre ces deux approches afin qu'elles puissent se renforcer l'une l'autre. L'approche envisagée par Heckman (2010) est plus nuancée que celle d'Acemoglu (2010) car elle ne met pas uniquement l'accent sur la validité externe de la randomisation, mais cherche à penser et à équilibrer aussi bien la validité interne qu'externe de la randomisation. En ce sens, elle pourrait permettre d'offrir le pont cherché par Duflo et Banerjee. Je développerai, tout d'abord, la causalité sous-jacente de la randomisation, ce qui me permettra de pointer les enjeux que souhaite soulever Heckman (4.2.1); puis, je présenterai son alternative (4.2.2).

4.2.1 Un seul maillon de la chaîne : la clause *ceteris paribus* ou la causalité sous-jacente de la randomisation

La randomisation n'estime, on l'a vu, qu'un paramètre d'intérêt - censé attraper les effets du programme. Comme l'explique Heckman, cela tend à superposer la définition de la causalité sous-jacente à une telle estimation et l'identification causale des paramètres :

« In this approach [program evaluation approach], the parameters of interest are defined as summaries of the outputs of experimental interventions. This is more than just a metaphorical usage. Rubin and Holland argue that causal inference only if an experiment can be performed. This conflation of the separate tasks of defining causality and identifying causal parameters from data is a signature feature of the

²¹⁰ On retrouve ici une position similaire à celle de Cartwright, lorsqu'elle explique que la randomisation permet de produire des lois causales, c'est à dire des affirmations de causalité valables dans un cadre précis et donné.

program evaluation approach. It is the consequence of the absence of clearly formulated economic models.
» (Heckman, 2010, 358)

En ne formulant pas clairement de modèle économique sous-jacent à leur approche, Banerjee et Duflo confondent la causalité inhérente à leur modèle et l'identification causale des paramètres d'intérêt. La notion de causalité sous-jacente à leur démarche est fondée sur le contrôle des variations (Heckman, 2010). La randomisation, en annulant les biais de sélections, contrôle tous les autres facteurs qui pourraient avoir un incident, et de ce fait ne laisse apparaître que l'effet du programme. En ce sens elle considère les autres facteurs fixés. Pour Heckman, c'est donc la clause *ceteris paribus* d'Alfred Marshall qui définit la causalité à l'œuvre dans leur approche. Autrement dit, la randomisation permet de rendre compte d'une causalité en équilibre partiel. L'identification causale du paramètre d'intérêt renvoie, on l'a vu, à la moyenne des effets de traitements, et en ce sens considère un individu moyen dont les différents éléments seraient résumés en lui. Dans cette perspective la randomisation laisse alors de côté de nombreuses dimensions causales qui pourraient être d'un intérêt central en termes de politique de lutte contre la pauvreté, en d'autres termes la randomisation « *rules out variety of alternative channels of identification of policy effects* » (Heckman, 2010, 360). La randomisation ne permet d'évaluer qu'un bout de la chaîne causale en supposant les autres éléments de cette chaîne fixés. Mais en évaluant qu'une partie de la chaîne causale, la randomisation peut-elle fournir un résultat utilisable ? Ce résultat permet-il d'appréhender réellement le phénomène étudié – l'effet du programme ?

Heckman tente de construire un pont entre l'approche structurelle et la randomisation. Il explique, tout d'abord, que l'évaluation de programme soulève trois problèmes. Le premier est d'identifier l'effet du programme dans un contexte donné, ce qui relève de la validité interne du résultat. La randomisation, on l'a vu, possède une forte validité interne et permet de ce fait de relever ce problème. Le deuxième problème est de savoir si le résultat trouvé dans un contexte particulier sera le même dans un autre contexte ; la randomisation possède une faible validité externe, alors qu'une approche structurelle a une plus forte validité externe. Le dernier problème est de parvenir à évaluer le bien être des individus, la randomisation ne peut en rendre compte alors que l'approche structurelle y parvient. La randomisation et l'approche structurelle semblent donc être complémentaires et pourraient alors s'enrichir l'une l'autre.

4.2.2 *Un pont entre la randomisation et les approches structurelles : la maxime de Marschak*

Heckman (2010) cherche à construire un pont entre ce qu'il appelle les « deux camps » – approche structurelle et randomisation. L'objectif est de concilier les points forts de ces deux approches afin qu'elles puissent se renforcer l'une l'autre. Pour cela, Heckman se réfère à Jacob Marschak qui a déjà, dans les années 50, tenté de trouver un juste-milieu entre ces deux approches. Pour Marschak afin d'évaluer l'impact d'un programme, il n'est pas nécessaire de se référer à une estimation structurelle pleinement explicitée. Heckman tire alors, ce qu'il appelle la maxime de Marschak : « *Marschak's Maxim suggests that economists should solve well-posed economic problems with minimal assumptions.* » (Heckman, 2010, 359) On est bien dans un niveau intermédiaire entre les deux approches, le problème se doit d'être clairement posé tout en minimisant le nombre d'hypothèses. Autrement dit, il n'est pas nécessaire de déterminer tous les paramètres qui déterminent la relation structurelle. Il existe un niveau de connaissance suffisant pour évaluer l'impact du programme, tout en étant capable de déterminer comment il pourrait être applicable ailleurs.

L'idée est que la différence des moyennes des effets du traitement dans le groupe traité et dans le groupe de contrôle ne soit pas déterminée par un unique paramètre d'intérêt mais par une combinaison de plusieurs paramètres qui ne prendraient pas en compte tous les paramètres. Heckman tente ensuite d'appliquer cette maxime au modèle de Roy, emblème de l'analyse contrefactuelle, utilisée par Duflo. Il est alors possible de favoriser la validité interne de l'approche structurelle en réduisant les hypothèses et le nombre de paramètres tout en améliorant la validité externe de la randomisation. Cette approche alternative n'a pas eu d'écho dans la randomisation utilisée par Duflo et Banerjee. Cela montrant qu'une telle approche privilégie sa validité interne au dépend de sa validité externe.

5. Conclusion

Le terme de « randomistas », associé aux chercheurs du J-PAL, est désormais double ; il soulève, à la fois, la forte validité interne proclamée par les chercheurs du J-PAL et sa faible validité externe, pointée par les critiques. C'est de cette double dimension que j'ai tenté de traiter. C'est à travers elle que j'ai essayé de rendre compte des différentes critiques adressées

à la randomisation de Duflo et de Banerjee. J'ai alors cherché à les unifier afin de faire ressortir la tension sous-jacente entre validité interne et validité externe.

Pour cela, j'ai pris pour point de départ le problème de l'hétérogénéité des effets de traitements ; en essayant de montrer que ce problème n'était, en fait, qu'un symptôme d'un problème plus large : une hétérogénéité indifférenciée. Les résultats produits par la randomisation de Duflo et Banerjee sont alors telle une boîte noire. Ils ne permettent pas de répondre explicitement à la question « pourquoi un programme fonctionne ». Cela m'a permis de questionner la validité externe de la randomisation. J'ai particulièrement cherché à mettre l'accent sur les critiques portant sur le second niveau de la validité externe, puisque le premier niveau avait été traité autour de la critique de Cartwright. Un arbitrage semble nécessaire entre la validité interne et externe de la randomisation du J-PAL. Son statut de *gold standard* tend alors à s'effondrer.

Chapitre 3. Expliquer la boîte noire : le rôle de la théorie dans l'approche du J-PAL

0. Introduction

« In fact, our preoccupation with accurate quantification often takes us away from the more important causes of a phenomenon and we concentrate on variables that are better measured but may socially and economically less significant, reminding one of the oft-repeated charge against economists, who look for the missing keys not in the dark place where they lost them but where there is more light. We tend to work with a thin conceptual menu and a large box of precision instrument. » (Bardhan, 2005, 4334)

La majeure partie des critiques adressées à la randomisation du J-PAL s'attache à sa dimension a-théorique. La forte validité interne du J-PAL, dû au fait que ses expériences minimisent les biais en stipulant très peu d'hypothèses, est contrecarrée par la faible portée des résultats du J-PAL. Cela tend à affaiblir très fortement la validité externe d'une telle méthodologie, et cela à deux niveaux : (1) il est difficile de transposer le résultat d'une expérience donnée à un autre contexte, (2) et il est compliqué d'envisager la mise en place d'un programme à une échelle plus large. L'absence de réflexion sur « pourquoi un programme fonctionne ? » explique la faible validité externe d'une telle approche. Comme le souligne Bardhan (2005), la randomisation utilisée par le J-PAL se concentre sur la précision de son instrument, mais n'utilise que très peu de concepts. Cela rejoint très fortement les critiques développées au chapitre précédent; les expériences du J-PAL, en ne se concentrant que sur un objet précis, perdent de vue l'explicitation des mécanismes à l'œuvre et se refusent à un cadre théorique défini. Il semble qu'en l'absence de théorie, une telle démarche ne puisse parvenir à saisir « pourquoi un programme fonctionne ? », perpétuant ainsi, une faible validité externe. L'objectif est ici de questionner le statut accordé à la théorie au sein de l'approche du J-PAL. Ce statut ne me semble pas se résumer à une simple absence. Le J-PAL accorde à la théorie un double rôle. En vue de garder une forte validité interne, une telle approche se refuse à penser toute théorie en amont ; mais envisage de créer une théorie unificatrice à partir des résultats obtenus à l'aide des expériences randomisées, afin d'améliorer sa validité externe. Le statut accordé à la théorie me semble pouvoir traduire la tension existante entre la validité interne et la validité externe d'une telle approche. De plus, questionner le statut de la théorie au sein de cette approche me permet de penser simultanément ses deux niveaux de validité externe.

Je développerai, tout d'abord, l'épistémologie que propose Deaton afin d'apprendre des mécanismes à l'œuvre en économie du développement. Tout comme au chapitre précédent, Deaton offre ici une des critiques les plus complètes quant au manque de théorie au sein de la randomisation du J-PAL. Il développe une démarche hypothético-déductive²¹¹, à partir de laquelle il devient possible de rendre compte des mécanismes centraux de l'économie du développement. L'approche envisagée par Deaton implique de formuler des hypothèses, de les tester, puis de les reformuler ; en exposant comment, à l'aide de la randomisation comme outil de test, il devient possible de rendre compte de mécanismes centraux et, *in fine*, de penser un cadre théorique. La démarche de Deaton vise explicitement à penser l'absence de théorie au sein des expérimentations du J-PAL.

Ensuite, je m'attacherai à montrer que la théorie a un double statut au sein de l'approche du J-PAL. En 2005, lors d'un symposium sur le rôle de la théorie en économie du développement, Banerjee répond à de nombreux économistes du développement de renom²¹² quant au manque de théorie au sein de son approche. La réponse de Banerjee à ces critiques, déjà développée dans de nombreux travaux communs avec Duflo, me semble mettre en évidence un double statut conféré à la théorie. Banerjee et Duflo se refusent tout d'abord à la définition d'une théorie pensée en amont de l'expérience, afin de garantir au mieux la validité interne de leurs résultats. Cependant, ils soulignent qu'un de leur principal défi est de parvenir à construire un nouveau cadre théorique englobant les différents résultats obtenus. Je distinguerai ces deux statuts, en considérant que Banerjee et Duflo se refusent à toute théorie *ex ante*, en vue de construire une théorie *ex post*. Je soulignerai la traduction politique de ce défi théorique : passer de la recommandation de politiques locales à des recommandations politiques d'ordre plus global ; ce qui semble délicat en l'absence d'un cadre théorique général.

Enfin, je montrerai que les résultats du J-PAL s'avèrent éclatés et difficiles à unifier, et que peu d'entre eux permettent d'offrir de réelles recommandations politiques. Ces résultats s'apparentent plus à une description extrêmement précise de la vie des pauvres. En ce sens, les expériences du J-PAL tendent à rendre explicites et à définir précisément les effets de la pauvreté, mais peinent à rendre compte des véritables causes de cette pauvreté. L'absence de théorie *ex ante* me semble jouer un rôle majeur au sein de cette difficulté, traduisant la peine à

²¹¹ J'ai développé cette démarche lors du chapitre 1 de cette partie.

²¹² Voir Bardhan (2005), Basu (2005), Kanbur (2005), et Mookherjee (2005) et Banerjee (2005b).

obtenir une certaine validité externe au sein de la perspective adoptée par le J-PAL. Cela me permettra d'englober les deux différentes critiques portant sur les deux niveaux de la validité externe des expérimentations du J-PAL.

1. Comprendre les mécanismes : l'approche hypothético-déductive de Deaton

Pour Deaton, la théorie pourrait rendre compte des mécanismes sous-jacents aux résultats obtenus par les expériences du J-PAL. Elle favoriserait donc la validité externe de la randomisation, permettant de comprendre pourquoi un programme fonctionne ou ne fonctionne pas. C'est pourquoi Deaton affirme :

« Learning about theory, or mechanisms, requires that the investigation be targeted towards that theory, towards why something works, not whether it works. » (Deaton, 2009, 30)

Dans la droite continuité de l'épistémologie de Cartwright, Deaton privilégie la mise en évidence des mécanismes qui sous-tendent les résultats obtenus par les expériences du J-PAL afin de pouvoir apprendre de la lutte contre la pauvreté et du développement, et d'être en mesure d'envisager une plus forte validité externe. Deaton adopte très clairement une démarche hypothético-déductive, comme celle définie par Nancy Cartwright, c'est à dire, un intermédiaire entre les « garantisseurs » et les « vérificateurs ». Deaton adopte alors un cadre Poppérien : c'est par un procédé d'essais et d'erreurs que l'on est en mesure d'apprendre, et donc de faire émerger de nouvelles connaissances. La théorie joue un rôle central dans ce procédé. L'aller-retour entre la théorie et l'empirie permet de montrer l'existence de certains mécanismes. En ce qui concerne l'économie du développement, Deaton montre que les connaissances se sont accumulées grâce à ce processus. Il montre cela à travers trois thèmes différents.

Je reprendrai chacun de ces thèmes afin de montrer, comment pour Deaton, il est possible d'apprendre des mécanismes en économie du développement. Pour cela, je développerai, tout d'abord, la première corrélation auquel il s'intéresse, celle entre l'épargne et la croissance. Il montre que l'existence et la définition de cette corrélation ne se sont faites qu'à travers le croisement de la théorie et de l'observation empirique (1.1). Je m'attacherai, ensuite, au lien entre le revenu et le prix des biens. Cela me permettra de mettre en évidence le rôle du test Poppérien des hypothèses dans le processus de falsification au sein de l'approche de Deaton (1.2). Puis je montrerai que, pour Deaton, en économie du développement, les connaissances

se sont constituées au sein d'un aller-retour perpétuel entre l'observation et la théorie; mettant au centre la notion de mesure, qu'il développe autour du lien entre revenu et alimentation (1.3). Enfin, je questionnerai l'épistémologie de Deaton (1.4).

1.1 La corrélation épargne et croissance : le croisement de la théorie et de l'observation empirique

Le premier domaine auquel s'intéresse Deaton est le lien entre la croissance et l'épargne. Il cherche à montrer comment, à travers un processus Poppérien de falsification, la recherche en économie progresse. Deaton illustre ceci par l'histoire de l'étude du lien entre l'épargne et la croissance. Son histoire commence par le célèbre article d'Arthur Lewis²¹³ dans lequel il montre qu'une économie développée passe d'un taux épargne de 5% à un taux d'épargne de 15%. La même année, Franco Modigliani et Richard Brumberg développent la théorie des cycles de vie²¹⁴, théorie qui explique le comportement d'épargne des individus. En 1970, Modigliani²¹⁵, s'appuyant sur le travail d'Hendrik Houthakker, montre que le taux d'épargne est plus important dans les économies en forte croissance. En effet, les individus épargnant principalement lorsqu'ils sont jeunes (et consommant leur épargne une fois retraités), alors la croissance de la population et du revenu par tête augmenteront le taux d'épargne national. Ce cadre d'analyse a prévalu pendant longtemps, jusqu'à ce que Christopher Carroll et Lawrence Summers, dans un article en 1991, (Carroll et Summers, 1991) nuancent la théorie du cycle de vie à l'aide d'observations empiriques. Pour eux, si la théorie du cycle de vie était vraie, alors l'âge des consommateurs devrait être corrélé positivement avec le taux de croissance de l'économie. Si l'on considère que les jeunes consommateurs ont plus de ressources que les consommateurs plus âgés, le ratio de la consommation des jeunes consommateurs et des consommateurs âgés devrait être plus important dans les économies connaissant une croissance plus rapide. Par ailleurs, comme les gens souhaitent maintenir leur niveau de consommation à tout âge, on ne devrait pas observer de modification de consommation liée à

²¹³ Voir : (Lewis, 1954).

²¹⁴ La théorie du cycle de vie de Modigliani, qui lui a valu le prix Nobel en 1985, prend en compte l'âge du consommateur dans la détermination de sa consommation. Les enfants consomment à travers leurs parents, les adolescents sont sensibles aux marques et aux effets de mode, au début de la vie d'adulte les revenus sont faibles donc l'épargne l'est aussi, durant sa vie active l'individu peut rembourser ses dettes et constituer une épargne qu'il utilisera quand il sera plus vieux. La principale conséquence au niveau macroéconomique de cette théorie est de lisser les consommations tout au long de la vie. De ce fait, cette théorie fournit une explication à la stabilité empirique de la propension moyenne à consommer sur le long terme (alors que la théorie keynésienne postulait sa décroissance à long terme).

²¹⁵ Voir : (Modigliani, 1970).

l'âge du consommateur. L'analyse de Carroll et Summers (1991) montre que l'âge de consommation est similaire dans une économie croissant faiblement ou rapidement ; ce qui revient à dire que la théorie du cycle de vie ne peut pas expliquer la corrélation positive entre épargne et croissance.

C'est donc la théorie qui a ici été remise en cause, via une simple observation empirique. Pour Deaton, cette histoire traduit une démarche hypothético-déductive, c'est-à-dire que si une prédiction est testée – la théorie du cycle de vie – et que le test échoue, la prédiction doit être revue et améliorée. Pour autant, ce n'est pas parce que le test échoue, que toute la théorie du cycle de vie est remise en cause. Ce qui est mis en évidence par Deaton, c'est que la théorie du cycle de vie, ne peut pas expliquer à *elle seule* la corrélation positive entre l'épargne et la croissance. La falsification de cette prédiction permet d'affiner la compréhension de la relation de l'épargne à la croissance :

“The falsification of the hypothesis as a general proposition has been fruitful in stimulating new research on behavioral approaches to saving in both rich and poor countries. So we are certainly no left with nothing, even if there are things that we thought we understood but no longer do. We have a more nuanced and qualified understanding of the life-cycle mechanisms and a useful agenda for future research.”
(Deaton, 2010b, 7)

Ce procédé de falsification permet de stimuler de nouvelles recherches et d'enrichir la connaissance, afin d'en apprendre un peu plus sur les mécanismes. Pour autant ici, les mécanismes ne sont pas définis, et Deaton ne dit pas explicitement ce qu'il entend par ces derniers. Il semble que les mécanismes soient pour lui les phénomènes qui permettent d'expliquer ou de rendre compte d'une corrélation particulière: ici, la croissance et l'épargne. Deaton obscurcit encore un peu plus l'ensemble de ces termes épistémologiques lorsqu'il loue l'intérêt du travail intermédiaire réalisé par Modigliani en 1970, et celui de Carroll et Summers en 1991 :

« I also want to emphasize the importance of the work that is intermediate between theory and data and that required to find specific and transparently testable predictions, like Modigliani's (1970) derivation of the link between saving, growth and the length of work and retirement spans, and Carroll and Summers (1991) work on age profile of consumption. This kind of intermediate work is not econometrics, nor is it theory as usually understood, but it is work that can only be done by those with a familiarity with the theory and an understanding of the evidence that is available or might be available » (Deaton, 2010b, 7, souligné par moi)

Ce travail intermédiaire renvoie-t-il à la recherche des mécanismes? Quel est le travail intermédiaire entre la théorie et les données ? Et si cela ne relève ni de l'économétrie, ni de la théorie au sens commun, quel est cet intermédiaire ?

1.2 Le lien entre revenu et prix des biens : l' « acid test »

La fluctuation du prix des biens mondiaux peut être un problème crucial pour les pays en développement, beaucoup des exportations de ces pays dépendant des matières premières. Si les prix fluctuent, cela peut engendrer beaucoup de troubles. C'est pourquoi en théorie, ces pays ont tout intérêt à épargner quand les prix sont élevés afin de se prémunir contre une baisse potentielle des prix. Deaton reconstruit là aussi une histoire de l'étude des mécanismes mis en évidence via le lien entre le revenu et le prix des biens des matières premières. Son histoire débute par le modèle spéculatif de Gustafson²¹⁶, qui met en évidence l'existence d'un processus sous-jacent à l'offre et la demande dans l'agriculture. Prenant pour point de départ un choc d'offre, l'effet de ce choc est filtré sur le marché par un comportement maximisateur de la part des agents qui achètent les matières premières lorsque le prix est faible, et qui les revendent lorsque le prix augmente. Deaton et Laroque (1992, 1996) ont regardé si cela coïncidait avec les données et ont conclu à son invalidité. Selon eux, lorsque le choc a lieu, les agents anticipent l'augmentation des prix. Cependant, lorsque les prix atteignent le niveau où les spéculateurs arrêtent de tenir leur inventaire, le prix anticipé par les agents ne change pas. Deaton et Laroque ont alors réalisé un test des prédictions du modèle de Gustafson (1958), permettant d'expliquer pourquoi il peut être difficile pour un gouvernement de lisser ses revenus²¹⁷? Selon Deaton – comme pour l'illustration de l'épargne et de la croissance –, on obtient ainsi un modèle plus satisfaisant. Mais ici encore, Deaton ne définit pas ce qu'il entend précisément par modèle, et c'est à l'aide d'un processus hypothético-déductif que la découverte de mécanismes a lieu. Deaton explique alors :

« One methodological lesson that I want to draw here is that the simulation and estimation of tightly specified models may reveal regularities and behaviors that are worth investigating as possible generalization and that can lead, in favorable cases, to “acid tests” of the theory. In effect, we are using the

²¹⁶ Voir : (Gustafson, 1958).

²¹⁷ La principale explication du fait que les gouvernements des pays en développement peuvent peiner à lisser les revenus est qu'ils sont souvent trop impatients pour anticiper et contrecarrer les fluctuations des prix des matières premières.

computer to think for us. Structural estimation is useful, not only for the estimates (whose credibility is often undercut by the panoply of supporting assumptions that are required to obtain them), but for understanding the empirical predictions of the theory. » (Deaton, 2010b, 9)

La simulation et l'estimation de modèles précis peuvent permettre de tester la théorie afin de déterminer si celle-ci est corroborée ou réfutée. On retrouve alors l'« *acid test* », le test Poppérien par excellence. Cela montre une fois de plus, que pour Deaton, c'est par un processus déductif de compréhension des mécanismes, puis de tests, que l'on arrive à faire avancer les connaissances en économie du développement.

1.3 Le lien entre revenu et alimentation : mesure et théorie

Cette dernière illustration offre ceci d'intéressant, pour Deaton, qu'à l'inverse des deux précédentes, elle ne part pas de la théorie, mais d'observations. Comme le soulignent Duflo et Banerjee dans leur dernier livre, lorsque l'on pense la pauvreté ou lorsque l'on imagine la pauvreté, on se représente des individus mourant de faim, au ventre gonflé et aux côtes apparentes. Un « pauvre » est un individu qui meurt de faim. C'est précisément à cet aspect que Deaton s'intéresse. Et il va alors s'efforcer de mettre en évidence deux principaux paradoxes. Le premier fut souligné par Simon Kuznets qui, dans un fameux article de 1976²¹⁸, montre que les dépenses des ménages augmentent quand la taille des ménages augmente – alors même que les dépenses par tête du ménage diminuent. Le second paradoxe est souligné par Deaton lui-même, qui explique que malgré un taux de croissance et un revenu par habitant plus important qu'avant en Inde, le nombre de calories consommées par personne a diminué d'un quart en un siècle. Alors qu'on aurait pu supposer qu'avec un revenu plus important, les individus se nourriraient mieux (soit de façon plus calorique²¹⁹), Deaton relie cette diminution de calories à la diminution de l'activité physique des pauvres du fait de l'amélioration des transports et des techniques (notamment dans l'agriculture). Cette explication part du postulat que les individus se nourriraient pour deux raisons : pour leur survie et pour réaliser des travaux physiques. Ce serait donc la part de nutrition relative à cette deuxième raison qui trouverait, selon Deaton, sa justification via les améliorations techniques et sociales. Même s'il n'existe pas de « preuve » de l'existence de ce mécanisme et que les recherches dans ce

²¹⁸ Voir : (Kuznets, 1976).

²¹⁹ Il peut être intéressant de noter que Duflo et Banerjee trouvent exactement les mêmes résultats avec leur méthode expérimentale, on développera ce point lorsque l'on s'intéressera aux différents résultats tirés des expériences du J-PAL dans la dernière section de ce chapitre.

domaine restent embryonnaires. Ici, à l'aide d'une démarche hypothético-déductive, Deaton prend pour point de départ des résultats empiriques (un revenu plus important n'implique une augmentation de calories consommées par les pauvres). Ensuite, Deaton cherche à expliquer ce phénomène au travers d'une théorie (la diminution de l'activité physique des pauvres expliquerait que le nombre de calories qu'ils consomment n'augmentent pas lorsque leur revenu augmente). Ici, il reste donc à tester cette théorie, afin qu'elle soit ou non falsifiée. Mais selon Deaton, la relation entre les données et la théorie sont les mêmes que dans les illustrations précédentes :

« we are trying to work back from the data to a theory, not the other way round. But the interplay between theory and data is the same and it makes no difference where we join the cycle. » (Deaton, 2010b, 12)

La relation des données à la théorie, et de l'intermédiaire qui vient s'y nouer, pose ici certaines questions. Comment le processus inverse des données vers la théorie peut-il être le même que le processus de la théorie vers les données ? Deaton propose deux façons de procéder : (1) partir de la théorie puis la tester empiriquement et (2) partir de données empiriques puis essayer de les expliquer au travers d'une théorie. Mais suivant sa démarche hypothético-déductive, cette théorie reste à être testée.

1.4 L'épistémologie de Deaton en question

Deaton met en évidence un processus hypothético-déductif, associé à un processus en termes d'essais et d'erreurs à l'initiative du progrès de la science, et détache deux dimensions centrales : la démarche épistémologique choisie (1), et l'idée que la science progresse (2). La première dimension met en évidence le choix méthodologique de Deaton vis-à-vis de son appréhension de la science. Selon lui, ce n'est qu'en essayant et en échouant parfois, que l'on sera à même de mettre à jour de nouvelles connaissances. Et ces connaissances vont en s'accumulant, en s'enrichissant. Il n'y a pas de rupture épistémologique comme chez Bachelard (1934) ou de succession de paradigmes comme chez Kuhn (1962). Pour Deaton, ce processus déductif est associé à l'idée d'une continuité et d'un progrès, grâce au lien qui unit la théorie à l'empirie :

« *My main concern with current practice in development economics is that these links are weak and that much empirical work makes no attempt to investigate mechanisms. Instrumental variables and randomized trials can play a role in uncovering the mechanisms of development. (...) Randomized trials have a powerful ability to isolate one mechanism from another; in particular, an experiment will often allow us to short the circuit the often difficult process of developing theoretical mechanism to the point where they can be convincingly tested on nonexperimental data. At the same time, the routine use of instrumental variable methods and of randomized controlled trials for project evaluation is often uninformative about why the results are what they are, and in such cases, nothing is learned about mechanisms that can be applied elsewhere.* » (Deaton, 2010b, 14)

La randomisation parviendrait à isoler des mécanismes en réduisant le cadre étudié, permettant de ce fait, d'étudier un cadre plus restreint duquel est détaché tout un processus théorique (mécanisme), souvent difficile à démêler. Deaton se situe ici, dans la lignée de Cartwright, pour qui la randomisation – si elle était tournée vers la détermination d'un modèle causal –, permettrait de mettre en évidence et de définir des capacités causales. Cependant, la randomisation du J-PAL ne pose pas clairement les causes qui font qu'un programme fonctionne. Ce qui altère – comme on l'a vu au précédent chapitre – les résultats des expériences du J-PAL, dont les mécanismes sont invisibles, telle une boîte noire.

2. Le rôle ambigu de la théorie dans l'approche du J-PAL

La théorie est ce qui semble faire défaut à l'approche du J-PAL. Sa dimension a-théorique fragilise clairement la validité externe des résultats produits. Les différentes critiques adressées à la randomisation du J-PAL s'accordent toutes sur cette absence. Pour autant, l'attitude de l'approche du J-PAL vis-à-vis de la théorie me semble être plus complexe. En 2005, Banerjee participe à un symposium sur le statut de la théorie et de l'empirie en économie du développement. Il répond aux critiques adressées par des grands noms en économie du développement (Dilip Moookherjee, Pranab Bardhan, Kaushik Basu, et Ravi Kanbur)²²⁰, sur le fait que ce domaine de connaissances tend depuis les quinze dernières années à devenir de moins en moins théorique mais de plus en plus empirique. La position de Banerjee sur cette question montre que l'approche du J-PAL rejette toute théorie *a priori*, mais souhaite construire une théorie *a posteriori* sur la base des résultats des expériences randomisées. Je choisis de distinguer explicitement ces deux statuts, en me référant, pour le premier à l'absence de théorie *ex ante* en vue de définir une théorie *ex post* qui unifierait

²²⁰ Voir Bardhan (2005), Basu (2005), Kanbur (2005), Mookherjee (2005) et Banerjee (2005b).

l'ensemble des résultats obtenus par les expériences randomisées. De plus, ce défi théorique me semble posséder sa traduction politique. Banerjee et Duflo proposent une perspective locale en vue d'obtenir des effets globaux sur la lutte contre la pauvreté. Cependant, le passage politique du local au global reste obscur dans leur approche. De la même façon, la construction d'une théorie qui unifierait les différents résultats du J-PAL paraît laborieuse. Cela traduit, ici encore, la faible validité externe d'une telle approche. L'objectif est ici de présenter le statut de la théorie dans l'approche du J-PAL ainsi que ses implications politiques. Pour cela, je montrerai, tout d'abord, que le J-PAL accorde deux statuts antagonistes à la théorie (2.1), puis je traduirai la difficulté politique pour l'approche du J-PAL de passer d'un niveau local à niveau global (2.2).

2.1 Théorie *ex ante* versus théorie *ex post*

Dilip Mookherjee s'inspire de la classification établie par Haavelmo²²¹, qui liste les principaux problèmes rencontrés en économie quantitative. Cette classification établit quatre différentes étapes : la construction d'un modèle théorique (1), le test des théories (2), le problème de l'estimation (3) et la question de la prédiction (4). Mookherjee modifie légèrement cette catégorisation pour y ajouter une étape précédant l'étape 1 – et qui constituerait la définition d'une « pré-théorie » – avant de faire fusionner l'étape 2 et 3 :

« Accordingly I would modify Haavelmo's classification to add a pre-theory stage prior to stage 1, and for purposes of brevity merge the second and third stages of testing and estimation. The modified classification would then constitute the following: Stage 1: empirical description of the relevant phenomenon, consisting of exploratory data analysis aimed at helping identify empirical regularities that need to be explained by a suitable theory, and in addition the nature of assumptions that such a theory can make without gross violation to the empirical patterns. Stage 2: the formulation of a relevant theory, including derivation of potentially observable (hence falsifiable) implications. Stage 3: The testing and estimation of theories, a stage which may lead back to modification or replacement of the previous theories, in an iterative back and forth with Stage 2. Stage 4: use of the least unsuccessful theory from the standpoint of empirical verification for purposes of prediction and policy evaluation. » (Mookherjee, 2005, 4330)

La version modifiée de la classification d'Haavelmo par Mookherjee identifie alors les étapes suivantes : la construction d'une première théorie (1), la formulation d'une théorie pertinente (2), le test et l'estimation de la théorie qui peut conduire à reformuler la première

²²¹ Voir (Haavelmo, 1944)

théorie posée à l'étape 1 (3) et la formulation de recommandations politiques (4). Selon Mookherjee, les expériences randomisées en économie du développement concernent principalement l'étape 4, laissant trop souvent de côté le test et la reformulation des théories. Alors même que la randomisation pourrait permettre de tester les théories, ces expériences ne sont pour autant ni envisagées, ni implantées dans ce dessein²²². L'objectif est de passer directement à l'étape 4, c'est à dire de laisser de côté la formulation explicite d'une théorie qui serait testée, afin de réduire au maximum les biais pour avoir une évaluation fiable du paramètre d'intérêt. Pour Mookherjee, les faibles efforts accordés à l'étape 3 contribue à fragiliser l'étape 4 : « *I am not quite sure that we made sufficient progress in Stage 3 to proceed comfortably to Stage 4.* » (Mookherjee, 2005, 4333) L'absence de réel test de théorie fragilise donc les prédictions politiques provenant des expériences randomisées. Cela tend alors à affaiblir la portée des résultats de telles expériences, et donc leur validité externe. Pour Pranab Bardhan, les expériences randomisées en économie du développement « *are much too microscopic, and there is a danger of missing the « forest for the trees ».* » (Bardhan, 2005, 4334) L'accent est porté sur la quantification précise en laissant de côté la théorie, les concepts permettant de saisir les phénomènes étudiés, et de les généraliser.

Lors de ce symposium, le manque de théorie au sein des expériences randomisées est fortement relevé et critiqué²²³. D'une part, la randomisation est trop peu utilisée pour tester des théories et les reformuler. D'autre part, cela renforce la dimension trop locale de telles expériences, en fragilisant leur validité externe. La théorie permettrait donc ici d'améliorer la validité externe de la randomisation. Cependant, en s'appuyant sur le fait que les expériences ont une portée nécessairement limitée et qu'il existe potentiellement des effets d'équilibre²²⁴

²²² « *Controlled experiments of course can be used for other purposes as well, such as in testing theories, as they are used frequently in natural science. With few exceptions, the wave of recent randomised evaluations in development economics are not designed to test theories, but proceed directly to the question of policy evaluation that constitutes Stage 4.* » (Mookherjee, 2005, 4332) Pour autant, Banerjee et Duflo souligne le rôle de la randomisation pour tester les théories : « les expérimentations peuvent être utiles et ont été utiles pour tester les théories » (Banerjee et Duflo, 2009, 719). Néanmoins, peu d'expériences sont conçues dans ce but.

²²³ Kaushik Basu, quant à lui, met l'accent sur le rôle de l'intuition dans la recherche, montrant qu'il n'existe aucun moyen de démontrer l'existence d'une causalité, comme il n'y a aucune raison de croire que la causalité existe en elle même : « *there is no real way for demonstrating causality. Indeed, there is no reason to believe that there is anything objective in nature called causality.* » (Basu, 2005, 4338) L'intuition est donc, pour Basu, un guide nécessaire. De plus, Basu, souligne qu'il est très difficile, à partir d'expériences randomisées, d'offrir des recommandations politiques ou des prédictions sur le futur : « *First, it must be recognised that the new method, at least by its own criterion of what constitutes correct inference, does not help us predict the future. Put differently, suppose we apply the strictness of criterion that this new method demands of empirical work, and then ask what we can predict based on this new method? It is important to understand that the answer is « very little ».* » (Basu, 2005, 4336)

²²⁴ J'ai défini ces effets d'équilibre lors du chapitre précédent et autour de la critique d'Acemoglu (2010).

(les deux niveaux de la validité externe), Banerjee considère que la théorie ne permettra pas de contrecarrer la critique de la validité externe de la randomisation :

« To apply these results, we need to have a theory that helps us decide whether a particular location (in space, or Professor Basu emphasizes, in time) is sufficiently similar to the location from where the results came. And in most cases this theory would not come close to meeting the standards of evidence that developments economics today aspires to, and even if it did, the problem of external validity would remain, just shifted back a layer. This is obviously correct. Indeed it seems to me to be a version of David Hume's justly famous demonstration of the lack of a rational basis for induction. » (Banerjee, 2005b, 4341)

Même muni d'une théorie qui permettrait de déterminer si un résultat peut être (1) transposé à une autre région (2) ou être généralisé à une échelle plus large, la question de la validité externe (et donc de ces deux niveaux) persisterait. Pour Banerjee, cela renvoie au problème Humien²²⁵ de l'induction ; qui, en ce sens apparaît être un problème insolvable. Selon lui, le seul moyen de consolider les résultats serait de réitérer l'expérience à d'autres contextes. La réplication serait donc la solution au problème de validité externe d'une telle approche²²⁶. La réplication offre alors deux attraits pour Duflo et Banerjee. Le premier est de rendre les résultats encore plus solides, au travers de plusieurs évaluations similaires. Le second est de permettre d'apprendre des mécanismes à l'œuvre, si les résultats de la réplication sont négatifs. Deaton (2009, 2010a, 2010b) et Easterly et Cohen (2007) avancent une objection commune à propos du critère de réplication : combien de répliques seront nécessaires avant qu'un programme puisse être généralisé?²²⁷ De plus, Dani Rodrik montre que sans une théorie à tester, il apparaît difficile de déterminer dans quelles conditions les expériences doivent être répliquées :

« Absent a full theory that is being put to a test, it is somewhat arbitrary to determine under what different conditions the experiment ought to be repeated. If we do not have a theory of which X_{it}^j 's matter, we cannot know how to vary the background conditions. Moreover, everyone is free to come up with an alternative theory that would enlarge the set of conditioning variables. » (Rodrik, 2007, 38)

²²⁵ David Hume pointe le fait que la répétition d'une même observation n'a aucune force logique sur la démonstration d'un énoncé, sinon qu'elle se fonde sur l'habitude. Par exemple, il est possible d'observer que le soleil se lève chaque matin, mais il est impossible d'en déduire logiquement qu'il se lèvera demain.

²²⁶ « *That being said, the only way to build trust in experimental and quasi-experimental results is to replicate them in several different locations.* » (Banerjee, 2005b, 4342)

²²⁷ Une seconde objection, cette fois-ci non plus épistémologique mais plus sociologique et déjà soulevée (Partie 1 – chapitre 2), est liée au biais de publication. Les revues scientifiques souhaitent souvent publier des résultats innovants et originaux ; les chercheurs ne sont pas incités à répliquer ces expériences.

La théorie semble donc avoir un rôle important à jouer. En corrélant la recherche de facteurs causaux sous-jacents aux résultats des expériences, la théorie permettrait d'améliorer la validité externe de telles expériences. Pour autant, il me semble que Banerjee et Duflo n'excluent pas totalement la théorie, lui conférant un double statut. Ils considèrent tout d'abord que la théorie peut être un mauvais guide :

« Ces études font clairement apparaître que notre intuition est un bien mauvais guide (ou la théorie économique en tant que telle) du choix entre des programmes concurrents qui visent le même objectif : comment aurions-nous pu deviner *a priori* que le traitement contre les parasites intestinaux est beaucoup plus efficace que le recrutement d'un professeur supplémentaire ? » (Banerjee et Duflo, 2009, 693)

Banerjee et Duflo semblent donc rejeter toute théorie *a priori*, soit toute théorie pensée en amont, avant l'expérience. L'objectif est que les expériences randomisées fassent émerger la théorie. En d'autres termes, c'est à partir des expériences qu'une théorisation est possible. Ils reconnaissent que leur principal défi est de transformer les résultats tirés de leurs expériences en théorie afin de leur offrir un cadre général :

« *the challenge is to turn these explanations into a theory – something that would have implications for other things that we care about and pose new questions for empirical research* ». (Banerjee, 2005b, 39)

Je distingue explicitement ces deux statuts accordés à la théorie, alors que Deaton a tendance à les envisager simultanément²²⁸. J'appelle le premier statut accordé à la théorie, c'est-à-dire son absence *a priori* : une absence de théorie *ex ante*. Et je définis son second statut, c'est-à-dire la volonté de créer une théorie à partir des résultats obtenus : une volonté de « théorie *ex post* ». La distinction de ces deux statuts accordés à la théorie permet de différencier deux étapes au sein de l'approche du J-PAL. La première est de collecter des résultats sans connaissance ou théorie *a priori*. La seconde étape est de constituer, à travers ces résultats, un cadre théorique capable de renouveler l'économie du développement. Autrement dit, la théorie *ex ante* renvoie à l'étape 1 défini par Mookherjee, et la théorie *ex post*, à l'étape 3. A l'inverse de Mookherjee qui regrette que trop peu d'efforts soient investis dans l'étape 3, il me semble que le manque d'efforts placés dans l'étape 3, provient du fait

²²⁸ Deaton utilise pourtant une fois le terme de théorie préalable (« *prior theory* » dans Deaton, 2009, 30), mais ne distingue cependant jamais explicitement les deux. De plus, Deaton centre principalement son analyse sur l'étude des mécanismes sans définir clairement ce qu'il entend par théorie.

que Banerjee et Duflo se refuse à toutes théorie *ex-ante* et donc peinent à l'étape 1. Ce que, d'une certaine manière, Banerjee soulève :

« *in Haavelmo's terms (...) we are back at stage 1: We are gathering impressions that would eventually allow us to build a set of new theories that, one day, will define a "newer" development economics.* »
(Banerjee, 2005b, 39)

Le principal défi de Banerjee et Duflo est donc de parvenir, à partir de leurs résultats, à envisager un cadre théorique cohérent. Autrement dit, d'essayer – sans avoir à procéder à l'étape 1, 2, et 3, mais uniquement à l'étape 4 – de créer une étape 5 qui serait la construction émergeant *a posteriori*. Pour autant, cette « théorie *ex post* » n'apparaît toujours pas très claire. Elle reste suspendue à une meilleure compréhension de son cadre théorique *ex ante*.

2.2 Du local ou global : quel passage (au) politique ?

Ce défi théorique implique un défi politique : l'impact des programmes locaux sur un développement plus global. Banerjee et Duflo envisagent une perspective uniquement locale, mais leur démarche vise à lutter contre la pauvreté globale²²⁹. Autrement dit, l'évaluation randomisée de programmes locaux vise une plus large échelle. Banerjee et Duflo – après avoir d'abord envisagé un simple impact local – prônent désormais un changement de perspective : passer des « Institutions » aux « institutions » :

« (...) les détails sont essentiels. Les institutions ne font pas exception à ce principe. Pour vraiment comprendre l'effet des institutions sur la vie des pauvres, il faut passer de l'examen des Institutions avec majuscule à celui des institutions avec minuscule : il faut adopter la 'perspective d'en bas' ». (Banerjee et Duflo, 2011, 366)

Ce changement de perspective doit s'opposer à ce qu'ils nomment l'« économie politique », qui assure la primauté du politique sur l'économie²³⁰. Cette vision serait prédominante en économie du développement. Cependant, Banerjee et Duflo ciblent explicitement le travail de Daron Acemoglu et de Jonathan Robinson²³¹, pour lesquels il est nécessaire de saisir l'histoire d'un pays et de ses institutions avant de comprendre son développement. De ce fait, selon eux, seules des transformations politiques profondes peuvent permettre ce développement. La révolution française aurait, par exemple, permis de tels changements structuraux, en retirant le pouvoir d'une élite. Pour Banerjee et Duflo, cette position est pessimiste, ne laissant place qu'à des changements d'envergures, que seule une révolution pourrait apporter :

« Le pessimisme de Daron Acemoglu et James Robinson vient notamment du fait que les changements de régime radicaux sont rares, que l'on passe rarement d'un régime autoritaire et corrompu à une démocratie au fonctionnement satisfaisant. En adoptant la perspective d'en bas, la première chose dont on s'aperçoit est qu'il n'est pas toujours nécessaire de changer radicalement les institutions pour que les dirigeants soient davantage tenus de rendre des comptes à la population, et pour limiter la corruption. » (Banerjee et Duflo, 2011, 366)

²²⁹ Comme en témoigne, par exemple, le titre de leur dernier livre : *Poor Economics, A Radical Rethinking of the Way to Fight Global Poverty*, (Banerjee et Duflo, 2011).

²³⁰ « L'« économie politique » est la vision (partagée, comme nous l'avons vu, par de nombreux spécialistes du développement) selon laquelle la politique prime sur l'économie : les institutions définissent et limitent la portée des politiques économiques. » (Banerjee et Duflo, 2011, 381).

²³¹ Voir par exemple, (Acemoglu et Robinson, 2012 ; 2006) et (Acemoglu, Johnson et Robinson, 2001).

A l'inverse d'Acemoglu et de Robinson, la perspective du J-PAL, est de montrer que des changements à la marge peuvent avoir d'importants effets, et qu'à partir d'un changement – même mineur – un processus vers le développement peut émerger. Par exemple, il est possible selon eux, de lutter contre la corruption sans pour autant transformer les institutions existantes. Ils reprennent ici une expérience d'un autre chercheur du J-PAL (Olken, 2007). Cette expérience cherche à évaluer l'effet d'un audit sur la corruption, et montre que la menace d'un tel audit diminue d'un tiers le vol de salaires et de matériaux par rapport aux villages qui n'ont pas subi cette menace. Duflo et Banerjee s'appuient aussi sur une de leurs expériences réalisées en Inde²³² cherchant à évaluer l'effet d'un contrôle sur la police. Selon eux, la police indienne est l'illustration parfaite d'une institution coloniale qui perdure encore aujourd'hui. Cependant, ce système n'apparaît pas si fermé pour Duflo et Banerjee. Afin de mettre en évidence cela, l'expérience organise des visites surprises par des observateurs (des habitants de la région participant à l'expérience), chargés de révéler à la police indienne leurs objectifs : chercher à enregistrer des infractions provenant de la police. Les résultats de cette expérience montrent que la peur d'avoir à faire à des observateurs a permis d'améliorer le travail de la police dans cette région. Ici, un changement à la marge à améliorer le travail de la police indienne, mais ne l'a pas pour autant transformé ni détaché de son passé colonial.

Banerjee et Duflo proposent de ne plus s'intéresser aux « Institutions » avec majuscule, mais aux « institutions » en lettres minuscules. D'une part, comme cela a été soulevé, car des changements à la marge peuvent avoir d'importants effets ; et d'autre part, car les institutions globales ne sont en fait constituées que d'institutions locales, pour Banerjee et Duflo²³³. J'ai cherché à regrouper les arguments de Banerjee et de Duflo qui incitent à modifier les institutions au niveau local – et au niveau local seulement –, et j'ai identifié six raisons principales : il n'est pas nécessaire de changer les institutions pour améliorer la vie des pauvres (1) ; il n'y a aucune garantie à ce qu'une « bonne institution » fonctionne bien (2) ; un petit changement peut avoir des effets importants (3), de « bonnes » institutions peuvent bien fonctionner dans un environnement politique « mauvais » (Banerjee et Duflo, 2011, 381) (4) ; de « mauvaises » institutions peuvent exister dans un « bon » environnement politique (5) ; il ne suffit pas d'être animé de « bonnes » intentions pour qu'une politique soit efficace (6). L'argument fondateur de ce changement de perspective tient donc au fait que,

²³² Voir (Banerjee, Chattopadhyay, Duflo, Keniston et Sinsh, 2012).

²³³ « Mais chaque Institution de ce type est incarnée, sur le terrain, par une multitude d'*institutions* locales particulières. » (Banerjee et Duflo, 2011, 365)

pour Banerjee et Duflo, il n'y a pas de corrélation entre un « bon » environnement politique et de « bonnes » institutions. Même dans un « mauvais » environnement politique, de « bonnes » institutions peuvent bien fonctionner, et inversement :

« De bonnes politiques publiques existent (parfois) dans de mauvais contextes. Et à l'inverse – et c'est peut être le plus important – des politiques déplorables peuvent (souvent) exister dans des environnements tout à fait acceptables. » (Banerjee et Duflo, 2011, 381-382)

Le principal problème d'une telle affirmation est qu'elle est nécessairement suspendue à ce que l'on entend par institution et par un « bon » environnement politique. Cependant, Banerjee et Duflo ne définissent à aucun moment ce qu'ils entendent par institution et par un « bon » ou « mauvais » environnement politique. Le passage entre « institutions » et « Institutions », est alors difficilement pensable sans une définition claire de ces dernières. En effet, de « bonnes » institutions locales mises en place grâce à la randomisation permettront-elles à de « bonnes » institutions globales d'émerger ? Assistera-t-on à ce passage du local au global ? Et, si tel est le cas, comment cela se produira-t-il ? On retrouve ici la traduction en termes politiques, de leur ambiguïté théorique. En n'envisageant aucun cadre de départ, il apparaît difficile d'entrevoir comment Banerjee et Duflo donnent une portée politique plus générale à l'évaluation de programmes de développement.

3. Les résultats du J-PAL : les effets versus les causes de la pauvreté

La problématique soulevée par l'ambiguïté du statut de la théorie dans l'approche de Duflo et Banerjee ne me semble pas concerner uniquement la généralisation de leurs résultats. Le choix de ne pas concevoir une théorie *a priori* me semble être, lui aussi, problématique dans l'utilisation des « preuves » données par la randomisation. La distinction entre une absence de théorie *ex ante* et une absence de théorie *ex post*, est ici particulièrement éclairante. Elle me permet de penser simultanément les deux niveaux de la validité externe. En ce sens, elle me permet de montrer, à travers les différents résultats du J-PAL, la difficulté pour une telle approche de produire des recommandations politiques claires. Je commencerai par présenter les résultats obtenus par le J-PAL (3.1), puis je tenterai de montrer que l'approche de Duflo et Banerjee propose certes une description extrêmement détaillée et rigoureuse de la vie des pauvres, mais peine à prescrire des recommandations politiques qui orienteraient l'aide au développement (3.2). Cela me permettra de montrer que l'approche du J-PAL peine à réaliser le second objectif qu'elle se fixe : guider la décision politique.

3.1 Les expériences du J-PAL : un « panorama » de la vie des pauvres

De nombreuses évaluations randomisées ont eu lieu ces dix dernières années. Cela a été noté, le J-PAL compte à ce jour plus de 305 évaluations en cours ou réalisées. Ces évaluations traitent de sujets différents et proposent des « preuves scientifiques » différentes. Dans leur dernier livre, Abhijit Banerjee et Esther Duflo (Banerjee et Duflo, 2011) tentent de donner un « panorama » de ces différents résultats²³⁴. Ils regroupent ces résultats autour de sept domaines : la faim, la santé, l'éducation, la population, le microcrédit, l'épargne et l'entrepreneuriat²³⁵.

En qui concerne la faim, comme on l'a vu autour de la critique de Deaton, Banerjee et Duflo partent du postulat qu'être pauvre c'est mourir de faim – la malnutrition étant souvent considérée comme un des phénomènes les plus importants dans les trappes à pauvreté. Cependant, Banerjee et Duflo montrent que les pauvres ne sont pas affamés, mais qu'ils mangent peu. Il n'y aurait donc pas de trappes à pauvreté dans ce cas. De plus, autour d'une expérience de Jensen et Miller (2008), on apprend que les pauvres ne privilégient pas les aliments caloriques qui leur fourniraient les apports nutritifs nécessaires à une meilleure productivité, préférant le goût des aliments. Cette explication alternative à celle proposée par Deaton permet à Banerjee et Duflo, d'établir une recommandation politique : l'urgence de repenser les politiques de la faim en se centrant davantage sur la qualité de la nourriture que sur sa quantité. En matière de santé, Banerjee et Duflo se demandent pourquoi les pauvres dépensent beaucoup d'argent dans des soins trop tardifs. Sur ce point, ils nuancent le point de vue de Jeffrey Sachs, en montrant que, si une trappe à pauvreté existe bel et bien, lorsque l'on donne aux pauvres la possibilité de s'en sortir, ils ne savent pas comment l'utiliser. Ils recommandent alors des dispositifs qui inciteraient les pauvres à dépenser en amont en matière de santé, et pointent la nécessité d'une politique publique plus générale²³⁶.

