

Introduction générale

Les controverses sur les expérimentations aléatoires dans le domaine du développement

Épistémologie, éthique et politique

Florent BÉDÉCARRATS, Isabelle GUÉRIN et François ROUBAUD

En octobre 2019, Abhijit Banerjee, Esther Duflo et Michael Kremer se sont vu remettre le 51^e prix de la Banque de Suède en sciences économiques en mémoire d'Alfred Nobel. Les trois chercheurs ont été récompensés « pour leur approche expérimentale de la lutte contre la pauvreté dans le monde » et pour avoir « transformé l'économie du développement – domaine qui étudie les causes de la pauvreté dans le monde et les meilleures façons de la combattre –, qui est maintenant un domaine de recherche florissant » (Royal Swedish Academy of Sciences, 2019 : 2). Contrairement à l'expérimentation en laboratoire, l'expérimentation sur le terrain sert à effectuer des tests grandeur nature sur les interventions, les comportements et la prise de décision dans le « monde réel », puis à démontrer des relations causales d'impact (*ibid.* : 3). En conséquence, selon le jury, « nous disposons désormais d'un grand nombre de résultats concrets sur les mécanismes spécifiques à l'origine de la pauvreté et sur les interventions spécifiques visant à l'atténuer » (*ibid.*). Les réalisations des lauréats dans les domaines de la santé, de la scolarisation, de l'égalité des sexes et de la politique, ainsi que du crédit sont de formidables illustrations de leur travail. Ce prix reconnaît le succès d'une méthode de longue date inspirée par le domaine médical, les évaluations par assignation aléatoire (*Randomized Controlled Trials* – RCT), et désormais appliquée aux questions de pauvreté et de développement. L'attribution du Nobel d'économie n'a pas vraiment été une surprise. Les RCT ont été introduites dans le domaine du développement dès le début des années 2000 et ont depuis connu un succès

croissant auprès des universitaires, des bailleurs de fonds et des praticiens du développement, à tel point qu'elles sont désormais considérées comme l'étalon-or de l'évaluation des politiques de lutte contre la pauvreté et de la compréhension de ses origines.

S'il y a bien des raisons de saluer l'attribution de ce prix (l'un des trois lauréats est une jeune femme¹ et le prix met en avant la question de la pauvreté et de la collecte de données primaires, longtemps négligée par l'économie du développement), il y a également lieu de s'interroger sur la pertinence et les répercussions du recours accru à cette méthode, que le prix pourrait encore stimuler. Quelle est la portée réelle des RCT ? Ont-elles vraiment « considérablement amélioré notre capacité à lutter contre la pauvreté dans le monde », comme le suggère le jury du prix de la Banque de Suède ? Quels types de questions les RCT sont-elles en mesure de traiter, ou non ? L'explication causale est-elle la seule approche permettant de comprendre la pauvreté, et les RCT parviennent-elles systématiquement à la fournir ? Enfin, et surtout, la prééminence de l'expérimentation dans le domaine de l'économie du développement, telle qu'elle est reconnue et saluée par le jury du prix Nobel, est-elle scientifiquement légitime et politiquement souhaitable ?

Le présent ouvrage propose de répondre à ces questions. L'initiative de ce projet éditorial est issue de la conférence European Development Network (EUDN) sur le *Malaise dans l'évaluation* organisée par l'Agence française de développement (AFD) à Paris en 2012 (AFD, 2012). Lors de cet événement, nous avons assisté à un véritable dialogue de sourds : alors que certaines voix critiques ont exposé les raisons de leurs doutes, ceux que nous appellerons les *randomistas*², pour reprendre un terme bien usité dans le milieu et par souci de facilité, exposaient avec assurance leurs convictions et leurs résultats en évitant d'engager le débat sur le fond.

Nous avons donc décidé d'analyser le succès des RCT sous trois angles (BÉDÉCARRATS *et al.*, 2013 ; 2019a ; 2021) : en développant des critiques théoriques fondées sur les questions classiques de validité interne et externe (les RCT « par le haut » ; *doing the maths*) ; en portant la critique sur le front empirique : comment les RCT sont menées sur le terrain ? (les RCT « par le bas » ; *doing the cooking*) ; enfin, en analysant l'économie politique des RCT en termes d'offre et de demande (les RCT comme « business » scientifique ; *doing the accounts, both financial and symbolic*). Autant le premier point avait déjà été largement étudié (notre contribution restant donc marginale), autant les deux autres dimensions étaient

1. Le prix Nobel d'économie, décerné pour la première fois en 1969, a depuis été remporté par 84 lauréats. Esther Duflo n'est que la deuxième femme lauréate. Au-delà du prix lui-même, l'économie en tant que science sociale est la discipline la plus marquée par la discrimination à l'égard des femmes (LUNDBERG et STEARNS, 2019).

2. Nous entendons par là les chercheurs qui défendent la supériorité de cette méthode sur toutes les autres. Sur le terme non péjoratif de « *randomista* », voir Ravallion (chap. 1, ce volume) et Ogden (chap. 4, ce volume). Voir aussi GIBSON (2019).

relativement inexplorées³. Nos propres analyses, issues de l'observation approfondie de deux RCT (microcrédit au Maroc [MORVANT-ROUX *et al.*, 2014] et micro-assurance au Cambodge [QUENTIN et GUÉRIN, 2013]), ont été largement corroborées par l'analyse de trois des RCT les plus emblématiques⁴. Ces RCT se sont finalement révélées très discutables, alors qu'elles avaient largement contribué à faire de cette méthode le véritable étalon-or.

Après ces recherches préliminaires, nous avons continué dans deux directions parallèles. Nous avons poursuivi notre travail sur le microcrédit rural au Maroc en menant une réplique. Les résultats ont non seulement corroboré l'hypothèse d'une contradiction entre les RCT en théorie et en pratique, mais ils ont aussi révélé de nouvelles dimensions de cette divergence (BÉDÉCARRATS *et al.*, 2019a ; 2021). Étendue à un ensemble de RCT sur le microcrédit, cette contradiction fait l'objet d'un des chapitres de cet ouvrage (chap. 7). Désireux de débattre de la question et de susciter une controverse scientifique, ou du moins une discussion, nous avons ensuite lancé ce projet de co-écriture d'un livre pour ouvrir la question à d'autres disciplines, voix et opinions, y compris des points de vue beaucoup plus positifs sur la méthode que les nôtres. Si certains disent que le débat est fatigant et lassant (DIMOVA, 2019 ; voir aussi Ogden, chap. 4, ce volume), nous considérons néanmoins qu'il est vital, tant sur le plan scientifique que démocratique, et ce pour des raisons que nous exposons plus avant.

Dans cet ouvrage, nous rassemblons certains des plus grands spécialistes du domaine, issus de divers horizons et disciplines (économie, économétrie, mathématiques, statistiques, économie politique, socio-économie, anthropologie, philosophie, santé globale, épidémiologie et médecine, élaboration des politiques), et examinons les principales faiblesses des RCT dans le domaine du développement, mais aussi quelques-uns de leurs points forts insoupçonnés. Nous prenons des exemples concrets pour expliquer le fonctionnement des RCT, ce qu'elles peuvent

3. Deux conclusions principales sont ressorties de nos analyses. Premièrement, si les RCT représentent un bon moyen d'estimer l'impact causal d'un certain nombre de projets délimités, cela n'est vrai que dans des conditions idéales définies en théorie, mais rarement observées sur le terrain. Dans ces conditions idéales, il est possible que les RCT puissent être utilisées pour quantifier statistiquement les impacts (significativité et ampleur), mais ils ne permettent pas d'identifier les mécanismes par lesquels ces impacts transitent (ce qui est paradoxal pour une méthode qui fait de l'analyse de la causalité son principe fondamental). Deuxièmement, trois des principales affirmations des *randomistas* sont sans fondement : les RCT sont supérieures à toute autre méthode ; la multiplication des RCT peut résoudre le problème de la validité externe, reconnu par tous comme une faiblesse intrinsèque (ce que nous avons qualifié de « projet hégémonique ») ; et les RCT peuvent fournir toutes les réponses sur « ce qui fonctionne et ce qui ne fonctionne pas dans le développement ».

4. La fameuse RCT associée au programme de transferts monétaires conditionnels (*Conditional Cash Transfers* – CCT) au Mexique (*Progresa*, rebaptisé *Oportunidades*, puis *Prospera*), que beaucoup considèrent comme le catalyseur de la ruée sur les RCT, ainsi que sur les CCT par la même occasion, mais dont la mise en œuvre et donc la validité interne sont contestées (FAULKNER, 2014) ; la RCT tout aussi célèbre sur les vers intestinaux au Kenya, réalisée par MIGUEL et KREMER (2004), dont les résultats ont été contestés par un groupe d'épidémiologistes (AIKEN *et al.*, 2015 ; DAVEY *et al.*, 2015 ; HUMPHREYS, 2015), ce qui est paradoxal car les *randomistas* ont fait des RCT en médecine le fer de lance du mouvement ; et enfin, la RCT sur le recrutement et la supervision des enseignants au Kenya (DUFLO *et al.*, 2015), dont BOLD *et al.* (2013) ont montré que la transposition à grande échelle *via* une politique nationale mise en œuvre par le gouvernement n'a produit aucun des résultats escomptés.

permettre de mesurer, pourquoi parfois elles échouent, comment elles peuvent être améliorées et pourquoi d'autres méthodes sont à la fois utiles et nécessaires. Nous abordons les questions de méthode, d'épistémologie, d'éthique, de théorie et d'idéologie. En mettant l'accent notamment sur la mise en œuvre des RCT *sur le terrain*, loin des conditions de laboratoire idéales, nous nous distinguons nettement des autres analyses critiques. Cela permet de révéler certaines utilisations et certains effets insoupçonnés de ces RCT, leurs utilisations et leurs fins politiques, mais aussi leur potentiel perturbateur (au sens positif du terme). Nous explorons la vision du monde implicite sur laquelle s'appuient de nombreuses RCT et qu'elles diffusent. Tout en examinant l'écart entre la portée limitée de la méthode et son succès dans le monde entier, nous proposons des domaines d'amélioration, ainsi que des méthodes alternatives. Sans contester la contribution des RCT à la science, nous mettons en garde contre leur prétendue supériorité et les dangers potentiels d'une utilisation inadaptée. Nous soutenons également que le meilleur usage des RCT n'est pas nécessairement celui qui vient immédiatement à l'esprit et que ses partisans promeuvent : comprendre certains comportements plutôt qu'évaluer les interventions.

Si le principe des RCT en science remonte à plus d'un siècle – leur utilisation dans le cadre du développement international est appelée la quatrième vague (JAMISON, 2017) –, leur utilisation à grande échelle dans les pays en développement est sans précédent (Ravallion, chap. 1, ce volume). Les RCT représentent une avancée indéniable pour l'économie du développement, et ce pour plusieurs raisons. Elles offrent une solution (parmi d'autres) à l'épineuse question de l'attribution causale (comment isoler l'effet d'une intervention de tous les changements qui se sont produits en même temps). Elles accordent une place centrale à la question de l'évaluation de l'aide et à la nécessité de rendre compte de l'aide. Les RCT donnent un nouvel élan à la collecte de données d'enquête de première main par les économistes du développement. Enfin, par le passé, les pays du Sud étaient marginalisés par la recherche économique en raison de leur déficit de données de qualité, en particulier longitudinales. La généralisation des RCT permet de placer la recherche économique sur ces pays au niveau des meilleurs standards internationaux. La nouvelle vague de RCT dans le domaine du développement se présente même comme un progrès méthodologique commencé au Sud et transféré vers le Nord (BÉDÉCARRATS *et al.*, 2019b).

Néanmoins et en dépit d'un champ d'application restreint (détailé ci-dessous et tout au long de l'ouvrage), les RCT sont aujourd'hui labélisées étalon-or de l'évaluation, à l'aune duquel il conviendrait de jauger toute approche alternative. Cette suprématie est par ailleurs susceptible d'être renforcée par l'attribution du prix de la Banque de Suède. Présentées par leurs adeptes comme une véritable révolution copernicienne en économie du développement⁵, on leur attribue en

5. « De la même manière que les évaluations randomisées ont révolutionné la médecine au xx^e siècle, elles ont le potentiel de révolutionner les politiques sociales au xxi^e » (DUFLO, GLENNERSTER et KREMER, 2004 : 29).

exclusive le qualificatif de « rigoureuses », voire de « scientifiques » (voir Ravallion, chap. 1, ce volume). Bien au-delà du champ méthodologique, l'ambition de certains défenseurs des RCT les plus médiatiques est de fournir une liste exhaustive des bonnes et des mauvaises politiques en matière de développement (Labrousse, chap. 8, ce volume). L'objectif invoqué est d'accumuler un nombre toujours plus grand d'études d'impact afin d'en tirer les enseignements généralisateurs. Il est cependant évident que la suprématie revendiquée des RCT en matière d'évaluation engendre un certain nombre d'effets pervers. Citons par exemple la disqualification et l'effet d'éviction des méthodes alternatives, la mobilisation toujours plus grande des ressources allouées, les positions de rentes, et la légitimation d'une vision spécifique et étroite du « développement » (ce que Lant Pritchett appelle dans le chap. 2 de ce volume le « développement fétichiste », *kinky development*). À cela, s'ajoute la disqualification des projets et des politiques de développement qui ne respectent pas les contraintes exigées par les protocoles de randomisation (Ravallion, chap. 1, ce volume ; Garchitorea *et al.*, chap. 5, ce volume ; Patnaik, entretiens, ce volume ; voir aussi ADAMS, 2016).

Nous ne sommes évidemment pas les premiers à formuler des critiques et de nombreuses voix se sont déjà élevées⁶. Les critiques de James Heckman et Angus Deaton (Deaton, prologue, ce volume ; DEATON, 2010a ; DEATON et CARTWRIGHT, 2018 ; HECKMAN, 1992 et chap. 12, ce volume) sont d'autant plus percutantes que tous deux ont également reçu le prix Nobel d'économie (Deaton en 2015 et Heckman en 2000). Si ces critiques sont désormais plus fréquemment reconnues par les membres du mouvement pro-RCT (Ogden, chap. 4, ce volume),