²³⁴Ce livre représente presque l'unique tentative des chercheurs du J-PAL à donner une vue d'ensemble de leurs résultats. La seule autre tentative est celle de Jacob Appel et Dean Karlan, voir (Appel et Karlan, 2011). Ce livre reprend un des articles de Banerjee et de Duflo (2007) qui cherchait aussi à rendre compte d'un tel panorama, ce livre s'inspire aussi de Banerjee, Benabou et Mookherjee (2006).

²³⁵ La classification en sept domaines offerte par Banerjee et Duflo se distingue légèrement de celle qu'offre le J-PAL pour classer ses publications académiques (partie 1 - chapitre 2), puisque Banerjee et Duflo ne traitent pas ici d'agriculture, ni de gouvernances, ni du marché du travail, ou encore de développement durable.

²³⁶ Je reviendrai très largement sur le point de vue de Banerjee et de Duflo en matière de santé dans la partie suivante de ce travail, puisqu'un chapitre entier sera consacré aux expériences de Pascaline Dupas sur le paludisme et qu'un autre est dédié à l'étude précise des recommandations politiques de Duflo dans ce domaine.

Concernant l'éducation, Banerjee et Duflo montrent que si les enfants apprennent mieux dans les écoles privées que dans les écoles publiques²³⁷, les parents envoient tous leurs enfants à l'école, alors qu'ils auraient tout intérêt à en cibler un. De plus, les parents ont tendance à douter des chances de la réussite scolaire de leur enfant, ce qui crée un piège à pauvreté :

« Les erreurs de perception peuvent avoir des conséquences dramatiques. En réalité, il ne devrait pas y avoir de piège de pauvreté lié à l'éducation : l'éducation a une valeur à quelque niveau que ce soit. Mais le fait que les parents croient que les bénéfices de l'éducation suivent une courbe en S les conduit à agir comme s'il y avait un piège de pauvreté et, par conséquent, à en créer un sans s'en rendre compte. » (Banerjee et Duflo, 2011, 149-150)

A ce titre, Banerjee et Duflo proposent d'envisager des politiques qui permettraient de renverser ces anticipations et de favoriser l'utilisation de nouvelles technologies. Lorsqu'ils s'intéressent à la fécondité des pauvres ainsi qu'à la taille des familles, ils partent du constat que les pauvres ont beaucoup d'enfants²³⁸, et montrent que cela représente un choix conscient et volontaire – les enfants étant souvent considérés comme une source de revenus potentiels et comme une assurance future. Ils recommandent alors des politiques qui permettraient d'alléger la nécessité pour les familles d'avoir beaucoup d'enfants ; en préconisant par exemple, l'implantation d'un filet de sécurité sociale²³⁹. En ce qui concerne l'entrepreneuriat des pauvres, Banerjee et Duflo montrent, qu'à l'inverse de ce que prône Muhammad Yunus, les pauvres ne sont pas des « entrepreneurs naturels » (Banerjee et Duflo, 2011, 316). Il leur est difficile, voire impossible, de transformer leurs mini-entreprises en entreprises prospères. C'est pourquoi, selon eux, les gouvernements devraient se substituer au marché afin de fournir une assurance aux pauvres²⁴⁰, leur permettant de prendre plus de risques et ainsi de faire prospérer leurs entreprises²⁴¹. De plus, Banerjee et Duflo affirment que si les pauvres se dirigent vers ce modèle d'entreprise, c'est d'abord parce qu'il n'existe pas d'alternative. Selon eux, les pauvres préféreraient posséder un emploi stable (par exemple être fonctionnaire), plutôt que de gérer une entreprise. Sur ce point, Banerjee et Duflo suggèrent que les gouvernements des pays en développement mènent des politiques permettant de créer des emplois stables.

²³⁷ Voir (Banerjee, Cole, Duflo, et Linden, 2007). Les élèves tendent à mieux apprendre dans les écoles privées que publiques, car au sein de ces dernières les enseignants sont fréquemment absents ; voir, par exemple, (Banerjee et Duflo, 2006), (Duflo, Dupas et Kremer, 2011b ; 2012).

²³⁸ Beaucoup de politiques de développement, de façon générale, visent à réduire la fécondité des pauvres.

²³⁹ Le rôle de l'éducation est aussi mis en avant, voir, par exemple, (Duflo et Breierova, 2004)

²⁴⁰ Voir (Duflo et Udry, 2004).

²⁴¹ Cette position est aussi développée dans (Duflo, 2006b)

Concernant le microcrédit, Duflo et Banerjee expliquent que peu d'évaluations ont eu lieu dans le but de mesurer ses effets. Le J-PAL a donc mené plusieurs expériences en vue de rendre compte des effets du microcrédit²⁴². Au vu des résultats obtenus, Banerjee et Duflo estiment que le microcrédit ne permet aucune grande transformation en matière de lutte contre la pauvreté. Ils se demandent alors si le processus pour sortir de la pauvreté ne serait pas de pouvoir emprunter au travers de microcrédits dans un premier temps, avant d'accéder par la suite, à une banque normale²⁴³. En s'intéressant à l'épargne des pauvres, Banerjee et Duflo montrent que s'il est vrai que les pauvres épargnent de façon extrêmement ingénieuse, soit « brique par brique », ces derniers épargnent bien trop peu, ce qui tend à créer une trappe à pauvreté. Ici, Banerjee et Duflo mettent en évidence le rôle de l'espoir et du désespoir dans le processus d'épargne des pauvres²⁴⁴. Là encore, la création d'un filet de sécurité sociale semble être une des solutions politiques pour favoriser l'épargne des pauvres. Enfin, Banerjee et Duflo insistent sur le poids croissant que la micro-épargne devrait prendre au sein de la micro-finance.

Les résultats présentés par Banerjee et Duflo semblent éclatés. En les classifiant par domaine, Duflo et Banerjee n'offrent pas de théorie d'ensemble, et aucune théorie *ex post*. Comme le soulignait Mookherjee, les expériences ainsi que leurs résultats s'apparentent à l'étape 4, mais les étapes précédentes semblent absentes. Et ce, même si comme j'ai essayé de le montrer, Banerjee et Duflo prennent pour point de départ ce que la littérature dominante en économie du développement a à dire. En ce sens, ils opposent souvent deux points de vue majeurs, sans se positionner. C'est au travers de l'expérience que leur position doit se former. Cependant, les différents résultats qu'ils exposent ne semblent pas leur permettre de produire une théorie unifiée. L'absence de théorie *ex ante* me semble fragiliser très fortement la possibilité d'une théorie *ex post*. Pour reprendre les termes de Mookherjee : sans la construction de l'étape 1, la formulation d'une première théorie qui sera ensuite testée (étape 3) ; il semble alors difficile d'apprendre des résultats expérimentaux, car aucune direction précise n'est définie. La construction d'une théorie *ex post* qui permettrait à Banerjee et Duflo

²⁴² Voir, par exemple, (Crépon, Devoto, Duflo et Pariente, 2011), (Banerjee, Duflo, Glennerster et Kinnan, 2010) et (Banerjee, Chandrasekhar, Duflo et Jackson, 2011).

²⁴³ Pour une analyse du crédit dans les pays en développement, voir (Banerjee et Duflo, 2010).

²⁴⁴ Je reviendrai très largement, dans le dernier chapitre, sur ce rôle de l'espoir et du désespoir dans la lutte contre la pauvreté chez Esther Duflo.

de définir une nouvelle économie du développement reste ainsi suspendue à la formulation d'une théorie *ex ante*.

3.2 La pauvreté selon le J-PAL : des effets sans explications causales ?

Au sein de quatre des sept domaines définis par Duflo et Banerjee, les recommandations politiques tirées par les auteurs sont d'ordre global et non local. En ce qui concerne la faim, l'épargne, la famille et l'entrepreneuriat, Banerjee et Duflo recommandent la création d'un filet de sécurité sociale. De plus, en matière d'entrepreneuriat, Banerjee et Duflo ajoutent une recommandation politique tout aussi globale, stipulant la création d'emplois par les gouvernements des pays en développement. Ces recommandations contredisent, d'une certaine manière, le positionnement initial de Banerjee et Duflo contre l'économie politique et la dominance des institutions sur l'économie, puisqu'ici Banerjee et Duflo en appellent finalement à un changement plus profond qu'à un changement à la marge.

En matière de santé, d'éducation, et concernant le microcrédit, les recommandations de Banerjee et de Duflo sont plus locales, même si elles ne sont pas explicitement définies. Au sujet de la santé, Banerjee et Duflo envisagent des dispositifs incitatifs, sans que ces derniers ne soient clairement explicités. Il en va de même pour l'éducation, domaine pour lequel Banerjee et Duflo proposent de renverser les anticipations des agents, mais sans expliquer comment. Enfin concernant le microcrédit, la recommandation de Banerjee et de Duflo s'apparente plus à un questionnement : existe-il un processus naturel qui assurerait un financement par microcrédit comme préalable à l'accès au circuit bancaire ordinaire ? Mais aucun élément de réponse n'est donné à cette question. Même l'existence d'un tel processus naturel n'est pas testée. En somme, les recommandations politiques de Banerjee et Duflo au niveau local ne sont pas claires. Les recommandations politiques d'ordre global sont, quant à elles, beaucoup plus évidentes, mais contredisent la perspective générale adoptée par Banerjee et Duflo. De plus, en l'absence de tests réalisés via les expériences randomisées, ces recommandations ne découlent pas de la mise en évidence d'un mécanisme. Elles ne sont pas le fruit d'un raisonnement théorique, et émergent de façon impromptue ; ce qui tend à fragiliser la perspective initiale du J-PAL.

Comment expliquer l'opacité des recommandations politiques offertes par Banerjee et Duflo ? L'explication de cette opacité me semble relever de la faible validité externe d'une

telle approche, et ce à un double niveau. Le premier niveau que l'analyse de Cartwright m'a permis de mettre en évidence relève de l'utilisation politique d'un résultat d'une situation donnée à une autre, et le second niveau considère le passage à une échelle plus large. Le second niveau traduit le passage politique du local au global, mais les recommandations politiques tirées des expériences du J-PAL ne permettent pas de rendre compte des effets d'équilibre potentiels ou des facteurs politiques sous-jacents au moment de l'extension d'un programme. Comme le soulignait Deaton, les expériences du J-PAL ne sont pas tournées vers la mise en évidence de mécanismes sous-jacents aux résultats qu'elles produisent. Comme ces expériences ne s'intéressent pas aux causes qui font qu'un programme fonctionne, il apparaît difficile de penser la généralisation d'un programme à une population plus large. De plus, le fait que Banerjee et Duflo proposent majoritairement des politiques d'ordre global accentue l'opacité du passage entre politique locale et politique globale.

Pour autant, ces recommandations politiques sont-elles utilisables dans un autre contexte, comme le questionne Cartwright ? Il s'agit ici du premier niveau de la validité interne. Les recommandations d'ordre local offertes par Duflo et Banerjee, à l'instar de celles proposant par exemple de renverser les anticipations en matière d'éducation, ont montré des effets positifs à Madagascar et en République Dominicaine – soit au sein de deux contextes distincts. Cependant, Banerjee et Duflo n'identifient dans ce cas aucune capacité causale. Et on ne peut savoir si cela fonctionnera aussi dans un énième contexte. La question de Cartwright (2010a) « *Will it Work for us ?* »²⁴⁵ persiste. Concernant le microcrédit, aucune recommandation politique n'apparaît clairement. Pour ce qui relève de la santé, l'efficacité des dispositifs testés, on le verra, est faible; fragilisant là aussi toute éventuelle recommandation politique, et son application dans un autre contexte. Tout comme le premier niveau de la validité externe, l'utilisation politique des résultats des expériences randomisées semble faible compte tenu de la non-explicitation des capacités causales à l'œuvre.

La randomisation du J-PAL, en se détournant de la recherche de capacités causales ou de mécanismes sous-jacents à ces résultats, ne permet pas – à partir de ses expériences – d'offrir des recommandations politiques claires. Les résultats du J-PAL apparaissent alors – dans leur

²⁴⁵ Cette question est le titre d'un article de Cartwright (2010a), Cartwright pose une question similaire dans (Cartwright, 2010c ; 2011c).

majorité²⁴⁶ – telle une description précise et « scientifique de la vie des pauvres » qui peine à produire des recommandations politiques claires. En ce sens, ces résultats tendent à rendre compte des effets de la pauvreté et non de leurs causes. Comme par exemple, lorsque Banerjee et Duflo s'intéressent à la faim et qu'ils décrivent précisément la vie des pauvres, montrant que ces derniers préfèrent privilégier le goût des aliments à leur apport calorique. Ou lorsqu'ils expliquent qu'une augmentation du revenu des pauvres financerait davantage l'achat d'une télévision que l'achat de nourriture²⁴⁷. Dans ce cas, Duflo et Banerjee mettent en évidence les effets de la pauvreté, et non la cause de cette pauvreté. Toute recommandation qui en découle demeure donc nécessairement obscure, puisqu'on ne sait pas sur quel mécanisme agir. A l'instar du chapitre précédent, on observe ici que les résultats des expériences du J-PAL se présentent comme une boîte noire. La faute en revient à l'absence de théorie *ex ante*, qui fragilise la possibilité d'une théorie *ex post* et empêche la mise en évidence de capacités causales ou de mécanismes.

La distinction entre théorie *ex ante* et théorie *ex post*, me permet de traduire clairement le statut de la théorie dans l'approche de Duflo et de Banerjee. D'une part, elle se retrouve pleinement dans les différentes étapes traduites par Mookherjee, et permet donc de séparer l'avant et l'après des expériences. D'autre part, la distinction permet de rapprocher les critiques de Cartwright et Deaton, cela me permet de rendre compte des deux niveaux de la validité externe simultanément. Cartwright comme Deaton reprochent à la randomisation de ne pas se demander pourquoi un programme fonctionne. Selon eux, la réponse à cette question se trouve dans la recherche et la mise en évidence des mécanismes ou des capacités causales sous-jacentes aux résultats. Il me semble difficile de rendre compte de ces mécanismes chez Deaton, sans formuler en amont certaines hypothèses, sans stipuler une première théorie : une théorie *ex ante*. Ce que, d'une certaine manière, le procédé hypothético-déductif suggéré par Deaton soulève. Cartwright redoutait que les phénomènes

²⁴⁶ Il est évident que certains résultats du J-PAL parviennent à fournir des recommandations politiques précises, comme cela peut être le cas pour l'expérience de Miguel et Kremer (2004) sur les médicaments contre les vers intestinaux. Le J-PAL ayant réalisé plus de 300 expériences, mon analyse ne peut tenir compte de l'ensemble de ces expériences. C'est pourquoi je ne peux traiter des expériences au cas par cas. Ma réflexion porte donc sur la majorité de ces expériences, mais n'excluent pas le fait que certaines expériences parviennent à fournir des recommandations politiques claires. L'idée importante étant ici, que même s'il existe des cas particuliers, l'approche du J-PAL peine à transformer ces résultats en recommandations politiques. Ce que je cherche à expliciter c'est la difficulté que les chercheurs du J-PAL ont à formuler des recommandations politiques locales précises et définies.

²⁴⁷ « Plus généralement, nous constatons souvent que, pour les pauvres, les choses qui rendent la vie moins ennuyeuse sont une priorité. Ce peut être la télévision ou quelque chose de bon à manger – ou juste une tasse de thé sucré. » (Banerjee et Duflo, 2011, 70)

sociaux soient trop holistiques pour pouvoir en déduire des capacités causales. La formulation d'une théorie *ex ante* permettrait de décrire en amont la dimension holistique des phénomènes sociaux, ce qui de ce fait, permettrait de tester à travers la randomisation, l'existence d'une capacité causale particulière. Les critiques adressées à la randomisation relèveraient donc, selon moi, du refus de définir une théorie *ex ante*. Celle-ci tendrait pourtant à améliorer la validité interne et externe de la randomisation. Une théorie *ex ante*, vouée à être reformulée, augmentée et améliorée, permettrait d'identifier les raisons de l'efficacité d'un programme particulier, et offrirait ainsi un moyen d'ouvrir la boîte noire des résultats produits par le J-PAL. Cela permettrait aux chercheurs du J-PAL de proposer des recommandations politiques plus claires, puisque la portée des résultats aurait été envisagée en amont. De plus, cela faciliterait la construction d'une théorie *ex post*, souhaitée par Banerjee et Duflo, éclaircissant le passage entre des politiques locales et des politiques globales.

4. Conclusion

Le développement d'une démarche hypothético-déductive, souhaitée par Deaton, pose très clairement les enjeux du rôle de la théorie au sein des expériences du J-PAL. D'une part, car Deaton montre que, sans hypothèses *a priori*, sans tests et sans reformulation, il apparaît difficile d'apprendre des expérimentations du J-PAL. D'autre part, tout en positionnant Deaton face à Cartwright, j'ai montré que l'approche de Deaton pointe la nécessité pour le J-PAL d'apprendre des mécanismes en vue de proposer des recommandations politiques plus générales. Développer la démarche proposée par Deaton m'a permis de poser les enjeux d'une absence de théorie et a offert un cadre d'analyse de départ éclairant. Pour autant, le statut de la théorie dans l'approche de Banerjee et Duflo ne peut se résumer à sa simple absence. Le statut accordé à la théorie dans une telle approche est ambigu. Si Banerjee et Duflo se refusent à toute théorie pensée en amont – compliquant, de ce fait, le test d'une première théorie ainsi que sa reformulation – ils pointent cependant la dimension essentielle de la théorie dans leur approche, en en faisant même l'un de leur défis principaux. Construire une théorie à partir des expériences et des résultats fait ainsi parti de leurs objectifs. Pour autant, la construction de cette théorie *ex post* apparaît difficilement envisageable sans une théorie pensée en amont. Ce défi théorique porte en lui la traduction du défi politique consistant à passer des recommandations politiques locales à des recommandations plus globales. Or, ce passage reste lui aussi suspendu à la création d'une théorie *ex post*.

A défaut, les résultats du J-PAL demeurent éclatés, difficiles à unifier. D'une part, il apparaît compliqué, à partir de ces derniers, de proposer des recommandations politiques d'ordre locales. Ces résultats s'apparentant plus à une description de la vie des pauvres, qu'à de réelles recommandations politiques. D'autre part, il semble encore plus compliqué, en l'absence de recommandations politiques locales, de penser des politiques plus générales. En ce sens, ces résultats traduisent majoritairement les effets de la pauvreté à défaut de ses causes. Le refus d'une théorie *ex ante* me semble être la principale explication à cela. L'ambiguïté du rôle accordé à la théorie dans l'approche du J-PAL permet donc de traduire les deux niveaux de la fragilité de la validité externe d'une telle approche. La distinction entre une théorie *ex ante* et une théorie *ex post* permet de penser simultanément la mise en évidence de capacités causales et des mécanismes visant à penser l'implantation à une échelle plus importante de programmes de développement.

Conclusion de la Partie 2.

L'analyse de Cartwright sur la randomisation m'a permis de définir un cadre épistémologique permettant de questionner la validité externe d'une telle méthode. Cette analyse met donc en évidence un premier niveau celui de l'utilisation des preuves fournis par les expériences par le J-PAL. Cette utilisation semble, encore, impossible car l'approche du J-PAL ne permet pas de mettre en évidence des capacités causales. En ce sens, il semble difficile de passer du contexte de l'expérience à un autre contexte. Le premier niveau de la validité externe d'une telle approche en est donc affaibli. Les différentes critiques adressées à la randomisation en économie s'attachent particulièrement au second niveau de la validité externe, questionnant la généralisation possible des programmes de développement évalués. Ces critiques mettent en évidence la tension entre la forte validité interne de la randomisation du J-PAL et sa faible validité externe, pressant les chercheurs du J-PAL à s'intéresser à une perspective plus générale et donc à stipuler certaines hypothèses. Sans cela, la validité externe d'une telle approche est vouée à rester fragile. L'ensemble de ces critiques, que ce soit celle de Cartwright ou celles des économistes du développement, s'accorde sur la dimension a-théorique de l'approche du J-PAL. C'est donc le rapport d'une telle approche à la théorie, qui me semble permettre d'englober toutes ces critiques. De plus ce rapport permet de penser simultanément les deux niveaux de la validité externe.

C'est pourquoi je me suis attachée à questionner ce rapport. La démarche hypothético-déductive de Deaton permet de pointer l'aller-retour que la théorie et que l'empirie doivent opérer afin de mettre en évidence les mécanismes à l'œuvre en économie du développement. Et en ce sens, cela met en évidence la nécessité pour l'approche du J-PAL de théoriser, ou du moins d'accorder un rôle à la théorie au sein de son approche. La démarche de Deaton m'a aussi permis de faire suite à l'analyse de Cartwright. Pour autant, il me semble que le statut accordé à la théorie dans l'approche du J-PAL est ambigu. Il ne se résume pas à une simple absence. Le J-PAL refuse de poser une théorie *ex-ante* mais entend créer une théorie *ex-post*. Mais en se refusant à toute théorie *ex-ante*, le J-PAL peine à produire une théorie *ex-post*. De plus, l'absence de cette théorie *ex-ante* fragilise aussi bien l'utilisation politique que la généralisation des résultats du J-PAL. Ce refus de théorie *ex-ante* s'explique par la volonté des chercheurs du J-PAL à ne pas perdre en validité interne. Mais cela leur fait perdre en validité externe. C'est ce refus de théorie *ex-ante* qui explique la tension persistante entre validité interne et externe au sein de l'approche du J-PAL. Cette tension entre validité interne

et externe ou refus de théorie *ex-ante* représente pour moi une faille épistémologique, empêchant les chercheurs du J-PAL de réaliser le second objectif qu'ils se fixent : guider la décision politique. En effet, une telle approche peine à rendre compte des causes de la pauvreté, elle peine à comprendre « pourquoi un fonctionne ou ne fonctionne pas ». Cela tend alors à rendre les recommandations politiques particulièrement floues. Néanmoins, leur approche traduit très précisément les effets de la pauvreté, et rend donc visible des faces de la pauvreté auparavant invisibles. Cela tend à donner une nouvelle direction, une nouvelle appréhension de la pauvreté extrêmement novatrice.

Partie 3. De l'échec de la causalité à un renouveau du paternalisme : les apports théoriques à l'économie du développement du J-PAL

Introduction de la Partie 3.

« Easterly ne voit pas comment des « experts » occidentaux pourraient juger que tel ensemble d'institutions est nécessairement bon ou mauvais dans tel contexte particulier. Quant à Sachs, il estime que les mauvaises institutions sont la maladie des pays pauvres, mais qu'on peut lutter efficacement – jusqu'à un certain point – contre la pauvreté même dans des environnements institutionnels défaillants, en s'attachant à mettre en œuvre des programmes concrets et mesurables. (...) Nous sommes d'accord avec l'un et l'autre : on a tort de se focaliser à ce point sur les grandes Institutions (...). Une compréhension précise des motivations et des contraintes de chacun (pauvres, fonctionnaires, contribuables, élus, etc.) peut conduire à mieux concevoir tant les politiques que les institutions (...). De tels changements ne peuvent être que progressifs, mais ils perdurent et s'additionnent les uns aux autres, et pourraient bien être à l'origine d'une révolution tranquille. » (Banerjee et Duflo, 2011, 397)

Banerjee et Duflo souhaitent proposer une « révolution tranquille ». Les résultats produits par les expériences randomisées en s'accumulant permettraient de comprendre les contraintes qui pèsent sur les différents acteurs de la lutte contre la pauvreté, et ce faisant d'envisager des politiques efficaces. Banerjee et Duflo se positionnent explicitement face à Jeffrey Sachs et William Easterly, les principaux tenants du débat sur l'aide au développement. Le premier étant en faveur d'une aide massive, alors que le second s'attache à en démontrer les méfaits. Ce débat représente la principale référence théorique face à laquelle Duflo et Banerjee souhaitent se situer et offrir une alternative. L'objectif de cette partie est d'évaluer l'apport théorique de l'approche du J-PAL à l'économie du développement. Cette dernière est considérée comme ayant transformé l'économie du développement, insufflant une nouvelle direction de recherche. En d'autres termes, il s'agira de questionner cette nouvelle direction, de déterminer son influence sur l'économie du développement. En effet, cette dernière partie soulève la seconde dimension épistémologique des expériences du J-PAL, un niveau d'analyse non plus méthodologique mais théorique. Cet apport théorique sera questionné à l'aune de ce que le J-PAL souhaite proposer, c'est-à-dire offrir une alternative au débat clivé de l'aide au développement.

Pour cela, le premier chapitre vise à rendre compte du bagage théorique auquel se réfère le J-PAL. En ce sens, ce chapitre ne traite pas de tous les débats récents en économie du

développement, mais se centre sur celui qui polarise l'attention des chercheurs du J-PAL, celui de l'aide au développement. Afin de saisir au mieux la teneur de ce débat et ses enjeux, il sera nécessaire de retourner brièvement à ce qui semble avoir ravivé la question de l'aide internationale : l'effondrement du « consensus de Washington ». Je développerai, ensuite, les positions de Jeffrey Sachs et d'Easterly. Le premier considère que les pauvres sont enfermés dans des trappes à pauvreté et que seule une aide extérieure massive permettra de les en sortir. Easterly cherche à montrer que l'aide internationale a contribué à perpétuer un impérialisme occidental. Easterly suggère de laisser les pauvres s'autonomiser. L'autonomisation des pauvres (ou le mouvement de l'*empowerment*), dont Muhammad Yunus est la figure centrale à travers le développement du microcrédit, est la principale réponse politique face à une aide massive et donc une importante assistance pour les pauvres. Je développerai alors aussi la position de Yunus, cela me permettra de montrer clairement qu'une position en faveur ou défaveur de l'aide donne lieu à des recommandations politiques nécessairement différentes. Cela rendra alors compte des différents enjeux qui sous-tendent le débat sur l'aide au développement.

Le deuxième chapitre s'attachera à évaluer l'apport théorique du J-PAL à ce débat ; à travers une analyse des expériences que Pascaline Dupas²⁴⁸ a menées sur le paludisme. La lutte contre le paludisme cristallise le débat sur l'aide au développement, et donc l'opposition entre Jeffrey Sachs et William Easterly. Le premier étant en faveur d'une subvention intégrale des moustiquaires, alors qu'Easterly suggère de les rendre même très faiblement payantes afin qu'elles soient utilisées à bon escient. Dupas teste chacune de ces propositions en vue de déterminer la plus efficace. Les expériences de Dupas, à un second niveau, sont particulièrement éclairantes car les retracer permet de traduire très clairement ce que le J-PAL propose et ce sur un sujet avec des enjeux définis en amont. Ces expériences me permettront de distinguer deux étapes au sein de l'approche du J-PAL. La première est caractérisée par le résultat surprenant de Cohen et Dupas (2010) : même gratuites les moustiquaires ne sont pas utilisées. La suite des expériences de Dupas ne vise qu'à tester des dispositifs incitant les pauvres à utiliser les moustiquaires. Ce faisant Dupas teste explicitement des dispositifs de *nudging*, s'apparentant au paternalisme libertarien développé par Sunstein et Thaler. Dupas interprète alors l'insuffisante utilisation des moustiquaires par les pauvres comme un manque de rationalité. Les dispositifs testés visent alors à contrecarrer ce manque de rationalité. Pour

²⁴⁸ Voir (Cohen et Dupas, 2010), (Dupas, 2009 ; 2010, 2011a), (Bhattacharya, Dupas et Kanaya, 2013) et (Cohen, Dupas et Schaner, 2013).

autant ces dispositifs semblent inefficaces à encourager les pauvres à utiliser les moustiquaires.

Le troisième chapitre cherchera à interroger le récent tournant d'Esther Duflo (2012a) en faveur d'un nouveau paternalisme. Duflo propose un « paternalisme démocratique », dépassant le paternalisme libertarien de Sunstein et Thaler, puisque les choix seront retirés des individus, alors qu'ils étaient orientés mais toujours présents chez Sunstein et Thaler. Je montrerai que ce tournant est dans la droite continuité de ce que les expériences de Dupas (2009) avaient laissé entrevoir en testant des dispositifs de *nudging*. Duflo, à l'instar de Dupas, invoque tout d'abord le paternalisme libertarien pour contrecarrer le manque de rationalité des pauvres et les aider à sortir de la pauvreté. Cependant, Duflo creuse plus précisément cette idée en montrant que les pauvres manquent tout autant de rationalité que les riches, à l'exception près que les riches vivent dans un monde beaucoup plus paternaliste et sont entourés d'incitations invisibles. Le paternalisme que souhaite proposer Duflo vise, *in fine*, à augmenter la liberté des pauvres. Pour cela, elle se fonde sur l'approche par les capacités de Sen. Néanmoins, le paternalisme que propose Duflo révèle certaines confusions et semble aller à l'encontre de l'approche développée par Sen. Les confusions du paternalisme de Duflo me semblent n'être que la traduction de la faille épistémologique de son approche développée lors de la partie précédente : l'incapacité à produire des capacités causales. L'apport théorique de l'approche du J-PAL s'en trouve alors réduit, même s'il tend à mettre en évidence un mécanisme particulier : le rôle de l'incohérence temporelle et du désespoir dans la perpétuation des trappes à pauvreté.

Chapitre 1 : La randomisation et ses spectres en économie du développement

0. Introduction

« A qui devons-nous faire confiance ? A ceux qui nous disent que l'aide internationale peut résoudre le problème ou à ceux qui nous disent qu'elle aggrave les choses ? Ce débat ne peut être tranché dans l'abstrait : il faut des faits précis. » (Banerjee et Duflo, 2011, 21)

Duflo et Banerjee cherchent à se positionner au sein d'un unique débat théorique : celui de l'aide au développement. C'est pratiquement le seul débat théorique que Banerjee et Duflo développent. Ce dernier est polarisé par les positions antinomiques de Jeffrey Sachs et de William Easterly²⁴⁹. Le premier adopte une vision globale du développement et montre qu'il est nécessaire de financer massivement les pays en développement afin de les aider à sortir de la pauvreté. Le second, Easterly, privilégie une vision locale du développement et s'attache à montrer tous les méfaits que l'aide internationale a eus sur les pays en développement. Il considère que l'aide internationale n'est que la continuité d'un impérialisme occidental et suggère d'aider les pauvres à trouver leurs propres moyens de lutter contre la pauvreté. Ces deux positions théoriques engendrent alors des préconisations politiques extrêmement différentes. La question de l'aide est une question à la fois théorique et politique. Banerjee et Duflo, à l'aide des expériences randomisées, cherchent à trancher ce débat en offrant des preuves d'efficacité des programmes d'aide au développement. L'objectif de ce chapitre est de définir ce débat tout en déterminant ses enjeux afin de pouvoir questionner ensuite l'apport des expériences du J-PAL à ce débat.

Pour cela, dans la première partie de ce chapitre, je reviendrai à ce qui a précédé ce débat : l'effondrement du « consensus de Washington » à la fin des années 1990. Ce dernier a fortement ravivé la question de l'aide internationale, donnant lieu ensuite, aux travaux de Sachs et d'Easterly. Expliquer l'effondrement du « consensus de Washington », me permettra alors de définir le cadre théorique et politique à partir duquel Sachs et Easterly ont développé leurs positions. La deuxième partie s'attachera alors à présenter les positions théoriques de Sachs et d'Easterly sur l'aide au développement. Je montrerai que leur désaccord concernant l'aide au développement traduit un désaccord plus profond sur la façon d'aborder le développement : ce dernier doit-il être appréhendé de façon locale ou globale ? Sachs tend à

²⁴⁹ Easterly et Sachs représentent deux camps distincts au sein de ce débat. D'autres économistes du développement s'ajoutent à leurs positions. Par exemple Paul Singer (2009) et Paul Collier (2007) partagent la vision de Sachs, et Dambisa Moyo (2009) partage celle d'Easterly. Banerjee et Duflo se centrent sur les figures de Sachs et d'Easterly, je ferais de même dans la suite de ce travail.

privilégier une vision globale du développement alors qu'Easterly favorise une vision locale de ce dernier. Dans la troisième partie de ce chapitre, je montrerai que ces deux positions conduisent à des implications politiques divergentes. Les opposants à l'aide internationale promeuvent l'autonomisation des pauvres et souhaitent rendre la lutte contre la pauvreté aux mains des pauvres. Alors que, par définition, les tenants de l'aide privilégient une assistance et un fort accompagnement vis-à-vis des pauvres.

1. Le fantôme du « Consensus de Washington »²⁵⁰

A la suite du choc pétrolier de 1973, les pays en développement sont les plus endettés auprès du Fonds Monétaire International (FMI). L'endettement étant très important, le FMI et la Banque mondiale choisissent de continuer à aider les pays en développement mais en durcissant leurs conditions de prêts. Ces nouvelles conditions sont les politiques d'ajustements structurels, ou ce que John Williamson (1990, 2003) a appelé le « consensus de Washington ». Ces politiques d'envergure, appliquées à plus d'une centaine de pays, n'ont pas eu les effets escomptés. Elles ont majoritairement contribué à affaiblir les économies en développement. L'échec de ces politiques a laissé place aux doutes en matière de lutte contre la pauvreté. Il n'y a plus de consensus - si un jour il y en avait eu un. Si le « consensus de Washington » est un échec, que faire pour lutter efficacement contre la pauvreté ? Quelles sont les alternatives à un tel échec ? Je commencerai par présenter les raisons de l'échec du « consensus de Washington » (1.1) ; puis, je développerai les différentes alternatives théoriques qui ont suivi cet échec (1.2).

1.1 Un échec politique : les politiques d'ajustements structurels

Les politiques d'ajustement structurel décidées par la Banque mondiale et le FMI avaient pour premier objectif de stabiliser les économies en développement en leur permettant de réduire leur déficit afin qu'elles puissent sortir d'une période de forte récession. Ces politiques visaient aussi à transformer la structure des économies en développement afin qu'elles puissent obtenir une croissance durable. Ces politiques stipulaient principalement une libéralisation importante des économies en développement. C'est cette importante libéralisation qui a largement été critiquée. Pour certains elle n'a pas été assez poussée et pour

²⁵⁰ Le terme est emprunté à John Williamson (2003, 1990).

d'autres elle a été trop loin. Cependant, la majorité s'accorde sur l'échec de ces politiques. Je commencerai par définir ces politiques d'ajustement structurel (1.1.1), puis j'expliquerai l'échec de ces dernières (1.1.2).

1.1.1 Cadre analytique du FMI et politiques d'ajustements structurels

Un des modèles théoriques de référence du FMI est celui défini par Jacques Polak²⁵¹ (1957). C'est une approche monétaire de la balance des paiements, dans laquelle on considère une petite économie ouverte sans marché financier. La demande de monnaie (M_d) y est déterminée par : le niveau général des prix (P), la vitesse de circulation de la monnaie (k) et le revenu réel (Y). Les prix nationaux sont fixés par la loi du prix unique²⁵². Il est alors possible d'exprimer les prix nationaux en fonction des prix internationaux de la façon suivante : $P = e.P_w$, où « e » est le taux de change et « P_w » les prix internationaux, en découle la première équation du modèle :

- $M_d = k.P.Y$, ou $M_d = k.e.P_w.Y$. (1)

L'offre de monnaie (M_s), quant à elle, est déterminée par : le crédit (celui des agents et de l'état) (C) et par les réserves de changes (R). La deuxième équation de ce modèle est alors :

- $M_s = C + R$. (2)

L'équilibre du modèle est défini par l'égalité entre l'offre et la demande de monnaie. Après transformation, il est possible d'obtenir une équation réduite de détermination de l'équilibre extérieur :

- $R = k.e.P_w.Y - C$. (3)

Selon ce modèle, l'accroissement des réserves de changes dépend de trois facteurs, maîtrisés par la politique économique : la dévaluation du taux de change, la restriction du crédit intérieur et l'augmentation de la production. La dévaluation du taux de change

²⁵¹ Jacques Polak (1914-2010) était un économiste, il faisait partie de la délégation néerlandaise de la conférence de Bretton Woods qui créa le FMI. Il commence à travailler pour le FMI à partir de 1947. Voir (Polack, 1957)

²⁵² Loi énoncée par Gustav Cassel et stipulant que dans un marché efficient un bien doit avoir le même prix en tout point de ce marché et dictant alors le principe d'unicité du pouvoir d'achat de chaque unité de monnaies.

engendrer un double effet : elle a un effet sur le revenu et un sur les prix. Cette dévaluation rend les prix plus compétitifs mais diminue le revenu. L'effet prix est considéré plus important par le FMI que l'effet revenu. Étant plus compétitifs, les pays exportent plus et de ce fait peuvent augmenter leurs réserves de change. Réduire le crédit au niveau national permet de ne pas réduire les réserves de changes. En augmentant la production, la croissance économique sera favorisée ce qui tend à augmenter les réserves de changes. L'ensemble des politiques structurelles imposées aux pays en développement sont fondées sur ces trois dimensions : dévaluation du taux de change (1), réduction du crédit national (2), et augmentation de la production (3).

Cet ensemble de politiques a été résumé par John Williamson (1990, 2003) autour de dix mesures que le FMI a imposées aux pays en développement. John Williamson (1990, 2003) les a regroupées sous l'appellation désormais fameuse du « consensus de Washington ». La première politique concerne le déficit de ces pays et stipule que ce dernier ne doit pas excéder 1 ou 2% du PIB (1). Il est préférable que leurs dépenses publiques se focalisent sur la santé et l'éducation, considérées comme du « capital humain », alors que les subventions doivent être éliminées car elles tendent à diminuer l'allocation des ressources (2). Il faut que le système fiscal ait l'assiette la plus large possible ainsi qu'un niveau de prélèvements modéré (3). Les taux d'intérêts doivent être déterminés par le marché afin de décourager la fuite des capitaux et de favoriser l'épargne (4). Le taux de change doit lui aussi être fixé par le marché afin d'atteindre un niveau compétitif (5). Il ne faut plus qu'il y ait d'entraves au commerce extérieur (6) ni aux investissements étrangers (7). Il faut que les entreprises soient privatisées (8) et que les marchés soient déréglementés (9). La propriété privée doit être protégée (10). Cet ensemble de dix politiques se situe dans la continuité du cadre théorique du FMI. La diminution du déficit et l'augmentation de l'assiette fiscale visent directement la diminution du crédit intérieur. La libéralisation du marché, du commerce et des investissements étrangers visent à augmenter l'offre de ces pays tout comme la protection de la propriété intellectuelle et la privatisation des entreprises. Cela part du postulat théorique que le marché laissé à lui-même permettra la meilleure allocation des ressources, et donc une production et une croissance importante. Et enfin la dévaluation du taux de change est le troisième outil envisagé par le FMI. Ces mesures ont commencé à être appliquées à la fin des années 1970, puis massivement au début des années 1980 dans quatre-vingt-dix pays du monde, majoritairement dans les pays sud.

1.1.2 Les effets des politiques d'ajustements structurels : l'aveu d'un échec ?

Afin de rendre compte de l'impact de ces politiques, le FMI les a fait évaluer²⁵³, retenant deux critères : l'un économique et l'autre social. Sur dix-neuf pays, choisis comme échantillon entre 1980 et 1986, on remarque que les résultats sont « faiblement positifs sur la croissance et les exportations ne stimulent pas l'entrée des capitaux privés étrangers et ont un effet négatif sur l'investissement » (Assidon, 2002, 69). L'effet des libéralisations apparaît dans cet échantillon plus positif pour les pays tels que la Turquie et les Philippines qui ont une industrialisation plus développée. Mais pour des pays tels que le Kenya ou le Malawi où l'industrialisation est tournée vers le marché intérieur, les résultats sont nettement moins positifs et ils le sont encore moins pour les pays en déclin à la spécialisation primaire tels que le Ghana. Autrement dit, moins le pays est développé moins les politiques d'ajustements structurels ont eu l'effet escompté. Les pays les plus pauvres sont alors considérés comme les pays les plus « inflexibles » à l'ajustement structurel. La Banque mondiale et le FMI considèrent que l'échec de ces politiques est dû aux manques de capacités institutionnelles de ces pays, considérées comme imparfaites. En matière économique, les politiques d'ajustements structurels n'ont pas été jugées d'un grand succès. Elles n'ont pas permis la croissance souhaitée. Et pire, dans les pays les plus pauvres, on y a observé un recul. La crise asiatique, dans les années 1990, a nettement sonné comme l'échec le plus frappant de ces politiques. Mais là où l'échec est le plus visible, c'est en matière sociale.

En 1991, l'UNICEF publie un rapport²⁵⁴ sur les effets sociaux de ces politiques. Le constat y est accablant, amenant les institutions de Bretton Woods à mettre en place des politiques correctrices avec « une dimension sociale de l'ajustement ». Un des effets de ces politiques a été d'augmenter les inégalités et la pauvreté. En d'autres termes, les effets redistributifs ont été désastreux. Le FMI et la Banque mondiale espéraient que ces politiques permettraient la croissance et que cette dernière aurait des retombées sociales positives. Mais cela n'a pas été le cas. C'est pourquoi, le FMI et la Banque mondiale ont été vivement critiqués. Joseph Stiglitz s'est fait le porte-parole le plus médiatique de cette critique. Pour ce dernier, toutes ces politiques ont été mises en place sans aucun accompagnement. De plus, elles ont été adressées à des pays non préparés à un tel choc :

²⁵³ Voir (Taylor, 1988).

²⁵⁴ Voir (Jolly, 1991) et (Cornia, Jolly et Stewart, 1987).

« Certes, on ne pouvait pas prévoir tous les effets négatifs sur les pauvres des politiques du consensus de Washington. Mais aujourd'hui ils sont clairs. La libéralisation du commerce *associée à des taux d'intérêts élevés* constitue, nous l'avons vu, une méthode presque infaillible pour détruire des emplois et répandre le chômage – aux dépens des pauvres. La libéralisation des marchés financiers *non associée à une réglementation appropriée* est un moyen à peu près sûr de créer l'instabilité économique – et elle risque de faire monter et non baisser les taux d'intérêts (...) La privatisation *sans stimulation de la concurrence et sans surveillance des abus du pouvoir de monopole* peut aboutir à une hausse et non à une baisse des prix pour les consommateurs. L'austérité budgétaire *appliquée aveuglément* dans une situation inadaptée peut faire monter le chômage et rompre le contrat social » (Stiglitz, 2002, 145-146).

Pour Stiglitz, ces politiques sont clairement un échec et la faute en revient au FMI et à la Banque mondiale²⁵⁵. La Banque mondiale, elle-même, reconnaît l'échec de ces politiques dans un rapport publié en 2005²⁵⁶. Ce rapport montre qu'à la suite des politiques d'ajustements structurels, l'Afrique subsaharienne a échoué à démarrer, que beaucoup de crises financières ont eu lieu, et que la reprise de l'Amérique latine n'a été que de court terme. Cependant Rodrik, (2006) note que les années 1990 n'ont pas été uniquement catastrophiques pour les pays en développement :

« Quite to the contrary. From the standpoint of global poverty, the last two decades have proved the most favorable that the world has ever experienced. Rapid Economic growth in China, India, and a few other Asian countries has resulted in an absolute reduction in the number of people living in extreme poverty. The paradox is that that was unexpected too! » (Rodrik, 2006, 975)

A l'heure du bilan, le constat est assez négatif, ce que le FMI et la Banque mondiale considéraient comme la solution aux problèmes de développement - les politiques d'ajustements structurels – s'est avérée être un important échec. Avec cet échec, c'est tout un consensus qui s'effondre.

1.2 L'après consensus de Washington

²⁵⁵ Pour Stiglitz (2002) tous les pays qui ont suivi les politiques d'ajustement structurel ont vu leur situation s'aggraver. Selon lui, cela est dû au fait que le FMI subit une importante pression de la part de la communauté financière internationale, le Trésor américain imposerait au FMI ses intérêts en étant le seul à disposer d'un droit de veto au sein du conseil d'administration du FMI. De plus, Stiglitz critique vivement le fait que le FMI se fixe comme objectif central la réduction de l'inflation. Il n'y aurait aucune preuve, selon lui que cet objectif permette d'augmenter la croissance des pays en développement ; alors qu'il apparaît clair, pour Stiglitz, que cet objectif favorise les intérêts financiers des marchés des pays industrialisés. Enfin, pour Stiglitz, les actions du FMI sont guidées par l'idée erronée que le marché fonctionne correctement et offre les meilleures solutions.

²⁵⁶ Voir (Banque mondiale, 2005).

L'effondrement du « consensus de Washington » laisse place à une importante confusion. Comme le souligne Dani Rodrik, « *The Washington Consensus of the 1990's has left lots of frustration and unrealized expectation in its wake.* » (Rodrik, 2010, 40) Cet effondrement donne la sensation que la lutte contre la pauvreté n'a plus de directions. Quelles politiques mettre en place pour favoriser le développement des pays pauvres ? Par où commencer ? Tout cela pose en réalité la question de ce qui va remplacer ce consensus. L'enjeu revient à trouver un autre consensus. A la fin des années 1990 et au début des années 2000, plusieurs alternatives au consensus de Washington se développent ; ce faisant, elles ouvrent la voie à un nouveau tournant en matière de lutte contre la pauvreté. Je commencerai par développer ce qu'il reste du « consensus de Washington » (1.2.1), puis je tâcherai de présenter les différentes alternatives qui lui ont succédé (1.2.2), au sein desquelles figurent la position de Sachs et d'Easterly.

1.2.1 Que reste-il du « consensus de Washington » ?

A la suite de cet échec la question lancinante est : pourquoi le consensus de Washington a échoué ? Il existe, en fait, peu de réponses à cette question. Dans son rapport, la Banque mondiale (2005) interprète cet échec par le fait que les réformes envisagées ont été trop obsédées par les pertes sèches et qu'elles ne se sont donc pas intéressées au processus de croissance. Elles n'ont pas cherché à comprendre les mécanismes à l'œuvre dans le processus de croissance, ce qui aurait permis d'aider plus efficacement les pays pauvres. Malgré tout, l'échec du consensus de « Washington » semble laisser place à un autre consensus : il n'existe pas de bon ensemble de politiques pour lutter contre la pauvreté. De plus, les agences internationales, telles que le FMI et la Banque mondiale perdent en crédibilité.

Rodrik explique que les objectifs des politiques d'ajustements structurels (une stabilité macroéconomique, une importante libéralisation, etc.) ne doivent pas nécessairement se penser simultanément. Au contraire, ces politiques sont autant d'ingrédients que l'on peut utiliser différemment. Autrement dit, il n'existe pas de bon ensemble de réformes, il existe seulement quelques ingrédients disponibles. Il n'y a pas de bonne pratique. Les politiques doivent être ciblées, et ne doivent pas provenir d'un ensemble. Pour Rodrik les politiques macroéconomiques ne sont pas mortes, mais il faut les détacher de leur ensemble, ne prendre que quelques éléments. Le rôle de la Banque mondiale et du FMI, à la suite de cet échec, est beaucoup discuté. Au cœur de ces discussions, il y a la remise en question de la crédibilité de

la Banque mondiale et du FMI en matière de la lutte contre la pauvreté. Des institutions internationales qui ne côtoient pas le terrain peuvent-elles penser des politiques efficaces ? On critique le fait que les politiques proviennent toujours d'institutions internationales et qu'on ne questionne pas le souhait des pauvres.

Avec l'effondrement du « consensus de Washington », on prend conscience que les pays pauvres sont différents et qu'ils ne sont pas pauvres pour les mêmes raisons. C'est pourquoi on met en évidence l'idée qu'il faut laisser la parole aux pauvres et donner la place au terrain. L'idée de contexte devient centrale. Il n'existe pas une pauvreté, mais plusieurs pauvretés ; qui sont liées à différents contextes économiques, sociaux et environnementaux. De plus, on prend aussi conscience qu'il n'existe pas de recettes miracles mais qu'il n'y a que des ingrédients. Tout cela conduit alors à une toute nouvelle façon d'appréhender le développement : *«Taking these conclusion at face value, what they entail is nothing less than a radical rethink of development strategies »*. (Rodrik, 2006, 977) Après l'effondrement du « consensus de Washington » l'idée est de repenser le développement. C'est sans doute la principale dimension qui reste de ce consensus²⁵⁷.

1.2.2 Quelles alternatives au consensus ?

Dani Rodrik (2006) définit et développe les différentes alternatives qui se sont développées pour pallier l'échec du « consensus de Washington ». Selon lui, à la suite de l'effondrement du « consensus de Washington », il existe deux principales alternatives : celle du « consensus de Washington amélioré » et celle de l'aide extérieure. Au sein de la première alternative, Rodrik distingue deux courants : le fondamentalisme du marché et le fondamentalisme des institutions (Rodrik, 2006, 979). Le premier courant, le fondamentalisme de marché, propose d'ajouter aux politiques d'ajustement structurel d'autres politiques²⁵⁸. Le deuxième courant met au cœur de son analyse les institutions. Il a rencontré

²⁵⁷ « *While the lessons drawn by proponents and skeptics differ, it is fair to say that nobody really believes in the Washington Consensus anymore. The question now is not whether the Washington Consensus is dead or alive; it is what will replace it.* » (Rodrik, 2006, 974)

²⁵⁸ Rodrik résume cette alternative de la façon suivante : « *do more of the same, and do it well.* » (Rodrik, 2006, 977). Au dix mesures distinguées par Williamson (1990, 2003), Rodrik (2006, 978) montre que ce « consensus amélioré » en ajoute dix autres. Ces dix mesures supplémentaires sont les suivantes : introduire une gouvernance semblable à celles des entreprises (1), mener des politiques contre la corruption (2), favoriser la flexibilité du marché du travail (3), respecter les accords de l'organisation mondiale du commerce (4), définir des codes financier (5), instaurer plus de prudence sur l'ouverture des comptes de capital (6), supprimer les régimes de changes intermédiaires (7), instaurer des banques centrales indépendantes dont l'objectif principal est la réduction de l'inflation (8), créer des filets de sécurité sociale (9), et cibler la réduction de la pauvreté (10).

un important succès au travers des thèses de Daron Acemoglu et de John Robinson, et leurs travaux fondateurs démontrant le lien entre une faible croissance et la colonisation²⁵⁹. Pour Acemoglu et Robinson, c'est par un changement profond des institutions qu'on parviendra à améliorer le développement des pays pauvres. Les institutions²⁶⁰ sont le levier de la croissance. Ces deux courants partagent le fait que le développement doit être appréhendé dans une perspective large, soit par les institutions soit par le marché. Dani Rodrik ne privilégie pas ces alternatives et les nuancent très fortement:

« First, the cross-national literature has been unable to establish a strong causal link between any particular design feature of institutions and economic growth (...) Second, we should not forget that Acemoglu, Johnson and Robinson (2001) work and other related focused on long-term economic performance. » (Rodrik, 2006, 980)

Rodrik ajoute que l'agenda qui sous-tend de telles perspectives est irréalisable. La seconde alternative définie par Rodrik, celle de l'aide extérieure, propose de financer une aide massive à destination des pays en développement en vue de les sortir de la pauvreté. La principale figure de ce mouvement est Jeffrey Sachs²⁶¹. Ce dernier montre que les pays pauvres connaissent une faible croissance car ils sont pris dans une trappe à pauvreté. Ici aussi, Rodrik est sceptique face à cette alternative. Selon lui, tous les pays pauvres ne connaissant pas nécessairement une faible croissance, ce qui tend à invalider la position de Sachs²⁶². C'est pourquoi Rodrik propose sa propre alternative. Comme les autres approches elle est centrée sur la croissance, mais vise à rendre compte de stratégies de croissance précises. Son approche se centre sur trois points : un diagnostic, une politique, et une institutionnalisation. Pour Rodrik il faut tout d'abord procéder à un diagnostic qui permettra d'identifier ce qui fait défaut à un pays donné pour qu'il ait une croissance durable. De ce diagnostic, il devient possible de définir une politique claire. Et enfin, afin que cette politique soit pleinement efficace, il faut l'institutionnaliser²⁶³. Les politiques menées par la Banque mondiale et le FMI ont fait office de consensus de la fin des années 1970 à la fin des années

²⁵⁹ Voir (Acemoglu, Johnson et Robinson, 2001) ou (Acemoglu et Robinson, 2006).