6. Voir par exemple HECKMAN, 1992 ; RODRIK, 2009 ; BARRETT et CARTER, 2010 ; DEATON, 2010a ; HARRISON, 2011 ; PRITCHETT et SANDEFUR, 2015 ; DEATON et CARTWRIGHT, 2018. Plusieurs ouvrages ont également contribué à cette discussion. Le premier, édité par COHEN et EASTERLY (2010), a mis le feu aux poudres et a déclenché la controverse naissante. Il ne contient qu'un seul chapitre consacré spécifiquement à ce sujet, avec un débat passionnant, bien que bref, entre Banerjee, Rodrik, Mulathain et Ravallion. Les autres chapitres abordent principalement la question de savoir si les politiques de développement fonctionnent et si oui, lesquelles, mais la question des RCT est présente en filigrane tout au long du livre. Le livre de OGDEN (2017) est le plus récent et porte essentiellement sur les RCT. Il se présente sous la forme de 20 entretiens avec des acteurs de premier plan dans ce domaine. Quatorze d'entre eux sont des figures actives du mouvement pro-RCT et quatre autres y sont plus modérément impliqués. Seules deux personnes se montrent critiques (Angus Deaton et Lant Pritchett). Malgré leur notoriété, leurs contributions sont assez courtes et copieusement réinterprétées et critiquées par les autres contributeurs. Troisièmement, le livre édité par TEELE (2014) est plus détaillé et plus nuancé. Il apporte une contribution majeure à notre compréhension du sujet, en comparant notamment des RCT menées dans des pays du Nord et du Sud par des politologues et des économistes. Néanmoins, il reste centré sur des considérations méthodologiques et épistémologiques. De nombreuses contributions ne font que reprendre des articles aujourd'hui dépassés, publiés ailleurs dans les années 2000, avant que les RCT ne connaissent un véritable essor. Dix ans plus tard, le présent livre actualise les ouvrages précédents en s'appuyant sur la littérature la plus récente et en adoptant une vision plus large, à la fois en termes d'angles disciplinaires et d'enjeux. Enfin, à la fin 2019, alors que nous finalisons notre manuscrit, la revue *World Development* a publié un numéro spécial sur les RCT dans le domaine du développement (paru début 2020). Profitant de l'attribution du prix Nobel d'économie à Banerjee, Duflo et Kremer, ce numéro spécial rassemble un peu plus de 50 courts articles (d'une ou deux pages) rédigés par un large éventail d'auteurs. Compte tenu du format condensé de ces contributions, elles ne peuvent évidemment pas fournir une analyse approfondie. Outre l'examen d'un large éventail de positions concernant les RCT (ce qui a bien fonctionné et ce qui n'a pas fonctionné), ce numéro spécial présente toutefois l'intérêt de proposer des pistes de recherche futures.

la question n'a toutefois jamais fait l'objet d'une véritable controverse scientifique. À défaut de cette controverse (les *randomistas* les plus éminents ont décliné notre invitation), ce livre instaure un dialogue entre les approches, les disciplines, les différents secteurs d'intervention, et, en définitive, entre les différents points de vue sur le rôle des RCT et leur potentiel.

Certains auteurs du livre considèrent que l'engouement pour les RCT constitue une « folie » (Pritchett, chap. 2, ce volume), que leur supériorité n'est autre qu'une « fiction » (Labrousse, chap. 8, ce volume), et que ce sont des « outils inefficaces en matière de redevabilité et d'apprentissage des organisations » et pas à proprement parler des évaluations (Picciotto, chap. 9, ce volume). D'autres considèrent qu'elles ont leur place dans la palette des méthodes d'évaluation, mais que leur supériorité autoproclamée relève « plus de la foi que de la science » et que, dans certaines situations et pour certaines questions, les études non expérimentales (*observational studies*) se révèlent nettement plus appropriées (Ravallion, chap. 1, ce volume). C'est ce que montrent également les analyses sectorielles portant sur la santé (Garchitorena *et al.*, chap. 5, ce volume), l'assainissement rural (Spears, Ban et Cumming, chap. 6, ce volume), le microcrédit (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume) et la gouvernance (Natarajan, entretiens, ce volume).

Un point de vue plus optimiste laisse entendre que les RCT ont pris en compte les critiques et que, sous leur forme actuelle, elles offrent de véritables réponses à de nombreuses questions de développement (Ogden, chap. 4, ce volume). Selon Jonathan Morduch, les RCT sont utiles non pas tant pour « évaluer », mais pour « explorer » les comportements en recourant à des manipulations des structures de prix, des contrats, des méthodes pédagogiques, etc. Les chercheurs peuvent exploiter la perturbation générée par les protocoles randomisés pour observer *in situ* les changements dans les interventions et les comportements, étudier leurs répercussions et en tirer des conclusions opérationnelles (Morduch, chap. 3, ce volume).

D'autres demandent qu'elles soient améliorées tant du point de vue éthique, lequel demeure un angle mort pour les protocoles d'enquête en économie du développement (Abramowicz et Szafarz, chap. 10, ce volume), que du point de vue de l'explication causale, qu'il soit question de mieux utiliser les *a priori* (*priors*) (Vivaldi, chap. 11, ce volume) ou des phénomènes de non-respect du protocole (*non compliance*) comme révélateurs des préférences des populations ciblées (Heckman, chap. 12, ce volume).

L'objectif de cette introduction, qui reflète uniquement le point de vue des éditeurs, n'est pas de réconcilier les auteurs ou de trouver un compromis, mais de donner au lecteur une image plus claire des enjeux du débat. Dans la première partie, nous exposons en détail les arguments épistémologiques, politiques et éthiques qui le sous-tendent. Dans la deuxième partie, nous nous efforçons de définir les politiques et les projets de développement qui pourraient se prêter aux spécificités des RCT. La troisième partie revient sur l'idée d'une controverse scientifique, que nous appelons de nos vœux, mais qui n'a malheureusement

pas encore eu lieu, en examinant les raisons de son inexistence. La conclusion propose des moyens d'améliorer les RCT, ainsi que des alternatives méthodologiques.

Les arguments du débat : épistémologie, politique et éthique

Nous ne reviendrons pas ici sur toutes les critiques formulées à l'égard des RCT, qui sont déjà énumérées dans différents chapitres (Ravallion, chap. 1, ce volume ; Ogden, chap. 4, ce volume ; voir aussi BÉDÉCARRATS *et al.*, 2019b), car nous pensons qu'il est plus utile de se pencher sur les dissensions épistémologiques, politiques et éthiques qui sous-tendent – souvent implicitement – bon nombre des divergences d'opinions concernant les RCT.

Loin d'être purement techniques, les débats autour des RCT renvoient à des conceptions de la connaissance et du savoir différentes et souvent bien difficiles à concilier. La recherche en sciences sociales sur les interactions humaines est-elle perçue de manière scientifique⁷ (PUTNAM, 2009), comme la recherche de la réponse ultime et universelle à un problème donné, ou comme un processus d'apprentissage continu pour trouver des réponses raisonnables, limitées dans le temps et l'espace, compte tenu de la diversité des connaissances, y compris celles des populations visées par le développement ? Les chiffres, les méthodes statistiques et économétriques appliquées aux sciences sociales ne sont-elles pour nous que des instruments et des techniques, fruits d'un progrès scientifique linéaire ? Ou les considérons-nous aussi comme une construction sociale et politique érigée par des conventions quelque peu arbitraires, inextricablement liées à une certaine conception de l'État et des politiques publiques, du marché, du pouvoir et de l'action collective (DESROSIÈRES, 2013b), qui façonnent en partie le monde qu'elles cherchent à représenter, à comprendre et à conseiller (MACKENZIE, MUNIESA et SIU, 2007) ? Cette deuxième acception de la connaissance ne rejette pas les preuves scientifiques, mais préconise leur ancrage dans des contextes sociaux et politiques particuliers. Par ailleurs, elle différencie clairement les connaissances scientifiques de la prise de décision politique, cette dernière impliquant de se référer à des valeurs pour choisir entre différentes options et évaluer leurs conséquences sociales, économiques et politiques (DRÈZE, 2018a).