²⁶⁰ Les institutions sont ici entendues au sens large, comme le droit de propriété, la démocratie, la décentralisation du pouvoir, etc.

²⁶¹ Je reviendrai très largement sur la position de Sachs dans la suite de ce chapitre, ainsi que sur les oppositions qu'elle a suscitées.

²⁶² « Or what do we make of the fact that economic growth is actually not uncommon among Sub-Saharan African nations themselves? The theory of poverty traps suggests that these countries are stuck in low-level equilibria from which they find it very hard to extricate themselves. The reality seems to be somewhat different. Most African countries have shown themselves capable of producing economic growth over nontrivial time horizons. » (Rodrik, 2006, 981)

²⁶³ Pour un exposé précis de l'alternative de Rodrik voir (Rodrik, 2010).

1990. Le constat de leur échec a laissé place à une importante confusion et à la nécessité de repenser le développement. Plusieurs alternatives sont alors disponibles et ouvrent la voie à de nouveaux courants de recherche, à de nouvelles perspectives.

2. Le débat sur l'aide au développement : aide massive ou « Fardeau de l'homme blanc » ?

Aux débuts des années 2000, la question de l'aide au développement²⁶⁴ est une des questions les plus brûlantes²⁶⁵. L'effondrement du « consensus de Washington » a fortement ravivé cette question. Les politiques de la Banque mondiale et du FMI ont montré que l'aide internationale pouvait avoir des effets négatifs. Les pays recevant l'aide du FMI et de la Banque mondiale n'ont pas connu de croissance importante. Est-ce dû à l'aide ? Ou est-ce dû aux politiques imposées par le FMI et la Banque mondiale ? Deux camps s'opposent. L'un montrant que l'échec n'est pas dû à l'aide internationale, et qu'au contraire cette dernière n'a pas été suffisante pour favoriser le développement des pays pauvres. Et le second camp mettant en cause très clairement l'aide au développement. Ces deux positions sont incarnées par les travaux de Jeffrey Sachs (2005) et de William Easterly (2006a)²⁶⁶.

Pour Jeffrey Sachs, économiste à l'université de Columbia, beaucoup de choses simples et peu coûteuses pour lutter contre la pauvreté pourraient facilement être mises en place. Par exemple, la malaria est particulièrement destructrice en Afrique, il serait pourtant possible de l'éradiquer et à faible coût à l'aide de moustiquaires imprégnées d'insecticides. C'est pourquoi pour Sachs, l'aide internationale doit subventionner ces moustiquaires. William Easterly critique vivement ce point de vue et montre que bien souvent les pauvres n'utilisent pas ces moustiquaires ou leurs donnent une tout autre utilité comme s'en servir de filet de pêche ou de voile de mariée. Pour Easterly, l'aide internationale n'est qu'une triste continuation du « fardeau de l'homme blanc » portant les stigmates de la colonisation et cherchant à faire perdurer un certain impérialisme occidental. Ce débat traduit donc deux dimensions essentielles : une dimension politique et une dimension théorique. Ce débat permet alors d'opposer deux nouvelles façons d'appréhender le développement ainsi que leurs enjeux.

²⁶⁴ Pour une définition précise de l'aide au développement, c'est à dire des pays en développement éligibles, des différents montants alloués, etc., voir (Charnoz et Severino, 2007).

²⁶⁵ Voir par exemple, (Kandur, 2006).

²⁶⁶ On peut aussi trouver d'autres critiques très virulentes D'Easterly contre Sachs, dans Easterly (2006b ; 2006c).

Esther Duflo et Abhijit Banerjee se positionnent principalement face à Jeffrey Sachs et William Easterly²⁶⁷. C'est au sein de cette dernière alternative au « consensus de Washington » qu'ils se situent. Il s'agit alors ici, de définir la position d'Easterly et de Sachs afin de saisir celle voulue par Duflo et de Banerjee, en vue de déterminer leurs apports en économie du développement (leur réponse à l'effondrement du « consensus de Washington »). Je commencerai par développer la théorie de Jeffrey Sachs (2.1), puis je développerai celle d'Easterly (2.2). Cela me permettra de mettre en évidence leurs points de désaccords.

2.1 Jeffrey Sachs et la fin de la pauvreté

Jeffrey Sachs est à l'initiative des objectifs du millénaire. Ces objectifs envisagent de réduire de moitié l'extrême pauvreté et la faim (1). Ils visent à favoriser l'éducation en assurant l'éducation primaire pour tous (2), en promouvant l'égalité des sexes et l'autonomie des femmes (3). Ils s'attachent aussi à la santé en cherchant à réduire la mortalité infantile (4), à améliorer la santé maternelle (5), à combattre le sida, le paludisme et toutes les autres épidémies (6). Ces objectifs visent également à assurer un environnement durable (7). Enfin ils doivent s'articuler autour d'un partenariat mondial pour le développement entre les pays riches et les pays pauvres²⁶⁸. Sachs les prolonge en promettant la fin de la pauvreté d'ici 2025 grâce à un plan d'aide massif²⁶⁹. Les pauvres sont, pour lui, enfermés dans un cercle vicieux ; ils sont pris dans des trappes à pauvreté :

« A large number of the extreme poor are caught in a poverty trap, unable on their own to escape from extreme material deprivation. They are trapped by disease, physical isolation, climate stress, environmental degradation, and by extreme poverty itself ». (Sachs, 2005, 19)

De façon tautologique, les pauvres sont pauvres parce qu'ils sont pauvres. Du fait d'être pauvres ils sont souvent en mauvaise santé, cette mauvaise santé les empêche d'être

²⁶⁷ Voir Par exemple (Banerjee et Duflo, 2011) et (Duflo, 2009 ; 2010a ; 2010b ; 2012a) ; et pour une recension du livre de Sachs (2005) par Banerjee, voir (Banerjee, 2005a).

²⁶⁸ Ces huit objectifs ont été adoptés par 193 pays états membres des Nations Unis en 2000 à New-York et doivent être atteints en 2015. Le rapport de l'ONU de 2013 (Voir, ONU, 2013) montre que de nombreux progrès ont été réalisés, mais que de nombreux objectifs ne seront sans doute pas atteints en 2015. Les objectifs d'accès à l'eau potable et de réduction de moitié des personnes vivant dans l'extrême pauvreté ont été réalisés. Les objectifs concernant le paludisme et le VIH sont considérés comme en bonne voie d'être atteints. Par contre les résultats concernant l'éducation, la lutte contre la faim, et le développement durable sont jugés décevants.

²⁶⁹ Pour une critique de ces objectifs du Millénaire, voir (Easterly, 2009).

productifs et donc de travailler ce qui les fait inévitablement rester pauvres. L'aide financière internationale permettrait de briser ce cercle et de sortir les pauvres de ces trappes. Selon Sachs, pour éradiquer la pauvreté d'ici 2025 il suffit d'apporter l'aide financière nécessaire aux pays en développement.

2.1.1 L'économie du développement comme économie « clinique »

Selon Jeffrey Sachs, il existe des similarités entre un « bon développement économique » et une « bonne médecine clinique » (Sachs, 2005, 75) : en vue de s'améliorer, l'économie du développement devrait s'inspirer plus largement de la médecine clinique. A l'inverse de la médecine clinique qui possède un caractère rigoureux, les institutions internationales connaissent peu de choses des pays bénéficiaires de l'aide au développement et semblent agir de façon approximative. L'économie du développement ressemble, pour Sachs, à la médecine du XVIIIe siècle : par manque de connaissances on y tuait plus de patients que l'on en sauvait. L'économie du développement doit se fonder sur la médecine clinique du XXe siècle et pour cela procéder à un diagnostic différentiel, permettant d'identifier les problèmes clefs. Le problème clef est, pour Sachs, l'existence de trappes à pauvreté, un diagnostic différentiel permettrait d'en définir les principales causes et ainsi d'orienter l'aide.

Cinq principales leçons de la médecine clinique peuvent être appliquées à l'économie du développement. En premier lieu, « le corps humain est un système complexe » (1) (Sachs, 2005, p.75), il est régi par différentes connexions, celles du système nerveux, mais aussi celles du système digestif ou encore celles du système respiratoire. Les maladies sont le résultat d'une combinaison complexe de facteurs. Afin d'y remédier il est nécessaire de comprendre la complexité de ces facteurs et leurs impacts. La compréhension de cette « complexité requiert un diagnostic différentiel » (2) (Sachs, 2005, 76). Son objectif est de déterminer la cause de la maladie. Dans ce but, le médecin ne questionne pas uniquement le patient, mais il interroge aussi la famille et cherche à connaître les antécédents familiaux (3). Une fois la pathologie clairement établie, il reste à déterminer le traitement approprié, afin que celui-ci permette la guérison. Sa surveillance et son évaluation sont primordiales (4). En dernier lieu, « la médecine est une profession » (5) (Sachs, 2005, 78) et comme toute profession, elle requiert des normes, un cadre éthique, et un code de conduite particulier.

L'économie du développement doit chercher à traduire ces différents aspects dans sa propre discipline. Elle doit envisager les économies et les individus comme un système complexe, procéder à un diagnostic différentiel, concevoir les politiques de développement d'un pays en lien avec celles des pays frontaliers. Ces politiques doivent être évaluées et suivies. De plus, la communauté internationale en charge du développement souffre d'un manque de normes et de cadres éthiques clairs qu'elle doit pallier.

Une économie clinique permettrait de déterminer la meilleure stratégie à adopter pour lutter contre la pauvreté. La pierre angulaire d'une telle stratégie est le diagnostic différentiel. Sachs définit très clairement les interrogations permettant d'aboutir à ce diagnostic, il en distingue sept. Ce diagnostic différentiel doit s'attacher à construire une carte de la pauvreté (1) en utilisant des données géographiques ou des études sur les ménages. Il doit aussi s'intéresser à la politique économique du pays afin de rendre compte de l'environnement des affaires, de la politique commerciale, des infrastructures existantes et du capital humain (2). Toujours dans un registre politique, il faut qu'il interroge la politique fiscale du pays et l'existence de trappe fiscale (3) ainsi que la géographie (4) et la géopolitique (5). Et enfin questionner ce qui relève de la gouvernance (6) et des barrières culturelles (7).

2.1.2 La fin de la pauvreté : aide massive et trappes à pauvreté

Le diagnostic différentiel permet de comprendre ce qui crée la trappe à pauvreté d'un pays particulier, afin d'envisager un plan d'aide ciblé. Les trappes à pauvreté sont dues à un manque de capitaux. Ce manque relève de six types de capitaux : le capital humain, le « capital d'affaires » (Sachs, 2005, 244), les infrastructures, le « capital naturel » (*Ibid.*), les infrastructures institutionnelles et les « connaissances » (*Ibid.*). Les deux premiers capitaux renvoient à ce dont un individu (capital humain) et une économie (capital d'affaires) ont besoin pour être productifs. Il faut y ajouter la « connaissance » qui représente le savoir-faire qui tend à améliorer la productivité d'une économie ou d'un individu. Le capital naturel comprend les facteurs environnementaux dont une économie a besoin pour bien fonctionner. Les infrastructures et les infrastructures institutionnelles sont composées des équipements routiers, ferroviaires, maritimes et du système judiciaire. Ce manque de capitaux enferme les pauvres dans une trappe, seul un investissement suffisant dans ces différents types de capitaux permettra de casser ces trappes :

« The poor start with a very low level of capital per person, and then find themselves trapped in poverty because the ratio of capital per person actually falls from generation to generation. The amount of capital per person declines when the population is growing faster than capital is being accumulated. Capital is accumulated in turn, in a balance of two forces, one positive and one negative. On the positive side is the capital accumulated when households save a part of their current income, or have a part of their income taxed to finance investments by the government. Household savings are either lent to businesses (often through financial intermediaries such as banks) or invested directly in family businesses or equities traded in the market. Capital is diminished, or depreciated, as the result of the passage of time, or wear and tear, or the death of skilled workers, for example because of AIDS. If savings exceed depreciation, there is positive net capital accumulation. If savings are less than depreciation, the capital stock declines. Even if there is positive net capital accumulation, the question for growth in per capita income is whether the net capital accumulation is large enough to keep up with population growth » (Sachs, 2005, 245).

Seule l'aide internationale, en finançant massivement l'épargne des ménages, peut permettre que l'accumulation du capital soit positive, rendant de ce fait la croissance possible et sortant ainsi les pauvres de leurs trappes. Le cœur d'une telle stratégie d'investissement est de mener un diagnostic différentiel afin de déterminer, pour un pays donné, dans quel type de capital il faut plus particulièrement investir pour que l'accumulation soit positive. Selon Sachs, pour que l'extrême pauvreté soit totalement éradiquée en 2025, chaque pays développé doit participer et financer une partie de cette aide. Un pacte pour le développement est nécessaire afin que les pays développés et les pays en développement s'engagent les uns vis-à-vis des autres. C'est dans cette même logique que les Nations Unies ont proposé un contrat mondial pour le développement au sein des objectifs du millénaire.

2.2 William Easterly et le « fardeau de l'homme blanc »

Pour William Easterly les pays en développement connaissent deux tragédies. La première est la pauvreté qui les frappe et la seconde « consiste en ceci que l'Occident au cours du dernier demi-siècle, a dépensé 2,3 milliards d'aide au reste de la planète ... sans parvenir à fournir aux enfants les médicaments à 12 cents qui auraient divisé par deux le nombre de morts de la malaria dans le monde » (Easterly, 2006a, 12). Cet échec est pour Easterly le résultat de politiques menées par des « planistes » (Easterly, 2006a, 14). Les « planistes » pensent qu'il existe une solution miracle à la pauvreté et dans cette logique proposent des plans d'envergure. Pour Easterly, l'illustration même du « planiste » est Jeffrey Sachs. A l'inverse les « essayeurs » (Easterly, 2006a, 15) tentent de répondre à un problème spécifique, ils expérimentent en vue de trouver des solutions particulières. Je commencerai par présenter

les critiques qu'Easterly adresse à la théorie de Sachs (2.2.1), puis je développerai la solution qu'il y propose (2.2.2). Cela me permettra de montrer le désaccord plus profond qui existe entre les théories de Sachs et d'Easterly : la façon d'appréhender le développement.

2.2.1 Faits stylisés contre mythes de l'aide massive

Pour Easterly, trois mythes justifient une aide au développement massive. Dans le but de montrer l'irréalité de ces mythes, Easterly confronte chacun d'entre eux à des faits stylisés. Le premier mythe, pour Easterly, part du postulat que « les pays les plus pauvres sont prisonniers d'un piège à pauvreté dont ils ne peuvent sortir sans un plan massif financé par l'aide internationale » (Easterly, 2006a, 54). Easterly rejette l'existence d'une trappe à pauvreté. Selon lui, si cette dernière existait réellement, les pays pauvres auraient connu une croissance stagnante, voire négative. A partir des données statistiques d'Angus Maddison²⁷⁰, Easterly montre que sur la période 1950-2001, les cinq pays les plus pauvres ont, en fait, connu une croissance positive. Easterly considère alors le premier mythe invalidé.

Le deuxième mythe considère que « si les pays pauvres connaissent une croissance déplorable, la faute en revient à un piège à pauvreté et non à une mauvaise gouvernance » (Easterly, 2006, 59). L'enjeu de ce mythe est important, comme le souligne Sachs : « si les pauvres sont pauvres à cause d'un gouvernement corrompu, (...) alors la coopération internationale n'est plus d'aucun secours ». (Sachs cité dans Easterly, 2006a, 59) Si ce mythe s'effondre, l'aide internationale massive devient un vaste gaspillage qui ne change rien à la situation des pauvres et ne sert qu'à enrichir un gouvernement corrompu. L'effet de l'aide deviendrait pervers. Afin de renverser ce mythe Easterly appuie son argumentation sur une étude réalisée à l'université du Maryland, « Polity IV »²⁷¹, qui évalue les différentes démocraties et la corruption. Cette étude permet à Easterly de comparer des données en termes de croissances et en termes de gouvernance. Easterly montre une corrélation entre une faible croissance et un pays « mal gouverné ». Cela tend à invalider, selon lui, le deuxième mythe de l'aide au développement.

Le troisième et dernier mythe avance l'idée que « l'aide internationale massive permettra aux pays pauvres d'atteindre une croissance autonome » (Easterly, 2006a, 62). Easterly part

²⁷⁰ Voir : (Maddison, 1995).

²⁷¹ Voir : (Polity IV Project, 2000).

du constat que, dans les années 1990, 15% des revenus africains provenaient de l'aide étrangère. Il compare ensuite les données de la croissance en Afrique sur la période 1970-1999 à celles de l'aide. Il montre une relation inverse entre la croissance et l'aide en Afrique sur cette période. En s'appuyant ensuite sur un article de Peter Boone (1996)²⁷², il montre que l'aide internationale a financé majoritairement la consommation et non les investissements - comme Sachs l'envisageait afin de revaloriser le capital des pays pauvres.

Ce mythe ne s'effondre pourtant pas complètement car Easterly en reprenant une étude de Burnside et Dollar²⁷³ montre, cette fois-ci, une corrélation entre l'aide et la croissance dans les pays disposant d'une « bonne gouvernance ». Le mythe de l'aide au développement se révèle alors plus compliqué qu'il n'y paraît, son effondrement étant dépendant des faits stylisés ou des études choisies.

2.2.2 La solution aux velléités impérialistes : l'absence de plan

William Easterly relie la question de l'aide à l'impérialisme occidental, reprenant ainsi le fameux poème de Rudyard Kipling, « fardeau de l'homme blanc »²⁷⁴. Ce poème représente les tâches que les occidentaux ont à accomplir envers les populations colonisées : les civiliser, les administrer et subvenir à leur besoins. Ce fardeau pose la supériorité du peuple occidental, et lui impose un devoir de civilisation envers les autres peuples. Pour Easterly, le « fardeau de l'homme blanc » d'aujourd'hui souffre des mêmes maux que ceux dont souffraient les anciens responsables coloniaux : « excessive assurance des bureaucrates, planification autoritaire descendante, connaissance biaisée et incomplète des spécificités locales, faibles remontées des populations locales sur l'efficacité des mesures engagées ». (Easterly, 2006a, 346). Les agences d'aides ne font que perpétrer le « fardeau de l'homme blanc », en

²⁷² Cet article est un des articles les plus fameux concernant l'aide au développement.

²⁷³ Voir (Burnside et Dollar, 2000). Cet article, tout comme celui de Boone (1996) est au centre des débats sur l'efficacité de l'aide internationale.

²⁷⁴ « O Blanc, reprends ton lourd fardeau : Envoie au loin ta plus forte race, Jette tes fils dans l'exil Pour servir les besoins de tes captifs;
Pour - lourdement équipé - veiller Sur les races sauvages et agitées, Sur vos peuples récemment conquis, Mi-diables, mi-enfants.
O Blanc, reprends ton lourd fardeau : Non pas quelque oeuvre royale, Mais un travail de serf, de tâcheron, Un labeur commun et banal.
Les ports où nul ne t'invite, La route où nul ne t'assiste, Va, construis-les avec ta vie, Marque-les de tes morts !
O Blanc, reprends ton lourd fardeau; Tes récompenses sont dérisoires : Le blâme de celui qui veut ton cadeau, La haine de ceux-là que tu surveilles.
La foule des grondements funèbres Que tu guides vers la lumière : "Pourquoi dissiper nos ténèbres, Nous offrir la liberté ?" ». (Kipling, 1899)

proposant uniquement des solutions d'envergure et en oubliant de s'interroger sur les besoins particuliers des différentes populations.

Selon Easterly, l'une des possibilités afin de rompre avec le « fardeau de l'homme blanc » consiste à appréhender le développement de façon locale. Dans cette perspective, les moyens sont nombreux et dépendent des caractéristiques de chaque pays. Le Japon, par exemple, a réussi à retrouver la croissance après la Seconde Guerre mondiale : « le Japon a réussi ce tour de force sans le Fardeau de l'homme blanc : l'Occident n'a jamais colonisé le Japon, qui s'est doté de ses propres essayeurs » (Easterly, 2006a, 422). Le principal problème de l'aide occidentale est de rechercher la recette idéale :

« L'excitation même que suscite la certitude de détenir le plan idéal est un symptôme de l'approche erronée que manifestent aujourd'hui comme hier, tant de politiciens pourtant soucieux d'aider les pauvres. Le plan idéal, c'est de n'avoir aucun plan. » (Easterly, 2006a, 13)

Pour Easterly, il n'y a pas de solution unique : il y a une solution propre à chaque contexte et ce n'est qu'en essayant, qu'en chassant les plans, que ces solutions se feront jour. Les « essayeurs » doivent prendre la place des « planistes ». Les solutions globales doivent être remplacées par des solutions locales. Pour réellement lutter contre la pauvreté, pour Easterly, il faudrait : rendre les acteurs de l'aide responsables (1), chercher ce qui fonctionne à travers des expériences de terrain (2), affiner les expériences avec les résultats des expériences précédentes (3), évaluer l'aide (4), récompenser les réussites de l'aide et punir ses échecs (5). Le débat sur l'aide au développement est un débat fortement clivé ; d'un côté par la position de Sachs qui promet la fin de la pauvreté à travers un plan d'aide massive et d'un autre côté par la position d'Easterly qui démontre l'échec de cette dernière. Le point de désaccord entre Easterly et Sachs semble être plus profond. Il est, selon moi, lié au fait qu'Easterly recommande une appréhension locale du développement alors que Sachs l'envisage de façon globale. La question de l'aide au développement est suspendue à une problématique plus générale : comment faut-il envisager le développement ? Faut-il le comprendre dans une perspective locale ou globale ? Jeffrey Sachs incarne une appréhension globale du développement, dans la lignée du *big push* de Paul Rosenstein Rodan²⁷⁵(1943). Easterly

²⁷⁵ La théorie du « *big push* » développée par Paul Rosenstein-Rodan (1943 ; 1961) montre qu'un programme d'investissement trop faible n'impactera pas la croissance des pays en développement, il faut pour cela que l'investissement soit important. Autrement dit, il faut que l'aide internationale finance un important investissement afin que celui-ci soit efficace. Si l'investissement est insuffisant l'aide internationale sera gâchée

critique très fortement cette appréhension du développement. Selon lui le *big push* a échoué²⁷⁶, tout comme les politiques d'ajustements structurels. Et avec l'échec de ces politiques c'est l'appréhension en terme global du développement qui a échoué. De plus, Easterly et Cohen (2007) soulignent que la littérature sur la croissance a aussi beaucoup été critiquée, fragilisant encore la perspective globale :

“ The growth literature was also criticized for its inability to address causality. In the absence of clear evidence that growth outcomes can be attributed to specific levers, development research has severely limited utility for policy.” (Cohen et Easterly, 2007, 3)

Banerjee et Duflo retrouvent la critique de Sachs sur la recherche de facteurs causaux. Selon eux, il ne faut pas s'intéresser à l'aide de façon générale, mais il faut s'intéresser à des problèmes concrets²⁷⁷, comme savoir s'il est plus efficace de financer les manuels scolaires en vue d'améliorer l'éducation ou financer des médicaments. Similairement, ils privilégiaient de ne plus s'intéresser aux « Institutions » en lettre majuscule mais aux « institutions » en lettre minuscule, ils proposent de s'intéresser à l'« aide » en minuscule et non à l'« Aide » (Banerjee et Duflo, 2011, 25). Le problème des trois « i » défini par Banerjee et Duflo et développé dans l'introduction de ce travail tend à rendre compte d'une paresse institutionnelle caractérisant l'aide : « *in many ways this episode captures very well one of the cores problems with delivering aid : institutional laziness.* » (Banerjee, 2007b, 7) Banerjee et Duflo proposent donc une voie alternative, déterminer à travers les expériences randomisées des preuves de l'efficacité de l'aide²⁷⁸.

3. La lutte contre la pauvreté dans la main des pauvres et l'aide au développement

Le débat sur l'aide au développement ne soulève pas uniquement la question d'un important soutien financier de la part des pays riches, ou la question de la façon dont le développement doit être appréhendé, ce débat conduit aussi à des positions politiques opposées et marquées. Comme Easterly le soulignait, une aide massive peut s'apparenter à une vision paternaliste de la part des pays riches envers les pays pauvres, une telle position (l'aide massive) suppose que les chercheurs occidentaux comme les agences internationales savent ce qu'il faut faire pour aider les pauvres. A l'inverse, les opposants à l'aide

²⁷⁶ Certains nuancent le point de vue d'Easterly en montrant que l'aide n'a pas été assez importante pour avoir un impact positif sur les investissements, voir par exemple (Singer, 2009).

²⁷⁷ Voir (Banerjee, 2008) et (Banerjee et Duflo, 2011).

²⁷⁸ Voir, par exemple, (Banerjee, 2007a ; 2006)

internationale privilégient l'idée de rendre la lutte contre la pauvreté aux pauvres, ce qu'on appelle aujourd'hui l'*empowerment*. Muhammad Yunus s'est fait le principal promoteur d'un tel mouvement. L'échec du consensus de Washington a contribué à promouvoir une telle proposition en soulevant que les institutions internationales avaient échoué à aider les pauvres, ces derniers étant les mieux à même de savoir ce dont ils ont besoin et donc les plus à même à se sortir de la pauvreté. Duflo et Banerjee prônent, là aussi, une vision alternative, intermédiaire à celle des tenants ou des détracteurs de l'aide ; ils prônent un *empowerment mandaté*. Je commencerai par montrer les liens entre l'*empowerment* et l'absence d'aide ainsi que les liens entre l'aide et l'impérialisme occidental, j'insisterai particulièrement sur la notion d'*empowerment* en développant la position de Yunus (3.1). Puis, je m'attacherai à développer la position alternative de Duflo et Banerjee pour un *empowerment mandaté* (3.2).

3.1 Impérialisme et aide massive vs *empowerment* et absence d'aide

Pour Duflo, les anti-aides seraient en faveur d'un *empowerment* pour les pauvres. L'aide internationale chercherait à imposer ce qui lui semble bon, car les pays occidentaux seraient considérés comme ceux qui savent ce qui est bien ou mauvais pour les pauvres :

« Critics claim that aid organizations impose on the poor a model of what is right or wrong that is based on the claimed superiority of the aid-providing civilization. Furthermore, the criticism goes, aid organizations attempt to substitute for both the local communities and for the national governments of the receiving countries. The motivation for aid may have nothing to do with the welfare of the poor, but everything to do with the economic interests of aid workers or their countries. » (Duflo, 2012a, 5)

L'aide internationale ne souhaiterait, en fait, pas aider les pauvres mais servir les intérêts économiques des pays donateurs, en se substituant aux communautés locales ou aux gouvernements des pays receveurs, considérés comme inefficaces. Muhammad Yunus qui a contribué - à travers le développement du microcrédit - à promouvoir l'*empowerment* des pauvres, retrouve pleinement cette idée²⁷⁹. Pour Yunus, l'aide internationale met l'accent sur

²⁷⁹ « Depuis 1972, le Bangladesh a bénéficié de quelque 30 milliards de dollars d'aide étrangère. Cette année il recevra près de 2 milliards de dollars. Mais où est allé tout cet argent ? (...) Près des trois quarts de toute l'aide étrangère destinée au Bangladesh sont dépensés dans le pays donateur ; en somme, l'aide est devenue un moyen pour les pays riches de donner du travail à leur population et de vendre leurs produits. Quant au dernier quart, la quasi-totalité en revient à une petite élite bangladaise de consultants, d'entrepreneurs, de bureaucrates et de fonctionnaires corrompus – lesquels dépensent leur argent en produits d'importation ou le transfert sur des comptes à l'étranger, ce qui n'apporte rien à notre économie. Le problème est le même partout dans le monde. » (Yunus, 1997, 36)

la quantité et non sur la qualité, les donateurs comme les receveurs se centrent sur la quantité d'aide offerte et non sur ces effets. L'aide n'est que la charité des riches, pour Yunus, elle leur permet de ne pas se questionner sur le problème de la pauvreté. On donne de l'argent et on n'y pense plus. Pour Yunus, cette charité a de très mauvais effets sur la pauvreté, elle la perdure, et ne sert qu'à donner bonne conscience aux riches :

« Je ne mets pas en cause la nécessité morale de l'aide, ni l'instinct qui nous pousse à aider les nécessiteux. Mais seulement la forme que revêt cette aide. Du point de vue du bénéficiaire la charité peut avoir des effets désastreux. (...) Dans tous les cas la mendicité prive l'homme de sa dignité. Le dispensant de subvenir à ses besoins, elle l'incite à la passivité. Ne suffit-il pas de rester assis là et de tendre la main pour gagner sa vie ? » (Yunus, 1997, 47)

L'aide internationale, au travers de la charité, incite donc, pour Yunus, à la passivité, et permet d'éviter de se demander pourquoi quelqu'un mendie. Yunus cherche alors à penser un projet qui permettra de contrecarrer le fait que les pauvres ne peuvent pas emprunter au sein des banques, et donc ne peuvent pas se sortir par eux-mêmes de la pauvreté. Il recevra en 2006 le prix Nobel de la paix pour ce projet, la *Grameen Bank*, au travers du développement du microcrédit. Le microcrédit n'est qu'une pièce d'un projet beaucoup plus large, pour Yunus. L'objectif est de redonner la lutte contre la pauvreté aux pauvres. Autrement dit, la position de Yunus incarne l'*empowerment*. Le microcrédit est donc un moyen d'autonomiser les pauvres. Ils ne peuvent pas emprunter au sein des banques par faute de solvabilité ; au lieu de faire la charité, Yunus propose de leur prêter une faible somme d'argent, avec un fort taux d'intérêt, compte tenu de l'absence de solvabilité. Dans le modèle pensé par Yunus, il n'est pas uniquement question de prêts à taux d'intérêts élevés, seize règles fondent la *Grameen Bank*, dont favoriser l'éducation de ses enfants, garder sa maison propre, boire de l'eau potable, cultiver des légumes et augmenter les investissements²⁸⁰. Cependant, deux aspects centraux ressortent de ce modèle. Le premier étant que les individus qui empruntent doivent emprunter par groupe de cinq, ces cinq individus sont liés par leurs emprunts, et si l'un ne rembourse pas alors les autres doivent le faire pour lui. Cela dans le but de renforcer les liens communautaires et l'entraide. La seconde dimension centrale est que le microcrédit est particulièrement tourné vers les femmes. Elles sont les principales destinataires de tels prêts. La première raison est, pour Yunus, que les banques sont sexistes (Yunus, 1997, 132), ce qui

²⁸⁰ On peut retrouver l'ensemble de ces règles sur le site de la *Grameen Bank* : http://www.grameen-info.org/index.php?option=com_content&task=view&id=22&Itemid=109 (dernière consultation le 16 août 2013).

l'a encouragé à cibler les femmes. Mais surtout, pour Yunus, lorsque l'argent est destiné aux femmes, il a tendance à être mieux utilisé :

« Il ne s'agissait pas seulement de donner aux femmes la place qui leur revenait, mais bien davantage de les considérer comme des acteurs privilégiés du développement. Plus j'avancais dans mon projet plus j'acquerrais la certitude que le crédit, lorsqu'il passait par les femmes, amenait plus rapidement des changements que lorsqu'il passait par les hommes. Relativement parlant, la faim et la pauvreté sont plus une affaire de femmes qu'une affaire d'hommes. » (Yunus, 1997, 132)

Les femmes sont donc le vecteur du développement²⁸¹. Ce dernier passe, pour Yunus par l'autonomisation des pauvres, et donc par l'autonomisation des femmes dans les pays en développement. L'*empowerment* des femmes, à l'inverse des effets de l'aide internationale est, pour Yunus, la recommandation politique à suivre. Sa position montre clairement le lien entre anti-aide et autonomisation des pauvres. Cela tend aussi à souligner que l'aide internationale suppose un certain paternalisme.

3.2 Pour un *empowerment* mandaté

Pour Yunus il suffit d'offrir un « petit coup de pouce » à travers le microcrédit pour que les pauvres développent leurs propres entreprises et sortent, par eux-mêmes de la pauvreté. Les pauvres possèdent de petites entreprises ; même un investissement faible au sein de ces dernières engendrerait d'importants rendements et rendrait alors leurs entreprises prospères. Banerjee et Duflo (2011) soulèvent que les effets du microcrédit, autrement dit de l'*empowerment* ont été très peu évalués. Plusieurs expériences du J-PAL ont alors cherché à évaluer ces effets²⁸². Ces expériences montrent que le microcrédit ne permet pas de grandes transformations dans la vie des pauvres. Elles montrent que les pauvres ne réalisent pas les investissements qui permettraient de rendre leurs entreprises prospères. En d'autres termes, pour Banerjee et Duflo, les pauvres ne sont pas des entrepreneurs naturels et s'ils deviennent tout de même entrepreneurs, c'est uniquement par qu'ils ne disposent pas d'autres opportunités²⁸³, Banerjee et Duflo montrent qu'en réalité, si les pauvres avaient le choix ils préféreraient avoir des emplois stables.

²⁸¹ Duflo a d'ailleurs écrit de nombreux articles cherchant à saisir ce lien, voir (Duflo, 2006a), (Duflo, 2012b) et (Duflo et Chattopadhyay, 2004).

²⁸² Voir par exemple, (Banerjee, Duflo, Glennerster, et Kinnan, 2010)

²⁸³ Cette étude a permis à Duflo et Banerjee de constituer une importante base de données sur 18 pays. Cette dernière est disponible sur le site internet accolé à leur dernier livre :

Un autre outil central mis en avant pour l'*empowerment* des pauvres, et que l'on retrouvait chez Yunus, est la participation communautaire. Afin, là aussi, de rendre compte des effets d'un tel outil dans la lutte contre la pauvreté, Banerjee et Duflo s'intéressent aux effets du contrôle de la communauté dans l'amélioration de l'éducation. La qualité de l'éducation dans les pays pauvres est souvent faible, le contrôle de la communauté sur ce service public pourrait donc améliorer la qualité de l'éducation :

« The basic theory here is that the families of the children who go to public schools, more than anyone else, see what is wrong with the school and the teaching, and have the interests of their children at heart. Therefore, if they only had the power to reward performance and punish negligence on the part of the teachers, things would work much better. » (Banerjee et Duflo, 2008, 338)

Les chercheurs du J-PAL au MIT, l'ONG Pratham ainsi que la Banque mondiale ont constitué un groupe de recherche afin de mener une étude à Jaunpur, un district du nord de l'Inde²⁸⁴. Cette étude vise à évaluer le rôle joué par la communauté dans le développement. Cette étude montre également que les différents habitants de ce district ne sont pas particulièrement engagés dans les institutions locales de gouvernance²⁸⁵. Durant les réunions au sein de ce district, l'éducation comme enjeu principal n'arrive qu'à la cinquième position. Banerjee et Duflo tentent d'expliquer ce phénomène. Une première possibilité, selon eux, est que les parents ne savent pas à quel point l'enseignement que reçoivent leurs enfants est mauvais. Une autre possibilité est que les parents ne sont pas conscients des bénéfices que peuvent tirer leurs enfants d'une bonne éducation. Cette seconde possibilité semble être la plus probable pour Banerjee et Duflo. Dans tous les cas le contrôle par la communauté ne semble pas efficace, et ne permet pas aux pauvres de lutter, par eux-mêmes, contre la pauvreté :

« If the main difference is in the relationship between public officials and the local community, then community control is unlikely to succeed in India without a radical realignment of the local political economy. If on the other hand, it is a matter of convincing parents that education is worth fighting for, then an entirely different approach may be called for before the problem can be thrown back to the community. Finally, if, after all, it comes down to the level of intervention (villages versus sub-district, say) or some

<http://pooreconomics.com/data/country/home>

²⁸⁴ Voir (Banerjee, Banerji, Duflo, Glennerster et Khemani, 2010)

²⁸⁵ « Only 14.2% of respondents know of a household member ever having been to a Gram Sabha (village meeting), which are mandatorily held in every village. Most (over 90%) said that they do not know when or where the Gram Sabha is held. » (Banerjee et Duflo, 2008, 338)

other detail of implementation, one could imagine community control working well, once the appropriate model has been figured out. » (Banerjee et Duflo, 2008, 340)

Si cette inefficacité s'explique par la relation entre les pouvoirs publics et les communautés locales, alors, pour Duflo et Banerjee seul un changement dans l'économie de la communauté locale permettra à cette dernière de pouvoir jouer un rôle clef dans la lutte contre la pauvreté. Par contre, si l'*empowerment* en faveur de l'éducation implique qu'il faille convaincre les parents du rôle positif de l'éducation, alors une tout autre approche est à envisager. Puis, si l'efficacité du contrôle de la communauté dépend du niveau d'implantation, et d'intervention, alors pour Duflo et Banerjee le bon modèle est encore à trouver. Le rôle de l'*empowerment* dans la lutte contre la pauvreté, n'apparaît donc pas être efficace, pour Banerjee et Duflo. Il leur semble déraisonnable de laisser la lutte contre la pauvreté dans l'unique main des pauvres, alors qu'ils ont déjà à s'occuper de tellement de choses avec si peu d'argent :

« Nevertheless, it seems unreasonable, and possibly even callous, to assume that they are raring to play a much greater role in pulling themselves out of poverty. After all, they already have to cope with so many challenge in living a life on very little money: what to do when you a loved one falls ill and there is no money to pay the doctor; how to explain to yourself you cannot have a break from work, though your back is breaking and it is hard to keep your eyes open. » (Banerjee et Duflo, 2008, 340)

Rendre la lutte contre la pauvreté aux pauvres est donc, d'une certaine manière, dangereux pour Banerjee et Duflo. De plus, les outils envisagés semblent inefficaces. Banerjee et Duflo suggèrent non pas de laisser la lutte contre la pauvreté dans l'unique main des pauvres, mais de les accompagner dans cette lutte en favorisant des outils dont l'efficacité est certaine.

4. Conclusion

Le « consensus de Washington » a dominé les politiques, en matière de lutte contre la pauvreté, de la fin des années 1970 à la fin des années 1990. Ce consensus fondé sur un modèle théorique particulier, le modèle de Polak, a conduit à mener des politiques macroéconomiques de stabilisation. Ces politiques ont été un échec pour les pays en développement qui les ont adoptées. Cet échec a engendré de nouveaux programmes de recherche tentant d'en tirer les principales leçons. La question de l'aide au développement est apparue au premier plan. Comment repenser l'aide ? Quel rôle doit-elle jouer dans le

développement ? Les positions sont ici marquées. J'ai principalement développé celles de Jeffrey Sachs et de William Easterly qui incarnent deux camps de ce débat. Ces deux positions m'ont permis de rendre compte des enjeux en matière de lutte contre la pauvreté d'un tel débat. Ces enjeux se retrouvent très largement dans ceux soulevés par l'échec du « consensus de Washington ». Cela pose alors la question de savoir comment il faut repenser le développement, s'il faut continuer à l'appréhender de façon globale ou s'il faut au contraire l'envisager de façon locale.

L'opposition entre une aide massive et une absence de plan (pour Easterly) conduit à des recommandations politiques extrêmement divergentes. Les opposants de l'aide, percevant cette dernière comme un impérialisme occidental, font la promotion de l'autonomisation des pauvres. L'objectif étant d'offrir aux pauvres des opportunités puis de les laisser lutter par eux-mêmes contre la pauvreté, ce qui tend à remettre les pauvres en première ligne. Le débat sur l'aide au développement cristallise deux principales visions du développement, à la suite de l'effondrement du consensus de Washington. En ce sens, il constitue le seul bagage théorique auquel Banerjee et Duflo font référence, et le seul face auquel leur approche tente d'apporter des réponses. Ce premier chapitre m'a alors permis de définir le cadre théorique auquel Banerjee et Duflo se réfèrent.

Chapitre 2 : La démarche du J-PAL à l'aune de ses expériences sur le paludisme

0. Introduction

« More than one million African children, and perhaps as many as three millions, succumb to malaria each year. This horrific catastrophe occurs despite the fact that the disease is partly preventable – through the use of bed nets and other environmental controls that do not reach the impoverished villages of Malawi and most of the rest of the continent – and completely treatable. There is simply no conceivable excuse for this disease to be taking millions lives each year. » (Sachs, 2005, 7)

« Mais si les moustiquaires constituent un remède si efficace pourquoi les planistes n'en ont-ils pas équipé les pauvres auparavant ? Hélas, parce que ni les célébrités ni les administrateurs de l'aide humanitaire ne savent comment faire parvenir ces moustiquaires à destination et que celles-ci sont souvent récupérées par le marché noir et disparaissent rapidement des hôpitaux – quand elles ne finissent pas comme filets de pêche ou voile de mariée ... » (Easterly, 2006a, 24)

En 2001, Jeffrey Sachs publie un article écrit en collaboration avec John Gallup (Gallup et Sachs, 2001) cherchant à montrer l'étroit lien qui unit le paludisme et la pauvreté. Les deux auteurs constatent que les pays les plus pauvres sont aussi ceux où le paludisme est le plus important. La question est alors de savoir si le paludisme cause la pauvreté. Sachs et Gallup concluent que le paludisme est l'une des causes de la pauvreté²⁸⁶. Lutter contre le paludisme revient donc à lutter contre la pauvreté. La lutte contre le paludisme est d'ailleurs un des objectifs du millénaire pour le développement²⁸⁷. L'un des moyens les plus efficace de prévenir le paludisme est l'utilisation de moustiquaires imprégnées d'insecticides²⁸⁸. Pour Jeffrey Sachs ces moustiquaires représentent un moyen peu coûteux et efficace de lutter contre la pauvreté : la communauté internationale doit donc financer massivement ce bien²⁸⁹. Dans la même lignée, l'Organisation Mondiale pour la Santé (OMS), publie en 2007 son

²⁸⁶ « This provides the most direct evidence of the continuing importance of malaria as a cause of poverty ». (Gallup et Sachs, 2001, 87)

²⁸⁷ La lutte contre le paludisme est la cible 3 de l'objectif 6 du Millénaire, les deux premières cibles concernent la lutte contre le VIH/SIDA. J'ai développé les différents objectifs du millénaire plus haut.

²⁸⁸ Plusieurs essais cliniques ont montré l'efficacité des moustiquaires imprégnées d'insecticides dans la prévention du paludisme. Voir par exemple (Lengeler, 1998 ; Schellenberg *et al.*, 2001) pour le lien entre moustiquaire imprégné d'insecticides et la survie des enfants, pour la prévention chez les femmes enceintes voir (Dolan *et al.*, 1993 ; D'Alessandro, 1994 ; Ter Kuile *et al.*, 2003)

²⁸⁹ Voir Sachs (2002)

premier rapport sur le paludisme²⁹⁰ dans lequel elle prône explicitement la subvention intégrale des moustiquaires pour tous les pays où le paludisme est endémique.

A l'inverse des positions de Sachs ou de l'OMS, comme cela a été évoqué dans le chapitre précédent, William Easterly s'élève contre le principe d'une subvention massive de ces moustiquaires, considérant que ces dernières finiront au marché noir, en voiles de mariées ou encore en filets de pêche. Selon lui, les moustiquaires doivent être en partie payantes afin que les individus accordent davantage de valeur à celles-ci et que de ce fait ils les utilisent au mieux. La question des moustiquaires est l'illustration la plus brûlante de ce débat sur l'aide au développement et d'une certaine façon le cristallise. Pour offrir des réponses à la question de l'aide au développement, Banerjee et Duflo expliquent qu'il ne faut pas rester dans l'abstraction, qu'il faut du concret. Par exemple, la question de la subvention des moustiquaires ne peut être tranchée que si on la divise en trois autres questions relevant de faits précis :

« Pour éclairer ce débat, il faut répondre à trois questions. Tout d'abord, si les gens doivent payer le prix de marché (ou presque) pour acheter une moustiquaire, ne préféreront-ils pas s'en passer ? Deuxièmement, si les moustiquaires leur sont distribuées gratuitement, ou à un prix réduit, les gens les utiliseront-ils, ou seront-elles gâchées ? Troisièmement, après avoir obtenu des moustiquaires à prix réduit, seront-ils plus ou moins prêts à en acheter une autre si les subventions à l'achat sont ensuite réduites ? » (Banerjee et Duflo, 2011, 27-26)

Pascaline Dupas, chercheuse au J-PAL, mène plusieurs expériences au Kenya²⁹¹ afin de répondre à ces différentes questions²⁹². Elle montre que même gratuites ces moustiquaires ne sont pas suffisamment utilisées par les pauvres pour éradiquer le paludisme. Elle teste, ensuite, des dispositifs incitatifs cherchant à favoriser l'utilisation des moustiquaires. Ces derniers n'ont aucun effet sur l'utilisation des moustiquaires par les pauvres. L'objectif de ce chapitre est de rendre compte de ces expériences afin d'évaluer la portée des évaluations du J-PAL face au débat sur l'aide au développement. Je choisis de me centrer sur les expériences

²⁹⁰Voir (OMS, 2007). L'OMS publie désormais un rapport chaque année sur le paludisme et a publié la seconde édition du rapport de 2007, voir (OMS, 2013).

²⁹¹ Voir (Cohen et Dupas, 2010), (Dupas, 2010), (Cohen, Dupas et Schaner, 2013), (Bhattacharya, Dupas et Kanaya, 2013) et (Dupas, 2011).

²⁹² Dupas souligne d'ailleurs la virulence de ce débat ainsi que les deux positions d'Easterly et de Sachs : « *The manufacture of ITNs is expensive, and the question of how much to subsidize them is at the center of a very vivid debate in the international community, opposing proponents of free distribution (Sachs 2005; World Health Organization [WHO] 2007) to advocates of cost-sharing (PSI 2003; Easterly 2006).* » (Cohen et Dupas, 2010, 3)

de Dupas car elles portent sur l'illustration de ce débat autour du paludisme. De plus, les expériences de Dupas concernant le paludisme me permettront de définir précisément la façon de procéder du J-PAL afin de déterminer comment sont collectées les preuves et comment elles sont utilisées.

Je commencerai, tout d'abord, par présenter la première expérience, réalisée par Cohen et Dupas (2010). Cette expérience pointe une énigme : même lorsqu'elles sont gratuites les pauvres n'utilisent pas suffisamment les moustiquaires. Cela me permettra de montrer que cette énigme retourne la question de l'aide posée par Sachs et Easterly. Ensuite, je développerai la deuxième expérience que mène Dupas (2009) en vue de tester un dispositif cherchant à amener les pauvres à utiliser les moustiquaires. Je montrerai que ce dispositif n'a aucun impact sur l'utilisation des moustiquaires. Puis, je détaillerai la troisième expérience que conduit Dupas (2010), expérience cherchant à tester un autre dispositif pour favoriser l'utilisation des moustiquaires. Ce dispositif est lui aussi inefficace. Enfin, je montrerai que les deux dispositifs testés ne permettent pas de comprendre pourquoi les pauvres n'utilisent pas les moustiquaires, et je développerai la principale explication offerte par Dupas (2011) sur la faible utilisation des moustiquaires : l'irrationalité des pauvres.

1. Paludisme et subvention : l'énigme

La première expérience randomisée menée par Jessica Cohen et Pascaline Dupas (2010) a pour objectif de savoir si afin d'être efficaces les moustiquaires doivent être subventionnées intégralement ou si elles doivent être payantes, même à un très faible prix. La subvention intégrale des moustiquaires est justifiée par le fait qu'elles possèdent une externalité positive. Faire payer même un faible prix permettrait, à l'inverse, de réduire le gaspillage de tels biens. Partager le prix des moustiquaires impliquerait trois effets positifs²⁹³. Le premier est un effet de sélection : les individus qui n'accordent pas de valeur à ce bien ne paieraient donc pas pour ce dernier, alors que ceux qui lui attribuent une utilité seraient prêts à payer. Il y a ainsi une sélection presque mécanique entre ceux qui vont réellement utiliser le bien et ceux qui vont le

²⁹³ « There are three possible effects of positive prices on the likelihood that people who acquire the product use it appropriately. First, a selection effect : charging a positive price could select out those who do not value the good and place it only in the hands of those who are likely to use it (Oster 1995 ; Population Services International [PSI] 2003 ; Ashraf, Berry, and Shapiro forthcoming). Second, a psychological effect : paying a positive price for a good could induce people to use it more if they exhibited « sunk cost » (Thaler 1980 ; Arkes and Blumer 1985). Third, higher prices may encourage usage if they are interpreted as a signal of higher quality (Bjorvell and Riordan 1991 ; Riley 2001). » (Cohen et Dupas, 2010, 2)

gaspiller (1). Un deuxième effet est un « effet psychologique » : payer un faible prix devrait inciter les individus à utiliser le bien (2). Enfin, le troisième effet traduit le fait que payer un prix plus élevé peut donner un signal en termes de qualité, ce qui inciterait les individus à utiliser davantage le bien (3).

L'objectif de Cohen et Dupas consiste donc à déterminer laquelle des deux propositions est la plus efficace dans l'utilisation des moustiquaires. Pour Cohen et Dupas les différents niveaux de subvention dépendent de quatre facteurs clefs :

- L'élasticité de la demande au prix
- L'usage des moustiquaires
- L'existence d'un effet psychologique
- L'impact d'une variation de prix sur la vulnérabilité, c'est à dire sur les besoins

C'est l'importance de ces quatre facteurs clefs que Cohen et Dupas (2010) cherchent à évaluer. C'est donc à travers ces derniers qu'elles cherchent à offrir une réponse à la question de l'aide au développement. Je commencerai par présenter le contexte et le cadre de l'expérience (1.1), puis je développerai les résultats de cette expérience (1.2), enfin je montrerai que cette première expérience soulève une énigme : la faible utilisation des moustiquaires par les pauvres (1.3).

1.1 Contexte et mise en place d'une première expérience : les moustiquaires doivent-elles être gratuites ou payantes ?

La première expérience que mènent Jessica Cohen et Pascaline Dupas (2010) a lieu au Kenya de mars à mai 2007. Le Kenya est un pays d'Afrique particulièrement touché par le paludisme²⁹⁴. En 2001, le gouvernement Kenyan lance un programme d'envergure en matière de lutte contre le paludisme et le ministère de la santé avec l'aide de 17 millions de dollars provenant du Fond mondial pour lutter contre le sida, la tuberculose et le paludisme, distribue plus de 3 millions de moustiquaires imprégnées d'insecticides²⁹⁵. L'expérience de Cohen et

²⁹⁴L'Afrique et l'Inde sont les deux endroits où le paludisme est le plus présent et le plus meurtrier. Au Kenya, le paludisme tue un enfant sur quatre, (Banque Mondiale, 2006)

²⁹⁵Une étude de Guyatt *et al.* (2002) montre qu'au Kenya, suite au programme lancé par le ministère de la santé, les moustiquaires arrivent bien à destination. Autrement dit, il n'y a pas de gâchis et l'aide arrive à destination ce qui était la principale critique adressée à l'aide au développement, notamment par Easterly (2006).

Dupas (2010) se concentre sur des femmes enceintes dans 4 districts de l'ouest du Kenya (Busia, Bungoma, Butere et Mumias). Les femmes enceintes sont particulièrement vulnérables face au paludisme car beaucoup de parasites peuvent s'accumuler dans le placenta, causant alors une importante anémie et une naissance prématurée. De plus, chez les enfants de moins de quatre ans la principale transmission du paludisme se fait de la mère à l'enfant pendant la grossesse. L'expérience cible la distribution des moustiquaires pour les femmes enceintes qui se rendent dans les cliniques des districts pour le suivi de leurs grossesses.

Ces quatre districts comptent 70 cliniques : 53 publiques et 17 privées. L'expérience ne retient que 20 cliniques (toutes publiques), elles sont choisies par Cohen et Dupas en fonction de leurs tailles, des services qu'elles offrent, de leurs situations géographiques, afin d'assurer une certaine homogénéité. Ces 20 cliniques sont assignées, de façon aléatoire, à cinq groupes distincts. Le premier groupe, groupe de contrôle, compte 4 cliniques (1). Cinq cliniques font partie du groupe où les moustiquaires sont subventionnées à 100% (2) et pour cinq autres la prise en charge financière des moustiquaires est de 97,5% (3). Dans ce groupe les femmes ont alors à payer 10 Shillings kenyans (0,12 dollar) pour bénéficier d'une moustiquaire. Au sein d'un quatrième groupe la part subventionnée est de 95% et cela pour 3 cliniques (4). Pour ces trois cliniques les femmes doivent payer 20 Shillings kenyans (0,23 dollar). Les trois dernières cliniques constituent le dernier groupe et au sein de ce dernier la part subventionnée est de 90% (5). Les femmes de ce dernier groupe doivent donc payer 40 Shillings kenyans (0,60 dollar).

Afin que le programme soit bien implanté et que le personnel soignant respecte les différents niveaux de subvention, Cohen et Dupas leur ont donné des incitations financières²⁹⁶. Chaque mois où le programme a été implanté et qu'aucune mauvaise gestion n'a été observée la clinique recevait une prime en espèce d'une valeur de 5000 Shillings kenyan (57 dollars) ou un équipement de son choix de la même valeur. Afin de vérifier que les différents niveaux de subvention ont bien été respectés les cliniques étaient visitées de façon aléatoire, tout comme les bénéficiaires des moustiquaires. Une mauvaise gestion a été

²⁹⁶ Cela soulève une question éminemment éthique inhérente à la randomisation. Il est difficile, dans un pays où tout le monde aurait besoin de moustiquaires, de refuser d'offrir gratuitement ces dernières. Cela est d'autant plus difficile pour le personnel soignant, qui à l'inverse du J-PAL n'est pas tenu par une volonté de rigueur scientifique. Lorsque Cohen et Dupas (2010) développent le rôle des incitations financières pour le personnel soignant, elles ne font absolument pas référence à cette question.

observée dans quatre des onze cliniques où le prix des moustiquaires était positif. Cette mauvaise gestion n'a pas impacté les différents niveaux de subvention et donc l'expérience, puisque le personnel soignant n'a pas fait payer un prix plus élevé ou inférieur aux femmes enceintes mais a fourni des moustiquaires à des personnes qui n'étaient pas éligibles pour l'expérimentation. Aucune mauvaise gestion n'a été observée dans le groupe de contrôle et dans le groupe où les moustiquaires étaient gratuites. Pour les cliniques où le prix était positif, un second niveau de randomisation a été ajouté cinq semaines après l'implantation du programme. Les femmes qui faisaient partie des groupes où les moustiquaires étaient payantes pouvaient participer à une loterie. Chaque femme devait piocher une enveloppe. Il y avait trois types d'enveloppes : une vide (1), une avec écrit « moustiquaire gratuite » (2), et une où le prix était toujours positif mais inférieur à celui que payaient les autres femmes enceintes (3).

1.2 Résultats : le rôle trouble de la subvention des moustiquaires

Les résultats de Cohen et Dupas (2010) s'articulent autour des quatre facteurs clefs qu'elles cherchent à évaluer. Le premier niveau, l'élasticité de la demande au prix (1), rend compte d'une demande très élastique au prix. Les auteurs concluent alors:

« We find that uptake of ITN's²⁹⁷ drops significantly at modest cost-sharing prices. Demands drop by 60 percent when the price is increased from zero to 40 Ksh (\$0,60). » (Cohen et Dupas, 2010, 4)

« The decrease in demand for an increase in price from 0 to 10 Ksh is estimated at 7 percentage points (...). An increase in price from 20 to 40 Ksh leads to a 43 percentage point drop in demand. » (Cohen et Dupas, 2010, 25)

La demande de moustiquaires est donc très élastique au prix: lorsque la part subventionnée passe de 100% à 97,5% la demande diminue de 7 points de pourcentage, et lorsque la part subventionnée passe de 95% à 90% la demande décroît de 43 points de pourcentage, enfin lorsque la part subventionnée passe de 100% à 90% la demande chute de 60%. La forte élasticité de la demande au prix, ici, joue en faveur d'une subvention à 100% des moustiquaires. Le résultat extrêmement frappant de cette expérience concerne le

²⁹⁷ITN's renvoie à *insecticide-treated bed nets* donc aux moustiquaires imprégnées d'insecticides.

deuxième facteur clef, l'utilisation des moustiquaires²⁹⁸ (2). Cette dernière est très faible, malgré la gratuité des moustiquaires, comme le soulignent Cohen et Dupas :

« Overall, one might be surprised that the level of net usage is not higher than 60%. »
(Cohen et Dupas, 2010, 30)

« Overall, our results suggest that, at least in the Kenyan context, positive prices do not help generate higher usage intensity than free distribution. » (Ibid.)

Les femmes enceintes sont très sensibles au prix, cependant même lorsque les moustiquaires sont gratuites elles sont très peu utilisées. De plus, les auteurs montrent que faire payer ne permet pas une plus grande utilisation du bien, ce qui conduit alors au troisième facteur clef défini par Cohen et Dupas (2010) : l'existence d'un effet psychologique (3). L'idée est que lorsque l'on paye pour un bien on aurait tendance à lui accorder davantage de valeur et donc, ici, à plus l'utiliser. Cohen et Dupas (2010) montrent qu'il n'y a pas d'effet psychologique²⁹⁹. Cela pointe une énigme : les raisons de la très faible utilisation des moustiquaires. L'absence d'effet psychologique annule une des explications économiques potentielles de cette faible utilisation. Le point de vue de William Easterly est donc infirmé, rendre les moustiquaires payantes ne favorise ni n'améliore leur utilisation. Cependant, le fait que les moustiquaires soient payantes évite-t-il, comme l'envisage Easterly, un important gaspillage ? Le fait que les moustiquaires soient payantes va-t-il permettre de cibler la population la plus vulnérable - ici les femmes enceintes ?

Si le prix de la moustiquaire est trop élevé les femmes, même les plus vulnérables, ne pourront pas en acheter, c'est pourquoi le niveau de la subvention doit quoi qu'il arrive absolument en tenir compte. C'est le quatrième facteur clef soulevé par Cohen et Dupas (4).

²⁹⁸ Afin de vérifier si les moustiquaires étaient utilisées, les chercheurs ont rendu visite à 246 femmes qui ont bénéficié gratuitement ou qui ont dû acheter les moustiquaires. Elles étaient alors interviewées et il leur était demandé de montrer les moustiquaires et de dire qui dormaient dessous : *« a random sample of 246 prenatal clients who had purchased/received a net through the program were selected to be visited at their homes three to ten weeks after their net purchases. All home visits were conducted within three weeks in July 2007 to ensure that all respondents faced the same environment (especially in terms of malaria seasonality) at the time of the follow-up. Of this subsample, 92% (226 women) were found and consented to be interviewed. During the home visits, respondents were asked to show the net, whether they had started using it, and who was sleeping under it. »* (Cohen et Dupas, 2010, 14)

²⁹⁹ *« We find no psychological effect on the price or the act of paying on usage, as expected from the earlier result that there is no overall effect of prices on usage. »* (Cohen et Dupas, 2010, 32)

Cohen et Dupas montrent qu'il n'y a pas d'effet de sélection³⁰⁰. Les femmes les plus vulnérables ne sont donc pas celles qui sont prêtes à payer le prix le plus élevé pour bénéficier des moustiquaires, ce qui réfute l'effet de sélection anticipé par Easterly. Enfin, Cohen et Dupas mènent une analyse coûts/bénéfices évaluant le prix de chaque stratégie (les différents niveaux de subvention) en termes de vies d'enfants sauvées. La conclusion de cette analyse est la suivante :

« The general conclusion of this cost-effectiveness exercise is thus that cost-sharing is at best marginally more cost-effective than free distribution, but the free distribution leads to many more lives saved. » (Cohen et Dupas, 2010, 41)

Autrement dit, il est moins coûteux de faire payer les moustiquaires et cela tend à une plus grande efficacité ; cependant moins de vies sont sauvées selon ce schéma. La recommandation politique est donc ambiguë et laisse place à un impératif moral, souvent sous-jacent à de telles analyses : quel est le prix de la vie - des vies sauvées ? La conclusion reste ici opaque, puisque Cohen et Dupas ne prennent pas partie.

1.3 La première pièce d'un « puzzle³⁰¹ » sur l'aide et la santé

Cette première expérience conduit à une énigme, et pose un problème que le premier débat théorique sur la gratuité ou non des moustiquaires n'avait pas soulevé : les moustiquaires sont insuffisamment utilisées. Cette expérience a permis de répondre en partie à la question qu'elle posait : les moustiquaires doivent-elles être gratuites ou payantes ? La réponse est double. La demande étant très élastique au prix il est plus efficace que les moustiquaires soient gratuites. Cela tend à confirmer la position de Jeffrey Sachs plutôt que celle de William Easterly. L'analyse coût/efficacité menée par Bhattacharya, Dupas et

³⁰⁰ « *Women who pay a higher price do not appear to be sicker than the average prenatal clients in the area.* » (Cohen et Dupas, 2010, 35)

³⁰¹ Cohen et Dupas (2010) parlent de « puzzle », j'ai traduit ce terme par énigme. Cohen et Dupas (2010) utilisent ce terme sans jamais faire référence à la théorie de Kuhn (1962). Pour Kuhn (1962), la science cherche à résoudre des énigmes (« puzzle-solving »). Il montre que la science fonctionne selon une succession de paradigmes. Un paradigme domine une période donnée, jusqu'à qu'un « puzzle », une énigme apparaisse. Cette énigme tend à être résolue par des scientifiques dissidents, ne s'apparentant pas au paradigme dominant, et de ce fait ils renversent le paradigme en imposant un nouveau. Cette analyse peut être particulièrement éclairante ici, puisque Cohen et Dupas (2010) font apparaître une énigme ; même, si l'on verra, qu'elles ne chercheront pas à y proposer d'explications, ni à proposer de nouveau paradigme. Je conserve, pourtant, pour le titre de cette sous-partie le terme de *puzzle* et non d'énigme. Cela, car je choisis ici de soulever l'ambiguïté du terme utilisé par Cohen et Dupas (2010) ; d'une part, pour ce que ce terme évoque de l'analyse de Kuhn (1962) et d'autre part, pour ce qu'il évoque en français - se référant ici explicitement au jeu du puzzle.

Kanaya (2013) opacifie cependant ce résultat, puisqu'elle montre qu'une subvention partagée est marginalement meilleure, mais que le nombre de vies d'enfants sauvées y est moindre qu'avec une subvention intégrale. Une subvention intégrale semble donc aller dans le sens d'une meilleure décision politique. Cependant, malgré une subvention intégrale les moustiquaires sont faiblement utilisées par les femmes enceintes. La dimension extrêmement intéressante et novatrice qu'offre l'approche du J-PAL au sein du débat sur l'aide au développement est ici parfaitement illustrée : il ne suffit pas seulement de savoir si l'aide internationale doit être massive, inexistante ou faible, ou si elle arrive bien à destination (ce qui était un des plus grands reproches adressé à l'aide). Mais que se passe-t-il lorsqu'elle arrive à destination ? Est-elle utilisée comme anticipé par les décideurs politiques ? L'expérience de Cohen et Dupas (2010) est emblématique, puisqu'elle permet de mettre cela en exergue. Elle montre qu'il est plus efficace que les moustiquaires soient gratuites pour les femmes enceintes mais que cela ne suffit pas à une « bonne » utilisation de celles-ci, au sens où cela n'a pas ou peu d'effets sur la réduction du paludisme. Le seul accès à des moustiquaires ne suffit pas à réduire le paludisme, comme cela était supposé par Sachs. L'expérience de Cohen et Dupas (2010) pose donc question. Elle introduit un fait non anticipé, non envisagé, non observé. C'est en ce sens qu'elle répond en partie à la question posée par le débat entre Sachs et Easterly sur le rôle de l'aide internationale dans la lutte contre le paludisme. Elle répond en partie car elle soulève une autre question : comment faire pour que les moustiquaires soient suffisamment utilisées afin d'éradiquer le paludisme dans les pays pauvres ? En ce sens elle ne donne pas de réelle réponse ou recommandation politique mais soulève un enjeu important.

Cette expérience me semble extrêmement représentative de ce que les chercheurs du J-PAL souhaitent faire. Elle illustre le rôle des expériences et le statut de la théorie voulus par Banerjee et Duflo. Les théories auxquelles l'expérience fait face (celles de Sachs et d'Easterly) ne permettent pas de tirer de conclusions valables concernant le comportement des pauvres. Les chercheurs du J-PAL ne veulent pas définir une première théorie, une théorie *ex-ante*. L'expérience doit faire émerger la théorie. Cela est rendu extrêmement visible par l'expérience de Cohen et Dupas (2010). Cette dernière représente un premier stade. Elle pose un problème – la faible utilisation des moustiquaires – que la théorie n'avait pas soulevé. Cohen et Dupas (2010) concluent d'ailleurs leur expérience sur ce thème :

« Our experimental results should thus be seen as one piece in the puzzle of how to increase uptake of effective, externality-generating health products in resource-poor setting. » (Cohen et Dupas, 2010, 43)

Cette expérience est donc une première pièce d'un « puzzle » plus large. Il apparaît ensuite nécessaire d'étudier l'utilisation de types de biens en matière de santé, comme les moustiquaires, qui ont des externalités positives. L'objectif est de trouver des moyens, des dispositifs permettant de favoriser, d'améliorer cette utilisation.

2. Un premier dispositif : le nudge– effet de cadrage et engagement oral

Pascaline Dupas émet plusieurs hypothèses sur la faible utilisation des moustiquaires montrée dans l'expérience qu'elle a menée avec Jessica Cohen :

« In this framework, the low take-up observed at relatively moderate prices by Cohen Dupas (2008) could be due to people underestimating the expected benefits of investing in prevention; or due to people being credit-constrained and unable to pay the cost up front. It is also possible that the standard model does not apply, because people have time inconsistency preferences or because they are uncertain about their own preferences and rely on external cues to resolve their own uncertainty when they need to make a decision. » (Dupas, 2009, 224)

L'insuffisante utilisation des moustiquaires peut être due à quatre facteurs :

1. la sous-estimation des bénéfices potentiels de l'utilisation des moustiquaires ;
2. la contrainte de crédit des individus ou le fait qu'ils n'étaient pas en mesure de payer, même à un très faible coût, les moustiquaires ;
3. l'incohérence temporelle des individus ;
4. l'incertitude des individus face à leurs préférences et aux décisions qu'ils doivent prendre.

Dupas (2009) évoque ensuite la littérature grandissante en psychologie³⁰² montrant que les décisions des individus sont souvent très influencées par le cadrage ou les signaux³⁰³ quand il s'agit d'investir dans un produit ou de l'utiliser. Afin d'augmenter l'utilisation des moustiquaires, Pascaline Dupas mène une seconde expérience, toujours au Kenya. Elle souhaite y évaluer l'impact de deux interventions visant à changer le comportement des individus (ici les amener à utiliser les moustiquaires) :

³⁰² Voir : (1) (Bertrand, Karlan, Mullainathan, Shafir et Zinman, 2008), (2) (Daryl, 1967), (3) (Snyder et Cunningham, 1975), (4) (Duncan, 1990).

³⁰³ La notion de cadre (*frame*) est développée par Kahneman et Tversky (1981). Ils montrent, à travers une expérience, que la façon dont sont présentées les problèmes joue un rôle important dans la prise de décisions. Dans leur expérience, il est demandé à un premier groupe d'étudiants de choisir entre sauver 200 personnes sur 600 et une chance sur trois de sauver les 600 personnes. Dans un second groupe, on propose aux étudiants de choisir entre laisser 400 personnes mourir sur les 600 personnes et deux chances sur trois de voir mourir 600 personnes. Les quatre possibilités possèdent la même espérance mathématique. Pour autant, dans le premier groupe les étudiants privilégient l'option « sauver 200 personnes » et dans le second groupe les étudiants choisissent d'avoir deux chances sur trois de voir mourir 600 personnes. La formulation du problème soit « sauver des vies » ou « laisser mourir » transforme la décision des étudiants. Il y a alors un effet de cadrage.

« This paper tests the effect on the take-up of a preventive health of two interventions based on behavioral models derived from psychology: varying the framing of the perceived benefits; and having people verbally commit to purchase the product. » (Dupas, 2009, 224)

L'objectif d'une telle expérience est donc de tester deux hypothèses pouvant expliquer la faible utilisation des moustiquaires observées chez les femmes enceintes dans l'expérience précédente menée par Dupas. Les deux hypothèses mises en avant sur les quatre envisagées par Dupas, sont que la faible utilisation des moustiquaires est due à une sous-estimation des bénéfices de la part des individus (1) et à l'incertitude des préférences des individus (2). Afin de contrecarrer ces comportements et donc d'amener à utiliser les moustiquaires, Dupas envisage deux dispositifs changeant le cadre (*frame*). Le premier dispositif cherche à contrecarrer la sous-estimation des bénéfices des individus. Dupas (2009) procède pour cela à un effet de cadrage. Le second dispositif vise à réduire l'incertitude des individus face à leurs préférences ; elle les fait donc s'engager oralement sur le fait qu'ils vont acheter une moustiquaire. Je commencerai par définir le cadre et la mise en place de cette expérience (2.1) ; puis, je développerai les résultats obtenus par Dupas (2009) (2.2) ; enfin, je montrerai que les dispositifs de nudging ayant échoué à favoriser l'utilisation des moustiquaires, l'énigme posée par la première expérience de Cohen et Dupas (2010) reste irrésolue et tend à en faire émerger une seconde (2.3).

2.1 Cadre expérimental et implantation

Cette seconde expérience a toujours lieu au Kenya et concerne plus de 1200 ménages entre le mois d'avril et le mois d'octobre 2007. Ces derniers sont sélectionnés à partir des registres scolaires. Ces différents ménages sont ensuite assignés aléatoirement à des niveaux de subvention différents. Le niveau de subvention peut varier de 40% à 100%. Le premier groupe est le groupe « marketing », le second le groupe « engagement oral », et le dernier le groupe « ciblé ». Au sein du premier groupe, les ménages reçoivent un message publicitaire concernant les moustiquaires. Dans le second groupe, les ménages s'engagent oralement à acheter les moustiquaires. Enfin dans le troisième groupe différents niveaux de subvention sont appliqués, afin de relier les résultats de cette expérience avec la précédente.

2.2 L'échec du Nudge

Dans les résultats obtenus par Dupas je distingue quatre dimensions : l'utilisation des moustiquaires par les ménages les plus aisés de l'expérience (1), l'élasticité de la demande au prix des moustiquaires (2), les effets des dispositifs de nudge sur l'utilisation des moustiquaires (3), et les effets des dispositifs de nudge sur le groupe ciblé – les femmes – (4). L'expérience montre que les ménages les plus éduqués et les plus riches dorment plus sous une moustiquaire que les autres ménages (1)³⁰⁴. En second lieu, la demande est toujours très élastique au prix mais l'est un peu moins que ne l'était celle de la première expérience menée par Cohen et Dupas (2):

« An increase in price from \$0 to \$1 leads to a drop of 35 percentage points in take-up, and an increase from \$1 to \$2 leads to further drop of 25 percentage points (...) The price elasticity observed here is also lower than that observed in Cohen and Dupas (2008), which found a drop in take-up of 60 percent when the price increases from \$0 to \$0.60. This is probably because households in this experiment had up to three months to redeem their voucher, and therefore time to save for it. The time taken to redeem the voucher increased with the price of the net: from 3 days on average in the free net group to 35 days in the 70-90 Ksh price group, and to 54 days in the highest price group. » (Dupas, 2009, 226-227)

L'explication donnée par Dupas (2009) pour la plus faible élasticité de la demande renvoie au fait que les ménages disposaient de trois mois pour acheter leur coupon afin de bénéficier des moustiquaires ensuite. Ces trois mois offraient alors la possibilité d'épargner un peu en laissant plus de temps, et donc d'être en mesure d'acheter les moustiquaires. Cette possibilité d'épargne expliquerait alors la plus faible élasticité de la demande, les ménages étant rassurés de savoir qu'ils pourraient réunir l'argent nécessaire à l'achat des moustiquaires sans crédit. Par contre les deux dispositifs testés dans le but de contrecarrer la sous-estimation des bénéfices et l'incertitude des individus face à leurs préférences n'ont pas fonctionné, ils n'ont eu aucun effet (3) :

« Neither of the two framing options (health or financial) had any impact at all (the coefficients are very close to zero, and the standard errors are also small, which rules out even small effects). Coefficients on the interaction between framing and gender targeting are also indistinguishable from zero (data now shown). Likewise, the verbal commitment treatment had no impact (...). Interestingly, the education level of the household members does not predict take-up. » (Dupas, 2009, 228)

³⁰⁴ « Wealthier and more educated households had a significantly higher share of their members sleeping under a bed net at the onset of the study. » (Dupas, 2009, 226)

Ces deux dispositifs n'ont donc pas permis d'accroître l'achat de moustiquaires. Cela revient-il à invalider l'hypothèse de Dupas (2009), selon laquelle les individus n'utiliseraient pas suffisamment les moustiquaires car ils sous-estimeraient les bénéfices qu'ils pourraient en tirer ou qu'ils seraient incertains face à leurs préférences? En d'autres termes, les résultats obtenus par cette expérience ne permettent pas de comprendre pourquoi les individus n'utilisent pas suffisamment les moustiquaires. Cette expérience ne permet pas non plus de savoir quel dispositif permettrait une plus grande utilisation puisque les deux testés n'ont aucun effet. Le dernier résultat développé par Dupas (2009) concerne le groupe où les femmes sont ciblées (4) :

« Targeting female heads somewhat increases the likelihood that a child sleeps under the net, but does not decrease the likelihood that a household heads sleeps under it, suggesting that women are more likely to share the net with a child than men (or more likely to report they do), though insignificantly. » (Dupas, 2009, 228)

Les femmes semblent être davantage enclines à partager leurs moustiquaires avec leurs enfants que les hommes et favorisent donc l'utilisation des moustiquaires pour les enfants. Cibler les femmes apparaît ainsi comme une stratégie efficace, pour favoriser le fait que les enfants dorment sous une moustiquaire. Dupas (2009) souligne que cela peut être nuancé par le fait que les femmes tendent à reporter, plus souvent que les hommes, partager leurs moustiquaires avec leurs enfants.

2.3 Une énigme toujours irrésolue ?

Cette deuxième expérience ne permet pas de répondre à l'énigme posée par la première. En d'autres termes, elle n'aide pas à comprendre pourquoi les femmes enceintes n'utilisent pas suffisamment les moustiquaires, et ne conduit pas à la mise en place d'un dispositif efficace pour améliorer l'utilisation de celles-ci. C'est donc une autre énigme qui apparaît :

« These results are not surprising given the (deadly) stakes involved in the decision to invest or not in malaria prevention, and suggest that liquidity constraints may be the main barrier to investments in malaria prevention, consistent with recent research in India showing that while 2 percents of households purchase ITN in cash, 59 percents purchase at least one when ITNs are offered on credit (Alessandro Tarozzi et al. 2009). Given the private returns to usage of LL-ITNs, it is surprising that households do not manage to overcome their liquidity constraint and invest in the product even at moderate prices. This puzzling

observation has been made for other human capital investments in other contexts (reviewed in AlakaHolla and Kremer 2008), and suggests that households are highly saving-constrained or highly present biased (or both), causing them to underinvest in health compare to what would be privately optimal. More research is needed to understand the respective role of these factors. » (Dupas, 2009, 230)

Le principal problème serait donc la contrainte de crédit, les individus pauvres n'achètent pas de moustiquaires car ils n'ont pas assez d'argent. Dupas (2009) souligne ensuite qu'il est surprenant, vu les effets positifs des moustiquaires, que les pauvres n'arrivent pas à dépasser cette contrainte de crédit. Le sous-investissement dans les moustiquaires serait dû à deux possibles raisons : les pauvres rencontrent des difficultés à épargner (1), et ils souffrent d'incohérence temporelle (2). Dupas (2009) suggère de continuer les recherches sur ces deux dimensions, ce qui permettrait alors d'améliorer l'utilisation des moustiquaires et pourrait donner une explication à cette faible utilisation.

3. Un second dispositif de *nudge* : effets d'apprentissage et effets sociaux

Dupas mène une troisième expérience toujours au Kenya, dans le but de savoir : « *Whether one-time subsidies for such technologies increase or dampen private investments in them in the long run* » (Dupas, 2010, 1). L'objectif est ici de déterminer s'il existe pour la population étudiée un effet d'apprentissage après avoir bénéficié gratuitement une première fois des moustiquaires. Dupas (2010) montre que deux principaux points de vue s'opposent face à cette question. Un premier point de vue est qu'il existe un effet d'apprentissage important³⁰⁵. En effet, lorsque les individus ont bénéficié une première fois d'un bien gratuitement, ils s'y sont faits, ils se sont habitués à ce dernier, auront plus de facilités à l'acheter la fois suivante et donc à l'utiliser sur le long terme. Cet effet d'apprentissage permettrait donc de contrecarrer la faible utilisation des moustiquaires en permettant de combattre la sous-estimation des bénéfices que supposent Cohen et Dupas (2010) chez les femmes enceintes de sa première expérience. Cependant, le fait d'avoir bénéficié d'une moustiquaire gratuitement une première fois peut aussi freiner l'achat ou l'utilisation de celle-ci la seconde fois. Ici Dupas (2010) évoque les travaux de Simonson et Tversky (1992) autour du concept « d'effet de contraste ». Ceux-ci montrent qu'un individu dispose d'une référence

³⁰⁵ « *Free or highly subsidized distribution of a product in the short run may increase demand in the long run if the product is an experience good. Beneficiaries of a free or highly subsidized sample will be more willing to pay for a replacement after experiencing the benefits and learning the true value of the product if they previously had underestimated these benefits. This learning might trickle down to others in the community (those ineligible for the subsidy) and increase the overall willingness to pay in the population as knowledge of the true value of the product diffuses. » (Dupas, 2010, 1)*

financière, c'est à dire que la première fois qu'il a bénéficié d'une moustiquaire sera sa référence financière pour ce bien. De ce fait, lorsqu'on lui propose le même bien à un prix plus élevé, il ne sera pas enclin à l'acquérir car le prix sera supérieur à sa référence financière. Dans cette perspective, ceux qui ont bénéficié de la subvention ont intégré la gratuité d'un tel type de bien et ne seront donc pas prêts à payer pour l'obtenir la seconde fois.

Afin d'évaluer la portée de ces deux propositions et donc de déterminer s'il existe un effet d'apprentissage positif grâce à la subvention des moustiquaires, Dupas (2010) construit dans un premier temps un modèle afin de comprendre le rôle des prix dans l'adoption de nouvelles technologies en matière de santé. Dans un second temps, Dupas (2010) souhaite tester ce même modèle à l'aide d'une expérience randomisée, sur une technologie particulière – les moustiquaires. Le principal trait caractéristique du modèle est que les individus évaluent dans l'incertain, et ce à un double niveau : ils ne connaissent pas l'efficacité du produit (1) (ils perçoivent des signaux de son efficacité mais cela peut prendre du temps), et ils ne connaissent pas le coût non monétaire du bien (2) (par exemple, ils ne savent pas si dormir sous une moustiquaire tient plus chaud que ne pas dormir sous une moustiquaire) – on ne peut mesurer ce coût qu'en utilisant le bien. Je commencerai par définir la structure et la mise en place de l'expérience (3.1) ; puis, je montrerai que l'énigme soulevée par Cohen et Dupas (2010) n'est toujours pas résolue même si les résultats de l'expérience de Dupas (2010) mettent en évidence un important effet d'apprentissage (3.2).

3.1 Structure expérimentale et implantation

L'expérience a été réalisée dans le district de Busia à l'est du Kenya en mai 2007 et portait sur 1120 ménages³⁰⁶. Ces 1120 ménages ont été sélectionnés à l'aide des registres scolaires : ont été sélectionnés tous les ménages avec des enfants. Ils ont ensuite été inscrits sur une liste qui a été utilisée pour assigner aléatoirement les ménages aux différents groupes. Chaque groupe bénéficiait d'un niveau de subvention distinct, allant de 100% à 40%, ce qui fixait le prix pour une moustiquaire de 0 Shillings kenyans à 250 (de \$0 à \$3.8 au moment de l'expérience). L'expérience se concentre sur six régions. Dans quatre d'entre elles, cinq groupes sont constitués. Ils se distinguent par le prix que leurs membres ont à payer pour acquérir une moustiquaire : 50 Shillings kenyan pour le premier groupe, 100 pour le

³⁰⁶ Ici Dupas se sert de sa deuxième expérience sur le paludisme et les moustiquaires afin d'évaluer les effets d'apprentissage. Elle reprend donc les données de (Dupas, 2009).

deuxième, 150 pour le troisième, 200 pour le quatrième et 250 pour le cinquième. Dans les deux autres régions la fourchette de prix est différente et permet de constituer quatre groupes distincts. Pour le premier groupe les moustiquaires sont gratuites, pour le deuxième elles coûtent 40 Shillings kenyan, 50 pour le troisième groupe et enfin les membres du quatrième groupe doivent payer 70 Shillings kenyan pour bénéficier d'une moustiquaire. Pendant un an les participants ont été suivis, des recenseurs rendaient visites aux différents ménages. Ils demandaient aux ménages s'ils utilisaient les moustiquaires et regardaient parfois si ces dernières étaient accrochées au-dessus des lits.

3.2 Importants effets d'apprentissage et faible utilisation des moustiquaires : une énigme toujours irrésolue ?

Pascaline Dupas cherche à évaluer deux effets : un effet d'apprentissage et un effet social³⁰⁷. La première évaluation vise à mesurer si en ayant bénéficié d'une moustiquaire durant la phase 1 il y a un effet d'apprentissage. On cherche ici à savoir si les individus par le fait qu'ils ont bénéficié de la moustiquaire apprennent des bénéfices de cette dernière et sont donc plus enclins à en acheter une sur la période suivante. La seconde évaluation vise à apprécier l'existence d'externalités positives, c'est à dire si les individus apprennent de leurs pairs dans le cas par exemple où une personne a une moustiquaire et qu'elle dit à ses voisins les bénéfices qu'elle peut tirer de cette dernière. Dupas s'intéresse donc à ces effets d'apprentissage. Pour cela, elle se concentre tout d'abord sur l'effet statique de la subvention et du processus d'apprentissage, et montre que :

« In this context, higher subsidies are likely to generate important learning-by-doing effects: higher subsidies lead to a much higher share of people experimenting with the product and thereby obtaining a signal about effectiveness. » (Dupas, 2010, 20)

Une importante subvention génère un effet d'apprentissage conséquent, d'ailleurs plus la subvention est importante plus les individus utilisent le produit et plus ils sont en mesure d'apprendre de son efficacité. Dupas se concentre ensuite sur les effets dynamiques de la subvention sur le processus d'apprentissage. Elle cherche donc à voir si les individus qui ont bénéficié d'une subvention totale ou très importante sont plus enclins à acheter une

³⁰⁷ L'effet d'apprentissage renvoie au fait qu'utiliser un bien permet d'apprendre sur ce dernier et donc de se rendre compte de ses bénéfices. L'effet social caractérise le fait que l'on peut apprendre d'un bien à partir du discours de personnes qui l'ont utilisé et donc apprendre des éventuels bénéfices de ce bien.

moustiquaire en phase 2. En d'autres termes elle cherche à savoir s'il y a un effet d'ancrage, si les individus qui ont bénéficié de la moustiquaire gratuitement sont prêts à payer pour cette dernière durant la seconde période ou s'ils ont intégré le prix nul et donc ne sont pas prêts à payer pour les moustiquaires durant la seconde période. Elle montre alors que:

« Overall, these results suggest that potential negative anchoring or entitlement effects of subsidies are at best limited in scope, and in any case overwhelmed by a positive effect. » (Dupas, 2010, 23)

Il n'y a donc pas réellement d'effet d'ancrage³⁰⁸ ou alors il est annulé par un effet positif, c'est à dire par le fait que les individus sont plus enclins à acheter des moustiquaires en phase 2. Autrement dit, un autre effet serait prédominant. Pour Dupas deux chaînes peuvent expliquer cela : la première est ce qu'elle cherche à évaluer, l'effet d'apprentissage, la seconde est l'effet revenu. Les moustiquaires ont une incidence positive dans la réduction de la transmission du paludisme. Cette incidence positive, en améliorant le niveau de santé des individus joue sur leur productivité, ce qui contribue à une augmentation des revenus. Cet effet revenu pourrait expliquer pourquoi les individus qui ont bénéficié gratuitement des moustiquaires sont prêts à payer pour ces dernières durant la seconde période. C'est pourquoi Dupas évalue cet effet et conclue:

« Overall, these results suggest that income effects played at best a limited role in the positive impact of LLIN subsidies on willingness to pay for LLINs observed in section 3.3 » (Dupas, 2010, 25)

Pour Dupas, il n'y a donc pas d'effet revenu *via* la santé. L'effet qui semble donc prendre le dessus est celui d'apprentissage. La première partie des résultats développés par Dupas conclue donc sur le rôle positif de l'effet d'apprentissage dans l'adoption des moustiquaires. Le second effet auquel s'intéresse principalement Dupas est l'apprentissage social. Autrement dit, savoir si le fait de voir bénéficier -par exemple ses voisins- de moustiquaires, d'échanger sur ce sujet, d'acquérir à travers autrui (et non comme l'effet d'apprentissage, à travers sa propre utilisation du bien) des connaissances sur un bien et ses bienfaits, permet une adoption plus large durant la seconde période des moustiquaires. Cet effet contrecarrerait là aussi le

³⁰⁸D'ailleurs lorsque Dupas évalue l'effet d'ancrage de façon directe elle conclue : *« Households who paid a positive price anchored somewhat around the offered price: at follow-up, the distribution of the gap narrows around zero for those in positive price group. This is not the case for households that received a free LLIN in Phase 1, however. For those, the density at zero is lower at the follow-up than at baseline, suggesting no anchoring at all ».* (Dupas, 2010, 24). Les personnes qui ont bénéficié gratuitement des moustiquaires n'ancrent donc pas le prix.

postulat que les individus puissent sous-estimer les bénéfices. Il semble, comme le souligne Dupas (2010), que cet effet joue un rôle dans l'adoption des moustiquaires :

« Overall, these results suggest that exposure through neighbors increased the likelihood that households bought at least one LLIN, but had no impact on the likelihood that households bought both LLINs. » (Dupas, 2010, 28)

Le fait que l'entourage possède une moustiquaire augmente la probabilité qu'un individu achète une moustiquaire mais n'a pas d'impact sur la probabilité qu'il en achète deux. Cet effet social est-il dû à un apprentissage ou bien est-ce le résultat d'une imitation ? En d'autres termes, les individus achètent-ils une moustiquaire parce que grâce aux échanges avec leurs voisins ils ont conscience des bénéfices d'un tel bien ou simplement parce que leurs voisins en possèdent une ? Dupas cherche donc à déterminer si cela est l'effet d'une imitation ou réellement d'un apprentissage social. Pour elle, il semble improbable que ce soit une pure imitation, c'est pourquoi l'apprentissage social lui semble être une explication beaucoup plus plausible. Elle conclut alors : *« Overall, while we cannot rule mimicry as a possible explanation for the social effects we observe, social learning appears the most likely factor. »* (Dupas, 2010, 29) Il y aurait donc un effet d'apprentissage des individus eux-mêmes lorsqu'ils utilisent le bien et un apprentissage social lorsque ces derniers échangent avec leur entourage. Les individus sont donc prêts à payer pour acheter une moustiquaire, même lorsque cette dernière a d'abord été intégralement subventionnée, et cela est dû pour Dupas à un double effet : un effet d'apprentissage et un effet social. Cela conduit Dupas à favoriser une importante subvention des moustiquaires³⁰⁹. Favoriser une subvention intégrale rejoint pleinement la recommandation de l'OMS faite en 2007 afin de lutter contre le paludisme³¹⁰.

4. Économie du développement et comportement de santé : un paradoxe insoluble ?

Pour Pascaline Dupas, les comportements des individus pauvres en matière de santé se caractérisent par deux faits stylisés : un important niveau de dépenses en soins curatifs (1) et un faible niveau de dépenses en matière de prévention (2). Cela semble constituer un important paradoxe, et conduire à une importante inefficacité. Un investissement en matière

³⁰⁹ « *The evidence provided in this paper suggests that, at least for some class of preventive health products, learning by doing and social learning are important channels through which short-term, targeted subsidies can translate into sustained levels of adoption. This provides a rationale for subsidies even for technologies that do not generate positive externalities.* » (Dupas, 2010, 30)

³¹⁰ Voir le rapport de l'OMS sur le paludisme, (OMS, 2007).

de prévention éviterait de nombreuses dépenses futures et surtout éviterait d'importantes maladies. Dans l'exemple développé tout au long de ce chapitre, la demande des moustiquaires est très élastique au prix ; cependant même à un prix nul ou très faible, les individus n'utilisent pas assez les moustiquaires alors que cela leur permettrait d'éviter le paludisme, et donc de fortes dépenses de soins potentiels et surtout la mort. Comment expliquer que les individus n'utilisent pas suffisamment ces dernières ? En d'autres termes comment expliquer l'énigme exprimée par la première expérience menée par Dupas ? Dans les expériences suivantes, Dupas (2009) tente d'évaluer l'impact de dispositifs incitatifs en vue d'améliorer cette utilisation : elle y pointe donc deux explications possibles à la faible utilisation des moustiquaires : la sous-estimation des bénéfices d'un tel bien par les individus (1) et leur incohérence temporelle (2). Les dispositifs évalués, on l'a vu, visent à contrecarrer ces deux biais.

Le premier fait stylisé s'explique, pour Dupas, à travers deux principales raisons : une maladie soignée trop tardivement conduit à une extrême souffrance rendant la demande de soins infiniment inélastique au prix (1), un accès limité à de bons diagnostics expliquerait aussi que les individus achètent trop tardivement les médicaments (2). De plus, les individus en matière de paludisme, ont tendance à prendre des médicaments dont ils n'ont pas besoin car ils ne sont pas malades, et ce souvent à cause de mauvais voire d'absence de diagnostics. Ces deux faits stylisés entraînent de nombreuses questions, en partie soulevées par Dupas :

« These two stylized facts discussed in this section raise many questions: Why do households buy medicines they do not need? Why do households not buy preventive health products, the private returns of which far outweigh their costs? These behaviors are inconsistent with the neoclassical model of economic behavior, in which fully informed households weigh benefits and costs when deciding on investments. In what ways does the situation in low-income countries depart from the neoclassical model? Do people lack basic information or the skills to process this information? Do they lack financial markets? Or are they simply not rational? » (Dupas, 2011a, 429, souligné par moi)

Les comportements mis en évidence par les expériences de Dupas (2009, 2010) et Cohen et Dupas (2010) contredisent le cadre économique standard, considérant qu'informer les individus opère un arbitrage efficace entre les coûts rencontrés et les avantages qu'ils peuvent tirer de ces mêmes coûts. Dupas (2011a) pose alors quatre questions :

1. Cette différence de comportements est-elle le propre des pauvres, et le modèle néoclassique serait-il insuffisant pour appréhender de tels comportements?

2. Cette inefficacité est-elle due à une absence d'information ou à une difficulté à la traiter de la part des pauvres ?
3. Ou encore, cette absence d'efficacité est-elle due au fait que les pauvres ont un accès difficile au marché financier ?
4. Enfin, cette inefficacité est-elle due à une irrationalité de la part des pauvres ?

Ces quatre questions rejoignent les deux explications offertes par Dupas (2011a) : la sous-estimation des bénéfices et l'incohérence temporelle. Le modèle néoclassique semble donc insuffisant pour rendre compte de ces deux biais (1), la sous-estimation des bénéfices est due à un manque d'information de la part des pauvres (2), l'incohérence temporelle explique en partie le problème financier des pauvres (3) et pointe clairement un biais de rationalité de ces derniers (4), reste cependant à savoir si c'est un trait propre aux pauvres, la question reste ouverte. Afin de contrecarrer ces biais de rationalité, Dupas (2009, 2011a) propose de jouer sur le comportement des individus³¹¹ en introduisant des dispositifs de *nudge*. Je commencerai par développer l'explication en termes de biais de rationalité qu'offre Dupas en vue d'expliquer le comportement des pauvres en matière de santé (4.1) ; puis, je présenterai les implications politiques qui en découlent, c'est à dire les différents dispositifs de *nudge* qu'envisage Dupas (2011a) (4.2).

4.1 Pauvreté et comportements irrationnels : sous-estimation des bénéfices et incohérence temporelle

Des biais de rationalité semblent, pour Dupas (2011a), expliquer en grande partie le comportement des pauvres. Les pauvres souffrent d'un manque d'information mais aussi d'une difficulté à traiter cette dernière. L'incohérence temporelle explique en grande partie la faible utilisation des biens en matière de prévention, par exemple les moustiquaires. La difficulté du traitement de l'information et l'incohérence temporelle traduisent un manque de rationalité de la part des pauvres et donc expliquent, selon elle, leurs comportements. Je commencerai par expliciter le manque d'information dont souffrent les pauvres ainsi que son impact (4.1.1) ; puis je m'attacherai à développer la seconde explication donnée par Dupas (2011a), c'est à dire l'incohérence temporelle (4.1.2).

³¹¹ Pour une analyse du lien entre les expériences du J-PAL et l'économie comportementale, voir (Davis, 2013).

4.1.1 Les pauvres et l'information

Le manque d'information est la principale raison invoquée pour un sous-investissement en termes de soins préventifs. Un simple accès à l'information permettrait donc d'améliorer les investissements en matière de prévention. Dupas confirme cette explication mais la nuance cependant en montrant que l'information compte mais que cette dernière doit être ciblée et à destination d'une population spécifique, pointant alors la complémentarité entre éducation et information, ainsi que le rôle de l'apprentissage social. Bhattacharya, Dupas et Kanaya (2013), en reprenant quatre études³¹² montrent l'important rôle de l'information. Cependant en reprenant une expérience qu'elle a menée avec Duflo et Kremer³¹³, toujours au Kenya, mais cette fois-ci en se centrant sur les jeunes filles et une politique de prévention en matière de VIH, Dupas (2011a) montre qu'il est plus efficace de fournir des messages préventifs ciblés expliquant précisément les risques plutôt que d'encourager un comportement spécifique³¹⁴. La population ciblée a une importance toute particulière dans l'impact de l'information. Par exemple, l'expérience fameuse de Miguel et Kremer (2004) sur les vermifuges montre que l'information transmise aux enfants n'a aucun impact alors que cela aurait pu être beaucoup plus efficace si l'information avait été destinée aux parents. Enfin, le genre peut aussi jouer dans le ciblage de l'information un rôle important ; Dupas montre en effet que les hommes et les femmes ont des comportements différents en matière d'acquisition de nouvelles technologies, comme les moustiquaires.

L'éducation joue un rôle important pour traiter l'information : « *There is evidence that education matters for health behavior through this cognitive-ability channel for developed countries.* » (Dupas, 2011a, 435) L'éducation permettrait donc de traiter l'information. Dupas (2011a) nuance cependant cela en montrant que les individus éduqués ont aussi tendance à se montrer sceptiques face à l'information, sans doute selon elle, parce qu'on leur a appris à l'être. L'apprentissage social, largement développé dans ce chapitre autour des moustiquaires et d'une expérience de Dupas (2009), semble jouer un rôle central dans l'information disponible pour les pauvres et être l'outil le plus efficace pour contrecarrer la sous-estimation des bénéfices supposés des pauvres. Ce qui conduit Dupas à conclure :

³¹² Les quatre études sont : (1) (Madajewicz *et al.*, 2007), (2) (Jalan et Somanathan, 2008), (3) Dupas, (2011b), et (4) (Cohen, Dupas et Schaner, 2013).

³¹³ Voir : (Duflo, Dupas et Kremer, 2011a).

³¹⁴ Celui évalué dans l'expérience est l'abstinence en vue de réduire la transmission du VIH chez les jeunes filles. Voir : (Dupas, 2011b).

« Overall, the evidence reviewed in this section suggests that rural communities in poor countries often lack basic information on the health returns to specific behaviors or products. Although simply providing information about health benefits is sometimes sufficient to change behavior, it is not always enough. In particular, when the healthy behavior comes at a cost (e.g., when it requires a costly tool or technology), people often need to experiment, or see others experiment, to be convinced about the health returns of that behavior or technology. » (Dupas, 2011a, 437)

L'information joue donc un rôle central dans le fait que les individus soient pleinement conscients des bénéfices qu'ils peuvent tirer de telles préventions. Pour autant, parfois l'information seule n'est pas suffisante, comme dans l'exemple des moustiquaires, il peut être important que les individus expérimentent et apprennent eux-mêmes du bien ou de l'expérimentation des autres.

4.1.2 Marché financier et incohérence temporelle

La contrainte de crédit à laquelle font face les pauvres empêche souvent cette expérimentation. Beaucoup de soins préventifs requièrent un investissement financier qui peut être impossible pour des individus vivant avec moins d'un dollar par jour. La question devient alors, pour Dupas : les pauvres n'ont-ils pas de moustiquaires (avant l'introduction de la subvention) parce qu'il leur est trop difficile d'emprunter ? Dupas (2011a) reprend une expérience menée par Banerjee, Duflo, Glennerster et Kothari (2010) cherchant à évaluer l'impact de la possibilité de souscrire à un prêt pour acheter une moustiquaire. L'expérience montre que seul 2% des individus du groupe de contrôle ont acheté en liquide une moustiquaire, alors que 52% des individus pouvant bénéficier d'un prêt ont acheté une moustiquaire à l'aide d'un emprunt. On est cependant face à une autre limite, il est souvent très difficile pour les pauvres d'emprunter. Dupas nuance cela :

« Even if they cannot borrow to invest in preventive health products, households should be able to save toward acquiring these products, even if over a very long time? So why do they choose not to? » (Dupas, 2011a, 439)

Même s'il leur est difficile d'emprunter, les pauvres devraient donc être capables d'épargner afin de pouvoir, au final, acheter une moustiquaire. Dupas se demande alors

pourquoi les pauvres n'épargnent pas? L'incohérence temporelle³¹⁵ et la procrastination qui la caractérise semblent expliquer, pour Dupas, cette faible épargne et donc ce faible investissement en matière de santé préventive :

« Another potential barrier to savings is that individuals might be present biased : Even though they would like to save in the long run, they might be subject to temptations on a daily basis that prevent from accumulating as much as they would like. » (Bhattacharya, Dupas et Kanaya, 2013, 440)

Les pauvres souffriraient donc d'un biais temporel, ils souhaiteraient épargner mais seraient « rattrapés » par les tentations du présent, tout en continuant de croire que ces tentations seront moins importantes dans le futur et qu'ils seront, à ce moment là, à même d'épargner. Les quatre enjeux soulevés par Dupas se retrouvent en un double problème : l'incohérence temporelle et la sous-estimation des bénéfices. Les pauvres n'ont aucune incitation à épargner s'ils doutent des bénéfices qu'ils vont tirer de l'utilisation d'un bien. Ce qui, cela a été souligné, justifie pour Dupas une subvention totale sur le court terme afin que les individus puissent expérimenter le bien et apprendre de ses bénéfices. Dupas montre d'ailleurs que concernant les moustiquaires cet apprentissage est important. Cependant, sur le long terme les investissements en matière de prévention sont faibles : l'explication est pour Dupas celle de l'incohérence temporelle. Elle soulève une seconde fois la question de savoir si cette incohérence temporelle en matière de prévention est le propre des pauvres : *« To what extent is time inconsistency also a factor behind the lowtake-up of preventive behaviors in low-income countries? »* (Dupas, 2011a, 440) Dupas (2011a) laisse la question ouverte. Le problème semble donc bien être une irrationalité, un comportement biaisé de la part des pauvres, mais à savoir si ce biais de comportement est le propre du pauvre, il n'y a pas de réponse.

4.2 Implications politiques : jouer sur les incitations et imposer des mandats

³¹⁵Le terme d'incohérence temporelle est, tout d'abord, introduit par Robert Strotz : *« An individual is imagined to choose a plan of consumption for a future period of time so as to maximize the utility of the plan as the present moment. His choice is, of course, subject to a budget constraint. Our problem arises when we ask: If he is free to reconsider his plan at later dates, will he abide by it or disobey it – even through his original expectations of future desires and means of consumption are verified? Our answer is that the optimal plan of the present moment is generally one which will not be obeyed, or that the individual's future behavior will be inconsistent with his optimal plan. »* (Strotz, 1956, 165) La définition standard de l'incohérence temporelle est reliée au fait que les individus valorisent plus le présent que le futur, ce qui les conduit à privilégier les actions plaisantes d'aujourd'hui au détriment des méfaits que cela peut entraîner demain. En conséquence, les individus ne prennent pas les décisions qui maximisent leur utilité.

Les implications politiques développées par Dupas visent à contrecarrer cette incohérence temporelle. Sans y faire explicitement référence, les implications politiques des expériences en matière de santé menées par Dupas s'inspirent très clairement du paternalisme libertarien³¹⁶ développé par Sunstein et Thaler³¹⁷. Ces derniers montrent que le paternalisme libertarien n'a rien d'un oxymore³¹⁸. Il est possible de mener des actions paternalistes sans pour autant entraver la liberté des individus. L'illustration phare de cela est l'exemple de la ligne de cafétéria développé par Sunstein et Thaler ; ces derniers montrent que de petits arrangements peuvent avoir d'importants effets. Afin de démontrer cela, ils imaginent une situation fictive dans laquelle une directrice d'école doit décider de l'organisation de la ligne de la cafétéria. Ils montrent qu'en changeant certains arrangements de la cafétéria, la directrice est capable de diminuer ou d'augmenter la consommation de certains biens alimentaires de 25%. Par exemple, en plaçant les frites sur un étal plus élevé qu'auparavant ou en disposant les légumes au premier niveau, de façon très accessible. Le premier arrangement permet de diminuer la consommation de frites et le second augmente la consommation de légumes. La directrice de l'école représente ce que Sunstein et Thaler nomment une architecte des choix : « *Carolyn is what we will be calling a choice architect. A choice architect has the responsibility for organizing the context in which people make decisions.* » (Sunstein et Thaler, 2008, 3) L'organisation qu'elle choisit permet d'orienter les décisions des individus. En ce sens, si la directrice d'école souhaite favoriser des comportements alimentaires sains, il lui est possible de procéder à un placement particulier des aliments afin d'orienter les individus. Ce paternalisme est libertarien au sens où aucun choix n'est retiré des individus, aucune décision n'est prise à leur place. On parle pourtant de paternalisme, car un agent extérieur a défini en amont ce qui lui semblait préférable. Ce paternalisme libertarien se traduit par l'implantation de dispositifs incitatifs orientant les individus vers un comportement efficace. Ces dispositifs sont appelés *nudge*. Ils jouent sur les incitations des individus de deux façons. La première est d'offrir une récompense à l'individu. La seconde est de sanctionner les comportements non jugés efficaces. Cela retrouve l'expression commune de « la carotte et du bâton », comme le souligne Dupas :

³¹⁶ Ce dernier est aussi appelé paternalisme libéral, voir (Ferey, 2011).