Les divergences d'opinions autour des RCT sont également fondées sur des notions différentes du développement, de la pauvreté et, plus largement, de la

7. Par scientisme, nous entendons l'idée que la science expérimentale est la seule source fiable de connaissances sur le monde et qu'elle est le meilleur moyen d'organiser l'humanité pour résoudre tous ses problèmes les plus urgents. L'expérimentation prétend se passer de tout raisonnement métaphysique, philosophique, éthique ou esthétique.

politique, considérée comme une conception du monde dans laquelle nous vivons et que nous nous efforçons d'atteindre. Le monde est-il un agrégat d'individus à la recherche d'autonomie ou bien un système complexe fait de dialectique, d'interactions multiples, de rétroactions et d'effets systémiques entre des êtres sociaux interdépendants et souhaitant le rester ? Devrions-nous considérer les « causes de la pauvreté comme un manque ou un besoin de ressources pertinentes ou comme un processus actif d'appauvrissement ou de perpétuation de la pauvreté » (SHAFFER, 2015 : 154) ? Si l'on appréhende les causes de la pauvreté sous l'angle des manques, il convient de mettre en place des politiques visant à « faire la différence » (pour faire face aux déficits en matière de santé, d'éducation, de nutrition, d'eau et d'assainissement, de crédit, etc.) ; et pour comprendre les effets de ces politiques, il faut un contrefactuel pour pouvoir isoler cette différence et attribuer l'impact à la politique en question. En revanche, si l'on conçoit la causalité de la pauvreté en termes de processus et de relations sociales, il convient d'adopter des politiques macro-économiques et structurelles (taux de change, politiques de contrôle des capitaux, mesures de protection sociale, etc.) et, pour comprendre les effets de ces mesures, il faut adopter une « approche fondée sur les mécanismes » qui explore la diversité et la complexité des processus causaux à l'origine de ces effets (SHAFFER, 2015).

Enfin, ces visions distinctes se traduisent par des conceptions divergentes du rôle des économistes. S'agit-il de « réparer » (*fix*) le monde et de se concentrer sur les détails pratiques de la mise en œuvre des politiques (DUFLO, 2017), à l'image d'un plombier ou d'un ingénieur qui répare des tuyaux fissurés ? Ou bien les économistes doivent-ils garder une distance critique par rapport au fonctionnement du système actuel, voire aller jusqu'à le remettre radicalement en question ?

Ces différentes positions épistémologiques (qui se présentent davantage comme un continuum plutôt qu'une opposition binaire) transparaissent dans les débats autour des RCT et se manifestent dans une série d'oppositions qui jalonnent les chapitres de ce livre : macro *versus* micro, biens publics *versus* biens privés, interventions sanitaires horizontales *versus* verticales, action publique *versus* marketing social, structure *versus* comportement individuel, attribution *versus* processus, etc.

L'épistémologie des RCT dans le domaine du développement

En théorie, les *randomistas* voient l'expérimentation précisément comme un antidote aux idées préconçues (RODRIK, 2009). Ce pragmatisme pourrait bien donner l'impression de rejeter le scientisme, mais le fait de revendiquer la supériorité d'une méthode reflète clairement une conception scientiste de la science (Picciotto, chap. 9, ce volume). Ce scientisme se manifeste de deux manières. Tout d'abord, les *randomistas* prétendent fournir des réponses universelles sur un grand nombre d'interventions de développement. En réponse à la question des particularités contextuelles, certains *randomistas*, comme Esther Duflo, soutiennent qu'il convient de considérer les RCT comme des « biens publics mondiaux » et de créer une instance internationale chargée de les multi-

plier (SAVEDOFF *et al.*, 2006 ; GLENNERSTER, 2012). Celle-ci constituerait ainsi une base de données universelle et jouerait le rôle de « chambre de compensation » apportant des réponses sur « ce qui fonctionne ou ne fonctionne pas » en matière de développement (BANERJEE et HE, 2008). Mais ce projet hégémonique (BÉDÉCARRATS *et al.*, 2019b) ne résout pas la question de l'hétérogénéité, qu'il s'agisse des pratiques ou des contextes d'intervention (voir notamment Spears, Ban et Cumming, chap. 6, ce volume).

Ensuite, ce scientisme se manifeste par un excès de confiance dans la technique, avec en quelque sorte un fétichisme du protocole théorique, censé garantir l'équilibre des échantillons et donc régler la question de l'attribution. La *mise en œuvre* du protocole sur le terrain est secondaire. Comme pour toutes les recherches – et plus encore pour les RCT, compte tenu des budgets en jeu, de la taille des échantillons, des contraintes de comparaison entre les groupes de contrôle et de traitement, et des risques de contamination –, la mise en œuvre des protocoles s'écarte nécessairement de ce qui est prévu en théorie et nécessite des ajustements, des aménagements et des compromis⁸. Dans de nombreux cas, la collecte de données des RCT va à l'encontre des hypothèses des théorèmes utilisés pour l'inférence statistique. Les ONG et les gouvernements qui travaillent dans le domaine du développement ne savent que trop bien que les interventions sur le terrain ne se déroulent jamais comme prévu (MOSSE, 2004 ; OLIVIER DE SARDAN, 1995 ; 2021). Pourquoi les expérimentations devraient-elles être différentes ? Comme le montrent les différents chapitres de cet ouvrage, des écarts entre le protocole et la mise en œuvre peuvent être observés tout au long de la chaîne de production des connaissances.

– Des écarts au niveau de la construction des échantillons avec trois types de difficultés. La première difficulté réside dans les multiples biais entre les groupes de traitement et de contrôle (Ravallion, chap. 1, ce volume). Cela se traduit par une focalisation sur des populations très spécifiques, sans que cette spécificité soit mise en évidence par les *randomistas* (voir, par exemple, WYDICK [2016] et BÉDÉCARRATS *et al.* [2019a] sur le microcrédit ; voir également BARRETT et CARTER [2014 : 75], MOATTI, entretiens, ce volume). La deuxième difficulté réside dans le taux d'adhésion (*take up*) trop faible et, par conséquent, dans une différence insuffisante pour l'exposition à l'intervention. Le manque de puissance statistique affaiblit la capacité à tirer des conclusions, un problème qui, pour être résolu, nécessiterait des échantillons de taille peu réaliste et donc des budgets eux aussi irréalistes (MCKENZIE, 2012 ; Spears, Ban, et Cumming, chap. 6, ce volume). Un *take up* (taux d'adhésion) insuffisant peut également entraîner une transformation artificielle de l'intervention (voir le point suivant). Enfin, la « virginité » des zones de contrôle, condition souvent nécessaire à la comparaison, s'avère particulièrement complexe et pose des problèmes d'éthique et de faisabilité (BÉDÉCARRATS *et al.*, 2019b).

8. Notre expérience de réplcation montre la difficulté qu'ont certains *randomistas* à reconnaître les difficultés pratiques que pose la réalisation d'une RCT idéale, dont l'équivalent n'existe pas dans la réalité (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume).

- Des écarts au niveau du type d'intervention, dont la mise en œuvre peut s'avérer très différente du « monde réel », comme le montrent Garchitorea *et al.* (chap. 5, ce volume) dans le domaine de la santé, ou qui risque même d'être transformée artificiellement pour favoriser un meilleur *take up* (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume).
- Des écarts au niveau de la collecte de données, puisque la priorité accordée aux considérations économétriques peut faire obstacle aux considérations statistiques. La statistique n'est pas seulement la science des chiffres, c'est avant tout une science de la collecte de données, dont la qualité est garantie par de multiples techniques (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume).
- Des écarts au niveau de l'interprétation des résultats qui, loin de se limiter à une comparaison de moyennes, comme le prétendent les *randomistas*, implique en réalité un éventail d'hypothèses implicites et tout un art de la rhétorique dont le pouvoir de persuasion est particulièrement manifeste (Labrousse, chap. 8, ce volume).

Au final, les contraintes de mise en œuvre de la méthode peuvent obliger les chercheurs à se concentrer sur des indicateurs intermédiaires, des périodes de courte durée, des populations ou des zones géographiques spécifiques et, ce faisant, à se limiter à un ensemble très restreint de questions ou à produire des résultats inutilisables (voir les chap. 5, 6 et 7 de ce volume sur différents secteurs avec de nombreux exemples ; voir aussi le cas de la santé publique [Moatti, entretiens, ce volume] et de la gouvernance [Natarajan, entretiens, ce volume]). L'importance disproportionnée accordée à la pureté théorique des protocoles et à la démonstration de la causalité au détriment de la *faisabilité* des protocoles et de la *qualité* des données est un point de friction majeur (bien que souvent implicite) dans les désaccords sur la hiérarchie des méthodes.

De notre point de vue, donner la priorité à la méthode plutôt qu'aux questions de recherche équivaut à « chercher ses clés perdues sous le lampadaire » parce qu'il n'y a que là que l'on voit quelque chose. D'une certaine manière, et pour paraphraser le titre d'un livre sur l'aide au développement (NAUDET, 1999), il s'agit de trouver des problèmes (les projets à évaluer) aux solutions (les RCT).

Les RCT et le « développement »

Comme le suggère Lant Pritchett (chap. 2, ce volume), le succès des RCT n'est que le symptôme d'une maladie plus grave : l'abandon, par une partie de la communauté de l'aide internationale, des politiques de développement transformatrices à grande échelle (nationale, régionale et même internationale), ainsi que des tentatives de transformer en profondeur les systèmes socio-économiques⁹. Il est donc utile d'examiner ces transformations dans le domaine de

9. Le jury du prix Nobel, dans son communiqué de presse, en témoigne : « Les lauréats de cette année ont introduit une nouvelle approche pour obtenir des réponses fiables sur les meilleurs moyens de lutter contre la pauvreté dans le monde. En résumé, elle consiste à diviser cette problématique en questions plus petites et plus faciles à traiter ».

l'aide pour mieux comprendre l'attrait des RCT et leur champ d'application. Le contraste entre la portée limitée des RCT et leur succès scientifique, médiatique et politique résulte à la fois d'une offre et d'une demande. Du côté de l'offre, nous avons déjà montré que les *randomistas* ont élaboré un *business model* scientifique inédit, dont le J-Pal est l'exemple le plus emblématique et le plus abouti, et qui associe des qualités qui se renforcent mutuellement : excellence académique (légitimité scientifique), effort de séduction en direction du public (visibilité médiatique et légitimité citoyenne) et des bailleurs de fonds (demande solvable), investissement massif dans la formation (offre qualifiée), et modèle d'entreprise performant (rentabilité financière) (BÉDÉCARRATS *et al.*, 2019b). Aussi efficaces qu'elles soient, ces stratégies supposent néanmoins qu'il existe une *demande*. Si certaines méthodes, théories et technologies se révèlent efficaces, ce n'est pas en raison de leur supériorité scientifique, mais parce qu'elles parviennent à « galvaniser et rallier durablement des acteurs et des intérêts prêts à produire et à utiliser [les technologies en question] » (CALLON, 2006a : 155).

Les RCT bénéficient ici d'un environnement qui leur est particulièrement favorable, et qu'elles entretiennent en retour. Elles n'auraient certainement pas eu le même succès à une autre époque. Le climat universitaire, notamment en économie, est propice à l'essor des RCT avec la défaite des écoles hétérodoxes centrées sur les structures sociales et les processus de domination, la recherche des fondements micros de la macro et le primat de la quantification et de l'économie dans les sciences sociales. La montée en puissance conjointe de l'économie comportementale et expérimentale, consacrée par l'attribution en 2002 du Nobel d'économie au psychologue Daniel Kahneman et à l'économiste Vernon Smith, puis à l'économiste Richard Thaler en 2017, illustre cette évolution de la discipline. Les RCT mobilisent très largement les préceptes de l'économie comportementale, et c'est d'ailleurs par leur intermédiaire que celle-ci s'est diffusée en économie du développement, jusqu'à y occuper aujourd'hui une place prépondérante (FINE *et al.*, 2016).

C'est également à la suite de transformations dans le domaine de l'aide que la demande de RCT est apparue. La fin de la guerre froide a favorisé une émancipation relative de l'aide publique au développement (APD) à l'égard du politique. Pendant cette période, la coopération technique et financière ne constituait souvent qu'un registre supplémentaire des rivalités entre blocs. Mais cette subordination de la coopération à la *realpolitik* a toutefois été battue en brèche après la chute du mur de Berlin. Dans le nouveau monde post-moderniste, les promoteurs de l'APD se sont retrouvés sous les feux de la rampe, sommés d'apporter la preuve de leur utilité dans un contexte de crise de l'aide, des Objectifs du millénaire pour le développement (OMD) et du *New Public Management* (NAUDET, 2006).

Le nouveau credo conjugue une focalisation des politiques de développement en faveur de la lutte contre la pauvreté et la mise en avant d'une gestion axée sur les résultats. Ces orientations, formulées dans la déclaration de Paris en 2005, ont été depuis systématiquement réitérées lors de grandes conférences internationales sur l'aide publique au développement, à Accra en 2008, puis à

Busan en 2011 et enfin à Addis Abeba en 2015. La montée en puissance du paradigme de l'*evidence based policies*, qui consiste à fonder toute décision publique sur des preuves scientifiques, réserve aux savants une légitimité nouvelle dans ces arènes politiques. Les RCT répondent en principe à toutes les conditions requises par ce tournant : empirisme agnostique, simplicité apparente (simple comparaison de moyennes), mobilisation élégante de la théorie mathématique (gage de scientificité) et concentration sur les pauvres (mobilisation compassionnelle et engagement moral ; Labrousse, chap. 8, ce volume). Leur (apparente) simplicité les rend aisément compréhensibles par les décideurs. Elles apparaissent donc comme un vecteur privilégié pour éclairer la décision publique. L'évaluation du programme *Progresa* au Mexique a constitué un prototype de cette méthode et un cas d'école de sa performativité¹⁰ (BÉDÉCARRATS *et al.*, 2019b).

La crise de l'aide est également une crise de l'aide *publique* au développement. Alors que les efforts de financement de l'APD s'essouffent, les investissements privés et les transferts de fonds internationaux prennent le relais (IFC, 2017). Les gouvernements ne sont plus qu'un organe parmi d'autres dans une « coalition d'acteurs » composée d'entreprises, d'ONG, et plus largement de la « société civile », de fondations et d'instituts de recherche. Renouant avec le philanthrocapitalisme de la période industrielle, les fondations jouent un rôle croissant, principalement dans le secteur de la santé, mais aussi dans l'innovation technologique, désormais présente dans la plupart, voire la totalité des secteurs du développement (voir aussi DE SOUZA LEÃO et EYAL, 2020). Ces nouveaux acteurs et bailleurs de fonds modifient les *outils* de l'aide au développement. Non seulement le retrait de l'État en tant que responsable de la planification et du développement conduit à « penser petit » (COHEN et EASTERLY, 2010), mais, lorsqu'il va de pair avec la résurgence de la philanthropie, il ouvre la voie à un développement qui juxtapose la privatisation (des interventions et des acteurs), la marchandisation (des biens et des services fournis), mais aussi la compassion.