³¹⁷ Voir, par exemple, (Sunstein et Thaler, 2003a ; 2003b ; 2008)

³¹⁸ Un de leur article est d'ailleurs nommé « *Libertarian Paternalism is Not An Oxymoron* », (Sunstein et Thaler, 2003a)

« In many cases, experimentation is really what people need to become convinced of the benefits of a given health technology. Moreover, in the presence of either credit constraints or externalities, experimentation levels might be inefficiently low. Two types of public policy interventions can redress this inefficiency: sticks, such as mandates, or carrots, such as price subsidies or financial incentives. » (Dupas, 2011a, 443, souligné par moi)

Dupas propose de jouer sur les incitations des individus pauvres afin qu'ils adoptent un comportement en matière de santé plus efficace. Elle envisage donc aussi bien des dispositifs incitatifs fondés sur les récompenses que sur les sanctions. En ce sens, elle s'inscrit complètement dans la perspective de Sunstein et Thaler. Je commencerai par présenter le rôle des dispositifs incitatifs envisagés par Dupas (4.2.1) ; puis, je développerai celui des mandats (4.2.2).

4.2.1 Jouer sur les incitations : la « carotte »

Un des premiers dispositifs envisagés par Dupas pour pousser les individus à investir en matière de prévention est qu'ils s'engagent dans ce processus. C'est exactement ce que teste Dupas, dans la deuxième expérience qu'elle mène au Kenya, lorsqu'elle teste un dispositif d'engagement oral. Une autre expérience, concernant toujours les moustiquaires, réalisée en Inde³¹⁹ évalue un autre dispositif d'engagement à travers le microcrédit. Dans cette expérience, il était nécessaire de ré-imprégner les moustiquaires d'insecticides tous les six mois afin que ces dernières soient efficaces. Deux types de contrats sont alors comparés à travers l'expérience. Un premier contrat laisse les individus libres d'acheter ou non la moustiquaire et de la ré-imprégner par eux-mêmes. Le second contrat est un contrat d'engagement, les individus achètent la moustiquaire et s'engagent à venir la ré-imprégner une première fois au bout de six mois et une seconde fois un an après l'expérience. La ré-imprégnation des moustiquaires a été plus importante dans le groupe bénéficiant d'un contrat d'engagement. S'engager permet alors, pour Dupas, de réduire d'une certaine façon l'incohérence temporelle³²⁰ :

³¹⁹ Voir (Tarozzi *et al.*, 2011).

³²⁰ Dupas évoque une autre expérience menée aux Philippines concernant le tabagisme. Voir : (Giné, Karlan et Zinman, 2010), dans cette expérience, les individus devaient s'engager à arrêter de fumer. Cet engagement a été considéré positif dans la réduction du tabagisme.

« Overall, these two studies suggest that people are quite sophisticated with regard to their time-inconsistent preferences and have a demand for products that enable them to commit to healthy behaviors. »
(Dupas, 2011a, 442)

Les individus pauvres auraient conscience de leur incohérence temporelle, et feraient donc appel à des dispositifs intérieurs sophistiqués pour la contrecarrer - Dupas ne dit pas lesquels. En ce sens, ils seraient demandeurs de dispositifs qui leurs permettraient de s'engager et donc de pallier à leur propre incohérence temporelle. Cependant, cela a été vu, ce dispositif d'engagement n'a pas fonctionné au Kenya. De plus, lorsque Dupas (2011a) évoque cette expérience en Inde, elle ne dit rien sur l'utilisation effective des moustiquaires. Dans la même lignée, le second dispositif majeur que développe Dupas (2011a) tente de jouer sur les incitations des individus. L'objectif est d'offrir aux individus une récompense lorsqu'ils poursuivent un comportement jugé positif pour eux et donc rationnel. En d'autres termes, l'idée est de leur donner une « carotte ». Dupas (2011a) reprend l'expérience menée par Banerjee, Duflo, Glennerster et Kothari (2010). Cette expérience part d'un constat identique à celui posé par l'énigme de la première expérience de Dupas sur les moustiquaires et donc au premier fait stylisé développé par Dupas, les individus pauvres ne font pas vacciner leurs enfants même lorsque la vaccination est gratuite. La raison invoquée est là aussi double, et toujours la même, les pauvres sont incohérents temporellement et sous-estiment les bénéfices de la vaccination. Le dispositif incitatif envisagé par Banerjee, Duflo, Glennerster et Kothari (2010) est d'offrir un kilo de lentilles à chaque ménage venant faire vacciner son enfant. L'impact de ce dispositif est jugé extrêmement positif pour contrecarrer l'incohérence temporelle et donc augmenter le nombre d'enfants vaccinés :

« The authors³²¹ interpret this as evidence that small incentives can help solve procrastination problems: By offering immediate gratification for the adoption of a behavior that has only long-run benefits for the parents, the incentives could overcome the natural tendency to delay a slightly costly activity. »
(Dupas, 2011a, 442)

Les incitations sous formes de récompenses sont alors considérées comme un dispositif efficace permettant de contrecarrer le premier fait stylisé développé par Dupas (2011a) – le faible investissement en termes de prévention. Ces incitations permettent de réduire l'incohérence temporelle ainsi que la sous-estimation des bénéfices. La distribution gratuite, par exemple, des moustiquaires donne une incitation à utiliser cette dernière et donc aux

³²¹ Banerjee, Duflo, Glennerster et Kothari (2010).

individus d'expérimenter le bien. Cette expérimentation, on l'a vu, permet à travers l'apprentissage social de réduire la sous-estimation des bénéfices de la part des individus. De plus, ces incitations permettent aussi de réduire la contrainte de crédit qui pèse sur les pauvres :

« All these results suggest that combining traditional antipoverty transfer programs with incentives in specific human capital investments might be highly effective: They can resolve the credit constraint, time inconsistency, and intra-household conflicts all at once by making the receipt of government transfers conditional on the adoption of a set of behaviors. » (Dupas, 2011a, 444)

La première implication politique que tire Dupas est donc d'introduire des mécanismes incitatifs au travers de récompenses afin de conduire les individus à investir dans les biens de prévention en matière de santé. Même s'ils sont dans l'ensemble considérés comme efficaces par Dupas (2011a), ces dispositifs se sont révélés inopérants dans la tentative de faire utiliser les moustiquaires aux individus kenyans (Dupas, 2009, 2010). Ces dispositifs ne permettent donc pas de répondre à l'énigme que son expérience (Cohen et Dupas, 2010) a soulevée concernant l'aide au développement.

4.2.2 Mandater : le « bâton »

La seconde implication politique développée par Dupas (2011a) est que les gouvernements requièrent légalement ces différents comportements, ce qu'elle appelle mandater. Elle prend alors l'exemple du fait qu'en France, par exemple, la vaccination des enfants est obligatoire, pointant le fait que si les enfants ne sont pas vaccinés les parents peuvent perdre leurs droits parentaux. Dupas (2011a) note cependant que ces types de mandats peuvent être compliqués, particulièrement pour les pays en développement :

« Enforcing these types of mandates might not always be easy, however, especially in low-income countries where governmental budgets and institutions are typically weak. For example, all public transportation vehicles in Kenya are required to have functioning seat belts, but rampant corruption among the police officers means that the law has essentially no bite – it is simply not enforced. » (Dupas, 2011a, 443)

L'exemple du Kenya montre que les institutions dans les pays en développement peuvent ne pas être à même d'appliquer ce genre de mandat, car la loi ne s'applique pas. De plus,

Dupas ne dit pas précisément quels types de mandats devraient être mis en place et comment. Cependant, elle souligne que de tels mandats ont été particulièrement efficaces dans certains pays en développement. Elle prend l'exemple de la Thaïlande et du VIH. En 1991, la Thaïlande lance un programme visant à ce que les travailleurs sexuels utilisent tous des préservatifs. La police fournit les préservatifs aux travailleurs sexuels, même si le travail de ces derniers est illégal en Thaïlande. Afin d'identifier ceux qui n'utilisaient pas de préservatifs, le personnel de santé se référait aux dossiers d'hommes suivant un traitement contre le VIH et leur demandaient de donner le nom de l'établissement où ils avaient été pour avoir des relations sexuelles. Ensuite, le personnel soignant se rendait dans l'établissement pour fournir plus d'informations. Ce programme est considéré comme un large succès, l'utilisation de préservatifs par les travailleurs sexuels étant passée de 14% en 1989 à 90% en juin 1992. Les gouvernements ont donc un rôle important à jouer en matière de prévention. Dupas souligne même que leurs actions seraient sans doute beaucoup plus efficaces que celles développées plus haut, c'est à dire au travers de dispositifs de « coup de pouce » poussant les individus à acquérir un comportement rationnel :

« Household behavior played a minor role. Although providing information, subsidies, or incentives might be effective, it might not be as cost-effective as large public health interventions. » (Dupas, 2011a, 445)

Ici, Dupas nuance très fortement les différentes recommandations politiques qu'elle suggère, en affirmant que les comportements des ménages ne jouent qu'un rôle mineur et qu'il est sans doute beaucoup plus efficace de mener une politique publique de grande envergure en matière de santé. Cependant, ces grandes politiques publiques ne peuvent être évaluées par le J-PAL et vont en grande contradiction avec la stratégie adoptée par celui-ci d'évaluer des politiques locales et de stipuler qu'un changement à la marge peut avoir des effets importants. Dupas ne résout pas l'énigme que sa première expérience a mis en évidence et pointe le fait qu'il faille aller plus loin à savoir s'interroger sur le rôle des gouvernements et institutions, afin que les individus utilisent réellement les moustiquaires et que la prévention en matière de santé soit conséquente dans les pays en développement. Cependant ce qu'il semble essentiel de retenir de toutes ces expériences est pour Dupas que :

« The important thing to take away from this review is that when it comes to health behavior in developing countries, there are a substantial number of deviations from the neoclassical model. First of all, people seem to lack basic information, and sometimes have limited ability to process information, because

of low education levels. Second, there are market imperfections and frictions, especially credit constraints, affecting people's ability to invest in health. Finally, there seem to be some deviations from the rational model, with, as has been widely shown in developed countries, a nontrivial share of people exhibiting time-inconsistent preferences as well as myopia. » (Dupas, 2011a, 445, souligné par moi)

L'énigme pointée par Dupas tend à montrer selon elle qu'en matière de santé, dans les pays en développement, le comportement des individus s'éloigne très fortement du modèle néoclassique, ou du « modèle rationnel ». Les pauvres semblent donc agir de manière irrationnelle, ce qui expliquerait qu'ils n'utilisent pas suffisamment les moustiquaires. Cependant, les différents dispositifs cherchant à retrouver cette rationalité semblent échouer ou se montrent beaucoup moins efficaces qu'une importante action publique. L'unique compréhension en termes de rationalité semble insuffisante pour répondre à l'énigme posée par la première expérience de Dupas.

5. Conclusion

Retracer chronologiquement les différentes expériences de Dupas m'a permis d'apprécier la portée théorique des évaluations du J-PAL face au débat sur l'aide au développement. Une première énigme a été mise en évidence, retournant la question de l'aide vers la question de l'utilisation des moustiquaires. Cette dernière s'expliquant pour Dupas par des biais de rationalité de la part des pauvres, l'incohérence temporelle et la sous-estimation des bénéfices. Cependant les dispositifs testés afin de contrecarrer ces deux biais n'ont eu que peu d'impact sur l'utilisation des moustiquaires. Les différentes expériences n'ont donc pas permis de répondre à l'énigme posée, elles ont comme tourné autour.

Aucunes des expériences n'a cherché à répondre explicitement à la question : pourquoi les femmes enceintes au Kenya n'utilisent pas suffisamment les moustiquaires ? Dupas (2009, 2010) a cherché à tester des dispositifs contrecarrant la faible utilisation, en posant a posteriori une explication potentielle à cette faible utilisation, mais ne l'a jamais testée ou évaluée. Aucun des mécanismes expliquant la faible utilisation des moustiquaires n'a été mis en avant. Dupas n'éprouve pas l'existence d'incohérence temporelle ou de sous-estimation des bénéfices mais la suppose. Ce qu'elle teste ce sont des dispositifs les contrecarrant. De plus, ces dispositifs s'avèrent non efficaces. Qu'est-ce que cela signifie ? Cela tend-il à infirmer l'hypothèse d'irrationalité des pauvres ? Dupas ne donne jamais la signification de l'échec de ces dispositifs et considère que de façon générale, les incitations sont une

implication politique à suivre. La suite s'opacifie encore plus lorsqu'elle souligne qu'une politique publique aurait un bien plus grand impact, contredisant alors la perspective locale du J-PAL. Deux étapes se distinguent alors. L'une est caractérisée par la première expérience de Cohen et Dupas (2010), autrement dit l'étape de l'énigme, que l'expérience fait émerger. Cela pointe le premier rôle accordé à la théorie par le J-PAL : la créer par l'expérience. La deuxième étape est caractérisée par les deux autres expériences de Dupas (2009, 2010), cherchant à contrecarrer le problème et non à y répondre. Sans théorie, sans volonté de mettre en évidence les mécanismes expliquant cette faible utilisation, les dispositifs de Dupas peinent à améliorer cette utilisation.

Chapitre 3. Du paternalisme libertarien au paternalisme démocratique : une boîte noire vide ?

0. Introduction

« We might begin looking for principles governing the acceptable use of paternalistic power in cases where it is generally agreed that it is legitimate. (...) What is it that justifies us in interfering with children ? The fact that they lack some of the emotional and cognitive capacities required in order to make fully rational decisions. It is an empirical question to just what extent children have an adequate conception for their own present and future interest but there is no much doubt that there are many deficiencies. »
(Dworkin, 1972, 76, souligné par moi)

En mai 2012, Esther Duflo donne une série de conférences³²² à Harvard sur les liens entre le paternalisme et la lutte contre la pauvreté. Lors de ces conférences Duflo adopte une position en apparence paradoxale, puisqu'elle propose d'être paternaliste vis-à-vis des pauvres. Elle définit un nouveau paternalisme, un « paternalisme démocratique » qui viserait à augmenter les libertés des individus en leur retirant certains choix. Pour cela, elle se fonde sur la définition de la liberté de Sen, envisageant cette dernière comme une augmentation des « capacités ». Promouvoir un paternalisme démocratique est contradictoire compte tenu de la perspective initiale du J-PAL – fonder la lutte contre la pauvreté sur des preuves. Ce paternalisme n'est le fruit d'aucune preuve. De plus, il apparaît difficile de l'évaluer à l'aide d'expériences randomisées. Il est d'autant plus surprenant de promouvoir un paternalisme, nécessairement lié à un cadre moral implicite ou explicite, lorsque les chercheurs du J-PAL se refusent à toutes théories, ou toutes connaissances *a priori*. Proposer un tel paternalisme représente donc un véritable tournant. L'objectif de ce chapitre est de questionner ce tournant. Ce dernier me semble être la traduction politique de la « faille épistémologique » que j'ai tâché de définir lors de la partie précédente de ce travail (*i.e.* partie 2 – chapitre 3).

Afin de rendre compte de ce tournant, je tâcherai, durant la première partie de ce chapitre, de définir le lien entre le paternalisme libertarien de Sunstein et Thaler et celui que Duflo cherche à développer. Je montrerai alors que ce dernier s'éloigne du paternalisme libertarien de Sunstein et Thaler, car celui-ci à travers l'implantation de dispositifs d'incitations s'avère insuffisant pour orienter et transformer le comportement des pauvres. Les pauvres comme les

³²² Cette série de conférences se situe dans le cadre des *Tanner Lectures on Human Values*, c'est l'une des conférences les plus prestigieuses de Harvard. Le lien pour la conférence : <http://mahindrahumanities.fas.harvard.edu/content/lecture-1-esther-duflo-human-values-and-design-fight-against-poverty> (dernière consultation, le 12 août 2013).

riches, pour Duflo, souffrent de biais de rationalité ; cependant, les riches vivent entourés d'incitations invisibles leur permettant d'adopter des comportements efficaces. Ces incitations n'existant pas dans les pays pauvres, cela laisse les pauvres dans un *statu quo* où ils ne peuvent que recourir à des comportements inefficaces. Je montrerai qu'à l'instar de Dupas (2011a), c'est un manque de rationalité qui justifie une intervention paternaliste pour Duflo. Et comme le soulignait Dworkin en 1972, l'incapacité à prendre des décisions de façon pleinement rationnelle justifie le fait que quelqu'un d'extérieur prenne ces décisions.

Dans la deuxième partie de ce chapitre, je tâcherai de définir précisément le nouveau paternalisme envisagé par Duflo. Elle légitime ce dernier par le fait que les choix sont coûteux. Les pauvres, vivant dans un monde non paternaliste, perdent de l'énergie à faire des choix. De ce fait, ils leur restent trop peu d'énergie pour se concentrer sur les choix importants ; ils finissent par prendre de mauvaises décisions, perpétuant ainsi un cercle vicieux. L'idée est, pour Duflo, à travers son nouveau paternalisme, de retirer certains choix aux pauvres, afin qu'ils puissent disposer de l'espace mental nécessaire pour réaliser les choix qui comptent vraiment pour eux. Pour Duflo, en retirant des mains des pauvres certains choix, cela augmentera leurs capacités. C'est en ce sens qu'elle entend fonder son paternalisme sur la notion Sennienne de liberté. L'objectif est donc, pour Duflo, de légitimer ce paternalisme, renvoyant ici au premier point soulevé par la citation de Dworkin.

Dans la troisième partie de ce chapitre, je montrerai que la proposition de Duflo recouvre deux principales confusions. La première confusion provient du fait que, la liberté, chez Sen, est à la fois instrumentale et substantielle. Dans cette perspective, il apparaît contradictoire de penser le paternalisme comme un instrument vers plus de libertés. Ici, Duflo semble confondre moyens et fins. La seconde confusion est définie par le fait que, Duflo souhaite imposer aux pauvres un panier de biens élémentaires - visant à supprimer certains choix -, mais il apparaît difficile de penser que l'imposition de biens élémentaires parvienne à augmenter les possibilités des individus. Cela revient à supposer qu'un bien, une donnée peut engendrer un processus. Je montrerai alors que ces deux confusions philosophiques ne sont que la traduction de la faille épistémologique de l'approche adoptée par le J-PAL. La forte validité interne d'une telle approche au détriment d'une certaine validité externe ne permet pas de comprendre les processus à l'œuvre, elle ne permet pas de produire de schéma causal (cf. partie 2 - chapitre 3). En ce sens, elle ne peut produire de recommandations politiques rendant compte d'un mécanisme ni de processus particulier. Cela me semble permettre

d'expliquer un tel tournant de la part de Duflo, tout autant que le flou philosophique de sa position.

1. Rationalité et paternalisme libertarien

Pour Duflo, tout comme pour Pascaline Dupas, c'est un manque de rationalité qui justifie le recours à des dispositifs de *nudging*. Duflo considère, aussi, que ce manque de rationalité se traduit par une sous-estimation des bénéfices et une incohérence temporelle (1.1). Duflo se demande, aussi, si ce manque de rationalité est le propre des pauvres. Pour Duflo, ce n'est pas le cas; la seule différence réside dans le fait que les riches sont entourés d'incitations invisibles orientant leurs comportements de façon efficace. En ce sens, elle légitime le paternalisme libertarien de Sunstein et Thaler. Ici, il s'agira donc de définir le « point de départ » ainsi que la justification de Duflo pour un nouveau paternalisme. Pour cela, j'expliquerai, tout d'abord, le manque de rationalité des pauvres pour Duflo (1.1). Puis, je montrerai en quoi ce manque de rationalité justifie, pour Duflo, le paternalisme libertarien de Sunstein et Thaler (1.2), ouvrant alors la voie à son nouveau paternalisme.

1.1 « Pauvres mais rationnels ? »³²³

A l'instar de Dupas, Duflo considère que les pauvres sous-estiment les bénéfices et qu'ils souffrent d'incohérence temporelle. Cependant, la définition qu'elle donne de la rationalité est ambiguë. Dans un de ses articles consacré à la question de l'efficacité et la rationalité des pauvres, Duflo (2006b) cherche à montrer que les pauvres n'ont pas un comportement efficace, comme le soulignait Théodore Schultz (Schultz, 1964)³²⁴, mais qu'ils ont un comportement rationnel. Elle illustre ce propos à partir de deux domaines d'expériences : l'assurance et l'agriculture. Au sein du premier domaine, elle montre que les pauvres auraient tout intérêt à s'assurer mais qu'ils ne le sont pas. Concernant l'agriculture, Duflo (2006c) explique qu'il serait efficace pour les fermiers d'utiliser des engrais, mais que ces derniers ne les utilisent pas. Afin de rendre compte précisément de cela, je me centrerai sur une des

³²³ Ce titre renvoie directement au titre de l'article de Duflo : « *Poor but Rational ?* » ; voir (Duflo, 2006c).

³²⁴ L'idée développée par Schultz est que les pauvres arrivent à avoir une utilisation efficace de leurs faibles ressources. Son travail se centre sur l'agriculture et contredit la proposition théorique qui prévalait dans ce domaine durant les années 1950. A cette période, la faible production agricole était attribuée au fait que l'agriculture des pays en développement était gouvernée par la tradition ou la culture ; l'idée était alors de transférer la main d'œuvre de l'agriculture vers l'industrie. A l'inverse, Schultz explique que compte tenu des circonstances dans lesquelles vivent les pauvres, ils font au mieux.

expériences menées dans le domaine de l'agriculture. Cette expérience conduite au Kenya par Duflo, Kremer et Robinson (2008) partent du constat que les pauvres n'utilisent pas les engrais, alors que ces derniers permettraient d'augmenter de façon conséquente la production agricole. Afin d'acheter les engrais, en vue de les utiliser, les fermiers ont besoin d'avoir épargné un peu avant. Cependant, très peu de fermiers disposent de l'épargne suffisante pour acheter les engrais au moment de les appliquer. En collaboration avec une ONG Kenyane ICS³²⁵, Duflo Kremer et Robinson (2008) évaluent un programme cherchant à contrecarrer ce problème au travers d'un dispositif de *nudging* : le SAFI programme³²⁶. Dans ce programme, un membre de l'ONG rend visite aux fermiers après la période de récolte, au moment où les fermiers disposent de l'argent nécessaire pour acheter les engrais ; le membre de l'ONG leur propose alors d'acheter les engrais en avance. Le programme est considéré efficace par les chercheurs :

« The researchers created a model to estimate which policy option delivers the highest welfare – i.e. which makes society as a whole better off. They find that a SAFI-style program improves welfare relative to taking no policy action, and it may provide larger welfare gains than heavy subsidies. » (J-PAL, Policy Briefcase 2011, 4)

Face à l'absence d'intervention ou d'un subventionnement total des engrais, le programme SAFI est jugé plus efficace³²⁷. Une première question apparaît : selon quel critère³²⁸ ce programme est-il jugé plus efficace que les autres ? Les auteurs jugent le programme SAFI plus efficace car il permet la plus grande augmentation de bien-être de la société dans son ensemble. Une première ambiguïté apparaît : quelle est la nature de bien-être ? Une production agricole plus importante ? Les auteurs ne donnent pas d'éléments de réponses. Si l'on le laisse de côté, une autre ambiguïté apparaît tout de même : qu'est ce qui constitue la société dans son ensemble ? Prend-on en compte le bien-être de chacun des individus constituant cette société ? Ou bien les auteurs s'intéressent-ils au bien-être de cette société en

³²⁵ « International Child Support » (ICS) est une ONG Kenyane visant à améliorer les conditions des enfants et de leurs familles dans les pays en développement.

³²⁶ « Savings and Fertilizer Initiative Program » (SAFI)

³²⁷ Avec la comparaison entre le groupe de contrôle, qui ne bénéficie d'aucun programme, et le groupe de traitement qui bénéficie du programme SAFI, les chercheurs sont en mesure de déterminer si le programme SAFI est plus efficace que l'absence d'intervention. Le coût d'une subvention totale des engrais par le gouvernement Kenyan n'est pas évalué ; cependant, les chercheurs concluent que le programme SAFI est plus efficace.

³²⁸ Pour juger de l'efficacité d'un programme, il est nécessaire de disposer d'un critère qui définit cette efficacité. Pour une étude des liens entre normativité et efficacité, voir, par exemple, (Picavet, 1999).

moyenne ? Là aussi, les auteurs ne donnent pas de réponses. La notion d'efficacité évoquée reste donc opaque.

Les auteurs montrent ensuite que le programme SAFI est efficace pendant la durée de l'évaluation, mais les fermiers arrêtent d'épargner pour acheter les engrais lorsque l'évaluation prend fin. Ce qui explique pourquoi Duflo, à l'inverse de Schultz, considère que les pauvres ne sont pas efficaces mais rationnels³²⁹. Une des principales raisons invoquées est qu'il était ressenti comme difficile pour les fermiers de constituer cette petite épargne, même si elle leur permettait d'espérer de meilleurs rendements. Pourquoi les fermiers avaient-ils du mal à constituer cette épargne ? L'explication invoquée par les auteurs est la procrastination :

« In the study region, two thirds of farmers who make plans to use fertilizer do not follow through with those plans. Even among those farmers who use fertilizer, many procrastinate: the vast majority (96-98 percent) purchases it only right before they need to apply it. » (J-PAL, Policy Briefcase 2011, 4)

Tout comme dans les expériences menées par Dupas, la raison invoquée pour la faible épargne est l'incohérence temporelle. Banerjee et Duflo définissent cette dernière de la façon suivante :

« Dans le présent, nous sommes impulsifs, dirigés en grande partie par nos émotions et nos désirs immédiats : de petites pertes de temps (comme faire la queue pour vacciner un enfant) ou de petits désagréments subis tout de suite (des muscles endoloris) paraissent bien plus désagréables sur le moment que lorsque nous y pensons sans d'immédiateté (...). De ce fait, nous sommes naturellement portés à remettre à plus tard les petits coûts, pour les faire payer à notre moi de demain plutôt que par celui d'aujourd'hui. » (Banerjee et Duflo, 2011, 113)

Toujours de la même façon, le dispositif incitatif qu'impliquait le programme SAFI pendant la durée de l'évaluation, visait à contrecarrer cette incohérence temporelle. On retrouve donc ici, très clairement, le système de *nudging* développé par Sunstein et Thaler (2008), auquel Duflo, Kremer et Robinson (2008) se réfèrent explicitement dans leur expérience, tout comme Dupas dans les siennes. Ce dispositif est efficace pendant la durée de l'évaluation, mais ces effets ne perdurent pas une fois l'expérience terminée. Un autre problème apparaît : Duflo (2006c) cherchait à montrer que les pauvres n'étaient pas efficaces

³²⁹ Duflo (2006c) ne définit ni les termes de rationalité ni d'efficacité. Sa position est donc en elle-même opaque. Je cherche donc ici, à travers les éléments qu'elle offre, à rendre compte de ce que sous-tendent pour elle ces termes.

mais rationnels, et cela, en se fondant sur l'expérience réalisée avec les fermiers Kenyans. Mais lorsque l'on développe la raison de l'inefficacité des fermiers, c'est à dire leur impossibilité à épargner, on tombe sur une explication en termes de biais de rationalité³³⁰, l'incohérence temporelle. Et le remède potentiel envisagé par les chercheurs est un dispositif visant à contrecarrer ce biais. Les pauvres seraient, en fait, inefficaces car ils ne seraient pas suffisamment rationnels.

L'ambiguïté entre rationalité et efficacité ne se retrouve pas uniquement dans ce domaine. Si l'on reprend les principaux résultats mis en avant par Duflo et Banerjee (2011) dans leur dernier livre³³¹, les comportements qu'ils illustrent peuvent parfaitement se retrouver dans la grille de lecture développée ici (inefficacité et biais de rationalité)³³². En matière de santé, par exemple, même lorsque les moustiquaires sont subventionnées, les pauvres ne les utilisent pas suffisamment pour lutter efficacement contre le paludisme. Leur comportement est là aussi non efficace. La raison invoquée par les auteurs est là-encore l'incohérence temporelle, ainsi que le fait qu'ils sous-estiment les bénéfices. La rationalité et l'efficacité n'étant cependant jamais clairement définies par Duflo et Banerjee, c'est comme si l'on s'accordait sur un sens commun donné à ces deux termes. Pourtant, la rationalité comme l'efficacité d'un individu ne peuvent être jugées que par rapport à une norme, à un critère explicite. Un individu est dit rationnel s'il respecte une norme de rationalité, parce qu'il répond à un critère particulier. Tout comme l'efficacité : le comportement d'un individu est perçu comme efficace selon un critère d'efficacité. Ici, ces critères ne sont pas définis, ils ne sont d'ailleurs même pas évoqués. C'est comme s'il y avait un accord tacite sur ce qui constitue un comportement rationnel ou efficace. De plus, même si les pauvres étaient considérés comme irrationnels, il faudrait comprendre les causes de cette irrationalité, qui ne pourrait être appréhendée comme une cause mais seulement comme un effet. De plus, à elle seule l'irrationalité ne pourrait expliquer l'inefficacité du comportement des pauvres à sortir de la pauvreté.

1.2 Le paternalisme et les riches

³³⁰ Même si Duflo et Banerjee ne réfèrent jamais à ce problème d'incohérence temporelle à un biais de rationalité. Cependant, le cadre économique standard le fait explicitement pour eux.

³³¹ J'ai discuté ces différents résultats dans la partie précédente de ce travail (cf. partie 2 – chapitre 3).

³³² Banerjee et Duflo (2011) relient l'incohérence temporelle à de nombreux domaines comme l'assurance, l'épargne, la santé, et l'éducation, voir (Banerjee et Duflo, 2011, 241 ; 297 ; 383).

La question devient donc de savoir, comme chez Dupas, si cette irrationalité est le propre des pauvres. Cependant, à l'inverse de Dupas, Banerjee et Duflo (2011) donnent des éléments de réponses :

« Les pauvres semblent enfermés dans le même type de problème que ceux qui nous affectent tous, notamment le manque d'information, des croyances infondées et la procrastination. Il est vrai que nous, qui ne sommes pas pauvres, sommes relativement plus instruits et mieux informés, mais la différence est faible car, au final, nous en savons en fait très peu, et certainement moins que nous ne l'imaginons. En réalité, notre avantage vient de tout ce que nous considérons comme évident. » (Banerjee et Duflo, 2011, 118)

Le manque de rationalité n'est donc pas le propre des pauvres, les riches souffrent des mêmes problèmes. La seule différence est ce qui semble évident pour les riches. Quelles sont ces évidences ? Par exemple, les riches ont directement accès à l'eau potable. Ils n'ont pas besoin de se rendre à des bornes ou de rajouter du chlore dans l'eau. Ils n'ont pas non plus à se demander comment se procurer leur prochain repas. Ils peuvent avoir confiance en certaines de leurs institutions, comme en matière de santé ou d'éducation. Autrement dit, ils n'ont pas besoin d'essayer de survivre chaque jour avec des ressources limitées. Ces évidences sont ce que Duflo et Banerjee (2011) appellent des « incitations invisibles », elles entourent les riches, et sans qu'ils s'en rendent compte les aident à agir de façon efficace :

« Les habitants des pays riches sont sans cesse entourés d'incitations invisibles, le but premier des politiques de santé dans les pays pauvres devrait être de rendre aussi facile que possible aux pauvres de bénéficier de soins préventifs et de réglementer la qualité de soins que les gens reçoivent. Etant donné l'extrême sensibilité des gens au prix, l'une des premières choses à faire est bien sûr d'assurer gratuitement des services de prévention, voire de récompenser les foyers qui y recourent et de faire si possible que cela devienne un choix naturel, une option par défaut. Il faut installer les distributeurs de chlore gratuits à proximité des sources d'eau ; récompenser les parents qui vaccinent leurs enfants ; donner aux enfants scolarisés et des compléments nutritionnels ; enfin, le réseau d'eau et les infrastructures sanitaires doivent faire l'objet d'investissements publics, en tout cas dans les zones densément peuplées. » (Banerjee et Duflo, 2011, 119, souligné par moi)

Pour Banerjee et Duflo, les riches sont entourés de dispositifs de *nudging*, qui leur facilitent grandement la vie mais l'ignorent. Surtout dans les pays riches, les choix par défaut sont devenus des choix naturels. Une option par défaut est une option qui sera choisie si le preneur de décision ne prend aucune décision. Banerjee et Duflo (2011) se réfèrent ici explicitement à Sunstein et Thaler (2008). Pour ces derniers, les individus font face à de

nombreuses options par défaut. L'exemple typique de ce genre d'option est le renouvellement automatique d'un abonnement à un magazine. Un individu souscrit à un magazine pour une année, et lorsqu'il s'abonne, il y a une option par défaut - s'il ne fait rien – l'année suivante il sera automatiquement réabonné. En effet, beaucoup de personnes ont tendance à être abonnées pendant plusieurs années à un magazine qu'elles ne lisent jamais. Le fait que l'option par défaut ne soit pas changée, tient pour Sunstein et Thaler du fait que les individus souffrent d'un biais : un biais de *statu quo*³³³, qu'ils définissent de la façon suivante : « *For lots of reasons, people have a more general tendency to stick with their current situation.* » (Sunstein et Thaler, 2008, 34) Les individus tendent à rester dans la même situation, ils ont tendance à ne pas changer. Par exemple, les élèves ont souvent des places définies en classe, et n'en changent pas. Pour Sunstein et Thaler, la raison d'une telle inertie est due à un manque d'attention. Les individus adoptent une attitude du peu importe (« *whatever heuristic* » - Sunstein et Thaler, 2008, 34) Par exemple, ils commencent à regarder une chaîne à la télévision, le soir, et n'en changent pas de la soirée.

Ce biais de *statu quo*, offre donc un rôle central aux options par défaut. Si les individus ont tendance à ne pas changer, à rester avec ce qui leur est donné ; l'option choisie par défaut sera très souvent celle choisie. Le secteur privé comme public se sert déjà beaucoup du rôle de l'option par défaut, comme l'illustre l'exemple du réabonnement automatique à un magazine. Dans leur perspective de *nudge*, Sunstein et Thaler (2008) montrent que jouer sur cette option par défaut peut s'avérer très efficace. Comme Banerjee et Duflo (2011) l'évoquent, cela appliqué à la pauvreté peut avoir beaucoup d'effets. Offrir explicitement le choix aux individus de mettre du chlore dans l'eau, afin que cette dernière soit potable, peut conduire à des résultats inefficaces. A l'inverse, introduire un dispositif de *nudging* permettant d'ajouter le chlore directement à l'eau dès que les individus tournent le robinet³³⁴ est beaucoup plus efficace. Dans la même lignée, donner des compléments alimentaires aux enfants est une option par défaut : s'ils ne les refusent pas, on leur en donne et s'ils les refusent, on cesse de leur en donner. Duflo et Banerjee recommandent, ici et explicitement - tout comme le faisait Dupas, un paternalisme libertarien ; de faire que ces options par défaut deviennent des choix naturels et qu'en ce sens, ils ne soient plus des choix mais des évidences. Ce sont ces évidences que possèdent les riches. Les options par défaut, dans les

³³³ Même si Sunstein et Thaler utilisent grandement le concept de biais de *statu quo*, c'est à William Samuelson et Richard Zeckhauser (1988) que l'on doit ce concept.

³³⁴ Kremer, Leino, Miguel et Peterson Zwane (2011) testent ce dispositif au Kenya et montrent que ce dernier est efficace.

pays riches, sont pensées de telle façon qu'elles sont bénéfiques aux riches. Alors que les pauvres disposent d'options par défaut, inefficaces et non tournées vers leurs intérêts. Transformer ces options permettrait de pallier le manque de rationalité des pauvres, comme cela est le cas dans les pays riches.

La position de Banerjee et de Duflo (2011) s'obscurcit lorsque ces derniers évoquent le rôle que les pouvoirs publics doivent jouer dans la création d'infrastructures. Et ce particulièrement lorsqu'ils discutent du réseau d'eau, en pointant que cela devrait relever des pouvoirs publics³³⁵. On est donc, comme chez Dupas (2011a), face à une ambiguïté. Le *nudging* est mis en avant constituant la principale recommandation politique. Cependant, Banerjee et Duflo (2011) soulignent le rôle des pouvoirs publics. L'implantation d'infrastructures ou de politiques d'envergure apparaît être beaucoup plus efficace qu'un dispositif de *nudging*. Pour autant, à l'encontre de Dupas, la position de Duflo sur la rationalité est ambiguë. Elle n'est pas directement liée aux pauvres, elle est liée aux manques d'évidences auxquelles font face les pauvres. C'est pourquoi Banerjee et Duflo affirment :

« Tout ceci peut paraître paternaliste – et, en un sens, ça l'est effectivement. Mais il est facile de discourir sur les dangers du paternalisme et la nécessité d'assumer la responsabilité de sa vie depuis notre canapé, à l'abri de nos confortables maisons. Nous qui vivons dans des pays riches, ne bénéficions-nous pas à chaque instant d'un paternalisme si ancré dans le système que nous ne le remarquons plus ? Non seulement il garantit que nous prenons soin de nous-mêmes mieux que si nous avions à décider seuls de tout, mais, en nous épargnant d'avoir à réfléchir à ces questions, il nous fournit l'espace mental nécessaire pour nous concentrer sur le reste de notre vie. Cela ne nous rend pas quittes pour autant de la responsabilité d'éduquer les gens à la santé. Nous devons effectivement à tout le monde, y compris aux pauvres, une explication aussi claire que possible de l'importance de la vaccination et des raisons pour lesquelles un traitement antibiotique ne doit pas être interrompu avant la fin. Nous devons cependant reconnaître que l'information à elle seule ne suffira pas et en assumer les conséquences. Cela tient à notre nature humaine, que nous soyons pauvres ou moins pauvres. » (Banerjee et Duflo, 2011, 120, souligné par moi)

Ce paternalisme libertarien, développé par Sunstein et Thaler, permettrait alors d'arriver à des choix naturels et donc à des évidences. Une des principales recommandations de Duflo et de Banerjee (2011) est de changer les options par défaut et de s'en servir, afin que le manque de rationalité qui nous est commun, que l'on soit pauvre ou riche, soit contrecarré. C'est pourquoi, pour Banerjee et Duflo (2011), la transmission seule de l'information (ce qu'avait

³³⁵ Voir (Banerjee et Duflo, 2011, 119, 401-403).

testé Dupas) n'est pas suffisante. Tous les dispositifs de *nudging* ne fonctionnent pas. Cependant, Duflo, tout comme Dupas, les placent sur le devant de la scène.

2. Pour un paternalisme démocratique : le récent tournant d'Esther Duflo

Les riches vivent dans un monde beaucoup plus paternaliste que les pauvres, est-ce que cela les rend pour autant moins libres ? C'est la question à laquelle Duflo (2012a) souhaite désormais répondre. Elle s'attache à montrer que les riches, même dans un monde paternaliste, sont en réalité plus libres, ont plus de chance de se réaliser. Afin de souligner cela, Duflo (2012a) développe un nouveau paternalisme, un paternalisme qui rend en fait les individus plus libres, un paternalisme démocratique. Dans ce but, elle entend fonder ce paternalisme sur la notion de capacités développée par Amartya Sen³³⁶. Duflo défend alors un paternalisme plus fort que celui de Sunstein et Thaler. La liberté de choix n'existe plus. Ce paternalisme n'est plus libertarien. La principale justification pour un tel paternalisme, pour Duflo, est que les choix auxquels font face les pauvres sont des bruits, qu'ils représentent un coût, empêchant les pauvres de se focaliser sur les décisions réellement importantes. Néanmoins, ce nouveau paternalisme soulève d'importantes implications philosophiques, que Duflo discute, d'ailleurs elle même. Je commencerai par définir le paternalisme souhaité par Duflo (2.1) ; puis, je développerai les limites que Duflo soulève à l'encontre de ce dernier (2.2).

2.1 Réconcilier le paternalisme et les libertés à travers les capacités

Esther Duflo définit le paternalisme de la façon suivante :

« Paternalism can be defined as the practice of providing a set of « basics needs » for people (the set may vary), typically without consulting them on what their needs actually are. This may involve the free or subsidized provision of housing, schooling, or health care etc., but also involves a set of mandates or restrictions on the shape that any assistance takes. » (Duflo, 2012a, 2)

Le paternalisme, pour Duflo, consiste à fournir des biens essentiels aux individus sans leur demander au préalable s'ils en ont vraiment besoin, et quels sont réellement leurs besoins. Dans le paternalisme développé par Thaler et Sunstein, la possibilité de choix était maintenue, alors qu'ici les individus ne sont pas consultés. Comment la liberté des individus peut-elle être garantie alors même que les individus ne sont pas consultés ? A juste titre, Duflo (2012) explique que l'on ne peut pas répondre à cette question sans définir la liberté. Elle emprunte à Amartya Sen sa conception de la liberté, dans le but de réconcilier le paternalisme et la

³³⁶ Voir, par exemple, (Sen, 1987, 1992, 1999, 2009)

liberté. Elle explique que Sen définit la liberté positivement³³⁷, comme ce que chacun peut réaliser. Duflo (2012a) reprend brièvement l'illustration de la famine au Bengale³³⁸, les Bengalais n'étaient pas libres car leurs faibles revenus ne leur permettaient pas d'acheter de la nourriture et donc de se prémunir contre la famine. Duflo relie ici la définition de la liberté de Sen à celle des capacités. La liberté nécessite un accès aux capacités :

« It would make no sense, for example to claim that people in Haiti are free because they can die of cholera if they to. The presence of cholera makes them less free: freedom to lose is not freedom. This does not mean that Sen equates freedom and physical well-being. Agency matters. Freedom does include the freedom to fast for someone who has enough money to buy food, and it does include the freedom to refuse a blood transfusion if it goes against one's beliefs. A person living in the rich West is capable of eating and capable of staying alive after a hemorrhage, even if they decide not to. » (Duflo, 2012a, 15)

Pour Duflo (2012a), la « capacité d'agir » ou la « puissance d'agir »³³⁹ est limitée par le pouvoir de l'inertie³⁴⁰, par les institutions et les lois. On l'a vu, les riches vivent dans un monde plus paternaliste mais sont, pourtant, plus libres que les pauvres. Ils sont amenés à prendre des décisions qui leur permettent d'avoir une meilleure santé (accès à l'eau potable, vaccination, etc.), et d'être moins vulnérables que ne le sont les pauvres. C'est pourquoi, ils sont en fait plus libres. Comme Duflo le souligne, un individu est plus libre dans une société où l'eau est directement potable, où le choix d'ajouter ou non du chlore ne se pose pas. Les riches sont donc plus libres que les pauvres, même dans un monde plus paternaliste. Cela est dû à deux principales raisons pour Duflo : plus les individus ont de choix à faire, plus ils ont de chance de prendre de mauvaises décisions (1) ; et les choix sont coûteux (2). Je m'attacherai, ici, à développer ces deux principales raisons. Je commencerai par montrer pourquoi, pour Duflo, lorsque les individus ont de nombreux choix à faire ils tendent à prendre les mauvaises décisions (2.1.1). Puis, j'explicitai en quoi, selon Duflo toujours, les choix sont coûteux (2.1.2).

2.1.1 Décisions et libertés de trébucher

³³⁷ « Sen defines freedom, not negatively, but positively, as the ability realize one's potential. » (Duflo, 2012a, 15)

³³⁸ Amartya Sen (1981) a montré que la famine au Bengale en 1943 n'était pas due à un manque de nourriture mais qu'elle était principalement due au fait que les individus avaient perdu leur droit d'accès au marché, et ne pouvaient donc plus acheter de quoi se nourrir. Les individus avaient perdu la liberté de pouvoir se nourrir.

³³⁹ Je traduis ici le terme « agency » par « capacité d'agir » ou « puissance d'agir ». Ce terme provient des théories de l'action et traduit la capacité d'agir sur le monde d'un individu.

³⁴⁰ Le pouvoir de l'inertie renvoie à la tendance, définie par Sunstein et Thaler plus haut, des individus à ne pas changer.

Devoir prendre un grand nombre de décisions, presque chaque jour, afin d'obtenir les éléments essentiels de la vie ne constitue pas, pour Duflo (2012a), une liberté. De plus, plus nombreux sont les choix plus grandes sont les chances de se tromper. Duflo reprend ici l'exemple du chlore et de l'eau³⁴¹. L'eau non potable est une des premières causes de mortalité dans les pays pauvres, ajouter du chlore dans l'eau suffirait à résoudre le problème. Tout comme pour le paludisme, qui, on l'a vu est intimement lié à la pauvreté ; il suffirait d'utiliser des moustiquaires imprégnées d'insecticides pour limiter la transmission du paludisme. Mais on l'a aussi vu, comme les riches les pauvres souffrent de biais de rationalité et en ce sens, ils prennent de mauvaises décisions :

« We often adopt a somewhat patronizing tone when thinking about the poor: why are they not boiling their water? Why are they not immunizing their children? Why are they not saving for a rainy day? Why don't they have a more caloric diet? We can see that all those “rights” choices are available to them, but we forget that they imply active choices, when the “wrong” choices are hardly available to us. The freedom here is just the freedom to stumble. » (Duflo, 2012a, 17, souligné par moi)

Le fait que les pauvres aient à faire face à beaucoup de choix ne les rend donc pas libres, et c'est en ce sens que les riches qui ont moins de décisions à prendre sont considérés plus libres. Car si les choix auxquels font face les pauvres leur donnent plus de libertés, c'est pour Duflo, une plus grande liberté de tomber.

2.1.2 Décisions et espace mental

La deuxième raison qui rend les riches plus libres, alors même qu'ils vivent dans un monde plus paternaliste, est que le choix peut être coûteux. Prendre des décisions, faire des choix demande du temps, de l'énergie. Et pour Duflo (2012a), les êtres humains ne disposent pas de ces ressources de façon illimitée. Passer du temps et de l'énergie sur un choix c'est le perdre pour un autre choix. C'est perdre de son espace mental. Ici, Duflo évoque une expérience en psychologie de Wegner *et al.* (1987). Dans cette expérience on cherche à évaluer ce qu'on pourrait appeler l'« effet Dostoïevski »³⁴² :

³⁴¹ Les expériences de Dupas (2009, 2010) et Cohen et Dupas (2010) pourraient aussi illustrer cela.

³⁴² Je choisis d'appeler cet effet, l'effet Dostoïevski.

« Whether one is trying not to think of a traumatic event, however, or is merely attempting to avoid the thought of food while on a diet, it seems that thought suppression is not easy. It is said, that when the young Dostoyevski challenged his brother not to think of a white bear, the child was perplex for a while. »
(Wegner *et al.*, 1987, 5)

Lorsqu'il est demandé à quelqu'un de ne pas penser à quelque chose, cela apparaît impossible. Wegner *et al.* (1987) cherchent à évaluer cet effet, à comprendre s'il est vraiment difficile de ne pas penser à quelque chose lorsque l'on nous le demande. Ils mènent alors une première expérience, dans laquelle les individus sont attribués à deux groupes différents. Il est intéressant de noter que l'assignation aux groupes se fait de façon aléatoire. Le premier groupe est le « groupe de suppression » et le second le « groupe d'expression ». Il est demandé à chacun des groupes d'exprimer toutes leurs pensées pendant cinq minutes, ces dernières sont enregistrées. Pour le premier groupe, on leur impose, avant l'enregistrement, de ne pas penser à un ours blanc, de supprimer cette pensée. Le second groupe commence l'enregistrement librement, ce n'est qu'au milieu de celui-ci, que l'on demande à ce groupe de ne pas penser à ce dit ours blanc. En somme, l'ordre des tâches est inversé pour ces deux groupes. Les résultats de cette première expérience montrent qu'il est extrêmement difficile de supprimer des pensées. Et même lorsque, ces dernières sont supprimées un court instant, elles reviennent ensuite de façon plus vive.

Durant la seconde expérience, un troisième groupe est ajouté aux deux premiers qui restent identiques. Comme pour le « groupe de suppression », il est demandé aux individus du troisième groupe de ne plus penser à l'ours blanc, avant l'enregistrement. La nouveauté est qu'on impose à ce groupe de remplacer la pensée d'un ours blanc par celle d'une Volkswagen rouge. Les résultats sont identiques à la première expérience. L'effet d'un transfert de pensée apparaît efficace pour supprimer une pensée, et empêche d'y repenser de façon trop vive après. Dans le but d'expliquer ces résultats, Wegner *et al.* (1987) excluent explicitement la théorie Freudienne sur le rôle de l'inconscient³⁴³, et privilégient la théorie psychologique de la réactance. Cette dernière développée par Brehm (1966) montre que lorsqu'un comportement est interdit à un individu, ce dernier se sent privé de liberté et va chercher à la retrouver, ce

³⁴³ Wegner *et al.* (1987) évoquent le concept de répression chez Freud. Ce concept traduit l'idée d'un aller-retour entre la conscience et l'inconscience. La répression traduit un désir provenant de l'inconscient arrivant à la conscience puis étant renvoyé dans l'inconscient. Ce concept se distingue très fortement du concept Freudien majeur, le refoulement ; puisque dans le refoulement le désir n'arrive pas à atteindre la conscience, il essaye mais n'y parvient pas et est directement renvoyé dans l'inconscient. Il n'y a donc pas d'aller-retour. Cependant, Wegner *et al.* (1987) soulignent que le concept de répression pourrait offrir des éclairages certains à leur analyse, mais qu'il ne parvient pas à s'appliquer réellement à leur cas.

qui est entraîné par un effet de réactance psychologique. Par exemple, un adolescent de moins de 16 ans aura tendance à avoir un plus grand désir d'aller voir un film interdit aux moins de 16 ans, qu'un film interdit aux moins de 12 ans ou qu'un adolescent de 17 ans, qui ne ressentirait aucun interdit. Brehm (1966) cherche à jouer sur cet effet de restriction de liberté et sur cette volonté de la retrouver en incitant les comportements.

Sans que Duflo (2012a) n'évoque jamais les résultats de ces expériences ni leurs procédures ou leurs buts, ces derniers retrouvent pleinement sa problématique - cela même si elle évoque cette expérience pour une unique dimension – la dimension de choix et d'espace mental. Le lien avec la théorie de la réactance psychologique, lui aurait, sans doute, permis d'affiner sa position par rapport aux libertés et aux comportements individuels. Puisqu'une telle théorie vise à jouer sur la sensation de libertés en cherchant à orienter les comportements. Cependant, l'expérience de Wegner *et al.* (1987) explicite un point important les actes de pensées prennent une importante place. Ces expériences permettent d'illustrer que lorsqu'un individu est pris par une pensée particulière, il lui est difficile de s'en défaire. Duflo, en s'appuyant sur d'autres expériences en psychologie³⁴⁴, montre qu'il est difficile pour un individu lorsqu'une tâche lui a été imposée (par exemple ne pas penser à l'ours blanc) de se concentrer. Elle considère que l'espace psychique est limité. De plus, elle souligne que prendre des décisions requiert de l'énergie, de l'espace mental et du contrôle de soi :

« If our minds have finite resources and if we are entirely occupied trying to figure out whether or not boiling water for before an infant drinks is a good idea, we may simply not have any time left to think about what is the right choice of school for the child's older sister (something American parents devote a lot of time to), or how to stock the shelves of the family business. » (Duflo, 2012a, 18)

Pour Duflo, avoir à se centrer sur un choix important de la vie quotidienne, comme chlorer l'eau afin qu'elle soit potable, réduit l'espace mental permettant de prendre d'autres décisions toutes aussi importantes, comme choisir une bonne école pour son enfant. Duflo se réfère ici explicitement au concept développé par Sendhil Mullainathan et Eldar Shafir (2013) le « problème de la mise en boîte » (« *packing problem* »). Les pauvres ont trop d'éléments à emballer et ce dans une boîte qui est trop petite. Autrement dit, ils ont trop de choses à

³⁴⁴ Voir : (Baumeister, Gailliot, DeWall et Oaten, 2006), et (Vohs, Baumeister, Schmeichel, Twenge, Nelson, et Tice, 2008).

penser. Banerjee et Mullainathan (2008) construisent alors un modèle sur cette base. Le fondement de ce modèle est que l'attention est une ressource rare et qu'elle joue un rôle important en termes de productif. Le but est donc de lier le concept d'attention à celui des trappes à pauvreté, afin de déterminer si le premier a un effet sur le second :

« Specifically, people may not be able to fully attend their jobs if they are also worrying about problems at home, and being distracted in this way reduces productivity. But not paying attention at home is also costly: early symptoms of a child's sickness may go unnoticed; water may run out at the end of the day; kerosene for lighting lamps at home might run and make it hard to do homework, etc. Finally, the extent to which home life distracts depends on the nature of home life. Specifically (e.g., a good baby sitter, a 24-hour piped water supply, a connection to a power supply grid) can reduce the extent of home life distraction. »
(Banerjee et Mullainathan, 2008, 489)

Comme les pauvres doivent gérer par eux-mêmes trop de choses, l'espace mental qui leur reste est faible. Cela enfreint leur productivité, puis réduit leur revenu et donc continue à les enfermer dans une trappe à pauvreté. L'approche de Banerjee et Mullainathan (2008) modélise cette idée. Les pauvres ne sont donc pas libres, trop de choix et une trop petite valise les empêchent de se réaliser, d'être productifs, d'augmenter leurs revenus. Ils font alors les mauvais choix, ou alors n'en font pas mais finissent tout de même avec le mauvais choix. Pour Duflo (2012a), il faut donc leur enlever certains choix, car ces derniers sont coûteux, et ne permettent pas aux pauvres de sortir de la pauvreté, au contraire. La solution est donc d'être paternaliste :

« We thus made the case that paternalism, far from being opposed to individual responsibility, may form a basis on which we might have freedom over what really matters in life. Most of the choices that the poor have to make are just pure "noise", which at best stand in the way of them making important choices and at worst lead them to make a wrong turn and fail to achieve the amenities needed for a decent life. Thus, it can be achieved, making it easier for them to obtain the basic constituents of a healthy and productive life is desirable. » (Duflo, 2012a, 23)

Un certain paternalisme peut être le fondement, pour Duflo, d'une plus grande liberté, élargissant les possibilités individuelles et permettant une vie meilleure, une vie digne. C'est dans sa finalité que ce paternalisme retrouve la conception de libertés développée par Sen, même si Duflo (2012a) ne le relève jamais. Et c'est ici qu'il vise la liberté, car il vise une augmentation des possibilités pour les pauvres, une augmentation de leur capacités. La

liberté est ici une fin et le paternalisme, un moyen d'y parvenir. Cela explique pourquoi Duflo le nomme paternalisme démocratique.

2.2 Un Paternalisme sans morale ?

Invoquer le paternalisme n'est pas sans implications philosophiques, et soulève nécessairement quelques questions. Celui proposé par Duflo n'y échappe pas. D'ailleurs c'est Duflo (2012a), elle même, qui expose les principales limites de son paternalisme. Elle développe alors trois majeures limites. Le premier enjeu est la légitimité de prendre des décisions à la place d'autrui. Le paternalisme libertarien prémunissait contre ce problème moral. L'imposition d'un panier de biens élémentaires, en vue de retirer certains « choix bruyants » des mains des pauvres, conduit à se demander précisément ce qui doit constituer ce panier de biens élémentaires, ainsi que les personnes qui doivent décider de sa détermination. Enfin, si les gouvernements doivent jouer un rôle dans ce paternalisme, peut-on réellement leur faire confiance ? La dernière question est donc institutionnelle. Je commencerai par expliciter la première difficulté que soulève Duflo : la légitimité de prendre des décisions à la place d'autrui (2.2.1). Je développerai ensuite la deuxième difficulté : que doit contenir le panier de bien que souhaite imposer Duflo (2.2.2)? Enfin, je traiterai de la dernière difficulté soulevée par Duflo : la possible confiance en l'état (2.2.3).

2.2.1 Paternalisme démocratique et absence de « confort philosophique »

La première limite que Duflo (2012a) adresse au lien qu'elle cherche à créer entre liberté et paternalisme est le lien que son paternalisme va entretenir avec le paternalisme libertarien de Sunstein et Thaler. La question est de déterminer que faire lorsque le paternalisme libertarien n'est pas possible et quelles implications cela engendre :

« First, when strong defaults with an exit option are not possible, (i.e. things need to be made mandatory with no libertarian paternalism “escape” of Thaler and Sunstein) can we still say that paternalism for the poor means more freedom? For example, would a universal health care insurance law increase or decrease freedom in poor countries? Is it legitimate to sometimes override (some) people's carefully considered decisions? We lose the relative philosophical comfort of having our cake and eating it too, when we have to choose between doing nothing and imposing a mandate. While I realize that this is more controversial, the arguments developed in this lecture suggest to me that there are circumstances where the mandates may actually be the freedom-enhancing option. » (Duflo, 2012a, 23)

Avec ce premier enjeu, Duflo rend encore opaque sa position face au paternalisme libertarien. D'une part, car tout au long de sa conférence, elle donne la sensation d'envisager un autre paternalisme. Comment se positionne-t-elle alors face au paternalisme libertarien ? Il semble que ce dernier soit, au départ, l'option privilégiée, car il ne restreint pas les choix, il les oriente. En ce sens, il ne retire aucune liberté aux individus. D'autre part, toujours tout au long de cette conférence, Duflo a envisagé de retirer certains choix aux pauvres, puisque ces choix sont coûteux et s'apparentent à des bruits. Au départ, Duflo semble privilégier le paternalisme libertarien. Mais en dernière analyse, au vu des résultats des expériences du J-PAL pointant l'échec des dispositifs de *nudging*, elle est conduite à envisager un élément plus ferme et donc un paternalisme plus fort. En privilégiant ce paternalisme, Duflo pose la question morale de savoir s'il est légitime de décider à la place de quelqu'un. Cette question ne se posait pas dans le paternalisme de Sunstein et Thaler. Dans celui-ci, la liberté de choisir n'était pas entravée, personne ne prenait de décisions à la place de quelqu'un d'autre. Lorsqu'il est impossible de jouer sur les options par défaut, Duflo s'éloigne du paternalisme libertarien. De ce fait, comme elle le souligne, elle perd son confort philosophique. Désormais elle ne peut échapper aux questions morales. Peut-on décider à la place de quelqu'un ? Comment décider à la place de quelqu'un ? Qui décide ? De quoi décider ?

2.2.2 Qui décide et que mettre dans les biens élémentaires?

Le deuxième problème, tout aussi moral, questionne les biens élémentaires, puisqu'afin qu'ils en disposent, il s'agit d'ôter une certaine liberté aux pauvres. La question est alors de savoir de quoi seront constitués ces biens élémentaires ? Comment les définir ? En d'autres termes, quels choix enlever aux pauvres ? Dans la perspective de Duflo, cela revient à se demander : quels choix faut-il supprimer pour permettre *in fine* d'augmenter la liberté des pauvres, à travers l'imposition d'un panier de biens élémentaires ? Duflo se réfère à la notion de liberté et de capacités chez Sen, pourtant elle ne met jamais en lien ces biens essentiels avec les capacités. La définition même de ces biens reste extrêmement floue. S'agit-il des biens premiers chez Rawls³⁴⁵, par exemple ? Ou ces biens renvoient-ils aux biens essentiels

³⁴⁵ Les biens premiers chez Rawls (1971) sont des biens élémentaires qu'un individu souhaitera avoir et qui lui permettront de se réaliser : « les biens premiers sont tout ce qu'on suppose qu'un être rationnel désirera quels que soient ses autres désirs. Quels que soient dans le détail les projets rationnels d'un individu, on suppose qu'il y a certaines choses dont il préférerait avoir plus que moins. » (Rawls, 1971, 122) Ces biens sont pour Rawls :

définis par l'Organisation internationale du travail³⁴⁶ ? S'agit-il ici de penser uniquement leur mise en place à travers un certain paternalisme ? Ou s'agit-il de définir ces biens essentiels ? Il semble, quand Duflo pose la question de leur détermination, qu'il s'agit de les définir. Cette détermination doit être le fruit d'un processus démocratique :

« Logically, the basic package should be decided as the outcome of a democratic process. But this requires the meaningful participation of all citizens, and that in itself requires the poor to have peace in mind. In my view while this is a real question, this is not the one that should stop us from getting started. A number of outcomes (avoiding infant mortality for example) should be uncontroversially desirable, and there is a fair amount of scientific evidence for how to achieve them. One could start with those issues and then let the democratic process play its role to change the rules over the time. » (Duflo, 2012a, 24)

C'est donc ensemble que les citoyens doivent activement définir les biens élémentaires. Pour que ce procédé démocratique s'accomplisse, il faut que les pauvres aient l'esprit libre. En d'autres termes, il faut que certains choix leur aient été déjà retirés, que le paternalisme ait déjà eu lieu. On est face à un problème de l'œuf et de la poule : la démocratie doit exister afin que les pauvres aient l'esprit libre, mais pour qu'ils aient l'esprit libre le paternalisme est nécessaire. Pour Duflo, la détermination de ces biens n'est pas le problème majeur. Elle considère qu'il peut exister un consensus sur ce qui est essentiel à la vie, et que cela n'est pas controversé comme dans le cas de la mortalité infantile. De façon morale évidente, personne n'est pour la mortalité infantile. Pour autant, ce que Duflo ne souligne pas c'est que cette position est avant tout morale. De plus, même s'il existe un consensus moral important contre la mortalité infantile, il n'existe pas de consensus sur la manière d'imposer les dispositifs qui la ferait disparaître. Ces dispositifs restent, d'ailleurs, à définir. Duflo tend à passer à côté de cela. La proposition de Duflo est donc suspendue à la première question qu'elle pose, c'est à dire la question morale du paternalisme. Même si Duflo écarte la question de la détermination des biens élémentaires en appelant au consensus, elle ne souligne pas que l'obstacle à ce qu'elle propose n'est pas la finalité (plus de liberté) mais le moyen (le paternalisme). En ne questionnant pas, en ne se positionnant pas face à l'outil qu'elle brandit, la position de Duflo est comme bloquée, statique ; elle n'est que position et non proposition.

les biens premiers naturels (comme la santé), les biens premiers sociaux (comme la liberté), et les avantages socio-économiques (les positions qui aident un individu à se réaliser).

³⁴⁶ L'Organisation internationale du travail introduit l'approche en termes de biens essentiels lors d'une conférence en 1976 (voir, Jolly, 1976). Ce terme désigne les ressources minimales nécessaires permettant le bien-être physique minimal pour un individu. Ces biens sont définis comme des biens de consommation et c'est d'eux dont dépend la définition du seuil de pauvreté. Ils recouvrent, par exemple, l'eau, les vêtements, etc.

2.2.3 Le paternalisme et rôle de l'état

La troisième question que soulève Duflo est de savoir si l'on peut faire confiance aux gouvernements des pays développés ou si ces derniers sont trop corrompus :

« Third, can developing country governments be trusted to be the stewards of this benevolent paternalism ? Or, will the responsibility to deliver more goods and services simply open the floodgates for corruption, graft, or vote buying ? That is, even if we assume that benevolent paternalism is desirable for freedom's sake, it is feasible ? In particular, do we suffer from another chicken and egg problem here ? If the citizens are too concerned with the problems of their daily lives, how can they effectively monitor the governments that are supposed to help them ? » (Duflo, 2012a, 24)

Duflo passe ici d'un paternalisme démocratique à un paternalisme bienveillant. Elle montre que même si l'on assume que ce type de paternalisme est dans l'intérêt de la liberté, on ne sait pas s'il est faisable. La question est celle de la faisabilité d'un tel paternalisme, de son opérationnalisation. Il semble que l'on retombe ici dans un problème du type de l'œuf et de la poule. Si comme le suppose Duflo les pauvres ne disposent pas d'assez d'espace mental, ils ne pourront pas se concentrer sur la vie civique et démocratique. En effet, ils ne pourront pas surveiller que l'Etat œuvre au mieux, qu'il n'y a pas de corruption, et que l'Etat ne transforme pas ce paternalisme bienveillant en paternalisme malveillant et auto-intéressé.

Le paternalisme envisagé par Duflo n'est donc pas sans soulever de problèmes. La légitimité même du paternalisme - la prise de décision par un tiers - reste ouverte, et semble ne pouvoir être réglée sans que les deux autres questions aient elles-mêmes trouvé une réponse. Autrement dit, la question de la légitimité du paternalisme ne peut être résolue que lorsqu'un cadre moral est établi, définissant quels biens élémentaires ce paternalisme doit imposer et comment il doit les imposer. Ces deux dernières questions, dans la proposition de Duflo, restent pourtant, elles aussi, ouvertes. Cependant, Duflo souligne que trouver le moyen d'offrir aux pauvres un espace mental plus large ainsi que les moyens de s'engager dans un processus politique permettrait de résoudre les deux dernières questions :

« Finding ways to give the poor the mental space and the tools to engage in the policy process may actually be an avenue for improving the quality of decision-making: this may solve our second and third questions together. » (Duflo, 2012a, 25)

Pour autant, les moyens permettant de donner aux pauvres un espace mental ainsi que la capacité de s'engager dans un processus politique restent à définir. Le tournant opéré par Duflo pour un nouveau paternalisme soulève donc certains problèmes philosophiques.

3. La double confusion du paternalisme démocratique de Duflo

La notion de capacités, que Duflo emprunte à Sen, n'est jamais explicitement définie par cette dernière. Alors qu'une définition précise de cette notion permettrait de répondre à certains des problèmes philosophiques qu'elle soulève. Cela permettrait aussi, de comprendre précisément la teneur du tournant qu'elle propose. Dans la seconde conférence qu'elle a donné dans le cadre des *Tanner Lectures on Human Values*, Duflo cherche à présenter le rôle que peut jouer une capacité particulière, l'espoir, dans le cercle vicieux des trappes à pauvreté. Ici, elle ne définit pas véritablement les capacités mais tente d'en illustrer une particulière, ce qui permet, d'une certaine façon, de matérialiser ce que Duflo entend par capacités. Cette matérialisation me permet alors de questionner plus précisément le paternalisme envisagé par Duflo. Pour cela, il est éclairant de retourner à l'approche initiale de Sen. Ce retour à la notion de capacités, telle que définie par Sen, pointe deux principales confusions philosophiques au sein du paternalisme de Duflo. L'objectif est donc ici de questionner le tournant de Duflo. Cela me conduira à mettre en évidence un « nœud philosophique ». Ce nœud me semble n'être que la traduction de la faille épistémologique, développée dans la partie précédente de ce travail. Je commencerai par développer la seule capacité que tente de définir Duflo : l'espoir (3.1). Afin de saisir au mieux les enjeux du paternalisme souhaité par Duflo, je retournerai à l'approche initiale de Sen des capacités (3.2). Enfin, je montrerai que le paternalisme proposé par Duflo soulève deux principales confusions (3.3).

3.1 L'espoir comme capacité

Duflo, cela a été souligné, ne donne pas de définition précise des capacités. Sa deuxième conférence peut alors offrir un éclairage à ce qu'elle entend par cette notion, puisqu'elle y traite directement des capacités. Elle cherche, tout d'abord, à déterminer un lien entre les

trappes à pauvreté et l'espoir, montrant que les pauvres souffrent de désespoir³⁴⁷ et que ce dernier tend à perpétuer les trappes à pauvreté. Duflo cherche, ensuite, à évaluer l'existence d'un tel mécanisme et ces effets dans le processus de décisions des pauvres. Cela conduit Duflo à mettre en évidence le rôle d'un mécanisme dans la persistance des trappes à pauvreté. Etudier cela, me permettra à la fois de matérialiser la position de Duflo face aux capacités, tout en positionnant Duflo face aux débats théoriques (développé au chapitre 1 de cette partie) sur les trappes à pauvreté. Je commencerai par définir le rôle que donne Duflo à l'espoir (3.1.1) ; puis, je montrerai comment Duflo questionne l'existence d'un tel mécanisme au sein des trappes à pauvreté (3.1.2).

3.1.1 *Trappe à pauvreté et espoir*

Lors de sa deuxième conférence, Duflo (2012a) se demande si le problème majeur des pauvres ne serait pas la perte d'espoir. Elle se questionne sur le fait de savoir si un déficit d'espoir pourrait causer les trappes à pauvreté ; et si, de façon réciproque, un regain d'espoir ne permettrait pas de casser le cercle vicieux des trappes à pauvreté. Il est étonnant de voir Duflo supposer, de cette façon, l'existence des trappes à pauvreté, alors que bon nombre des expériences qu'elle a menées, cherchent à évaluer leurs existences. De plus, on l'a vu, dans leur dernier livre, Banerjee et Duflo invalident l'existence de trappes à pauvreté dans certains domaines, mais reconnaissent dans d'autres l'existence de telles trappes. Duflo souligne, d'ailleurs, elle-même cette ambiguïté³⁴⁸, mais en sort très habilement en montrant que la notion de trappe à pauvreté qu'elle utilise, peut être appliquée de façon logique à toutes les situations où il y a un seuil à franchir avant d'être productif. On retrouve ici, un lien très fort avec le modèle de Banerjee et Mullainathan (2008) où la notion de trappes à pauvreté se fonde uniquement sur la notion de productivité. Sortir des trappes à pauvreté, c'est devenir productif. La définition que retient ici Duflo des trappes à pauvreté est en lien direct avec les capacités de Sen. C'est donc la seconde utilisation que Duflo fait de cette notion :

³⁴⁷ Banerjee (2004) montre aussi que les pauvres souffrent de désespoir ; Banerjee (2004) ajoute, cependant, une seconde dimension car il considère qu'en plus d'être désespérés les pauvres sont vulnérables. Ces deux dimensions définissent ce qu'il appelle « les deux pauvretés » (qui est d'ailleurs le titre de son article).

³⁴⁸ « In "Poor Economics", we argue that there is no obvious evidence that most of the poor are really trapped in this sort of nutrition-based poverty trap, in part because we don't see the kind of behavior that it would suggest. However, the basic logic applies to any situation where there is a threshold to cross before investment can be productive. » (Duflo, 2012a, 33)

« Poverty traps occur whenever there is a very steep relationship between income today and income in the future over some range. For the very poor, the rate of return from investing (in food, in education, in a business) is so low that, with the resources they have, they cannot invest enough to improve their lot: they tend to become poorer and poorer. Beyond a given threshold, however, investment becomes productive, such that if someone is sufficiently rich that the outset, their income can increase in each period. » (Duflo, 2012a, 30)

« The notion of the poverty trap is related to that of capabilities, as developed by Amartya Sen briefly discussed in our first lecture: poverty deprives individuals of central capabilities, limiting their ability to achieve. That deprivation in turn limits how much income those individuals can earn, keeping them both poor and with low capabilities. » (Duflo, 2012a, 30)

Duflo illustre, tout d'abord, le rôle des trappes à pauvreté à travers le micro-crédit et des pauvres. Pour des petites entreprises, comme celles des pauvres, où il n'y a pas d'employés, un faible investissement peut avoir des rendements extrêmement élevés. Pour autant, les différentes expériences menées par le J-PAL montrent que les pauvres ne font pas ces petits investissements. Cela paraît, au premier abord, irrationnel. Cela ne répond pas au cadre standard³⁴⁹. Pour autant, Duflo considère ici les pauvres rationnels :

« Note that the poor in this case behave rationally, given a realistic perception of their possible outlook. It is trying incredibly hard despite all odds that would be irrational (or require a perspective on life which would be too rosy. » (Duflo, 2012a, 35)

Ici les pauvres ne sont pas rationnels mais agissent de façon rationnelle. Il y a donc chez Duflo, une différence entre l'individu rationnel du cadre néoclassique et le comportement des individus. Il est donc possible d'être irrationnel et d'agir de façon rationnelle. Compte tenu des difficultés auxquelles font face les pauvres, il serait irrationnel qu'ils voient la vie plus rose qu'elle n'est, et envisagent toujours positivement leur futur. Ici pour Duflo, c'est le manque d'espoir et la peur d'échouer qui freinent les pauvres pour investir et donc qui réduisent leur capacité de réaliser leur potentiel. En d'autres termes, le manque d'espoir et la peur réduisent les capacités des individus. Pour Duflo, l'espoir est une capacité, mais l'espoir permet aussi de développer d'autres capacités. Pour autant, Duflo ne définit pas précisément de quelles capacités elle parle, et le sens qu'elle y met.

³⁴⁹ Dupas (2011) soulignait aussi que le cadre standard ne permettait pas de penser les comportements irrationnels des pauvres.

3.1.2 Le mécanisme de l'espoir dans la lutte contre la pauvreté

Pour Duflo, l'espoir permet d'encourager les aspirations et ces aspirations tendent à affecter le comportement des individus. Ce qui conduit Duflo à mettre en évidence un mécanisme spécifique :

« In all these examples, we see the same mechanism at work: the existence of a step that is too high to climb creates a rational temptation to hold back, to avoid trying too hard. Having not tried, individuals may never discover what they are capable of. This worsens the poverty trap, or in some cases creates one where there was not one in the first place. One implication is that interventions that remove these steps (help with access to credit for example) can have a double dividend. Furthermore, in some cases, better information about what opportunities lies ahead may be sufficient. » (Duflo, 2012a, 41)

Le mécanisme à l'œuvre est la tentation rationnelle d'abandonner lorsque la marche à grimper semble trop élevée, impossible à surmonter. Cela crée ou empire les trappes à pauvreté. Comme les individus n'essayent pas, ils ne savent pas de quoi ils sont capables, le cercle vicieux est maintenu et empiré. Même si ce cercle n'existait pas, la non-action et l'abandon des individus tendraient à le créer. L'implication politique est donc de casser ce cercle en retirant les marches trop élevées à monter. Cela peut se faire, en facilitant l'accès au crédit des pauvres, en leur fournissant une meilleure information (afin de réduire la sous-estimation des bénéfices), en jouant sur des dispositifs de *nudging* à travers les incitations, ou encore en imposant un paternalisme démocratique. C'est sans doute ici que le lien entre l'espoir, les capacités et le paternalisme se comprend. Le paternalisme démocratique permettrait de redonner espoir aux pauvres, de transformer leurs comportements et donc de casser les trappes à pauvreté.

Duflo cherche ensuite à évaluer l'existence de ce mécanisme. Elle cherche désormais à comprendre si les capacités sont influencées par l'espoir dans le processus de décisions. Après avoir traité la décision comme résultat, Duflo cherche désormais à questionner le rôle du désespoir dans le processus de décision. Ce sont les liens entre psychologie et économie qu'elle cherche à interroger. Elle se réfère au problème d'emballage défini par Mullainathan et Shafir (2013) (les pauvres ont trop de choses à mettre dans une trop petite valise)³⁵⁰. C'est à travers ce concept que Duflo justifiait des politiques paternalistes afin que

³⁵⁰ Voir aussi (Bertrand, Mullainathan, Shafir, 2004).

les pauvres aient moins de choix à faire et qu'ils aient donc moins de choses à mettre dans leurs valises. Ce que cherche à comprendre ici Duflo est comment ce mécanisme opère, s'il opère ou si le désespoir n'est qu'un phénomène. Duflo, en reprenant sa première conférence montre qu'il existe un lien entre les infrastructures de base et le stress. Un manque d'infrastructure de base, comme l'accès à l'eau, en rendant la vie plus difficile est source de stress. Duflo explique, ensuite, que le stress renforce le fait que les individus soient incohérents temporellement. Le stress peut alors jouer un rôle non négligeable dans le processus de prise de décisions. Ce qui amène Duflo à conclure :

« These studies, nevertheless, suggest the possibility that hope and confidence may be closer to having physical manifestations (similar to proper nutrition) than we have acknowledged until now. There could be both functional and intrinsic value generating such hope. » (Duflo, 2012a, 46)

L'espoir serait l'un des mécanismes pouvant expliquer l'existence et la persistance des trappes à pauvreté, au même titre que la malnutrition. Il y aurait pour Duflo un apport intrinsèque et fonctionnel à comprendre le rôle de l'espoir. D'une part, le manque d'espoir permet d'appréhender les décisions des pauvres sous un prisme particulier. D'autre part, le rôle de l'espoir peut être fonctionnel, redonner de l'espoir aux pauvres peut avoir pour fonction de diriger leurs comportements vers des attitudes plus efficaces. Le manque d'espoir ou le désespoir permettent, pour Duflo, d'expliquer des comportements conservateurs de la part des pauvres. Ces comportements se traduisent par l'inaction ou le *statu quo* :

« An alternative interpretation of the result is that the poor are wary of engaging in any activity where they might lose anything relative to the status quo. It is conceivable that a similar explanation can account for low adoption of new technologies, as long as those technologies entail any perceived risk (a new crop, a new type of fertilizer, a new activity, a new school for child, even an insurance product). The current situation is not pleasant, but it is what it is: being pessimistic about the possibility that anything can change may lead to large losses due to extreme conservatism. » (Duflo, 2012a, 48)

La faible adoption de nouvelles technologies en matière de santé, comme les moustiquaires, s'explique donc pour Duflo, par le fait que les pauvres, par manque d'espoir sont méfiants et donc ne bougent pas, ne changent pas de situations. Si ces nouvelles technologies semblent comporter un risque, les pauvres ont peur de tout perdre et donc ne font rien. En somme, être pessimiste crée un cercle vicieux, les pauvres pensent que rien de bien ne va leur arriver, et quand quelque chose se propose à eux, ils ne l'adoptent pas de peur que

les pertes soient trop grandes. Dans la même lignée, l'inaction due au pessimisme s'exprime par le fait que les pauvres peuvent avoir peur d'échouer, et qu'ils craignent de céder aux tentations qui se trouveront sur leur chemin. Et les tentations sont beaucoup plus fortes pour les pauvres que pour les riches :

« For example, if sugary tea is the archetypal temptation good, then the rich are unlikely to be troubled by it – not because they are not tempted but because they can already afford so much tea (or other substitutes for tea that do not have to worry about their hard-earned savings being frittered away on extra cups of tea. This may not be true for the poor: tea is more substantial part of the budget so they know a non-negligible part of money they save today will be wasted tomorrow. In that case, they might as well give in to the temptation today if all they are going to do is give in to it tomorrow, especially since they know they'll enjoy it today, but do not look forward to waiting until tomorrow. » (Duflo, 2012a, 49-50)

Le fait de ne pas être serein vis-à-vis du futur, comme on a pu l'entrevoir plus haut, rend l'incohérence temporelle et ses méfaits encore plus criants. Les pauvres savent que leur route sera semée d'embûches, que les tentations seront nombreuses et qu'il sera dur d'y rester. Duflo prend alors l'exemple du thé, elle montre que les riches ne seront pas tentés par l'achat d'une tasse de thé, car ils peuvent s'en offrir plusieurs sans que cela n'ait de conséquence sur leur revenu futur. Alors que pour les pauvres, cette tasse de thé est une véritable tentation. Le prix de cette dernière est une réelle part du budget des pauvres, la dépenser c'est donc perdre cette part pour autre chose. D'autre part, si les pauvres sont pessimistes, ils auront tendance à avoir une vision négative du futur, et de ce fait privilégieront les tentations d'aujourd'hui au détriment de l'inattendu de demain. Le rôle de l'épargne, que l'on a beaucoup discuté lors du chapitre précédent avec les expériences de Dupas, est tout aussi influencé par l'espoir que l'on place en demain. En d'autres termes, les comportements d'épargne dépendent beaucoup de la façon dont les individus anticipent le futur :

« Saving behavior crucially depends on what the people expect will happen in the future. Poor people who feel that they will have opportunities to realize their aspirations will have strong reasons to cut down on their “frivolous” consumption and invest in that future. Those who feel that they have nothing to lose, by contrast, will tend to make decisions that reflect that desperation. This may explain not only the differences between rich and poor but also the differences between different poor people. » (Duflo, 2012a, 50)

Si les individus sont désespérés, cela aura tendance, selon Duflo, à se répercuter dans leur processus de prise de décision. Autrement dit, leurs décisions ne seront que le reflet d'un désespoir ou d'un optimisme. Comme cela a été vu, particulièrement au travers des

expériences de Dupas, l'épargne joue un rôle explicatif crucial. Elle est la question centrale, et les chercheurs du J-PAL tendent d'y répondre à travers l'analyse du comportement des pauvres. Le désespoir permettrait d'expliquer cette faible épargne. Il engendrerait un mécanisme néfaste, redonner de l'espoir aux pauvres permettrait de casser ce cercle. L'espoir est donc pour Duflo une capacité et cette capacité, lorsqu'elle est introduite permet de créer d'autres capacités. Autrement dit, elle engendre un processus positif. Ici, elle permettrait aux pauvres d'épargner afin d'investir dans leurs petites entreprises et d'obtenir des rendements importants qui leur donneraient la possibilité de sortir des trappes à pauvreté.

3.2 Des réelles notions de capacités et de libertés chez Sen à l'opacité philosophique de Duflo

L'approche par les capacités, développée par Sen, peut être comprise comme une alternative à celle des biens premiers de John Rawls (1971). L'approche par les capacités cherche à intégrer l'idée qu'il peut être difficile voire impossible de transformer un ou des moyens en fin - ce qui semblait pour Sen absent de la théorie de Rawls. La notion de capacité permet de penser la possibilité pour un individu d'utiliser ses moyens. Cette notion permet d'appréhender la difficile conversion des moyens en fin. Elle relève alors, pour Sen, d'une liberté substantielle. Pour autant, ces capacités ne peuvent s'acquérir qu'au travers de la liberté. Autrement dit, l'approche par les capacités de Sen accorde à la liberté un rôle à la fois substantiel et instrumental. De plus, Sen s'est toujours refusé à endosser une liste de capacités essentielles, considérant que son approche se devait d'être pluraliste, et que la hiérarchisation des capacités ne pouvait avoir lieu qu'au travers d'un processus démocratique, laissant les individus concernés déterminer eux-mêmes les capacités centrales. L'objectif est ici de rendre compte de l'approche par les capacités développée par Sen, en vue de disposer des éléments de définition d'une telle approche, pour ensuite être à même de saisir pleinement ce que recouvre le paternalisme envisagé par Duflo. Je commencerai par définir l'approche des capacités de Sen (3.2.1) ; puis, je montrerai que la liberté est à la fois instrumentale et substantielle chez Sen (3.2.2) ; enfin, j'insisterai sur le fait, que pour Sen, ce sont les individus eux-mêmes qui doivent déterminer quelles sont les capacités qui leur sont essentielles (3.2.3).

3.2.1 La notion de capacité chez Sen

Avant de définir la notion de capacités, Sen définit ce qu'il appelle les dotations. Ces dernières sont ce dont un individu dispose. Par exemple, un ouvrier dispose de sa force de travail, un propriétaire terrien de ses terres. Les dotations sont alors les ressources des individus, les moyens dont ils disposent. Ces moyens sont différents d'un individu à un autre. Ces dotations sont comme une sorte de paniers de bien différents, selon les individus. Les individus possèdent des dotations, mais ils peuvent surtout s'en servir. Ils peuvent en faire usage. Ainsi, l'ouvrier dispose de sa force de travail, mais il a aussi la possibilité ou non d'en faire usage. Certains individus ont la possibilité de convertir leurs dotations, de les utiliser et certains individus n'ont pas cette possibilité. L'usage des dotations permet, pour Sen, un droit d'accès à l'échange. C'est ici que l'on convertit ses moyens (dotations) en fin (droit d'accès à l'échange). Par exemple l'ouvrier utilise sa force de travail, il utilise donc sa dotation, et de cette utilisation il tire un droit d'accès à l'échange, qui lui permet d'obtenir autre chose qu'il désire et donc de se réaliser. Pour Sen, l'ouvrier s'est ici accompli. L'ensemble de ces dotations est appelé fonctionnements par Sen. Ils traduisent les différentes possibilités offertes aux individus, et sont extrêmement variés :

« Les fonctionnements pertinents peuvent aller de l'élémentaire – avoir suffisamment à manger, être en bonne santé, échapper aux maladies évitables et à la mortalité prématurée, etc. – au plus complexe être heureux, rester digne à ses propres yeux, prendre part à la vie de la communauté, etc. » (Sen, 1992, 65)

Ces fonctionnements, au-delà de constituer ce qu'un individu peut réaliser, contiennent aussi, par leur variété, ce qu'un individu devrait être en droit de pouvoir atteindre. Un accomplissement est donc un fonctionnement qui a abouti, un droit d'accès à l'échange. Les fonctionnements reflètent les possibilités offertes aux individus, ce qu'ils « peuvent aspirer à faire ou être » (Sen, 1999, 105). Sen ajoute le dernier concept clé permettant de comprendre celui de capacité : la liberté d'accomplir. C'est le passage entre fonctionnement et liberté d'accomplir qui définit les capacités :

« L'ensemble des capacités, représente, quant à lui, sa liberté d'accomplir, c'est-à-dire les combinaisons de fonctionnements possibles, à partir desquelles l'individu peut choisir. Si l'on suit cette « approche par les capacités », on peut alors faire porter l'évaluation, soit sur les fonctionnements réalisés (ce qu'un individu est en mesure d'accomplir) soit sur l'ensemble de capacités à sa disposition (ses opportunités réelles). On en retire deux séries différentes d'informations – dans le premier cas sur ce qu'une personne accomplit, dans le second sur ce qu'elle est libre d'entreprendre. » (Sen, 1999, 106)

Avec l'approche par les capacités, on est face à deux types de possibilités. La première, celle des fonctionnements, traduit ce qu'un individu est capable d'atteindre, compte tenu de ses caractéristiques propres et sociales. La seconde, est une possibilité en termes de choix. Un individu, au travers de sa liberté d'accomplir, choisit ou non de réaliser tel ou tel fonctionnement. Par exemple, un gréviste de la faim choisit de ne pas réaliser sa possibilité de se nourrir ; alors qu'un enfant malnutri n'a pas la possibilité de se nourrir, il ne dispose pas de ce fonctionnement. Les capacités sont donc la liberté d'accomplir des fonctionnements : « il s'agit donc d'une forme de liberté, c'est-à-dire de la liberté substantielle de mettre en œuvre diverses combinaisons de fonctionnements (ou pour le dire de façon plus concrète, la liberté de mener des modes de vie divers) » (Sen, 1999, 106). Une capacité est la possibilité de choisir parmi ses possibilités. Les capacités sont par essence des libertés ; ou comme le dit Sen une liberté substantielle.

3.2.2 Liberté substantielle et instrumentale

La notion de capacité a donc un lien plus qu'étroit avec la liberté, ces deux notions se définissent mutuellement. Duflo définit la liberté chez Sen uniquement comme une liberté positive. Sauf que chez Sen la liberté est double : « une définition adéquate de la liberté devra inclure à la fois la conception positive de la liberté et sa conception négative, toutes deux ayant de l'importance. » (Sen, 1999, 49) Sen reprend ici la distinction d'Isaiah Berlin (1958) entre liberté positive et négative. La liberté négative renvoie à la liberté en relation avec autrui, elle « a trait à la manière dont l'intervention d'un tiers peut rendre une personne « non libre » d'agir » (Sen, 1999, 47). La liberté négative dépend donc des autres, l'autre peut imputer notre liberté. En d'autres termes, cette liberté ne dépend pas de l'individu lui-même, mais de l'influence des autres individus sur sa liberté. A l'inverse la liberté positive correspond à la liberté qui traduit les caractéristiques propres des individus. La liberté n'est plus ici sous l'influence d'autrui mais sous l'unique influence de l'individu qui tente d'exercer sa liberté. En d'autres termes, la liberté positive n'est rien d'autre que la liberté d'accomplir, que nous avons défini quand nous avons cherché à définir les capacités.

De plus chez Sen, la liberté est à la fois substantielle et instrumentale. En d'autres termes, la liberté est à la fois pensée de façon positive et négative ; de plus, elle est à la fois substance et instrument. La liberté substantielle, c'est-à-dire ce qu'elle a d'essentiel, est définie par Sen au travers des capacités élémentaires :

« Par libertés substantielles, j’entends l’ensemble des « capacités » élémentaires, telles que la faculté d’échapper à la famine, à la malnutrition, à la morbidité évitable et à la mortalité prématurée, aussi bien que les libertés qui découlent de l’alphabétisation, de la participation politique ouverte, de la libre expression, etc. » (Sen, 1999, 56)

Sen évoque des capacités élémentaires, même s’il ne les liste pas.

3.2.3 Les capacités, une approche nécessairement pluraliste

L’approche par les capacités est nécessairement pluraliste, puisque les capacités sont plurielles, différentes d’un individu à l’autre. Pour Sen c’est un des principaux atouts de cette approche. Les capacités sont donc vastes, elles permettent de penser de nombreuses situations. Pour autant l’approche par les capacités n’a aucun caractère opérationnel. C’est pourquoi Sen propose de procéder par dominance afin de hiérarchiser partiellement les différents fonctionnements permettant les capacités. Il y a donc certains fonctionnements qui sont jugés pertinents, plus importants que d’autres. Par exemple, il est communément accordé qu’il est préférable de pouvoir « bien se nourrir » que de pouvoir « posséder une belle voiture ». On peut ici faire une mise en ordre partielle, comparative. Mais pour Sen il faut aussi « accepter l’incomplet » (Sen, 1992, 74). Sen n’a jamais listé les capacités, il n’a jamais défini de mise en ordre même partielle des fonctionnements. Martha Nussbaum (2000b, 41-42) a cependant proposé une liste de dix capacités essentielles^{351 352}:

1. La vie : être capable de vivre une vie d’une durée normale (ne pas mourir prématurément)
2. La santé du corps : être capable d’avoir une bonne santé (pouvoir se reproduire, être bien nourri, ...)
3. L’intégrité physique : être capable d’être libre de ses mouvements, vivre en sécurité
4. Les sens, l’imagination et la pensée : être capable d’utiliser ses sens, d’imaginer, de penser, et de raisonner
5. Les émotions : être capable d’être attaché aux choses et aux êtres en dehors de soi-même
6. La raison pratique : être capable de former une conception du bien et d’engager une réflexion critique sur les plans de la vie

³⁵¹ Nussbaum souligne que cette liste est ouverte, et qu’elle est sujette à des changements : « *I consider the list as open-ended and subject to ongoing revision and rethinking, in the way that any society’s account of its most fundamental entitlements is always subject to supplementation (or deletion).* » (Nussbaum, 2003, 42)

³⁵² Il est intéressant de souligner, que Duflo aurait pu s’inspirer de cette liste afin de déterminer le panier de biens nécessaires à l’imposition de son paternalisme, même si Nussbaum s’est toujours refusée au paternalisme, voir par exemple (Nussbaum, 2000a).

7. L'affiliation : être capable de vivre avec et vers les autres (a), avoir les bases sociales du respect de soi-même, être capable d'être traité comme une personne digne (b)
8. Les autres espèces : être capable de vivre avec des animaux, des plantes, et le monde de la nature
9. Jouer : être capable de rire, jouer, de profiter d'activités de loisir
10. Le contrôle sur son environnement : être capable de participer effectivement aux choix politiques (a), être capable de détenir des biens matériels (b)

La position de Sen face à cette liste est double. D'une part, il considère que les capacités sont absolument centrales et non-négociables, ce qui irait dans le sens de Nussbaum. Et d'autre part, il a toujours refusé d'endosser une liste de capacités. Ce refus provient du fait, que cette liste doit émerger, selon lui, d'un processus de discussion publique et démocratique. Les définir en amont, ce serait donc aller à l'encontre de ce procédé démocratique qui est au fondement de l'approche par les capacités pour Sen. Ce sont les individus eux-mêmes, au travers d'un processus de discussion et d'échange public, qui doivent déterminer cette liste. C'est pourquoi l'approche par les capacités est nécessairement pluraliste pour Sen et qu'une liste ne peut être fixée en amont. Sen met clairement, tout comme Rawls, l'accent sur la liberté. La liberté est considérée indépassable³⁵³. Les capacités sont donc doublement traversées par la liberté, car à la fois définies par cette dernière et promues par elle. Cela nécessite alors une pluralité de capacités, impossible à fixer en amont.

3.3 Le « nœud » philosophique comme faille épistémologique

Le paternalisme envisagé par Duflo semble s'éloigner de l'approche par les capacités de Sen. Les trois limites soulevées par Duflo elle-même ; me semblent être, en fait, une et même question : la question de l'imposition. La première question que soulevait Duflo est celle de la légitimité du paternalisme ; autrement dit, peut-il être légitime d'imposer quelque chose à autrui ? La deuxième était celle de ce qui devait être imposé, de la définition même du panier de biens. La dernière question concernait l'implantation d'une telle imposition et relevait de la

³⁵³ Même si comme le souligne Nussbaum cette conception de la liberté chez Sen peut recouvrir un certain trouble et n'est pas sans poser de questions : « *First of all, it is unclear whether the idea of promoting freedom is even a coherent political project. Some freedoms limit others. The freedom of rich people to make large donations to political campaigns limits the equal worth of the right to vote. The freedom of businesses to pollute the environment limits the freedom of citizens to enjoy an unpolluted environment. The freedom of landowners to keep their land limits projects of land reform that might be argued to be central to many freedoms for the poor. And so on. Obviously these freedoms are not among those that Sen considers, but he says nothing to limit the account of freedom or to rule out conflicts of this type. Indeed, we can go further: any particular freedom involves the idea of constraint: for person P is only free to do action A if other people are constrained from interfering with A.* » (Nussbaum, 2003, 44)

personne ou de l'institution qui imposerait ce panier de biens. Il apparaît donc explicite que la principale question morale, que soulève Duflo, sans jamais la nommer ainsi est la question de l'imposition (cette dernière prenant plusieurs formes). L'imposition d'un panier de biens élémentaires n'est pensée que comme un instrument pour plus de liberté. Cet instrument est supposé augmenter les capacités des individus, les rendre plus libres. Pour autant, chez Sen ce n'est que par la liberté que ces capacités peuvent être augmentées. Cela pointe une première confusion au sein du paternalisme développé par Duflo (3.3.1). De plus, en souhaitant imposer un panier de biens élémentaires, Duflo s'éloigne du cœur de la notion de capacités où l'objectif est d'offrir aux individus plus de possibilités et non plus de biens. Cela traduit une seconde confusion (3.3.2). Ces deux confusions philosophiques me semblent n'être que la traduction de la faille épistémologique de l'approche développée par le J-PAL, et que j'ai développée dans la partie précédente de ce travail (3.3.3). L'objectif est ici de rendre compte du flou philosophique de la position de Duflo afin de montrer que celui-ci n'est que le reflet philosophique d'une faille épistémologique. Pour cela, je montrerai qu'il est contradictoire de fonder un paternalisme sur la notion de liberté de Sen compte tenu du rôle à la fois instrumental et substantiel de cette dernière (3.3.1). Je m'attacherai ensuite à montrer que Duflo confond les notions de fonctionnements et de capacités chez Sen (3.3.2). Enfin, je tenterai de montrer que ces deux confusions philosophiques ne sont que la traduction des problèmes épistémologiques que rencontre la randomisation du J-PAL (3.3.3).

3.3.1 La question de l'opérationnalisation : le paternalisme comme instrument politique

La liberté est à la fois le moyen et la fin chez Sen. La position de Duflo, en pensant le paternalisme comme un moyen, est contradictoire avec celle de Sen. Elle ne partage que la finalité d'une telle perspective, lorsque le moyen utilisé est tout aussi central. C'est, d'une certaine manière, ce que Sen lui répond³⁵⁴, lorsqu'il lui dit que la vraie question est celle de la définition de la liberté. Il s'accorde avec Duflo sur le fait que la liberté a à voir avec les choix. Cependant, il nuance très fortement la position de Duflo, en montrant que les pauvres peuvent accepter qu'on leur retire certains choix, mais n'accepteront sans doute jamais qu'on leur retire la valeur de ces choix. C'est à dire que les pauvres souhaitent choisir eux-mêmes les choix qui vont leur être retirés, autrement dit que la définition du panier de biens élémentaires

³⁵⁴ Sen était, avec Angus Deaton, le discutant de la conférence de Duflo à Harvard.

leur revienne. Cela nuance très fortement l'instrument qu'est le paternalisme. Duflo fait l'impasse sur le double rôle accordé à la liberté par Sen. De ce fait, elle rend sa position confuse.

Cette opacité entre moyen et fin est renforcée lorsque Duflo évoque des décisions consensuelles, comme être contre la mortalité infantile. Il est, sans doute, vrai que tout le monde s'accorde, moralement, sur le fait que la mort des enfants est inacceptable. En effet, c'est un fait moral qui semble consensuel. Cependant, ce qu'elle met en question avec sa proposition de paternalisme, ce n'est pas le consensus sur ce fait, mais le consensus sur le moyen pour aboutir à une réduction de la mortalité infantile. Autrement dit, il y a consensus sur la finalité (une réduction de la mortalité infantile) mais il ne semble pas exister de consensus sur un instrument paternaliste pour y aboutir. Cela me permet de pointer précisément une des confusions philosophiques de Duflo, celle entre le moyen et la fin. Un autre exemple va dans ce sens. Duflo a montré que les pays riches étaient beaucoup plus paternalistes que les pays pauvres. L'exemple le plus frappant en est, pour Duflo, l'importante présence d'infrastructures dans les pays riches, comme l'existence d'un filet de sécurité sociale. Ces infrastructures ont beau réduire les choix des riches, elles ne leur ont pas été toutes imposées, ils ont choisi ces infrastructures par le biais de procédés démocratiques. Autrement dit, ces infrastructures ont peut-être une finalité paternaliste, mais leur implantation ne l'a pas été. Le paternalisme, dans cet exemple, à l'inverse de ce que prône Duflo, n'est pas un instrument politique, il est une fin choisie. Ou comme le lui disait Sen, la valeur du choix n'a pas été retirée, les riches ont choisi de s'imposer certaines décisions. Ils sont donc toujours restés libres, et c'est au moyen de la liberté qu'ils se sont retirés certaines décisions. Il semble que l'on ne peut pas confondre le moyen et la fin politique, sans rendre sa proposition opaque.

3.3.2 la définition des capacités : confusion entre processus et résultat

Duflo (2012a) limitait elle-même, on l'a vu, la détermination du panier de biens élémentaires, montrant qu'il serait compliqué de définir ces biens et de savoir qui les imposerait. Cela a aussi été noté, on ne sait pas bien à quoi renvoient ces biens pour Duflo. On ne sait pas s'il s'agit de capacités, de biens premiers, ou encore des biens élémentaires définis par la Banque Mondiale. Cependant, si je regroupe certains des exemples qu'elle évoque comme le chlore dans l'eau, la vaccination, les moustiquaires, ou encore les options

par défaut, il semble clair que Duflo ne se réfère pas aux capacités mais plutôt aux fonctionnements. Elle n'envisage pas un processus d'expansion des possibilités mais des biens, des dotations. Cela retrouve la définition des fonctionnements de Sen. L'objectif est donc d'imposer un certain nombre de fonctionnements en vue d'augmenter les capacités des individus. Imposer des fonctionnements revient à augmenter les dotations des individus. La question qui s'impose est alors : est-ce qu'une augmentation des dotations permet une augmentation des capacités ? Cela a été vu, les fonctionnements renvoient, chez Sen, à quelque chose de donné, à ce que l'on a. Tandis que les capacités renvoient à la possibilité d'utiliser ces données, ces dotations. Les fonctionnements sont figés, ils sont une donnée ; alors que les capacités sont un processus. Les capacités traduisent une possibilité de choix, elles traduisent la possibilité de traduire ces moyens (ces dotations) en fins (ce que l'on souhaite réaliser).

La question est alors de savoir si une donnée, quelque chose de figé, peut donner lieu à un processus. L'imposition de fonctionnements permettrait-elle de créer les possibilités individuelles au cœur de la notion de capacités chez Sen ? Sen distingue explicitement les fonctionnements des capacités en vue de penser la difficile conversion des moyens en fins. La double dimension de possibilités définissant les capacités permet, on l'a vu, de penser cette conversion et ses difficultés. Sen différencie explicitement les dotations (les données) du processus, et cette distinction est au cœur de la notion des capacités. Il est difficile de penser qu'un nombre plus important de fonctionnements puisse augmenter les capacités. Prenons un exemple concret, si une personne est amputée des deux jambes, et qu'elle dispose d'un vélo (ce vélo fait partie de ces fonctionnements) elle n'a pas la capacité de l'utiliser, elle n'a pas la possibilité d'utiliser ce fonctionnement puisqu'elle n'a pas de jambes. Augmenter les fonctionnements de cette personne, en lui offrant d'autres vélos, n'aura donc aucun effet sur ces capacités. Dans cet exemple, c'est la caractéristique individuelle d'un individu qui le prive d'une capacité, qui l'empêche de faire usage de ses fonctionnements.

Prenons un autre exemple, celui d'une femme vivant dans un environnement social hostile à sa condition (cette femme ne peut y être pleinement libre, elle est donc privée de certaines capacités), lui imposer un certain nombre de fonctionnements, ne permettra pas de changer sa condition, cela ne changera rien à ses capacités. Autrement dit, le développement des capacités ne peut se faire qu'au travers de la liberté d'accomplir. C'est la liberté d'accomplir qu'il faut mettre en place et non des fonctionnements. Ces derniers ne pourront pas être

utilisés si cette liberté n'existe pas. Cela renvoie explicitement aux travaux fondateurs de Sen (1982) sur les famines au Bengale en 1943. La nourriture était présente mais les individus avaient perdu leurs droits d'accès, la possibilité de se fournir en nourriture. L'environnement politique n'étant pas libre, le gouvernement n'avait alors aucune incitation à nourrir la population. C'est ici que Sen donne un rôle central et instrumental à la liberté de la presse en vue d'augmenter les capacités :

« Aucun pays démocratique, même le plus pauvre, n'a jamais subi de famines. La raison en est que la prévention de ce fléau ne présente aucune difficulté pour autant que les gouvernements aient la volonté de la mettre en œuvre. Bien entendu, une démocratie multipartite, dotée de médias libres et d'un système électoral, constitue, pour un gouvernement une forte incitation politique à prendre les mesures préventives. »
(Sen, 1999, 76)

Le rôle substantiel et instrumental de la liberté est ici saillant. La proposition de Duflo revient à augmenter le nombre de dotations, mais si les individus ne peuvent pas se servir de ces dotations, cela ne produira, sans doute, aucun effet. C'est ici que la position de Duflo face aux capacités, qu'elle met pourtant au cœur de son paternalisme, est confuse. Elle s'éloigne du fondement de cette notion. Dans la continuité de la confusion entre moyens et fins, Duflo confond une donnée avec un processus. Pour transformer un processus, il faut jouer sur ce processus. Il faudrait donc imposer un processus et non un panier de biens.

3.3.3 De la confusion philosophique à la faille épistémologique

Afin d'agir sur les capacités, de favoriser la liberté des pauvres et donc leur permettre de sortir des trappes à pauvreté, il faudrait imposer non pas un panier de biens, mais des opportunités ou des processus permettant de favoriser les capacités. Pour être en mesure de déterminer des outils qui favoriseront les capacités des pauvres, il faut d'abord comprendre l'environnement personnel et social qui les entoure. Il est nécessaire de s'intéresser au contexte des individus, aux institutions qui les entourent, à l'environnement social qui les définissent et qui les traversent. Autrement dit, il faut comprendre, déterminer de quelles capacités ils sont privés, si l'on veut agir dessus. Dans la perspective de Duflo, il ne faudrait pas seulement rendre compte de leurs mauvais choix, ou analyser seulement leur irrationalité ; mais tenter de comprendre le processus de décision auquel ils font face. Il faut rentrer dans la boîte noire.