En faisant des pauvres des entrepreneurs aux pieds nus, le microcrédit, avec sa promesse d'un double résultat – réduction de la pauvreté et rentabilité ou au moins durabilité financière – a été un pionnier de la marchandisation. Cette marchandisation s'est ensuite développée sous le nom de *Bottom of the Pyramid* (BoP – bas de la pyramide en français), version *low-cost* de la théorie du *trickle down*, avec l'idée que la consommation des pauvres finira par constituer un moteur de croissance et de redistribution (ELYACHAR, 2012).

À cette motivation économique s'ajoute une « raison humanitaire » (FASSIN, 2010). Face à des infrastructures publiques considérées comme moribondes,

10. Il est cependant édifiant de constater que ce programme a été un puissant outil de contrôle social et politique, rongé par le népotisme et la corruption (CRUCIFIX et MORVANT-ROUX, 2018 ; KIDD, 2019). Par ailleurs, les lacunes de son évaluation expérimentale, notamment en termes de validité interne (FAULKNER, 2014), ont précisément été les arguments utilisés par le nouveau gouvernement mexicain pour annoncer son retrait début 2019 (ENCISCO, 2019).

délabrées ou utopiques, et aux souffrances et besoins qu'elles engendrent, un devoir moral d'agir est en train d'émerger. Animés par un sentiment de compassion et d'urgence, les financiers et les praticiens – mais aussi les chercheurs – s'associent pour concevoir et tester un ensemble d'interventions à micro-échelle : ces « biens humanitaires », pour reprendre l'expression de REDFIELD (2012), tentent de répondre au mieux, de manière ponctuelle et temporaire, aux besoins considérés comme les plus urgents et les plus criants. Ces biens humanitaires visent à pallier les défaillances des gouvernements, et ils s'inscrivent dans cette double logique compassionnelle et économique, même si le volet économique n'exclut pas les mesures de redistribution (voir ci-dessous).

Dans cette nouvelle configuration, et même si le financement public des grandes infrastructures continue de représenter une part importante de l'aide internationale, les pouvoirs de décision et de planification des gouvernements cèdent progressivement la place aux fonds verticaux¹¹, aux fondations¹², aux entreprises privées¹³ et aux nouveaux mécanismes financiers tels que les obligations à impact social. Les fondations, nouveaux acteurs en pleine croissance, sont vouées à jouer un rôle de plus en plus important (voir également Pritchett, chap. 2, ce volume). À l'instar de la fondation Ford, qui a soutenu l'essor des expériences aux États-Unis dans les années 1960, de nombreuses fondations jouent aujourd'hui un rôle moteur dans la généralisation des RCT dans le domaine du développement (à commencer par la mise en place de J-Pal ; JATTEAU, 2016 : 230). Le principe même des obligations à impact social, dont le remboursement aux investisseurs est conditionné par l'obtention de résultats sociaux précis, favorise une tendance similaire. Enfin, dans ce processus de privatisation du développement (privatisation des interventions et des acteurs), les ONG occupent une place de choix en tant que partenaires de mise en œuvre.

Loin des ambitions réformatrices et parfois idéalistes des générations précédentes d'acteurs du développement, les biens privés, marchands et humanitaires ont le mérite d'être réalistes et concrets, et d'offrir une solution pragmatique à des besoins considérés comme urgents. Leur déploiement n'est pas exempt de critiques – les plus connues étant sans doute les débats sur l'alimentation thérapeutique vue comme une pratique commerciale déloyale ayant un impact sur les systèmes agricoles locaux. Pourtant, du point de vue de leur objectif – résoudre un problème temporaire et individuel –, ils fonctionnent bel et bien (REDFIELD, 2012). Comme le soulignent plusieurs chapitres, et nous y reviendrons plus tard, ce sont précisément ces types de biens, en raison de leur ciblage individuel et de leur nature à court terme, qui se prêtent le mieux aux contraintes

11. Comme le Fonds mondial et Gavi, l'Alliance du vaccin (anciennement l'Alliance globale pour les vaccins et l'immunisation) dans le domaine de la santé.

12. La fondation Bill et Melinda Gates reste leader dans de nombreux sous-secteurs de la santé, mais aussi dans tout ce qui concerne les nouvelles technologies. Les fondations bancaires telles que Citi et Mastercard sont des acteurs de premier plan dans le domaine de l'inclusion financière.

13. Comme Nutriset pour les aliments thérapeutiques destinés à traiter la malnutrition et Vestergaard Frandsen pour les filtres à eau, les moustiquaires contre la mouche tsé-tsé et les moustiquaires imprégnées d'insecticides.

des essais randomisés. De même, les ONG restent les partenaires opérationnels de choix pour les *randomistas*, car elles sont plus flexibles, moins bureaucratiques, plus ouvertes à l'innovation et plus fiables que les gouvernements (COHEN et EASTERLY, 2010 ; WEBBER et PROUSE, 2018). Si les *randomistas* expriment le souhait de travailler davantage avec les gouvernements (BANERJEE, 2013), ils ont toutefois du mal à le faire (Pritchett, chap. 2, ce volume). En Inde, terrain d'étude privilégié pour les RCT, les témoignages apportés dans les entretiens de cet ouvrage par un haut fonctionnaire indien (Natarajan) et un ancien conseiller économique principal du gouvernement indien (Patnaik) suggèrent que l'impact des RCT sur l'élaboration des politiques est non seulement négligeable, mais aussi contre-productif, car les RCT détournent l'attention des problèmes réels et sapent le métier des économistes.

Cette transformation du domaine du développement étant intervenue bien avant les RCT, il serait exagéré de dire qu'elles en sont responsables (Morduch, chap. 3, ce volume), même si les effets d'éviction sont à prendre au sérieux (voir plus bas la section « La controverse peut-elle vraiment être évitée compte tenu des effets d'éviction ? »). Mais faut-il pour autant condamner sans réserve ces changements en dénonçant l'abandon de toute perspective réelle de réforme, la non-viabilité des interventions individuelles ponctuelles et l'illégitimité des acteurs privés qui ne sont pas démocratiquement responsables ? Ou faut-il apprendre à vivre avec eux de manière rationnelle, en considérant que, même si les tuyaux ont été mal conçus au départ ou sont sur le point de se casser – pour reprendre la métaphore du plombier/de l'ingénieur – cela vaut toujours la peine de réparer les fuites ? La réponse à cette question (rarement énoncée) éclaire de nombreux désaccords autour des RCT, ainsi que les différentes positions que l'on trouve dans cet ouvrage.

Éthique et RCT

La question éthique est récurrente avec les RCT, non seulement dans le domaine du développement, mais aussi en général (surtout en médecine). Si tout le monde s'accorde sur la nécessité d'aborder cette question de front, du moins sur le principe, ces mises en garde n'ont pas encore été suivies d'effet (dans ce volume, Abramowicz et Szafarz, chap. 10 ; et aussi Ravallion, chap. 1 ; Ogden, chap. 4 ; Bédécarrats, Guérin et Roubaud, chap. 7 ; Picciotto, chap. 9 ; Patnaik, entretiens, ce volume). Parmi les *randomistas*, cette prise de conscience reste marginale¹⁴, comme si la foi dans les progrès scientifiques que les RCT peuvent apporter – et leurs retombées automatiques sur le plan des politiques et de l'amélioration du bien-être – suffisait à exempter les chercheurs de toute considération éthique.