La perspective du J-PAL ne donne pas accès à ce processus. Elle ne donne pas accès à ces différents environnements. En privilégiant une très forte validité interne aux dépens d'une validité externe, elle peine à produire (cela a été vu) des capacités causales. En ce sens, elle peine à mettre en évidence des mécanismes causaux. Comme je l'ai montré elle souffre de l'absence d'une théorie. En l'absence de cette théorie, il apparaît difficile de comprendre l'environnement des pauvres, leurs contextes, les institutions qui les entourent. Les expériences du J-PAL ne montrent pas les processus à l'œuvre dans les décisions des pauvres. Les expériences du J-PAL donnent à voir des résultats, créent des données solides sur la vie des pauvres, mais elles offrent une image figée de la vie des pauvres, et ne permettent pas de comprendre ce qui se cache derrière cette image. Aucun schéma causal n'est déduit. Pour cela, il faudrait, soit comme Deaton le propose procéder par hypothèses et erreurs, soit comme Cartwright l'envisage, utiliser la randomisation en vue de tester une capacité causale particulière. Dans ce dessein, il est nécessaire de construire un modèle causal, de dessiner une histoire causale. Dans les deux cas, que ce soit dans les propositions de Deaton ou de Cartwright, il faut poser des hypothèses, il faut supposer une théorie *ex ante*. Sans cette théorie les expériences randomisées du J-PAL sont vouées à produire des données fiables sur la vie des pauvres, mais ne peuvent produire de recommandations politiques utilisables dans différents contextes.

Cette insoluble tension entre validité interne et externe traduit la confusion philosophique du paternalisme démocratique souhaité par Duflo. Les expériences du J-PAL ne donnant accès qu'à des données et non à des processus, il peut sembler logique que Duflo propose un panier de biens avec des éléments donnés, fixes. Avec l'évocation de ce paternalisme démocratique, Duflo souligne implicitement la nécessité de s'intéresser non plus uniquement à l'image de la vie des pauvres mais à ce qui se cache derrière l'image. Cela pointe la nécessité d'une plus forte validité externe. Une autre étape est désormais à envisager, celle de la compréhension de l'image. Dans sa seconde conférence, au travers des capacités et de l'espoir, c'est ce que, d'une certaine manière, Duflo essaye de faire : mettre en évidence un mécanisme. Elle met en évidence le fait que les pauvres n'épargnent pas car ils souffrent de désespoir. L'approche du J-PAL permet donc ici de pointer un mécanisme, mais ne permet pas de voir un schéma causal. Elle ne permet pas d'intégrer ce mécanisme à un processus plus large. Pour autant, elle développe un mécanisme qui n'avait pas été mis en lumière

auparavant. Elle invite à se re-questionner. Elle invite à « Repenser la pauvreté »³⁵⁵, elle invite à voir la vie des pauvres, elle invite à déplacer les débats. Elle invite à comprendre, que par définition la vie des pauvres, les difficultés auxquelles font face les pauvres transforment leurs choix. Elle invite à balayer certains *a priori*. Elle montre, qu’une fois de plus, l’appréhension de la pauvreté est un phénomène extrêmement complexe. Dans cette perspective, il n’y a pas une méthode mais plusieurs et elles devraient se nouer.

4. Conclusion

Le récent tournant d’Esther Duflo en faveur d’un paternalisme démocratique est dans la droite ligne de ce qui était déjà sous-jacent dans les recommandations politiques testées et évoquées par Pascaline Dupas sur le paludisme. Les dispositifs de *nudging* peinant à contrecarrer les comportements inefficaces des pauvres en vue de sortir de la pauvreté. Ces dispositifs semblent inefficaces car ils ne rendent pas compte des mécanismes sous-jacents de la pauvreté, mais cible une unique dimension sans rentrer dans la boîte noire des résultats, sans comprendre ces derniers. Le paternalisme invoqué par Duflo s’avère souffrir de limites semblables.

Duflo envisage un paternalisme permettant d’augmenter la liberté des pauvres et envisage donc le développement comme le fait Sen (1999), c’est à dire comme un processus d’expansion des libertés. Cette idée est extrêmement intéressante et novatrice. Elle pose de façon directe et explicite la question de l’implantation de politiques permettant d’éliminer l’extrême pauvreté. En ce sens, elle pose la question de l’opérationnalisation, de la mise en place concrète de l’approche par les capacités. Pour autant, la position de Duflo est confuse à un double niveau. Elle semble confondre le rôle instrumental et substantiel de la liberté, tout en s’écartant du fondement même de la notion de capacités. Ces limites se retrouvent pleinement : à privilégier la validité interne de leurs expériences les chercheurs du J-PAL ne peuvent appréhender les processus sous-jacents aux phénomènes dont ils tentent de rendre compte.

Pour parvenir à rendre compte de ces processus, il faudrait pouvoir mettre en évidence des capacités causales et donc définir des mécanismes afin de produire un schéma causal. A elle

³⁵⁵ « Repenser la pauvreté » est le titre en française de son dernier livre avec Banerjee ; voir Banerjee et Duflo (2011).

seule, la randomisation, malgré une forte validité interne, n'y parvient pas. Afin de proposer de tels schémas causaux il faudrait rentrer dans la boîte noire des résultats, se demander pourquoi un programme fonctionne ou ne fonctionne pas. Il faudrait stipuler certaines hypothèses. Il faudrait supposer une théorie *a priori* dont la randomisation pourrait tester les mécanismes clefs.

Conclusion de la Partie 3.

Les chercheurs du J-PAL ont pour ambition d'offrir une alternative au débat sur l'aide au développement – débat qui devient central à la suite de l'effondrement du « consensus de Washington ». L'objectif est d'offrir une troisième voie à ce débat en mettant en évidence des preuves d'efficacité de l'aide internationale. Ce débat se matérialise particulièrement autour de la question du paludisme, les « pro-aides » promouvant la subvention intégrale des moustiquaires pour lutter contre cette maladie, quand les opposants à l'aide suggèrent qu'elles soient payantes même à un faible prix. Pascaline Dupas mène plusieurs expériences au Kenya afin de tester la validité de chacune de ces positions. La première expérience qu'elle réalise incarne parfaitement les objectifs du J-PAL, poser non pas la question de l'aide de façon générale, mais la poser de façon concrète à travers la question du prix des moustiquaires. Cette expérience vise donc explicitement à trancher le débat entre les positions de Sachs et d'Easterly. Elle vise à offrir des preuves scientifiques concernant l'efficacité de l'aide.

La première expérience de Cohen et Dupas (2010) met en évidence une énigme : les pauvres n'utilisent pas suffisamment les moustiquaires, même lorsque ces dernières sont gratuites. Le résultat de cette expérience offre alors une position intermédiaire, une troisième voie, entre les points de vue de Sachs et d'Easterly. Il est plus efficace, compte tenu de l'élasticité de la demande, que les moustiquaires soient gratuites, cependant elles restent insuffisamment utilisées. Elle pointe donc, comme l'envisageait le J-PAL, un phénomène que la théorie n'avait pas vu, pas anticipé. En ce sens, elle peut être considérée comme un apport important. A la suite de cette première expérience, Dupas (2009) tentera de trouver une solution pour contrecarrer la faible utilisation des moustiquaires. Elle teste deux dispositifs de *nudging* visant à inciter les pauvres à utiliser les moustiquaires. Ces deux dispositifs n'ont, pourtant, aucun effet. Dupas explique alors que les pauvres souffrent d'un manque de rationalité. Ils souffrent d'incohérence temporelle et sous-estiment les bénéfices qu'ils peuvent tirer des biens de prévention en matière de santé. C'est donc, en toute logique, qu'elle teste des dispositifs s'apparentant au paternalisme libertarien de Sunstein et Thaler. Ces dispositifs visent à transformer la présentation des choix afin de rendre ces dernières efficaces. Face à l'échec de tels dispositifs, Dupas évoque cependant le rôle de politique de l'offre, c'est-à-dire provenant des gouvernements, elle souligne l'importance de politique d'envergure. Cette position laisse alors entrevoir le début d'un paradoxe : invoquer des politiques d'envergures, quand la position initiale du J-PAL est de s'opposer à ce type de

politiques et de privilégier des interventions locales à la marge. Retracer les expériences de Dupas sur le paludisme permet de mettre en évidence une première explication latente à ce paradoxe. Dupas teste des dispositifs cherchant à contrecarrer un comportement, elle suppose l'irrationalité des pauvres, sans pour autant jamais la tester, et sans jamais chercher, à travers des expériences, à comprendre la faible utilisation des moustiquaires. Aucun élément de compréhension, face à la première énigme posée n'est mis en évidence.

Dans la droite ligne des expériences de Dupas, Duflo considère que les pauvres souffrent d'incohérence temporelle et qu'ils sous-estiment les bénéfices. Cependant, pour Duflo il est clair que les riches, comme les pauvres, souffrent des mêmes biais de rationalité. Même si sa position face à la notion de rationalité est un peu opaque puisque Duflo se réfère, en fait, à un manque d'efficacité. Mais à l'inverse des pauvres, les riches vivent dans un monde rempli d'incitations invisibles, remplis de *nudge*, les aidant à opter pour un comportement efficace. C'est donc ici la principale justification, pour le paternalisme libertarien de Sunstein et Thaler. Pour autant, comme le présageait l'évocation par Dupas de politiques d'envergures, Duflo en appelle à un paternalisme démocratique. La principale différence entre les riches et les pauvres, pour Duflo, est que bon nombre de décisions sont déjà prises à la place des riches tel n'étant pas le cas pour les pauvres, alors même qu'ils ont à vivre avec si peu d'argent. Le paternalisme démocratique de Duflo entend se fonder sur l'approche par les capacités de Sen et lui emprunter sa définition de la liberté. Ce paternalisme est pensé comme un instrument pour davantage de liberté. Duflo, elle-même, soulève les enjeux moraux d'un tel instrument. Pour autant, en creusant l'approche par les capacités de Sen, le paternalisme invoqué par Duflo admet deux principales confusions. Duflo tend à confondre le rôle instrumental de la liberté chez Sen avec son rôle substantiel. De plus, le paternalisme de Duflo tend à s'éloigner fortement du cœur de la notion de capacités, définie par Sen, en négligeant la double possibilité qu'elle traduit. Ces deux confusions sont ce que j'ai appelé le nœud philosophique de la proposition de Duflo. Ce nœud me semble n'être que la traduction de la faille épistémologique de l'approche défendue par le J-PAL. En mettant au premier plan la validité interne de leurs résultats, les chercheurs du J-PAL ne peuvent rendre compte ni des processus ni des mécanismes à l'œuvre. En l'absence d'une théorie *ex ante*, les résultats obtenus par les expériences randomisées du J-PAL peinent à mettre en évidence les mécanismes sous-jacents aux comportements dont ils rendent compte.

Pour autant, les résultats obtenus par de telles expériences offrent une nouvelle voie au débat sur l'aide au développement. Ils soulèvent des questions qui étaient, auparavant, absentes du débat. Ils tendent à rendre visible la vie des pauvres et à soulever des enjeux nouveaux. En creusant la notion de capacités, au travers de l'espoir, Duflo renouvelle la vision des trappes à pauvreté développées par Sachs. Elle tend à appréhender ces dernières à travers les comportements des pauvres, en les individualisant. Elle redéfinit en réalité la notion de trappe à pauvreté. Le rôle de l'épargne semble être un des éléments majeurs mis en avant par les chercheurs du J-PAL. Cela était déjà visible dans les expériences de Dupas. A travers ce qu'elle nomme le désespoir, Duflo cherche à mettre en évidence un tel mécanisme au même titre que le rôle de la malnutrition dans les trappes à pauvreté. En ce sens, il est indéniable que l'approche du J-PAL a contribué à renouveler le débat sur l'aide au développement, même si les objectifs souhaités ne sont pas tous atteints encore. Concernant l'implication politique sous-jacente à un tel débat, en invoquant un paternalisme démocratique, Duflo limite très fortement le rôle de l'*empowerment* dans la lutte contre la pauvreté. Dans cette perspective elle met en évidence que la lutte contre la pauvreté ne peut être uniquement l'affaire des pauvres, mais doit unir les efforts de chacun. En ce sens, elle pointe très clairement la question de l'opérationnalisation de l'approche par les capacités de Sen. Elle met au premier plan le rôle des actions concrètes dans la lutte contre la pauvreté, ce qui avait été auparavant largement négligé. Elle donne alors un nouveau souffle à la lutte contre la pauvreté, montrant aussi que ce dernier ne peut engendrer un réel tournant si les expériences randomisées restent le seul outil.

Conclusion Générale

« Il est évident que toutes les sciences, d'une façon plus ou moins importante, ont une relation à la nature humaine, et que, si loin que l'une d'entre elles peut sembler s'en écarter, elle y revient toujours d'une façon ou d'une autre. Même les *mathématiques*, même la *philosophie naturelle* et la *religion naturelle* dépendent dans une certaine mesure de la science de l'HOMME, car elles tombent sous la connaissance des hommes et sont jugées par leurs pouvoirs et leurs facultés. Il est impossible de dire quels changements et quelles améliorations nous pourrions faire dans ces sciences si nous connaissions entièrement l'étendue et la force de l'entendement humain. » (Hume, 1739, 4)

La randomisation du J-PAL a-t-elle constitué un réel tournant épistémologique ? Cette approche a, de façon certaine, insufflé une nouvelle manière d'appréhender la pauvreté et certaines de ses manifestations : elle a rendu ces dernières mesurables et visibles et a placé l'évaluation au centre des débats. La pauvreté serait désormais une histoire d'évaluation. Le J-PAL a construit une véritable « machine d'évaluation » en formant plus de 800 personnes à sa méthode, en étant présent sur chaque continent, en multipliant le nombre de ses partenaires, à la fois de ses partenaires de terrain et de ses partenaires financiers, en publiant dans les meilleures revues académiques, en cherchant à étayer un lien toujours plus étroit avec les diverses agences internationales, et en essayant d'être au plus près et au plus accessible pour les décideurs politiques.

En ce sens comme le soulignaient Angrist et Pischke (2010), la randomisation du J-PAL constitue un véritable tournant empirique, voire une révolution empirique. Quant à savoir s'il s'agit d'une « révolution totale », c'est à dire à la fois empirique, théorique et politique, la réponse est moins catégorique. L'apport théorique du J-PAL à la lutte contre la pauvreté reste nuancé et peu clair, les preuves d'efficacité voulues par le J-PAL ne sont pas explicites, elles ne permettent pas de quitter un registre abstrait. Savoir si l'aide internationale doit être massive ou inexistante reste une question à laquelle l'approche du J-PAL ne permet pas de répondre totalement. La théorie qui devait émerger des évaluations apparaît finalement opaque et comme suspendue à la formulation d'une théorie *ex-ante*, qui offrirait une ligne

directrice aux différents résultats obtenus mais qui, en l'absence de cette dernière, restent éclatés.

De ce fait, la « révolution douce » souhaitée par les chercheurs du J-PAL et constituée d'une multitude de changements à la marge semble encore lointaine. La formulation théorique apparaît comme nécessaire à la fois pour unifier les différents résultats et pour penser une action politique contre la pauvreté qui soit plus explicite. Le J-PAL parvient à fournir des preuves : des preuves car la fiabilité de leurs résultats est solide. Néanmoins, cette fiabilité semble avoir un coût : ces preuves sont éclatées et difficilement utilisables par la sphère politique, définissant une la faille épistémologique. A la question qui a tenu lieu de fil rouge tout au long de ce travail (dans quelle mesure la faille épistémologique de l'approche du J-PAL explique-t-elle sa difficulté à fournir des recommandations politiques claires de lutte contre la pauvreté ?), la réponse est qu'il existe une tension inhérente à de l'approche du J-PAL entre la volonté de produire des résultats fiables et solides et celle de les rendre accessibles aux décideurs politiques. Cette approche mettant l'accent sur la production de preuves, le fait au détriment de leur portée politique.

J'ai essayé, tout au long de ce travail, de démontrer ce point. Pour ce faire, j'ai tenté de conduire une analyse épistémologique construite autour de deux principales dimensions : une dimension méthodologique et une dimension théorique.

La première partie de cette thèse s'est attachée à la dimension méthodologique de l'analyse autour de la question de la validité interne de la randomisation. J'ai retracé l'histoire de la randomisation et montré qu'une telle méthode avait toujours été utilisée et promue pour la forte validité interne qu'elle offrait. La première introduction d'une dimension aléatoire au sein d'un protocole expérimental, par Peirce, chargée de contredire la loi de Fechner, vise à renforcer la fiabilité des conclusions obtenues. C'est autour de la conclusion de Peirce, évoquant la télépathie, que plusieurs milliers d'expériences, comprenant une dimension aléatoire, sont réalisés à Stanford un siècle plus tard. Néanmoins, il faut attendre Ronald Fisher pour que le protocole expérimental et le cadre statistique de la randomisation soient clairement définis. Fisher souhaite que ces derniers soient extrêmement rigoureux – et la randomisation est considérée comme la garantie de la fiabilité des résultats. Fisher offre ainsi à la randomisation sa dimension de *gold standard* méthodologique. Pour autant, le protocole envisagé par Fisher ne connaît pas son succès le plus important dans le domaine pour lequel il avait été initialement pensé, l'agriculture. C'est la médecine qui offre à la randomisation son utilisation la plus intense au travers des essais cliniques, considérés comme méthodologie

fiable. L'histoire de la randomisation m'a ainsi permis de pointer l'une des dimensions centrales de cette approche, sa forte validité interne. Et c'est cette dimension que le J-PAL importe, lui ajoutant le rôle des expériences en sciences sociales, c'est à dire l'évaluation de politiques publiques, cela même si celles menées par le J-PAL se réalisent à une échelle plus restreinte. Ces deux histoires, celle de la randomisation au travers de l'introduction d'une dimension aléatoire dans le procédé expérimental, et l'histoire des expériences sociales de terrains pointent les deux objectifs du J-PAL. La première histoire permet d'illustrer le premier objectif du J-PAL : produire des preuves, et la seconde histoire traduit la volonté du J-PAL de guider la décision politique.

Le J-PAL parvient à fournir des preuves, en reprenant le cadre défini par Fisher et en se fondant sur le modèle de Rubin, ce qui lui permet d'annuler tout biais de sélection. Ces preuves sont ensuite destinées aux décideurs politiques en vue d'éclairer leurs décisions. Le J-PAL s'inscrit ainsi dans le mouvement d'*evidence-based policy*, dont l'objectif est de fonder les décisions politiques sur des preuves expérimentales. Ce mouvement s'inspire de *l'evidence based medicine*, où l'objectif est de fonder les décisions médicales sur ce même type de preuves. Afin de favoriser la diffusion de ces résultats vers les décideurs politiques le J-PAL constitue des « malles politiques » dans lesquelles les différents résultats sont regroupés par thème et synthétisent les analyses coûts-bénéfices que le J-PAL a conduites.

L'histoire croisée de la randomisation et des expériences sociales de terrain m'a donc permis d'illustrer les deux objectifs du J-PAL. De plus, elle m'a aussi permis de situer historiquement les expériences voulues par le J-PAL, en définissant sa structure institutionnelle, son fonctionnement et son histoire interne qui montre qu'un tel laboratoire connaît un essor fulgurant depuis 2007, avec une antenne sur chaque continents et des évaluations toujours plus nombreuses à travers le monde.

La deuxième partie de ce travail a continué à s'attacher à la dimension méthodologique de l'analyse mais, s'est concentrée sur le second objectif du J-PAL, pour en questionner la validité externe. J'ai ainsi posé la question de l'utilisation des résultats du J-PAL dans la sphère politique. Et cela à un double niveau. Le premier niveau interroge l'utilisation d'un résultat d'une expérience donnée vers un autre contexte, et le second questionne la généralisation d'un résultat à une échelle plus large.

Pour le premier niveau, je me suis appuyée sur le cadre analytique de Nancy Cartwright qui distingue la production de la preuve de son utilisation. Cartwright montre que la randomisation possède une importante validité interne. En ce sens, elle parvient à produire

des preuves, ou ce qu'elle appelle des lois causales. Mais pour être utilisées dans la sphère politique, ces lois causales doivent pouvoir se transformer en capacités causales. Lorsque ces dernières sont mises au jour, il devient possible d'envisager si elles perdureront dans un autre contexte. Ces capacités causales prennent sens au sein d'un « modèle causal », qui doit être rendu explicite afin que les capacités causales en jeu puissent être précisées. Vincent Guillin interroge la mise en évidence de capacités causales par la randomisation du J-PAL. Il montre que, pour l'instant, la randomisation du J-PAL ne permet pas d'en rendre compte. Mais, à l'inverse de Cartwright, il souligne que ce programme de recherche est encore très récent et qu'il peut progresser, Cartwright elle doutant de la possible mise en évidence de capacités causales dans le monde social et soulevant l'argument selon lequel ce dernier est sans doute trop holistique. Le premier niveau de validité externe en est donc fortement affaibli.

Il en va de même pour le second niveau. Les différentes critiques adressées à la randomisation du J-PAL ont en commun de soulever le problème de la validité externe de cette approche : problème qui est traduit par le fait que l'approche du J-PAL ne se demande pas « pourquoi un programme fonctionne ». On passe d'une hétérogénéité des effets de traitement à une hétérogénéité indifférenciée. La première soulève le fait que la randomisation donne accès à la moyenne des effets du programme sur le groupe de traitement et sur le groupe de contrôle. Cette moyenne peut renvoyer à de nombreuses situations différentes. De plus, en ne s'interrogeant pas sur la question de savoir « pourquoi un programme fonctionne », les résultats du J-PAL se transforment en boîte noire. Ils apparaissent difficilement utilisables dans un autre contexte puisqu'aucun élément n'est mis en évidence afin de penser cette utilisation. En d'autres termes, ces résultats sont valables en équilibre partiel mais rien ne dit qu'ils seront toujours valables si on les étend à une échelle plus large, dès lors qu'il existe de nombreux effets d'équilibre qui peuvent transformer les résultats obtenus en équilibre partiel. On retrouve alors ici, un problème épistémologique plus large et toujours irrésolu : le problème de l'induction, traduisant le difficile passage des faits à des principes déductifs.

De plus, le J-PAL tend ainsi à perdre de vue la « grande image » - et donc à oublier le rôle central que peuvent jouer certains facteurs politiques. Le second niveau de la validité externe d'une telle approche s'en voit donc fortement fragilisé. Pourtant, selon Acemoglu, il est possible de renforcer cette validité externe en reliant la randomisation à des approches structurelles. Mais cela suppose de perdre en validité interne afin de gagner en validité externe, ce à quoi les chercheurs du J-PAL semblent se refuser.

De l'analyse de Cartwright aux critiques qui émanent de l'économie du développement, ou aux critiques économétriques, cet ensemble renvoie à la dimension a-théorique de l'approche du J-PAL.

Deaton soulève la nécessité de s'intéresser à la recherche de mécanismes : il propose une méthode hypothético-déductive, déjà évoquée par Cartwright. Cette méthode permet de rendre compte des mécanismes à l'œuvre dans le processus d'un aller-retour entre la théorie et l'empirisme. Deaton l'illustre à partir de différents liens centraux en économie du développement. Développer la position de Deaton permettrait d'offrir un point de départ et un cadre à une réflexion sur le rôle de la théorie dans l'approche du J-PAL. Le terme de théorie permet de penser simultanément la réflexion de Deaton sur les mécanismes et celle de Cartwright sur les capacités causales. Bien évidemment, cela suppose un raccourci, tout comme cela suppose d'entendre le terme « théorie » de façon large : comme la recherche et la définition de mécanismes ou de capacités causales.

Cela m'a conduit à ne plus distinguer entre les termes capacités causales et ceux de mécanismes, et donc à m'éloigner des épistémologies spécifiques de Deaton et de Cartwright. Lors d'un symposium tenu en 2005, la question du rôle de la théorie dans l'approche du J-PAL a été explicitement posée. La réponse de Banerjee à cette question, déjà présente dans de nombreux travaux co-écrits avec Duflo, définit bien les deux statuts qu'ils accordent à la théorie. Le premier concerne ce que j'ai appelé le refus d'une théorie *ex-ante* ; le second est caractérisé par la volonté de construire une théorie *ex-post*. Mais, la construction de cette théorie *ex-post* apparaît compromise par l'absence d'une théorie *ex-ante*. De plus, en l'absence de théorie *ex-ante*, l'utilisation des résultats du J-PAL dans la sphère politique est faible, tout comme la possibilité d'en élargir le spectre d'application. C'est l'absence d'une théorie *ex ante* qui empêche la mise en évidence de certains mécanismes ou de certaines capacités causales, et en ce sens, qui fragilise à un double niveau la validité externe de l'approche du J-PAL.

Le refus de théoriser en amont est en lien direct avec le premier objectif que se fixe le J-PAL, à savoir produire des preuves. Théoriser en amont ferait perdre à la randomisation une part de sa validité interne, ce qui entraînerait des résultats moins fiables, et donc n'apporteraient plus de preuves à proprement parler. C'est là la faille épistémologique d'une telle approche.

La deuxième partie de cette thèse a permis de questionner la validité externe de l'approche du J-PAL sous un angle méthodologique. J'ai tenté de centrer l'analyse sur la méthode de la

randomisation, et la façon dont elle est utilisée en économie du développement. Ce questionnement a mis en évidence la tension existante entre validité interne et validité externe de la randomisation du J-PAL.

La troisième partie de la thèse a été consacrée à la dimension théorique de l'analyse. Elle a donc continué à questionner la validité externe du J-PAL, mais du point de vue de sa portée théorique, l'objectif étant d'évaluer l'apport du J-PAL à l'économie du développement. Comme on l'avait esquissé dans la partie précédente de ce travail, la portée théorique du J-PAL est apparue comme étroitement liée à sa portée politique.

Le J-PAL se réfère, presque uniquement, au débat sur l'aide au développement. C'est face à ce dernier qu'il souhaite se positionner et offrir des réponses. J'ai présenté ce débat, afin d'être à même d'y situer les apports du J-PAL. Dans ce débat fortement clivé, le rôle choisi par le J-PAL consiste à le trancher concrètement, à l'aide de preuves expérimentales. De plus, ce débat trouve une traduction politique, tout aussi clivée. Les opposants à une aide massive ont tendance à favoriser l'autonomisation des pauvres – *l'empowerment* –, au nom d'une analyse de l'assistance extérieure comme forme d'impérialisme occidental. Cette autonomisation vise à faire des pauvres les premiers acteurs de leur pauvreté. À l'inverse, les partisans d'une aide internationale massive acceptent toute forme d'assistance.

Le débat s'est alors cristallisé autour de la question de la distribution de moustiquaires en vue d'éradiquer le paludisme dans les pays en développement. Les tenants de l'aide insistent sur la nécessité de subventionner intégralement et de distribuer massivement ces moustiquaires, compte tenu de leur efficacité dans la lutte contre le paludisme et de leur faible prix. L'autre camp pointe le fait que rendre les moustiquaires payantes, même à un faible prix, poussent les individus à valoriser ce bien et donc à l'utiliser pour sa fonction. J'ai détaillé toutes les expériences de Pascaline Dupas portant sur cette question. D'une part, ces expériences permettent d'illustrer précisément, sur un thème particulier, l'approche du J-PAL ; d'autre part elles permettent d'évaluer concrètement l'apport théorique du J-PAL au débat sur l'aide au développement.

La première expérience menée par Pascaline Dupas est extrêmement intéressante car elle éclaire singulièrement le débat sur l'aide au développement et les objectifs du J-PAL. Cette expérience montre que la demande est très élastique par rapport au prix, confirmant ainsi la position de Sachs. Mais surtout, cette expérience pointe une énigme : même si elles sont intégralement subventionnées, les moustiquaires ne sont pas suffisamment utilisées pour lutter efficacement contre le paludisme. La première expérience de Dupas montre qu'il ne

suffit pas uniquement de savoir si les moustiquaires doivent être gratuites ou payantes, comme il ne suffit pas de déterminer si ces dernières arrivent à destination. En ce sens, elle pose une nouvelle question à l'aide au développement : comment favoriser l'utilisation des moustiquaires ?

Les autres expériences que Dupas mène visent à contrecarrer cette faible utilisation, en testant des dispositifs de *nudging* incitant les individus à utiliser davantage les moustiquaires. Pour cela, ces dispositifs de *nudging* cherchent à pallier l'incohérence temporelle des pauvres et leur possible sous-estimation des bénéfices. En testant ces dispositifs, Dupas suppose l'incohérence temporelle des pauvres tout comme la sous-estimation de leurs bénéfices, mais ne les teste jamais. Les expériences ne posent jamais la question de savoir « pourquoi les moustiquaires sont-elles si peu utilisées » ? Tout comme elles ne demandent pas « pourquoi les dispositifs de *nudging* sont-ils inefficaces » ? De ce fait, l'énigme posée par la première expérience persiste. Dupas souligne alors le rôle important des politiques de l'offre et des politiques publiques à grande échelle en matière de santé.

Tout comme Dupas, Duflo tente d'expliquer l'inefficacité du comportement des pauvres, par un biais de rationalité. Ce dernier est, pour Duflo, propre à tout un chacun, à l'exception des « riches » qui vivent dans un monde entouré d'incitations invisibles qui leur permettent d'adopter un comportement efficace. De plus, les riches disposent de nombreuses infrastructures, et beaucoup d'institutions décident à leur place. Ainsi, la différence réside dans le fait que les riches vivent dans un monde paternaliste lorsque les pauvres vivent dans un monde où toutes les décisions leur reviennent. Duflo propose alors d'être paternaliste vis-à-vis des pauvres, au nom d'un paternalisme dépassant le paternalisme de Thaler et Sunstein.

Le paternalisme qu'envisage Duflo a pour objectif de favoriser la liberté des individus pauvres. Il entend se fonder sur la notion de « libertés » développée par Sen. Cette dernière se comprend au travers du concept de capabilité, qui traduit l'idée d'une double possibilité : celle des fonctionnements et celle de la possibilité d'accomplir ces fonctionnements. La liberté est donc, pour Sen, à la fois instrumentale et substantielle chez Sen. Duflo, en pensant le paternalisme comme un instrument pour plus de liberté, oublie la dimension instrumentale de la liberté chez Sen. En outre, en envisageant l'imposition d'un panier de biens, Duflo semble s'éloigner de la notion de capabilité et elle tend à privilégier les fonctionnements au détriment des capacités. La position de Duflo est alors confuse à un double niveau. En se concentrant sur la production de preuves, dans le but de développer des politiques à la marge en vue de produire d'importants effets politiques, cette approche perd de vue la nécessité d'envisager la pauvreté et la lutte contre la pauvreté comme un processus.

Le refus d'une théorie *ex ante*, ou de la recherche de capacités causales ou de mécanismes, permet de saisir ce flou philosophique. Celui-ci n'est que la traduction d'un problème épistémologique. Ne parvenant pas à produire de recommandations politiques claires à travers la seule randomisation, Duflo s'en remet au paternalisme. Mais ce dernier n'échappe pas au problème inhérent à son approche.

La double confusion de Duflo montre que la philosophie sous-jacente à une telle approche n'est pas encore clairement définie, ce qui tend à fragiliser l'action – et ce que traduisent les différentes critiques adressées au J-PAL.

Il reste qu'en plaçant l'action en son centre, une telle approche a offert à la lutte contre la pauvreté un nouveau souffle. Elle a mis en lumière le rôle du terrain. Elle a redonné aux pauvres une place centrale dans la lutte contre la pauvreté, sans pour autant les définir comme seuls acteurs de leur situation. Se voyant comme des « plombiers », Duflo (2009, 27) et le J-PAL ont cherché à définir les contours de la pauvreté, à faire de cette dernière, non plus un phénomène abstrait mais un phénomène concret. Ainsi, en traduisant explicitement les effets de la pauvreté, le J-PAL contribue à redéfinir la pauvreté. Le phénomène n'est plus abstrait mais visible. En 2000, la Banque mondiale mène une importante enquête en demandant à 20000 pauvres d'expliquer leur situation, d'expliquer ce qu'est pour eux la pauvreté. L'un d'eux a répondu : « *Poverty is like heat ; you can not see it ; you can only feel it ; so to know it you have to go through it.* » (Banque Mondiale, 2000, xvii) L'approche du J-PAL a permis de rendre visible un phénomène qu'on pouvait uniquement sentir mais qui restait invisible. C'est en cela qu'elle a offerte, à mon sens, une nouvelle voix à la lutte contre la pauvreté.

Cette recherche a alors permis de s'intéresser à de nombreux champs : elle a été à la croisée de l'*evidence-based policy*, de l'économie expérimentale et de l'économie du développement, ces trois champs s'unissant à travers la notion de preuves. L'*evidence-based policy* s'attache à l'aide, par exemple, de l'économie expérimentale à développer des méthodes, toujours plus fines, en vue de déterminer des preuves solides et fiables. L'économie du développement emprunte, ici, à l'économie expérimentale la possibilité de produire ces preuves et à l'*evidence-based policy* leur utilisation par la sphère politique. De façon plus générale, cela montre l'intérêt toujours croissant porté à la détermination de « causalités locales », et l'abandon des explications sous-jacentes à ces causalités. Cela pointe alors une tension épistémologique en sciences sociales concernant le lien entre ces causalités

ciblées et leurs explications. Une enquête épistémologique plus générale pourrait s'intéresser à cette tension, ainsi qu'à ses répercussions en sciences sociales.

De plus, l'économie expérimentale, en se concentrant sur l'étude d'un unique paramètre d'intérêt et non plus sur plusieurs paramètres, tend à transformer, de façon plus générale, l'analyse économique : la théorie y est reléguée au second plan, l'accent est porté sur le comportement individuel, l'économétrie utilisée n'est plus structurelle. De ce fait, les recommandations politiques qui en découlent se centrent presque uniquement sur le comportement des individus. Ces recommandations cherchent à produire des incitations afin que les individus parviennent à répondre aux critères de l'analyse économique standard, au centre desquelles les seules notions d'efficacité et de rationalité. Une recherche tournée vers l'étude de ces transformations peut être envisagée, ici de façon plus large.

Enfin, l'économie du développement et ses pratiques, on l'a vu tout au long de ce travail, en ont été transformées. L'accent a, là aussi, été porté sur les comportements individuels, mais ces derniers peuvent aussi être regardés non plus uniquement au travers d'expériences, mais aussi à l'aide d'une réelle anthropologie. Le terrain offert par les expériences du J-PAL pourrait s'alimenter et s'appuyer de cette anthropologie. Cela peut alors représenter un nouveau pan à étudier.

La croisée de ces champs disciplinaires insiste sur la nécessité de mener et de poursuivre un important travail s'intéressant aux transferts disciplinaires traversant l'économie, cela offrirait un éclairage plus que lumineux sur toutes les questions épistémologiques d'ordre plus général soulevées plus haut.

Bibliographie

ACEMOGLU, Daron et James ROBINSON, (2006), *Economics Origins of Dictatorship and Democracy*, Cambridge University Press, New York.

_____, (2012), *Why Nations Fail, The Origins of Power, Prosperity, and Poverty*, Crown Business, New-York.

ACEMOGLU, Daron, Simon JOHNSON, et James ROBINSON, (2001), “The Colonial Origins of Comparative Development: an Empirical Investigation”, *American Economic Review*, 91(5) : 1369-1401.

ANGRIST, Joshua et Jörn-Steffen PISCHKE, (2010), “The Credibility Revolution in Empirical Economics : How Better Research Design Is Taking Con out of Econometrics”, *Journal of Economic Perspectives*, 24(2) : 3-30.

APPEL, Jacob et Dean KARLAN, (2011), *More than Good Intentions: Improving the Ways the World Poor Borrow, Save, Farm, Learn, and Stay Healthy*, Dutton, New-York.

ARCAND, Jean-Louis et Eric DJIMEU WOUABE, (2010), “Teacher Training and HIV/AIDS Prevention in West Africa : Regression Discontinuity Design Evidence from the Cameroon”, *Health Economics*, 19 : 36-54.

ARMATTE, Michel, (2004), *Histoire de probabilités et de statistiques. La théorie des erreurs (1750-1820), enjeux, problématiques, résultats*, Ellipses, Paris.

ARMITAGE, Peter, (2003), “Fisher, Bradford Hill, and Randomization”, *International Journal of Epidemiology*, 32(6) : 925-928.

ASHRAF, Nava, Dean KARLAN et Wesley YIN, (2006), “Tying Odysseus to the mast: Evidence from a Commitment Savings Product in the Philippines”, *Quarterly Journal of Economics*, 121 (2) : 635–672.

ASSIDON, Elsa, (2002), *Les théories économiques du développement*, La découverte, Paris.

AUVERT, Bertran *et al.*, (2005), “Randomized, Controlled Intervention Trial of Male Circumcision for Reduction of HIV Infection Risk : The ANRS 1265 Trial”, *PLoSMED*, 2(11) : e298.

AVVISATI Franceso, Marc GURGAND, Nina GUYON et Eric MAURIN, (2010), « Quels effets attendre d’une politique d’implication des parents d’élèves dans les collèges ? Les enseignements d’une expérimentation contrôlée », *Rapport pour le Haut Commissariat à la jeunesse*.

BACHELARD, Gaston, (1934), *Le nouvel esprit scientifique*, Alcan, Paris.

BAILEY, Robert, *et al.*, (2007), “Male Circumcision for HIV Prevention in Young Men in Kisumu, Kenya : a Randomised Controlled Trial”, *The Lancet*, 369(9562) : 643-656.

BARDHAN, Pranab, (2005), “Theory or Empirics in Development Economics”, *Economics and Political Weekly*, 40(40) : 4333-4335.

BANERJEE, Abhijit, (2001), "The Uses of Economic Theory: Against a Purely Positive Interpretation of Theoretical Results", *Mimeo MIT*.

_____, (2004), "The Two Poverties", *Nordic Journal of Political Economy*, 26(2) : 129-41.

_____, (2005a), "Making Aid Work: Review of J.Sachs *The End of Poverty: Economic Possibilities for Our Time*", *The New Leader*, 88(2) : 20-21.

_____, (2005b), "'New Development Economics' and the Challenge to Theory", *Economic and Political Weekly*, 40(40) : 4340-4344.

_____, (2006), "Making Aid Work: How to Fight Global Poverty Effectively", *Boston Review*, 31(4) : 7-9.

_____, (2007a), "Inside the Machine: Toward a New Development Economics", *Boston review*, 32(2) : 12-18.

_____, (2007b), *Making Aid Work*, MIT Press, Cambridge.

_____, (2008), "Big Answers for Big Questions: The Presumption of Growth Policy", in *What Works in Development? Thinking Big and Thinking Small*, Jessica COHEN et William EASTERLY (ed.), Brookings Institution Press, Washington DC.

BANERJEE, Abhijit et Esther DUFLO, (2005), "Growth Theory through the Lens of Development Economics", in *Handbook of Economic Growth*, Steve DURLAUF et Philippe AGHION (ed.), Elsevier Science, Amsterdam.

_____, (2006), "Addressing Absence", *Journal of Economic Perspectives*, 20(1) : 117-132.

_____, (2007), "The Economic Lives of the Poor", *Journal of Perspectives*, 21(1) : 141-167.

_____, (2008), "Mandated Empowerment. Handing Antipoverty Policy Back to The Poor?" in "Reducing the Impact of poverty on Health and Human Development", *Annals of New York Academic Sciences*, 1136 : 333-341.

_____, (2009), « L'approche expérimentale en économie du développement », *Revue d'économie politique*, 119(5) : 691-726.

_____, (2010), "Giving Credit Where it is Due," *Journal of Economic Perspectives*, 24(3): 61-79.

_____, (2011), *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, Public Affairs, Boston ; traduction française par Julie Maistre, *Repenser la pauvreté*, (2012), Le seuil, Paris.

BANERJEE, Abhijit et Sendhil MULLAINATHAN, (2008), "Limited Attention and Income Distribution", *American Economic Review*, 98(2) : 489-493.

BANERJEE, Abhijit Roland BENABOU, et Dilip MOOKKHERJEE, (2006), *Understanding Poverty*, Oxford University Press, New York.

BANERJEE, Abhijit, Esther DUFLO et Rachel GLENNESTER, (2008), “Putting a Band-Aid on a Corpse : Incentives for Nurses in the Indian Public Health Care System”, *European Economic Association*, 6(2-3) : 487-500.

_____, (2011), “Is Decentralized Iron Fortification a Feasible Option to Fight Anemia Among the Poorest?”, in *Explorations in the Economics of Aging*, David A. Wise (ed.), University of Chicago Press, Chicago.

BANERJEE, Abhijit, Suraj JACON et Michael KREMER, (2001), “Promoting School Participation in Rural Rajasthan : Results from Some Propsective Trials”, *Mimeo MIT*.

BANERJEE, Abhijit, Shawn COLE, Esther DUFLO, et Leigh LINDEN, (2007), “Remedying Education: Evidence from Two Randomized Experiments in India”, *Quarterly Journal of Economics*, 122(3):1235-1264.

BANERJEE, Abhijit, Arun CHANDRASEKHAR, Esther DUFLO et Matthew JACKSON, (2011), “The Diffusion of Microfinance”, *NBER Working Paper*, 17743 : 1-52.

BANERJEE, Abhijit, Esther DUFLO, Rachel GLENNERSTER, et Cynthia KINNAN, (2010), “The Miracle of Microfinance? Evidence from a Randomized Evaluation”, *BREAD Working Paper*, 278 : 1-52.

BANERJEE, Abhijit, Esther DUFLO, Rachel GLENNERSTER, et Dhruva KOTHARI, (2010), “Improving Immunization Coverage in Rural India: A Clustered Randomized Controlled Evaluation of Immunization Campaigns with and without Incentives”, *British Medical Journal*, 340:c2220.

BANERJEE, Abhijit, Esther DUFLO, Daniel KENISTON, et Nina SINGH, (2010), “Making Police Reform Real : The Rajasthan Experiment”, *Yale Working Paper*.

BANERJEE, Abhijit, Esther DUFLO, Raghavendra CHATTOPADHYAY, Daniel KENISTON, et Nina SINGH, (2012), “Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police”, *NBER Working Paper*, 17912 : 1-51.

BANERJEE, Abhijit, Rukmini BANERJI, Esther DUFLO, Rachel GLENNERSTER, et Stuti KHEMANI, (2010), “Pitfalls of Participatory Programs: Evidence from a randomized evaluation in education in India”, *American Economic Journal: Economic Policy*, 2 (1): 1-30.

BANKS, Gary, (2009), “Evidence-based Policy Making: What is it and How do We get it ?”, *ANU Public Lecture Series*, Productivity Commission, Canberra.

BANQUE MONDIALE, (1990), *World Development Report: Poverty*, World Bank, Washington DC.

_____, (2005), *Economic Growth in the 1990's: Learning From a Decade of Reforms*, World Bank, Washington DC.

_____, (2006), *Human Development Report*, World Bank, Washington DC.

_____, (2007-2008) *Human Development Report*, World Bank, Washington DC.

BARDET, Francis, et Roser CUSSÓ, (2012), « Les essais randomisés contrôlés, révolution des politiques de développement? Une évaluation par la Banque mondiale de l'empowerment au Bangladesh », *Revue française de socio-économie*, 2(10) : 175-198.

BARDHAN Pranab, (2005), "Theory or Empirics in Development Economics", *Economic and Political Weekly*, 40(40) : 4333-4335.

BARETT, Christopher, et Michael CARTER, (2010), "The Power and the Pitfalls of Experiments in Development Economics: Some Non-random Reflections", *Applied Economic Perspective and Policy*, 32(4) : 515-548.

_____, (2012), "Retreat from Radical Skepticism: rebalancing Theory, Observational Data and Randomization in Development Economics", in *Field Experiments and Their Critics*, Dawn TEELE (ed.), Yale University Press, New Haven, à paraître.

BARDSLEY, Nicholas, *et al.*, (2010), *Experimental Economics: Rethinking the Rules*, Princeton University Press, Princeton.

BASLOW Joel, Naihua DURAN et Richard KRAVITZ, (2004), "Evidence-Based Medicine, Heterogeneity of Treatment Effects, and The Trouble With Average", *The Milbank Quarterly*, 82(4) : 661-687.

BASU Kaushik, (2005), "The New Empirical Development Economics: Remarks on Its Philosophical Foundations", *Economic and Political Weekly*, 40(40) : 4336-4339.

BAUMSEISTER, Roy, Matthew GAILLOT, Nathan DEWALL et Megan OATEN, (2006), "Selfregulation and personality : How Interventions Increase Regulatory Success, and How Depletion Moderates the Effects of Traits on Behavior", *Journal of Personality*, 74(6) : 1773-1802.

BEAMAN, Lori, Esther DUFLO, Rohini PANDE and Petia TOPALOVA, (2012), "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India", *Science Magazine*, 335 (6068).

BERLIN, Isaiah, (1958), *Two Concepts of Liberty*, Clarendon Press, Oxford.

BERNARD, Claude, (1865), *Introduction à l'étude de la médecine expérimentale*, Flammarion, Paris.

BERTRAND, Marianne, Esther DUFLO et Sendhil MULLAINATHAN, (2004), "How Much Should We Trust Difference in Differences Estimates?", *Quarterly Journal of Economics*, 119(1): 249-275.

BERTRAND, Marianne, Sendhil MULLAINATHAN, et Eldar SHAFIR, (2004), "A Behavioral-Economics View of Poverty", *The American Economic Review*, 94(2) : 419-423.

BERTRAND, Marianne, Dean KARLAN, Sendhil MULLAINATHAN, Eldar SHAFIR et Jonathan ZINMAN, (2008), "What's Advertising Content Worth? Evidence from Consumer Credit Marketing Field Experiment", non publié.

BHATTACHARYA Debopam, Pascaline DUPAS et Shin KANAYA, (2013), "Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets", *NBER Working Paper*, No.18833 : 1-57.

- BREHM, Jack, (1966), *A Theory of Psychological Reactance*, Academic Press, Waltham.
- BOONE, Peter, (1996), "Politics and the Effectiveness of Foreign Aid", *European Economic Review*, 40(2) : 289-329.
- BOUVENOT, Gilles, et Muriel VRAY, (2006), *Essais cliniques, théories, pratique et critique*, Flammarion, Paris.
- BURNSIDE Craig, et David DOLLAR, (2000), "Aid, Policies, and Growth", *American Economic Review*, 90(4) : 847-868.
- CAMPBELL, Donald, (1957), "Factors Relevant to the Validity of Experiments in Social Settings", *Psychological Bulletin*, 54(4) : 297-312.
- CAMPBELL, Donald, Thomas COOK et William SHADISH, (2002), *Experimental and Quasi-Experimental Designs for Generalized Causal Inferences*, Houghton Mifflin, Boston.
- CARROLL, Christopher et Lawrence SUMMERS, (1991), "Consumption Growth Parallels Income Growth: Some New Evidence", in *National Savings and Economics Performance*, Douglas BERNHEIM et John SHOVEN (ed.), Chicago University Press, Chicago.
- CARTWRIGHT, Nancy, (1989), *Nature's Capacities and Their Measurement*, Clarendon Press, oxford.
- _____, (2007a), "Are RCT's the Gold Standard?", *Biosocieties* 2(1): pp. 11-20.
- _____, (2007b), "Causal Claims: Warranting Them and Using Them", Chapter 3 in Part 1 of Nancy Cartwright's book *Hunting Causes and Using Them*, Cambridge
- _____, (2008), "Evidence for Evidence-based Policy", présenté à "Home Office Seminar on Criminology and Evidence". University Press, New-York.
- _____, (2009a), "Evidence-based Policy: What's To Be Done About Relevance?", in *Philosophical Models, Methods and Evidence: Topics in The Philosophy of Science* (proceeding for the *Thirty-Eight Oberlin Colloquium in Philosophy*), Special Issue of *Philosophical Studies*.
- _____, (2009b), "Evidence-based Policy: Where is our Theory of Evidence", *Journal of Children 's Services*, 4(4): pp. 6-14.
- _____, (2009c), "What is This Thing Called Efficacy?", *Philosophy of the Social Science. Philosophical Theory and Scientific Practice*, Mantzavinos (ed.), Cambridge University Press, Cambridge.
- _____, (2010a), "Predicting 'It Will Work For Us'; (Way) Behond Statistics", in *Causality in the Sciences*, P.M. Illari, F. Russo, J. Williamson édition, Oxford University Press, New-York.
- _____, (2010b), "What Are Randomized Controlled Trials Good For?", *Philosophical Studies*, 147(1): pp.59-70.
- _____, (2010c), "Will This Policy Work For You? Predicting Effectiveness Better: How Philosophy Helps", présenté à "PSA Presidential Adress, Montréal", à paraître dans *Philosophy of Science*.

_____, (2011a) “A Philosopher’s View of the Long Road From RCT’s to Effectiveness”, *The Lancet*, (377): pp. 1400-1401.

_____, (2011b), “Evidence, External Validity and Explanatory Relevance”, in *Philosophy of Science Matters: The Philosophy of Peter Achinstein*, G. Morgan édition, Oxford University Press, New-York.

_____, (2011c), “Predicting What Will Happen When We Act: What Counts as Warrant?”, *Preventive Medicine*, 53(4): pp. 221-224.

CARTWRIGHT, Nancy, et Sophia EFSTATHIOU, (2008), « Evidence-Based Policy and Its Ranking Schemes: So, Where's Ethnography? », présenté à la conférence « the Association of Social Anthropologists, *The Pitch of Ethnography* ».

CARTWRIGHT, Nancy et MUNRO Eileen, (2010), “The Limitations of RCT’s in Predicting Effectiveness”, *Journal of Evaluation in Clinical Practice*, 16(2): pp. 260-66.

CARTWRIGHT, Nancy, GOLDFINCH Andrew et HOWICK Jeremy (2008), “Evidence-based Policy: where is our theory of Evidence?”, *Philosophy: Foundations and Applications*, édité par A; Beckermann, H. Tetens and S. Walter , Paderborn: mentis-Verlag.

CHAMBERS, Robert, Deepa NARAYAN, Meera SHAH et Patt PETESCH, (2000), *Crying Out for Change : Voices of the Poor*, Banque mondiale, Washington DC.

CHARNOZ, Oliver et Jean-Michel SEVERINO, (2007), *L’aide publique au développement*, La découverte, Paris.

COHEN, Jessica et Pascaline DUPAS, (2010), “Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment”, *Quarterly Journal of Economics*, 125(1) : 1-45

COHEN, Jessica et William EASTERLY, (2007), *What Works in Development Economics, Thinking and Thinking Small*, Brookings Institution Press, Washington DC.

COLLIER, Paul, (2007), *The Bottom Billion. Why The Poorest Countries Are Failing and What Can Be Done About It*, Oxford University Press, Oxford.

CORNIA, Giovanni, Richard JOLLY et Frances STEWART, (1987), *Adjustment With a Human Face: Protecting the Vulnerable and Promoting Growth*, Oxford University Press, Oxford.

COT, Annie et Jérôme LALLEMENT, “The Historiography of Economics. A methodological Approach”, in *Joseph A. Schumpeter, Historian of Economics. Perspectives on the History of Economics Thought*, Laurence MOSS (ed.), Routledge, Londres.

CREPON, Bruno, Florencia DEVOTO, Esther DUFLO et William PARIENTE, (2011), “Impact of Microcredit in Rural Areas of Morocco: Evidence from a Randomized Evaluation”, *MIT Working Paper*.

CRUMP Richard, joseph HOTZ, IMBENS Guido et Oscar MITNIK, (2008), “Nonparametric Tests for Treatment Effect Heterogeneity”, *Review of Economics and Statistics*, 90(3) : 389-405.

D’ALESSANDRO, Umberto, (1994), “Nationwide Survey of Bednet Use in Rural Gambia”,

in *Bulletin de l'Organisation Mondiale de la Santé*, 72 : 391–394.

DARIL, Bem, (1967), “Self-Perception : An Alternative Interpretation of Cognitive Dissonance Phenomena”, *Psychological Review*, 74(3) : 183-200.

DAVIES Huw, Sandra NUTLEY, et Peter SMITH (ed.), (2000), *What Works? Evidence-based Policy and Practice in Public Services*, The Policy Press, Bristol.

DAVIS, John, (2013), “Economics Imperialism under the Impact of Psychology. The case of Behavioral Development Economics”, *Oeconomia*, 3(1) : 119-138.

DEATON, Angus, (2001), “Health, inequality, and economic development”, *NBER Working Paper*, No. 8318.

_____, (2009), “Instruments of Development: Randomization in the tropics, and the search of Elusive Keys to Development Economics”, *British Academy Lectures 2008*, 162 : 123-160, Oxford University Press, New York.

_____, (2010a), “Intruments, Randomization, and Learning About Development”, *Journal of Economic Literature*, 48 : 424-455.

_____, (2010b), “Understanding the Mechanisms of Economic Development”, *Journal of Economic Perspectives*, 24(3): pp. 3-16.

DEATON, Angus, et Olivier DUPRIEZ (2011), « Purchasing Power Parity for the Global Poor », *American Economic Journal : Applied Economics*, 3(2) : 137-166.

DEATON, Angus, et Guy LAROQUE, (1992), “On the Behavior of Commodity Prices”, *Review of Economic Studies*, 59(1) : 1-23.

_____, (1996), “Competitive Storage and Commodity Price Dynamics”, *Journal of Political Economy*, 104(5) : 896-930.

DESROSIÈRES, Alain, (2013), « La mesure du développement : un domaine propice à l'innovation méthodologique », *Revue Tiers Monde*, 1(213) : 23-32.

DEVOTO, Florencia, Esther DUFLO, Pascaline DUPAS, William PARIENTE, et Vincent PONS, (2012), “Happiness on Tap: Piped Water Adoption in Urban Morocco”, *American Economic Journal*, 4(4) : 68-99.

DHALIWAL, Iqbal, Esther DUFLO, Rachel GLENNERSTER, et Caitlin TULLOCH, (2011), “Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries”, *MIT Working Paper*.

DOLAN G. *et al.*, (1993), “Bed nets for the Prevention of Malaria and Anaemia in Pregnancy”, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 87(6) : 620–626.

DOLL, Richard, et Austin Bradford HILL, (1950), “Smoking and Carcinoma of the Lung. Preliminary Report”, *British Medical Journal*, 2(4682) : 739-748.

DUFLO, Esther, (2004), “Scaling Up and Evaluation,” in Francois Bourguignon et Boris Pleskovic, (eds.), *Accelerating Development*, World Bank and Oxford University Press, Washington DC, pp. 342-367.

_____, (2005), « Évaluer l'impact des programmes d'aide au développement : le rôle des évaluations par assignation aléatoire », *Revue d'économie du développement*, Volume 19, Numéro 2, pp. 185-226.

_____, (2006a), "Egalité des sexes et développement" in *le livre noir de la condition des Femmes*, Christine OKRENT (ed.), XO, Paris.

_____, (2006b), "Field Experiments in Development Economics" in Richard Blundell, Whitney Newey, Torsten Persson, *Advances in Economics and Econometrics: Theory and Applications*, Cambridge University Press, Cambridge.

_____, (2006c) "Poor but Rational?", in *Understanding Poverty*, Abhijit BANERJEE, Dilip MOOKHERJEE et Roland BENABOU (ed.), Oxford University Press, Oxford.

_____, (2009), *Expérience, savoirs et lutte contre la pauvreté*, Fayard, Paris.

_____, (2010a), *Lutter contre la pauvreté, le développement humain*, tome 1, La république des idées, Paris.

_____, (2010b), *Lutter contre la pauvreté, la politique de l'autonomie*, tome 2, La république des idées, Paris.

_____, (2012a), "Human Values and the Design of the Fight Against Poverty", *The Tanner Lectures on Human Values*, Harvard University.

_____, (2012b), "Women's Empowerment and Economic Development", *Journal of Economic Literature*, 50(4) : 1051-1079.

DUFLO, Esther et Lucia BREIEROVA, (2004), "The Impact of Education on Fertility and Child Mortality: Do Fathers Really Matter Less Than Mothers?", *NBER Working Paper*, 10513 : 1-30.

DUFLO, Esther et Raghabendra CHATTOPADHYAY, (2004), "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India", *Econometrica*, 72(5): 1409-1443.

DUFLO, Esther et Michael KREMER, (2005), "Use of Randomization in the Evaluation of Development Effectiveness", in *Evaluating, Development Effectiveness*, George PITMAN, Osvaldo FEINSTEIN, et Gregory INGRAM (ed.), NJ: Transaction Publishers, New Brunswick.

DUFLO, Esther et Rohini PANDE, (2007), "Dams", *Quarterly Journal of Economics*, 122(2) : 601-646.

DUFLO, Esther et Christopher UDRY, (2004) "Intrahousehold Resource Allocation in Cote d'Ivoire: Social Norms, Separate Accounts, and Consumption Choices", *NBER Working Paper*, 10498 : 1-50.

DUFLO, Esther, Pascaline DUPAS et Michael KREMER, (2011a), "Education, HIV and early fertility: experimental evidence from Kenya", non publié.

DUFLO, Esther, Pascaline DUPAS, et Michael KREMER, (2011b), "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya", *American Economic Review*, 101(5) : 1739-1774.

DUFLO, Esther, Pascaline DUPAS, et Michael KREMER, (2012), "School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools", NBER Working Paper, 17939 : 1-48.

DUFLO, Esther, Rachel GLENNESTER et Michael KREMER, (2007) "Using Randomization in Development Economics Research: A Toolkit" in *Handbook of Development Economics*, Paul SCHULTZ, John STRAUSS (ed.), Elsevier Science, North Holland, 4 : 3895-62.

DUFLO, Esther, Rema HANNA, et Stephen RYAN (2012), "Incentives Work: Getting Teachers to Come to School," *American Economic Review*, 102(4) : 1241-1278.

DUFLO, Esther, Michael KREMER, et Jonathan ROBINSON, (2008), "How High are Rates of Return to Fertilizer? Evidence from Field Experiments in Kenya", *American Economics Review*, 98 (2): 482-88.

_____, (2011), "Nudging Farmers to Use Fertilizer: Evidence from Kenya", *American Economic Review*, 101 (6): 2350-2390.

DUFLO, Esther, Pascaline DUPAS, Michael KREMER, et Samuel SINEI, (2006), "Education and HIV/AIDS Prevention: Evidence from a Randomized Evaluation in Western Kenya", *World Bank Policy Research Working Paper*, 4024.

DUNCAN, Thomas, (1990), "Intra-Household Resource Allocation : An Inferential Approach", *Journal of Human Resources*, 25(4) : 645-664.

DUNCAN, Thomas, (1997), "Sailors, Scurvy and Science", *Journal of the Royal Society of Medicine*, 90 : 50-54.

DUPAS, Pascaline, (2009), "What Matters (and what does not) in households' decisions to invest in malaria prevention?", *American Economic Review*, 99(2) : 224-230.

_____, (2010) "Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence from a Field Experiment", *NBER Working Paper*, No. 16298 : 1-46.

_____, (2011a), "Health Behavior in Developing Countries", *Annual Review of Economics*, 3: 425-449

_____, (2011b), "Do Teenagers respond to HIV Risk Information? Evidence from a Field Experiment in Kenya", *American Economic Journal : Applied Economics*, 3 : 1-34.

DUPAS, Pascaline, Jessica COHEN, et Simone SCHANER, (2013), "Price Subsidies, Diagnostic Tests, and the Targeting of Malaria Treatment Externalities With Application to Anti-Malaria Bednets". *NBER Working Paper*, No.17943 : 1-44.

DEHUE, Trudy, (1997), "Deception, Efficiency, and Random Groups: Psychology and the Gradual Origination of the Random Group Design", *Isis*, 88(4): 653-673.

_____, (2000), "From Deception Trials to Control Reagents, The Introduction of the Control Group About a Century Ago", *American Psychologist*, 55(2) : 264-268.

DWORKIN, Gerald, (1972), « Paternalism », *The Monist*, 56 : 64-84.

EASTERLY, William, (2006a), *The White Man's Burden: Why the West's Efforts to Aid the Rest Have Done So Much Ill and So Little Good*, The Penguin Press, New York.

_____, (2006b), "Planners vs. Searchers in Foreign Aid", *Asian Development Review*, 23(2) : 1-35.

_____, (2006c), "The Big Push Déjà Vu: A Review of Jeffrey Sachs's The End of Poverty: Economic Possibilities for Our Time", *Journal of Economic Literature*, 44(1): 289-318.

_____, (2009), "How the Millennium Development Goals Are Unfair to Africa", *World Development*, 37(1): 26-35.

EDGEWORTH, Francis Ysidro, (1885), "Calculus of Probabilities Applied to Psychic Research", *Proceedings of Society for Psychic Research*, 3 : 190-199.

EL-SERAG Hashem, (2007), "Comments from the Editors", *Gastroenterology*, 123 : 8-10.

EVIDENCE-BASED MEDICINE GROUP, (1992), "Evidence-based Medicine. A New Approach to Teaching the Practice of Medicine", *Journal of the American Medical Association*, 268 : 2420-2425.

FECHNER, Gustav, (1860), *Elemente der Psychophysik*, Zweiter Theil, Leipzig.

FERBER, Robert et Werner HIRSCH, (1982), *Social Experimentation and Economic Policy*, Cambridge University Press, Cambridge.

FEREY, Samuel, (2011), « Paternalisme libéral et pluralité du moi », *Revue Economique*, 62(4) : 737-750.

FISHER, Ronald, Aymler, (1926), "The Arrangement of Field Experiments", *Journal of the Ministry of Agriculture of Great Britain* ; 33 : 503-513.

_____, (1960) [1935], *The Design of Experiment* (7^{ème} édition), Hafner Publishing Company, New York.

FOUCAULT, Michel, (1963), *Naissance de la Clinique*, Presses Universitaires de France, Paris.

FRIEDMAN, Milton, (1962), *Capitalism and Freedom*, University of Chicago Press, Chicago.

GALLUP, John et Jeffrey SACHS, (2001), "The Economic Burden of Malaria", *The American Journal of Tropical Medicine and Hygiene*, 64(1) : 85-96.

GERTLER, Paul et Simone BOYCE, (2001), "An Experiment in Incentive-Based Welfare : The Impact of PROGRESA on Health in Mexico", Haas School of Business, University of California, Berkeley.

GINE, Xavier, Dean KARLAN et Jonathan ZINMAN, (2010), "Put Your Money Where Your Butt Is : A commitment Savings Account for Smoking Cessation", *American Economic Journal : Applied Economics*, 2(4) : 213-235.

GODLONTON, Susan, Alister MUNTHALI et Rebecca THORNTON, (2012), “ Response to Information ? Circumcision, Information and HIV Prevention”, *University of Michigan Working Paper*.

GRAY, R, *et al.*, (2007), “Male Circumcision for HIV Prevention in Men Rakai, Uganda : a Randomised Trial”, *The Lancet*, 369 (9552) : 657-666.

GREENBERG, David et Mark SHRODER, (2004), *The Digest of Social Experiments, Third Edition*, The Urban Institute Press, Washington DC.

GUALA, Francesco, (1999), “The Problem of External Validity (or 'Parallelism') in Experimental Economics”, *Social Science Information*, 38: 555-573.

_____, (2003), “Experimental Localism and External Validity”, *Philosophy of Science*, 70: 1195-1205.

_____, (2005), *The Methodology of Experimental Economics*, Cambridge University Press, Cambridge.

GUALA, Francesco et Luigi MITTONE, (2005), “Experiments in Economics: External Validity and the Robustness of Phenomena”, *Journal of Economic Methodology*, 12(4): 495-515.

GUSTAFSON, Robert, (1958), *Carryover Levels for Grains*, Technical Bulletin 1178, U.S. Department of Agriculture.

IMBENS, Guido et Donald RUBIN, (2008), “Rubin Causal Model”, *The New palgrave Dictionary of Economics*.

GUILLIN, Vincent, (2013), « De quoi les essais contrôlés randomisés sont-ils capables ? Evaluation, mécanismes et capacités en sciences sociales », *Cahiers Philosophiques*, 133(2) : 79-102.

GUYATT, Gordan *et al.*, (1992), “Evidence-Based Medicine. A New Approach to teaching the Practice of Medicine”, *The Journal of the American Medical Association*, 268(17) : 2420-2425.

GUYATT, Helen, Sam OCHOLA et Robert SNOW, (2002), “Too Poor To Pay : Charging for Insecticide Bednets in Highland Kenya” », *Tropical Medicine and International Health*, 7(10) : 846-850.

HAARELM, Trygve, (1944), “The Probability Approach in Econometrics”, *Econometrica*, 12 : 1-115.

HACKING, Ian, (1988), “Telepathy : Origins of Randomization in Experimental Design”, *Isis*, 79(3) : 427-451.

HALL, Nancy, (2007), “R.A. Fisher and His Advocacy of Randomization”, *Journal of the History of Biology*, 40(2) : 295-325.

HARRISON, Glenn, (2011), “Randomization and its Discontents”, *Journal of African Economies*, 20(4) : 626-652.

HARRISON, Glenn, et John LIST, (2004), “Field Experiments”, *Journal of Economic Literature*, 42(4) : 1009-1055.

HEAD, Brian, (2009), "Evidence-based Policy : Principles and Requirements", in *Strengthening Evidence Based Policy in the Australian Federation, Volume 1 : Proceedings*, Productivity Commission, Canberra.

HECKMAN, James, (1991), "Randomization and Social Evaluation", *NBER Working Paper* : 107.

_____, (2010), "Building Bridges Between Structural and Program Evaluations Approaches to Evaluating Policy", *Journal of Economic Literature*, 48(2) : 356-398.

HECKMAN, James, et Jeffrey SMITH, (1995), "Assessing the Case for Social Experiments", *The Journal of Economic Perspectives*, 9(2) : 85-110.

HECKMAN, James, Nancy CLEMENT, et Jeffrey SMITH, (1997), "Making the Most Out of the Programme Evaluations and Social Experiments : Accounting for Heterogeneity in Programme Impact", *The review of Economic Studies*, 64(4) : 487-535.

HECKMAN, James, Hidehiko ICHIMURA, Jeffrey SMITH et Petra TODD, (1998), "Characterizing Selection Bias Using Experimental Data", *Econometrica*, 66(5) : 1017-1098.

HEUKELOM, Floris, (2009), "Origin and Interpretation of Internal and External Validity in Economics", *NiCE Working Paper 09-111*.

HUME, David, (1739), *A Treatise of Human Nature*, White Hart, Londres.

JACOBSON, Lionel, Adrian EDWARDS, Stephen GRANIER et Christopher BUTLER, (1997), "Evidence-based Medicine and General Practice", *British Journal of General Practice*, 47 : 449-452.

JALAN, Jyotsa et E. SOMANATHAN, (2008), "The Importance of Being Informed : Experiment Evidence on Demand for Environmental Quality", *Journal of Development Economics*, 87(1) : 14-28.

JATTEAU, Arthur, (2013), *Les expérimentations aléatoires en économie*, La découverte, Paris.

JENSEN, Robert et Nolan MILLER, (2008), "Giffen Behavior and Subsistence Consumption", *American Economic Review*, 98(4) : 1553-1577.

JOLLY, Richard, (1976), "The World Employment Conference: The Enthronement of basic Needs", *Development Policy Review*, A9(2) : 31-44.

_____, (1991), "Adjustment with a Human Face : A UNICEF Record and Perspective on the 1980s", *World Development*, Volume 17, Numéro 12, pp. 1807-1821.

J-PAL, (2007), "Cheap and Effective Ways to Change Adolescents' Sexual Behavior", *Policy Briefcase No.3*.

_____, (2011), "A well-Timed Nudge", *Policy Briefcase*.

KANDUR, Ravi, (2005), "Goldilocks Development Economics", *Economic and Political Weekly*, 40(40) : 4344-4346.

_____, (2006), "The Economics of International Aid", in *Handbook of the Economics of Giving, Reciprocity and Altruism*, Elsevier, Amsterdam, pp.1559-1588.

KAHNEMAN, Daniel et Amos TVERSKY, (1981), “The Framing of Decisions and the Psychology of Choice”, *Science*, 211 : 453-458.

KREMER, Michael et Edward MIGUEL, (2004), “Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities”, *Econometrica*, 72(1) : 159-217.

KREMER, Michael, Jessica LEINO, Edward MIGUEL et Alix PETERSON ZWANE, (2011), “Spring Cleanings: Rural Waters Impacts, Valuation and Property Rights Institutions”, *The Quarterly Journal of Economics*, 126(1) : 145-205.

KUHN, Thomas, (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.

KUZNETS, Simon, (1976), “Demographic Aspects of the Size Distribution of Income: An Exploratory Essay”, *Economic Development and Cultural Change*, 25(1): 1–94.

LABROUSSSE, Agnès, (2010), « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d’un outil de preuve et de gouvernement », *Revue de la régulation*, 1(7) : 1-32.

LAURENT, Catherine *et al.*, (2009), « Pourquoi s’intéresser à la notion « d’Evidence-based Polciy » ? », *Revue Tiers Monde*, 200 : 853-873.

LEAMER, Edward, (1983), “Let’s Take Out the Con of Econometrics”, *The American Economic Review*, 73(1) : 31-43.

_____, (2010), “Tantalus on The Road to Asymptopia”, *Journal of Economic Perspectives*, 24(2) : 31-46.

LENGELER, Christian, (1998), “Insecticide-treated Bednets and Curtains for Preventing malaria”, *Cochrane Library*, CD000363.

LEVITT, Steven et John LIST, (2008), “Field Experiments in Economics : The Past, The Present, and The Future”, *NBER Working Paper*, 14356.

LEWIS, Arthur, (1954), “Economic Development with Unlimited Supplies of Labour”, *The Manchester School of Economics and Social Studies*, 22(2) : 139-197

LIND, James, (1753), *A Treatise of Scurvy*, A. Millar, Londres.

MADDISON, Angus, (1995), *Monitoring the World Economy 1820-1992*, Publication de l’OCDE, Paris.

MACKIE, John, (1988), *The Cement of the Universe : A Study in Causation*, Clarendon Press, Oxford.

MADAJEWICZ, Malgosia, *et al.*, (2007), “Can Information Alone Change Behavior? Response to Arsenic Contamination of Groundwater in Bangladesh”, *Journal of Development Economics*, 84(2) : 731-754.

MARKS, Harry, (1997), *The Progress of Experiment : Science and Therapeutic Reform in the United-States, 1900-1990*, Cambridge University Press, Cambridge.

MARMOT, Michael, (2004), *Status Syndrome: How Your Social Standing Directly Affects Your Health and Life Expectancy*, Bloomsbury, Londres.

MASQUELET, Alain, (2006), *Le raisonnement médical*, Presses Universitaires de France, Paris.

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

McGOVERN, Dermot, VALORI Roland, William, SUMMERSKILL, et Marcel LEVI, (2001), *Evidence-based Medicine*, BIOS Scientific Publishers, Oxford.

MELDRUM, Marcia, (2000), “A Brief History of the Randomized Controlled Trial”, *Hematology/Oncology Clinics of North America*, 14(4) : 745-760.

MILL, John Stuart, (1843), *A system of Logic, ratiocinative and inductive : being a connected view of the principles of evidence, and methods of scientific investigation*, John W. Parker, West Strand, Londres.

MODIGLIANI, Franco, (1970), “The life-cycle hypothesis and intercountry differences in the saving ratio,” in *Induction, growth, and trade: essays in honour of Sir Roy Harrod*, W. A. Eltis, M. FG. Scott, and J. N. Wolfe (ed.), Oxford University Press, Oxford.

MOOKHERJEE Dilip, (2005), “Is There Too Little Theory in Developent Economics Today?”, *Economic and Political Weekly*, 40(40) : 4328-4333.

MOYO, Dambisa, (2009), *Dead Aid. Why Aid Is Not Working and How There Is Another Way For Africa*, Allen Lane, Londres.

MULLAINATHAN, Shendil et Eldar SHAFIR, (2013), *Scarcity: Why Having Too Little Means So Much*, Times Books, New-York.

NAYLOR, Christopher, (1995), “Grey Zones of Clinical Practice: Some Limits to Evidence-Based Medicine”, *The Lancet*, 345(8953) : 840-842.

NUSSBAUM, Martha, (2000a), “Women’s Capabilities and Social Justice”, *Journal of Human Development*, 1(2) : 219-247.

_____, (2000b), *Women and Human Development. Capabilities Approach*, Cambridge University Press, New-York.

_____, (2003), “Capabilities as Fundamental Entitlements: Sen and Social Justice”, *Feminist Economics*, 9(2-3) : 33-59.

OAKLEY, Ann, (2000), “A Historical Perspective on the Use of Randomized Trials in Social Science Settings”, *Crime and Delinquency*, 46(3) : 315-329.

OLKEN, Benjamin, (2007), “Monitoring Corruption : Evidence from a Field Experiment in Indonesia”, *Journal of Political Economy*, 115(2) : 200-249.

ORGANISATION DES NATIONS UNIES, (2013), *Objectifs du Millénaire pour le développement*, rapport de 2013.

ORGANISATION MONDIALE DE LA SANTÉ (OMS), (2007), *World Malaria Report*.

- _____, (2013), *World Malaria Report*.
- PARIENTE, William, (2008), « Analyse d'impact : l'apport des évaluations aléatoires », *DIAL STATECO*, 103 : 1-14.
- PAWSON, Ray et Nick TILLEY, (1997), *Realistic Evaluation*, Sage, Londres.
- PEIRCE, Charles Sanders, et Joseph JASTROW, (1885), « On Small Differences in Sensation », *Memoirs of the National Academy of Sciences*, 3 : 73-83.
- PICAVET, Emmanuel, (1995), *Approches du concret. Une introduction à l'épistémologie*, Ellipses, Paris.
- _____, (1996), « Qu'est-ce que l'épistémologie ? Le rapport entre épistémologie et science », *Recherches en soins infirmiers*, 50 : 7-13.
- _____, (1999), « De l'efficacité à la normativité », *Revue Economique*, 50(4) : 833-858.
- PIOT, Peter, (2010), *L'épidémie du sida, mondialisation des risques, transformation de la santé publique et développement*, Fayard, Paris.
- POLAK, Jacques, (1957), "Monetary analysis of income formation and payments problems", *FMI Working Paper*.
- POLITY IV PROJECT, (2000), *Polity IV Dataset: Political Regimes Characteristics and Transitions, 1800-1999*, College Park, University of Maryland, Center for International Development and Conflict Management.
- RAVALLION, Martin, (2009a), "Evaluation in The practice of Development", *World Bank Research Observer*, 24(1) : 29-53.
- _____, (2009b), "Should the Randomistas Rule ?", *The Economist' Voice, Berkeley Electronic Press*, 6(2) : 1-6.
- RAVALLION, Martin, Shaohua CHEN et Prem SANGRULA, (2009), "Dollar a Day", *The World Bank Economic Review*, 23(2) : 163-184.
- RAWLS, John, (1971), *A Theory of Justice*, Harvard University Press, Cambridge ; traduction française par Catherine Audard, *Théorie de la Justice*, (1987), Le Seuil, Paris.
- REILEY, David et John LIST, (2007), "Fields Experiments", *The New Palgrave Dictionary of Economics*.
- RICHET, Charles, (1885), « La suggestion mentale et le calcul des probabilités », *Revue Philosophique de la France et de l'étranger*, 18 : 609-674.
- RIVLIN, Alice, (1974), "How Can Experiments Be More Useful ?", *The American Economic Review*, 64(2) : 346-354.
- RODRIK, Dani, (2006), "Goodbye Washington Consensus, Hello Washington Confusion ? A Review of the World Bank's 'Economic Growth in the 1990s : Learning from a Decade of Reform'", *Journal of Economic Literature*, 44(4) : 973-987.

_____, (2007), “The New Development Economics : We Shall Experiment, But How Shall We Learn ?”, in *What Works in Development, Thinking Big and Thinking Small*, Jessica COHEN et William EASTERLY (ed.), Brookings Institution Press, Washington, DC.

_____, (2010), “Diagnostics Before Prescription”, *Journal of Economic Perspectives*, 24(3) : 33-44.

ROSENSTEIN-RODAN, Paul, (1943), “Problems of Industrialization of Eastern and South-Eastern Europe”, *Economic Journal*, 53(210-211) : 202-211.

_____, (1961), “International Aid for Underdeveloped Countries”, *The review of Economics and Statistics*, 43 (2) : 107-138.

ROTHWELL, Peter, (2005), “External Validity of Randomised Controlled Trials : To Whom Do the Result of This Trial Apply ?”, *The Lancet*, 365 : 82-93.

ROY, A.D., (1951), “Some Thoughts on the Distribution of Earnings”, *Oxford Economic Papers*, 3(2) : 135-146.

RUBIN, Donald, (1974), “Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies”, *Journal of Educational Psychology*, 66(5) : 688-701.

RUDD, Kevin, (2008), “Address to Heads of Agencies and Members of Senior Executive Service”, *texte présenté au Parlement Australien*, Canberra.

SACHS, Jeffrey, (2002), “A New Global Effort to Control Malaria”, *Science*, 298(5591) : 122-124.

_____, (2005), *The End of Poverty: Economic Possibilities for Our Time*, Penguin Press, New York.

SACKETT, Dave et al., (1995), “Evidence-based Medicine”, *British Medical Journal*, 310 : 1085-1086.

_____, (1996), “Evidence based-Medicine : What it is and What it isn’t”, *British Medical Journal*, 312 : 71- 72.

SALZBURG, David, (2002), *The Lady Testing Tea : How Statistics Revolutionized Science in the Twentieth Century*, Holt Paperback, New-York.

SAMUELSON, William et Richard ZECKHAUSER, (1988), «“ Status Quo in Decision Making ”», *Journal of Risk and Uncertainty*, 1 : 7-59.

SHELLENBERG, Joanna, et al., (2001), “Effect of Large-Scale Social Marketing of Insecticide-Treated Nets on Child Survival in Rural Tanzania”, *The Lancet*, 357 : 1241–1247.

SCHULTZ, Paul et John STRAUSS, (2007), *Handbook of Development Economics*, Elsevier Science, Amsterdam.

SCHULTZ, Theodore, (1964), *Transforming Traditional Agriculture*, Yale University Press, New-Heaven.

SEN, Amartya, (1981), *Poverty and Famines : An Essay on Entitlements and Deprivation*, Clarendon Press, Oxford.

_____, (1987), *Commodities and Capabilities*, Rashtriya, Delhi.

_____, (1992), *Inequality Reexamined*, Harvard University Press, Cambridge ; traduction française par Paul Chemla, *Repenser l'inégalité*, (2000), Le Seuil, paris.

_____, (1999), *Development as Freedom*, Oxford University Press, Oxford ; traduction française par Michel Bessières, *Un nouveau modèle économique. Développement, justice, liberté*, (2003), Odile Jacob, Paris.

_____, (2009), *The Idea of Justice*, Allen Lane, Londres.

SERRA, Daniel, (2012a), « Principes méthodologiques et pratiques de l'économie expérimentale : une vue d'ensemble », *Revue de philosophie économique*, 13 : 21-78.

_____, (2012b), "The Experimental Method in Economics : Old Issues and New Challenges", *Revue de philosophie économique*, 13 : 3-19.

SIMONSON, Itamar et Amos TVERSKY, (1992), "Choice in Context : Tradeoff Contrast and Extremeness Aversion", *Journal of Marketing Research*, 29(3) : 281-295.

SINGER, Peter, (2009), *The Life You Can Save: Acting Now to End World Poverty*, The Random House Publishing Group, New York.

SMITH, Davey et Matthis EGGER, (1998), "Incommunicable Knowledge ? Interpreting and Applying the Results of Clinical Trials and Meta-analyses", *Journal of Clinical Epidemiology*, 51 : 289-295.

SNYDER, Mark et Michael CUNNINGHAM, (1975), "To Comply or Not Comply : Testing the Self-Perception Explanation of the Foot in the Door Phenomenon", *Journal of Personality and Psychology*, 31(1) : 64-67.

STIGLER, Stephen, (1992), "A Historical View of Statistical Concepts in Psychology and Educational Research", *American Journal of Education* 101(1) : 60-70.

STIGLITZ, Joseph, (2002), *Globalisation and its Discontents*, W.W. Norton & Company, New-York ; traduction française par Paul Chemla, (2002), *La grande désillusion*, Fayard.

STROTZ, Robert, (1956), "Myopia and Inconsistency in Dynamic Utility Maximization", *The review of Economic Studies*, 23(3): 165-180.

SUNSTEIN, Cass et Richard THALER, (2003a), "Libertarian Paternalism is not an Oxymoron", *The University of Chicago Law Review*, 70(4) : 1159-1202

_____, (2003b), "Libertarian Paternalism", *The American Economic Review*, 93(2) : 175-179.

_____, (2008), *Nudge: Improving Decisions about Health, Wealth and Happiness*, Yale University Press, New-Heaven.

TAROZZI, Alessandro, *et al.*, (2011) "Micro-loans, Insecticide Treated Bednets and Malaria: Evidence from a Randomized Controlled Trial in Orissa (India)", non publié.

TAYLOR, Lance, (1988), *Varieties of Stabilisation Experience : Towards Sensible Macroeconomics in the Third World*, Oxford University Press, Oxford.

TEIRA, David, (2013), “Blinding and the Non-Interference Assumption in Medical and Social Trials”, *Philosophy of the Social Sciences*, 43(3) : 358-372.

TEIRA, David et Julian REISS, (2013), “Causality, Impartiality and Evidence-Based Policy”, in *Mechanism and Causality in biology and Economics*, édité par H-K Chao, Chen et R. Millstein, Springer, New-York.

TER KUILE, Feik, *et al.*, (2003), “Reduction of Malaria during Pregnancy by Permethrin-Treated Bed Nets in an Area of Intense Perennial Malaria Transmission in Western Kenya”, *American Journal of Tropical Medicine and Hygiene*, 68 : 50–60.

THISTLETHWAITE, Donald et Donald CAMPBELL, (1960), “Regression-Discontinuity Analysis : An Alternative to the Ex Post Facto Experiment”, *Journal of Educational Psychology*, 51(6) : 309-317.

VOHS, Kathleen, *et al.*, (2008), “Making Choices Impairs Subsequent Self-Control : A Limited Account of Decision Making, Self-Regulation, and Active Initiative”, *Journal of Personality and Social Psychology*, 94(5) : 883-898.

VORMS, Marion, (2009), *Théories, mode d'emploi. Une perspective cognitive sur l'activité théorique dans les sciences économiques*, Thèse de doctorat : philosophie, Université Paris 1 Panthéon-Sorbonne.

WEGNER, Daniel, Samuel CARTER, Teri WHITE, et David SCHNEIDER, (1987), “Paradoxical Effects of Thought Suppression”, *Journal of Personality and Social Psychology*, 53(1) : 5-13.

WILLIAMSON, John, (1990), “What Washington means by Policy Reforms”, in *Latin American Adjustment: How Much Has Happened?*. Institute for International Economics, Etats-Unis.

_____, (2003), “The Washington Consensus and Beyond”, *Economic and Political Weekly*, 38(15) : 1475-1481.

WORALL, John, (2002), «“ What Evidence in Evidence-Based Medicine ”», *Philosophy of Science*, 69(3) : 316-330.

YUNUS, Muhammad, (1997), *Vers un monde sans pauvreté*, Jean Claude Lattès, Paris.

Table des matières

INTRODUCTION GENERALE	1
1. Rendre la lutte contre la pauvreté efficace.....	2
2. Mener une analyse épistémologique de l'approche du J-PAL	6
2.1 <i>La dimension méthodologique de l'approche du J-PAL : les critères de validité interne et externe</i>	<i>6</i>
2.1.1 La notion de « validité ».....	8
2.1.2 La validité interne.....	9
2.1.3 La validité externe	11
2.2 <i>La dimension théorique (et politique) de l'approche du J-PAL.....</i>	<i>12</i>
3. La problématique de la thèse	13
4. La méthode de la thèse.....	15
5. Structure et organisation de la thèse	18
5.1 <i>La première partie : la force du J-PAL, son importante validité interne.....</i>	<i>18</i>
5.2 <i>La deuxième partie : la faiblesse du J-PAL, son insuffisante validité externe..</i>	<i>19</i>
5.3 <i>La troisième partie : les apports théoriques du J-PAL, entre paternalisme et irrationalité.....</i>	<i>20</i>
 PARTIE 1. LA RANDOMISATION DU J-PAL ET LE ROLE DE LA PREUVE	21
INTRODUCTION DE LA PARTIE 1.	22
CHAPITRE 1 : L'HISTOIRE DE LA « RANDOMISATION ».....	25
0. Introduction.....	26
1. Les origines de la randomisation : de la sensation à la télépathie	27
1.1 <i>La psychologie comme discipline originelle de la randomisation</i>	<i>28</i>
1.1.1 Sensations et seuil : les expériences de Fechner	28

1.1.2	Absence de seuil et randomisation	29
1.2	<i>Randomisation et télépathie : de la psychologie à la parapsychologie</i>	31
1.2.1	Télépathie et probabilités : les expériences de Richet.....	32
1.2.2	Télépathie et randomisation : les expériences de Coover	34
2.	Naissance d'une structure expérimentale rigoureuse et son utilisation par la médecine.....	35
2.1	<i>Ronald Fisher ou la naissance de la structure expérimentale de la randomisation</i>	36
2.1.1	Un cadre statistique rigoureux, une illustration : « the lady tasting tea » ...	37
2.1.2	La randomisation : le cœur d'une structure expérimentale rigoureuse	38
2.2	<i>Randomisation et application : les premiers essais cliniques</i>	40
2.2.1	Genèse de l'essai clinique : l'essai de James Lind.....	40
2.2.2	Premier essai clinique et structure expérimentale	41
3.	Expériences sociales et de terrain	43
3.1	<i>Des expériences de laboratoire aux expériences de terrain en économie</i>	44
3.1.1	Qu'est-ce qu'une expérience de terrain en économie ?.....	44
3.1.2	Évolution des expériences de terrain en économie	45
3.2	<i>Des expériences de terrain aux expériences sociales</i>	46
3.2.1	Les expériences sociales comme des expériences à vocation politique.....	47
3.2.2	Les expériences sociales, des évaluations à grande échelle : une illustration	48
4.	Conclusion	50
CHAPITRE 2 : RANDOMISATION, ECONOMIE DU DEVELOPPEMENT ET LE J-PAL		51
0.	Introduction.....	52
1.	Le cadre institutionnel du J-PAL	53
1.1	<i>Histoire du J-PAL</i>	53
1.1.1	Les débuts (2003-2007).....	54
1.1.2	L'essor du J-PAL (2007-2011)	55
1.2	<i>Structure du J-PAL</i>	56
1.2.1	Le J-PAL Global et ses différentes antennes.....	57

1.2.2	La Spécificité du J-PAL Europe.....	58
1.3	<i>La randomisation et le terrain : une articulation autour de trois partenaires</i> .	59
1.3.1	Les partenaires de terrain	59
1.3.2	Les financiers	60
1.3.3	Les centres de recherches	61
2.	La randomisation : le <i>gold standard</i> ?	63
2.1	<i>Le modèle de Rubin</i>	63
2.1.1	Intuitions.....	65
2.1.2	Application statistique.....	65
2.2	<i>Économie du développement et transposition du modèle de Rubin</i>	66
2.2.1	Contrefactuel et économie du développement	67
2.2.2	La notion de biais de sélection	69
2.3	<i>Les autres méthodes permettant de réduire les biais de sélection</i>	71
2.3.1	La méthode d'appariement et d'appariement par score de propension.....	71
2.3.2	La méthode de double différence	72
2.3.3	La méthode de discontinuité	73
3.	La randomisation en pratique	74
3.1	<i>Enjeux rencontrés avant la mise en place de la randomisation</i>	74
3.1.1	Collecte de données.....	74
3.1.2	Le niveau de randomisation	77
3.1.3	Structure de la randomisation.....	77
3.2	<i>L'introduction de la randomisation</i>	78
3.2.1	« Oversubscription »	79
3.2.2	« Encouragement design »	79
3.2.3	« Phase-In »	80
3.2.4	Randomisation au sein des groupes	80
3.3	<i>Limites (internes) de la randomisation</i>	81
3.3.1	Les effets Hawthorne et John Henry	81
3.3.2	Tout le monde ne veut pas participer	82
4.	Conclusion	83
	CHAPITRE 3 : L'APPROCHE DU J-PAL ET LA MEDECINE : DE L'ESSAI CLINIQUE A L'EVIDENCE-BASED POLICY	84

0. Introduction.....	85
1. Essais cliniques <i>versus</i> randomisation en économie du développement	86
1.1 <i>Essai clinique en médecine</i>	87
1.1.1 L'essai clinique : une partie d'un essai plus large.....	87
1.1.2 La méthode d'un essai clinique	88
1.1.3 Sida et circoncision : illustration d'un essai clinique médical	89
1.2 <i>Une nouvelle forme d'essai clinique : l'évaluation des programmes d'aide au développement</i>	90
1.2.1 La méthode des essais cliniques en économie du développement	91
1.2.2 Nouvelle forme d'essai clinique et pratique.....	94
2. De l' <i>evidence based medicine</i> à l' <i>evidence based policy</i>	96
2.1 <i>Le paradigme de la médecine fondée sur les preuves</i>	96
2.1.1 Contrecarrer une faible utilisation des connaissances scientifiques en unissant la médecine de ville et la médecine scientifique.....	97
2.1.2 Les quatre étapes constitutives de la médecine fondée sur les preuves	98
2.1.3 La hiérarchisation des preuves au sein de l' <i>evidence-based medicine</i>	100
2.1.4 Critiques de l' <i>evidence-based medicine</i>	100
2.2 <i>L'evidence based Policy ou la politique de la preuve</i>	101
2.2.1 Histoire de l' <i>evidence-based policy</i>	102
2.2.2 Les différents « ingrédients » qui fondent l' <i>evidence-based policy</i>	103
2.2.3 Pluralité des méthodes, communauté des caractéristiques.....	104
2.2.4 Types et niveaux de preuves	104
3. Le J-PAL et l' <i>evidence-based policy</i>	106
3.1 <i>Publications académiques et biais de publication : le rôle des agences politiques</i>	106
3.1.1 Un classement des publications autour de sept domaines.....	107
3.1.2 Le biais de publications.....	108
3.1.3 Le rôle des agences internationales	109
3.2 <i>Les analyses coût/efficacité</i>	110
3.2.1 Les analyses coût/efficacité : quelques éléments de définition.....	110
3.2.2 Quantifier l'impact	111
3.2.3 Quantifier les coûts.....	112
3.3 <i>Les « mallettes » politiques</i>	114

3.3.1 Les comportement sexuels à risques : illustration d'une « mallette politique »	114
3.3.2 Résultats des différents programmes.....	115
3.3.3 Les « leçons » politiques	117
4. Conclusion	117
CONCLUSION DE LA PARTIE 1.....	119
PARTIE 2. DE LA VALIDITE INTERNE A LA VALIDITE EXTERNE DE LA RANDOMISATION DU J-PAL : LA FIN DU GOLD STANDARD	121
INTRODUCTION DE LA PARTIE 2.	122
CHAPITRE 1 : LA RANDOMISATION A LA LUMIERE DE L'EPISTEMOLOGIE DE NANCY CARTWRIGHT.....	125
0. Introduction.....	126
1. La randomisation « idéale » de Nancy Cartwright	128
1.1 <i>La justification de la preuve</i>	128
1.1.1 Les méthodes qui offrent des « garantisseurs ».....	129
1.1.2 La démarche hypothético-déductive : un « intermédiaire » chez Cartwright	130
1.1.3 Les méthodes qui offrent des « vérificateurs »	131
1.2 « Randomisation idéale » et théorie probabiliste de la causalité	131
1.2.1 Une théorie probabiliste de la causalité.....	132
1.2.2 Ce que garantit la randomisation à l'aide des probabilités.....	133
2. Rendre les preuves utiles ou la nécessité d'un modèle causal.....	133
2.1 <i>Une théorie manquante de la preuve ?</i>	134
2.1.1 La pertinence de la preuve.....	135
2.1.2 Intégration de la preuve et théorie de la preuve	136
2.2 <i>Un modèle causal : éléments de définition</i>	137
2.2.1 Le modèle causal : une illustration.....	137
2.2.2 Rendre le modèle causal explicite ou produire une histoire causale.....	138
2.3 <i>Les capacités comme fondement d'un modèle causal</i>	139
2.3.1 Distinction loi causale et capacités.....	140

2.3.2	La notion de capacité.....	141
2.3.3	Modèle causal et capacités	142
3.	La randomisation du J-PAL à l'épreuve de l'épistémologie de Cartwright	143
3.1	<i>Le passage de l'efficacité réelle à l'efficacité potentielle de la randomisation : les trois ingrédients essentiels</i>	<i>144</i>
3.1.1	De la confusion entre l'efficacité réelle et l'efficacité potentielle dans la randomisation.....	145
3.1.2	Les trois « ingrédients » nécessaires à l'efficacité potentielle	145
3.2	<i>De la « randomisation idéale » de Cartwright à la randomisation de Duflo et de Banerjee</i>	<i>146</i>
3.2.1	Un raisonnement théorique patent dans les expériences du J-PAL.....	147
3.2.2	La randomisation du J-PAL : des capacités à l'œuvre ?	148
4.	Conclusion	149
 CHAPITRE 2 : LES « RANDOMISTAS » OU LE DIFFICILE ARBITRAGE ENTRE VALIDITE INTERNE ET EXTERNE DE LA RANDOMISATION DU J-PAL		151
0.	Introduction.....	152
1.	L'absence de la distribution des effets de traitement : un premier symptôme de la faible validité externe de la randomisation du J-PAL	154
1.1	<i>L'hétérogénéité des effets de traitement : l'illustration de la médecine</i>	<i>155</i>
1.2	<i>L'hétérogénéité des effets de traitement en économie.....</i>	<i>156</i>
1.3	<i>De l'importance de la distribution des effets de traitement en matière de lutte contre la pauvreté.....</i>	<i>158</i>
2.	D'une hétérogénéité des effets de traitement à une hétérogénéité indifférenciée : la question du pourquoi un programme fonctionne	159
2.1	<i>De Bombay à Vadodara : rentrer dans la boîte noire.....</i>	<i>160</i>
2.2	<i>L'absence de réponses au « pourquoi un programme fonctionne »</i>	<i>162</i>
3.	Le problème de la généralisation des résultats ou la faible validité externe de la randomisation	164
3.1	<i>La question de la généralisation des résultats.</i>	<i>165</i>
3.1.1	La question de la validité externe de la randomisation en médecine : efficacité réelle vs efficacité potentielle.....	165
3.1.2	La question de la validité externe en économie du développement.....	166

3.2	<i>Les effets négligés de la randomisation du J-PAL</i>	168
3.2.1	Effets d'équilibre : le passage d'un équilibre partiel à un équilibre général	168
3.2.2	Économie politique : la « grande image »	169
4.	L'impossible arbitrage entre validité interne et validité externe ou la fin du gold standard	170
4.1	<i>l'estimation de paramètres structurels au sein de la randomisation : une plus grande validité externe pour une plus faible validité interne</i>	171
4.1.1	La possibilité de traduire la forme réduite en forme structurelle	171
4.1.2	Un gage de validité externe pour une plus faible validité interne ?	173
4.2	<i>Un pont possible entre l'approche structurelle et la randomisation ?</i>	175
4.2.1	Un seul maillon de la chaîne : la clause ceteris paribus ou la causalité sous-jacente de la randomisation.....	175
4.2.2	Un pont entre la randomisation et les approches structurelles : la maxime de Marschak.....	177
5.	Conclusion	177
CHAPITRE 3 : EXPLIQUER LA BOITE NOIRE : LE ROLE DE LA THEORIE DANS L'APPROCHE DU J-PAL		179
0.	Introduction.....	180
1.	Comprendre les mécanismes : l'approche hypothético-déductive de Deaton	182
1.1	<i>La corrélation épargne et croissance : le croisement de la théorie et de l'observation empirique</i>	183
1.2	<i>Le lien entre revenu et prix des biens : l'« acid test »</i>	185
1.3	<i>Le lien entre revenu et alimentation : mesure et théorie</i>	186
1.4	<i>L'épistémologie de Deaton en question</i>	187
2.	Le rôle ambigu de la théorie dans l'approche du J-PAL	188
2.1	<i>Théorie ex ante versus théorie ex post</i>	189
2.2	<i>Du local ou global : quel passage (au) politique ?</i>	194
3.	Les résultats du J-PAL : les effets versus les causes de la pauvreté.....	196
3.1	<i>Les expériences du J-PAL : un « panorama » de la vie des pauvres</i>	197
3.2	<i>La pauvreté selon le J-PAL : des effets sans explications causales ?</i>	200

4. Conclusion	203
CONCLUSION DE LA PARTIE 2.....	205
PARTIE 3. DE L'ECHEC DE LA CAUSALITE A UN RENOUVEAU DU PATERNALISME : LES APPORTS THEORIQUES A L'ECONOMIE DU DEVELOPPEMENT DU J-PAL	207
INTRODUCTION DE LA PARTIE 3.	208
CHAPITRE 1 : LA RANDOMISATION ET SES SPECTRES EN ECONOMIE DU DEVELOPPEMENT	211
0. Introduction.....	212
1. Le fantôme du « Consensus de Washington »	213
1.1 <i>Un échec politique : les politiques d'ajustements structurels</i>	213
1.1.1 Cadre analytique du FMI et politiques d'ajustements structurels	214
1.1.2 Les effets des politiques d'ajustements structurels : l'aveu d'un échec ?.	216
1.2 <i>L'après consensus de Washington</i>	217
1.2.1 Que reste-il du « consensus de Washington » ?	218
1.2.2 Quelles alternatives au consensus ?	219
2. Le débat sur l'aide au développement : aide massive ou « Fardeau de l'homme blanc » ?	221
2.1 <i>Jeffrey Sachs et la fin de la pauvreté</i>	222
2.1.1 L'économie du développement comme économie « clinique »	223
2.1.2 La fin de la pauvreté : aide massive et trappes à pauvreté	224
2.2 <i>William Easterly et le « fardeau de l'homme blanc »</i>	225
2.2.1 Faits stylisés contre mythes de l'aide massive	226
2.2.2 La solution aux velléités impérialistes : l'absence de plan	227
3. La lutte contre la pauvreté dans la main des pauvres et l'aide au développement	229
3.1 <i>Impérialisme et aide massive vs empowerment et absence d'aide</i>	230
3.2 <i>Pour un empowerment mandaté</i>	232
4. Conclusion	234

CHAPITRE 2 : LA DEMARCHE DU J-PAL A L'AUNE DE SES EXPERIENCES SUR LE PALUDISME 236

0. Introduction.....	237
1. Paludisme et subvention : l'énigme	239
1.1 Contexte et mise en place d'une première expérience : les moustiquaires doivent-elles être gratuites ou payantes ?	240
1.2 Résultats : le rôle trouble de la subvention des moustiquaires	242
1.3 La première pièce d'un « puzzle » sur l'aide et la santé.....	244
2. Un premier dispositif : le <i>nudge</i> – effet de cadrage et engagement oral	247
2.1 Cadre expérimental et implantation.....	248
2.2 L'échec du <i>Nudge</i>	248
2.3 Une énigme toujours irrésolue ?	250
3. Un second dispositif de <i>nudge</i> : effets d'apprentissage et effets sociaux.....	251
3.1 Structure expérimentale et implantation	252
3.2 Importants effets d'apprentissage et faible utilisation des moustiquaires : une énigme toujours irrésolue ?	253
4. Économie du développement et comportement de santé : un paradoxe insoluble ? 255	
4.1 Pauvreté et comportements irrationnels : sous-estimation des bénéfices et incohérence temporelle.....	257
4.1.1 Les pauvres et l'information	258
4.1.2 Marché financier et incohérence temporelle	259
4.2 Implications politiques : jouer sur les incitations et imposer des mandats	260
4.2.1 Jouer sur les incitations : la « carotte »	262
4.2.2 Mandater : le « bâton ».....	264
5. Conclusion	266

CHAPITRE 3 : DU PATERNALISME LIBERTARIEN AU PATERNALISME DEMOCRATIQUE : UNE BOITE NOIRE VIDE ?..... 268

0. Introduction.....	269
----------------------	-----

1. Rationalité et paternalisme libertarien	271
1.1 « <i>Pauvres mais rationnels ?</i> »	271
1.2 <i>Le paternalisme et les riches</i>	274
2. Pour un paternalisme démocratique : le récent tournant d'Esther Duflo.....	279
2.1 <i>Réconcilier le paternalisme et les libertés à travers les capacités</i>	279
2.1.1 Décisions et libertés de trébucher.....	280
2.1.2 Décisions et espace mental.....	281
2.2 <i>Un Paternalisme sans morale ?</i>	285
2.2.1 Paternalisme démocratique et absence de « confort philosophique »	285
2.2.2 Qui décide et que mettre dans les biens élémentaires?	286
2.2.3 Le paternalisme et rôle de l'état	288
3. La double confusion du paternalisme démocratique de Duflo	289
3.1 <i>L'espoir comme capacité</i>	289
3.1.1 Trappe à pauvreté et espoir	290
3.1.2 Le mécanisme de l'espoir dans la lutte contre la pauvreté.....	292
3.2 <i>Des réelles notions de capacités et de libertés chez Sen à l'opacité philosophique de Duflo</i>	295
3.2.1 La notion de capacité chez Sen.....	295
3.2.2 Liberté substantielle et instrumentale.....	297
3.2.3 Les capacités, une approche nécessairement pluraliste	298
3.3 <i>Le « nœud » philosophique comme faille épistémologique</i>	299
3.3.1 La question de l'opérationnalisation : le paternalisme comme instrument politique	300
3.3.2 la définition des capacités : confusion entre processus et résultat.....	301
3.3.3 De la confusion philosophique à la faille épistémologique.....	303
4. Conclusion	305
CONCLUSION DE LA PARTIE 3.....	307
CONCLUSION GENERALE	310

BIBLIOGRAPHIE	319
Table des matières	337

RESUME

Cette thèse porte sur la méthode expérimentale utilisée par les chercheurs du J-PAL en économie du développement. La volonté des chercheurs du J-PAL est double : (1) produire des preuves d'efficacité des programmes de développement, (2) afin de guider la décision politique. L'objectif du présent travail est de mener une analyse épistémologique de l'approche du J-PAL. Cette analyse s'offre d'étudier une double dimension : une dimension méthodologique et une dimension théorique.

Tout d'abord, la dimension méthodologique vise à interroger la méthode utilisée par le J-PAL : la randomisation. Elle permet de questionner alors les modes d'inférence que la randomisation permet de produire, la fiabilité des résultats obtenus par la randomisation, ainsi que la possible utilisation par la sphère politique de ces résultats. Dès lors, deux principales questions guident cette analyse méthodologique : (1) la nature de *gold standard* méthodologique de la randomisation, (2) et la possible transposition des résultats obtenus par le J-PAL dans la sphère politique. La seconde dimension (théorique) questionne l'apport du J-PAL aux débats théoriques qui ont traversé l'économie du développement ces dix dernières années. Cette dimension s'interroge sur le tournant que le J-PAL a ou non opéré en économie du développement. La question est de savoir s'il s'agit d'un unique tournant empirique ou s'il est aussi question d'un tournant théorique et politique. S'intéresser à ces deux dimensions permet d'examiner l'approche du J-PAL dans son ensemble.

La thèse montre alors que cette approche offre des résultats dont la validité interne est importante mais dont la validité externe est faible, rendant difficile l'utilisation de tels résultats dans la sphère politique. Cette tension entre la validité interne et la validité externe montre l'antagonisme des deux objectifs que se fixe le J-PAL et définit un problème épistémologique : le refus de théorie ainsi que l'absence de mise en évidence des mécanismes qui sous-tendent les résultats obtenus par le J-PAL empêche ce dernier de produire des recommandations politiques claires.

DISCIPLINE : Sciences Economiques

MOTS-CLES : Randomisation – Economie Expérimentale – Pauvreté – Philosophie économique – Epistémologie – Esther Duflo – Abhijit Banerjee.