14. Par exemple, parmi les 22 pages du chapitre « *Concerns about experiments* » de BANERJEE et DUFLO (2014), aucune n'aborde la question éthique, si ce n'est pour dire, en réponse au fait que la randomisation n'est pas une façon équitable de répartir le programme (car considérée comme un problème méthodologique, mais pas comme une question éthique), que « les responsables de la mise en œuvre pourraient trouver que la façon la plus simple pour [...] présenter [le programme] à la communauté est de dire que [son] expansion [...] est prévue pour les zones de contrôle de l'avenir » (*ibid.* : 101).

Alors que toute recherche soulève des questions éthiques, les RCT sont encore plus concernées que les études non expérimentales en raison de leur principe même (TEELE, 2014), car elles présentent généralement une forme de manipulation de l'environnement de recherche (elles « tirent la queue du lion », pour reprendre l'expression utilisée par DEATON et CARTWRIGHT [2018 : 18]).

Les analyses critiques font également abstraction de ces considérations éthiques. Elles se contentent souvent de faire allusion au problème en ne donnant que bien peu de détails. Le chapitre d'Abramowicz et Szafarz (chap. 10, ce volume) fait figure d'exception en explorant en profondeur les implications du principe d'*équipoïse*, c'est-à-dire l'exigence éthique selon laquelle une expérience impliquant des sujets humains doit révéler « un état de réelle incertitude de la part de l'investigateur clinique quant aux mérites thérapeutiques comparatifs de chaque bras de l'essai » (FREEDMAN, 1987 : 141, cité par Abramowicz et Szafarz, chap. 10, ce volume). Les auteurs se demandent pourquoi les expérimentateurs en économie ignorent presque systématiquement ce principe, alors qu'il s'agit d'un pilier essentiel de la science médicale. Ils fournissent une série de pistes pour répondre à la question. Ravallion aborde également ce sujet (chap. 1, ce volume), en insistant sur l'importance de bien évaluer les risques et les informations déjà disponibles, et en montrant que le principe d'*équipoïse* prend différentes formes selon les divers cas et types de randomisation (raisonnement inévitable des traitements, randomisation conditionnelle, et essais d'équivalence). Il évoque également la proposition d'*expérience adaptative* avancée par NARITA (2018) pour établir un optimum de Pareto entre les potentiels effets positifs et négatifs sur les participants en fonction des connaissances disponibles.

Cette quasi-négation des considérations éthiques par les *randomistas* est d'autant plus contestable qu'il existe différentes normes de bonnes pratiques, tant pour les RCT dans le domaine médical que pour la plupart des RCT en sciences sociales menées dans les pays du Nord. Les principes éthiques destinés à régir les essais randomisés impliquant des sujets humains ont été édictés dans des normes reconnues, en particulier la déclaration d'Helsinki en 1967 (WMA General Assembly, 2014), le rapport Belmont en 1974 et les lignes directrices internationales d'éthique du Conseil des organisations internationales et des sciences médicales (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979 ; Council for International Organizations and Medical Sciences, 2002). Ces normes établissent des principes clairs : le consentement éclairé, le principe « *do no harm* » (ne causer aucun préjudice), la fourniture d'une protection spécialement envisagée pour les populations vulnérables, l'analyse des risques et la surveillance réactive, pour n'en citer que quelques-uns.

Ces principes fondamentaux sont rarement respectés dans le domaine du développement (Abramowicz et Szafarz, chap. 10, ce volume). Comme BARRETT et CARTER (2014), nous détaillons quatre exemples de RCT qui illustrent les effets néfastes de cette négligence éthique. Dans le premier exemple, la RCT visait à démontrer les mécanismes de la corruption dans le cas de l'obtention du permis de conduire en Inde (BERTRAND *et al.*, 2010). L'un des bras du traitement consistait à offrir une prime aux candidats qui obtiennent leur permis. Barrett et Carter

montrent que cette RCT a enfreint le code éthique du « *do no harm* » (ils parlent même de « conception irresponsable de la recherche ») de deux façons : non seulement le dispositif encourageait la corruption, mais il mettait également en danger la vie d'autrui en laissant le volant à des conducteurs potentiellement imprudents, puisque l'expérience a montré que le groupe traité prenait moins de leçons de conduite. Le second exemple concerne une RCT mise en place au Kenya pour tester « l'effet Rockefeller » (qui affirme que trop de ressources nuisent plus qu'elles ne sont bénéfiques) *via* un projet d'assistance à des groupes de femmes (GUGERTY et KREMER, 2008). Les effets du projet se sont avérés négatifs (les femmes les plus pauvres ont été exclues des postes à responsabilité), confirmant ainsi l'hypothèse de Rockefeller. Cette RCT a causé un préjudice au sujet de l'expérience, un préjudice qui aurait pu être prévu, au moins comme étant une possibilité. Les femmes auraient donc dû être informées de cette éventualité afin qu'elles puissent décider de participer ou non à l'expérience (principe du consentement éclairé). Le troisième cas est une RCT impliquant des élèves de l'enseignement secondaire en République dominicaine. Il s'agissait de tester si des informations concernant les rendements de l'éducation sur le marché du travail, supérieurs à l'idée que s'en faisaient les élèves *a priori*, pouvaient les conduire à prolonger leurs études (JENSEN, 2010). Ici, le problème éthique est le suivant : les informations fournies aux élèves du secondaire (estimées à partir de données non expérimentales) – qui sont à la fois surestimées (étant donné les biais d'endogénéité) et calculées en moyenne sans tenir compte des caractéristiques des élèves et des écoles – ont probablement conduit certains des élèves, sans doute les plus pauvres, à « surinvestir » dans l'éducation en se basant sur les bénéfices attendus. Sans compter l'effet de l'augmentation de l'offre des diplômés susceptible de peser sur les rendements futurs (effet d'équilibre général). Enfin, le traitement dans le cadre de la quatrième RCT consistait à accorder des crédits aux personnes rejetées par un prestataire de microcrédit en raison de la forte probabilité de défaut de paiement identifiée par son modèle de notation (KARLAN et ZINMAN, 2009)¹⁵. Outre que cette stratégie faisait courir au groupe traité le risque d'être incapable de rembourser les prêts (avec les pénalités associées) et de se retrouver dans une situation potentielle de surendettement, le fait de ne pas l'en avoir informé est contraire au principe du consentement éclairé. Cette situation soulève un dilemme délicat puisque, si les participants avaient été informés du risque, leur comportement aurait probablement changé, compromettant ainsi la validité interne de la RCT. Aucune de ces failles n'a empêché ces quatre RCT d'être publiées dans des revues universitaires de premier plan, ce qui soulève également des questions de responsabilité des revues économiques dans le non-respect des normes éthiques (Abramowicz et Szafarz, chap. 10, ce volume).

D'autres exemples sont cités dans le présent ouvrage (Ogden, chap. 4 ; Abramowicz et Szafarz, chap. 10 ; Patnaik, entretiens, ce volume). Le recours

15. Une approche similaire consistant à inclure des sujets initialement jugés insolvable figure également dans AUGSBURG *et al.* (2015) ; elle est examinée par Bédécarrats, Guérin, et Roubaud, chap. 7, ce volume.

accru aux RCT, notamment par des institutions moins visibles et donc encore moins contrôlées sur le plan éthique, pourrait finir par ébranler les principes de base. Le cas d'une RCT en cours illustre ce point. Cette RCT, commandée par des bailleurs de fonds, a été mise en place pour tester la manière dont l'information affecte le comportement migratoire dans les zones rurales du Mali. Les participants ont visionné un court-métrage retenu parmi quatre films sélectionnés au hasard, illustrant différents scénarios de migration et de non-migration (vers l'Europe) : une migration réussie ; la souffrance, les mauvais traitements et en fin de compte une tentative de migration ratée ; une non-migration réussie ; et une comédie n'ayant rien à voir avec la migration à titre de placebo. Outre les problèmes liés à la communication aux participants des enjeux de ce test et à l'obtention de leur consentement éclairé, aucun des volets du test ne peut être considéré comme bénéfique pour les participants (violation du principe de *bienfaisance*). Les préférences des individus peuvent simplement être modifiées en fonction de ce que les commanditaires de la RCT considèrent comme étant bon pour eux (ou pour eux-mêmes). Le principe du « *do no harm* » n'est pas non plus respecté, car, après avoir visionné les films, certains participants pourraient décider d'émigrer au risque de mourir en Méditerranée ou d'être torturés dans les prisons libyennes. Enfin, la motivation politique du commanditaire semble évidente : freiner la migration africaine vers l'Europe. Il semble donc que cette RCT ait été conçue sans considération éthique sérieuse.

Au vu de la multitude d'exemples, il semblerait que la création de comités de protection des personnes dans de nombreuses institutions universitaires (*Institutional Review Boards*) n'ait rien fait pour remédier aux lacunes éthiques observées, ou du moins pas suffisamment (BARRETT et CARTER, 2020). On peut avancer deux raisons qui se renforcent mutuellement. La première est la difficulté de garantir simultanément la protection des sujets de l'expérience et la validité interne du protocole. La seconde est la compréhension imparfaite et le manque d'intérêt manifeste des *randomistas* pour le sujet. Face à ce que l'on pourrait appeler un dilemme éthique, ils penchent trop souvent du côté de l'impératif méthodologique. Pourtant, les garanties éthiques sont d'autant plus nécessaires dans les pays du Sud. Premièrement, le fait de ne pas informer les participants (principe du consentement éclairé), voire de mal informer délibérément les sujets humains pour assurer une stratégie d'identification théoriquement pure, est contraire au principe d'appropriation promu par les politiques de développement. Deuxièmement, les participants sont généralement des individus vulnérables, tant sur le plan économique (pauvres) que politique (sans voix), à qui il est plus facile d'imposer l'essai, quitte à les induire délibérément en erreur. Cette asymétrie est d'autant plus forte que les RCT sont de plus en plus souvent associées à des jeux de laboratoire grandeur nature, supervisés par de jeunes étudiants et des assistants de recherche rattachés à des universités de pays du Nord. Nous devons également nous pencher sur le choix de ces populations, notamment lorsque nous testons une hypothèse comportementale ou une théorie avancée par certains partisans des RCT (BANERJEE et DUFLO, 2011 ; voir Morduch, chap. 3, ce volume). Hormis l'argument selon lequel les pauvres des pays du Sud ont une rationalité spécifique, les arguments de

moindre coût et de moindre capacité à refuser de participer (problème récurrent avec les RCT menées dans les pays du Nord) en raison d'une méconnaissance de leurs droits et de rapports de force déséquilibrés (y compris vis-à-vis des expérimentateurs) semblent être des explications crédibles (Patnaik, entretiens, ce volume ; voir aussi TEELE, 2014), comme cela a déjà été observé lors de la « délocalisation » des essais cliniques médicaux (PETRYNA, 2007). Sans aller jusqu'à demander un « moratoire sur l'expérimentation » dans les pays du Sud (HOFFMANN, 2020), le sujet devrait au moins être traité en priorité.

Si les *randomistas* invoquent comme argument éthique l'amélioration à long terme du bien-être des populations grâce aux progrès scientifiques rendus possibles par les RCT, cette hypothèse est néanmoins loin d'être démontrée (Ravallion, chap. 1, ce volume). En définitive, outre la foi inébranlable dans la théorie de la technique au détriment de sa faisabilité (comme on l'a vu dans la section « L'épistémologie des RCT dans le domaine du développement »), il semble que trop souvent une hiérarchie de valeurs difficilement acceptable privilégie les résultats scientifiques plutôt que le bien-être des populations.

Quel est le champ d'application des RCT ?

Après avoir examiné de près les nombreuses limites des RCT, tant en termes de validité interne qu'externe, DEATON et CARTWRIGHT (2018) considèrent qu'elles restent néanmoins valables dans deux domaines : d'une part, tester une théorie ; d'autre part, évaluer ponctuellement et dans un contexte donné un projet ou une politique particulière, mais à condition que les problèmes potentiels de validité interne soient résolus et en tenant compte du fait que l'explication des résultats obtenus est souvent inappropriée. Les chapitres de ce livre confirment et développent cette analyse. Les évaluations randomisées ne sont possibles que pour un champ d'intervention très restreint, qui concerne le plus souvent des biens privés, marchands et humanitaires. Les RCT peuvent également être utilisées pour tester la théorie économique concernant les réactions comportementales aux interventions, et remettre ainsi en question certaines idées préconçues. Mais, en fin de compte, elles ne répondent ni à la question de l'*impact* tel qu'il est depuis longtemps défini dans le domaine de l'aide au développement, ni à la question de l'*explication* des effets mesurés.

Biens privés, marchands et humanitaires

Les conditions requises par les protocoles des méthodes randomisées les limitent à un spectre étroit que BERNARD *et al.* (2012) appellent les programmes « tunnels ». Ces derniers se caractérisent par des impacts à court terme, des *inputs* et *outputs* clairement identifiés, facilement mesurables, des liens de causalité unidirectionnels (A cause B), linéaires et enfin non soumis à des risques

de faible participation de la part des populations visées. Ils rejoignent les suggestions de WOOLCOCK (2013) : les projets qui se prêtent à la randomisation doivent avoir une « *low causal density* », être peu contingents aux capacités de mise en œuvre (*implementation capability*) et comporter des résultats prédictibles.

Ce type de méthode n'est donc applicable qu'à des interventions simples ou ponctuelles, de courte durée, ciblant des individus. Concrètement, ces micro-interventions concernent essentiellement des biens et services privés, c'est-à-dire rivaux et exclusifs (voir Ravallion, chap. 1 ; Pritchett, chap. 2 et Picciotto, chap. 9, ce volume).

Dans le domaine de la santé, elles concernent les actions de prévention et de traitement de certaines maladies. Il peut également être question de filtres à eau, de moustiquaires, de formations et de systèmes de primes pour les professionnels de la santé, de consultations gratuites, de conseils médicaux par *Short Message Service* (SMS) et de micro-assurances. Cependant, les RCT ne répondent pas à la question de la gestion des systèmes de santé, qui sont par nature complexes et systémiques, et qui impliquent une main-d'œuvre qualifiée et motivée, une infrastructure, la fourniture de médicaments, etc. (Garchitorena *et al.*, chap. 5, ce volume). En matière d'assainissement, ces micro-interventions touchent à la distribution, la construction et l'utilisation de latrines. Là encore, les RCT ne répondent pas à la question de la gestion des flux de déchets humains : quel type de réseau d'assainissement ou de nettoyage utiliser, quel genre d'infrastructure, et quel type de régulation (Spears, Ban et Cumming, chap. 6, ce volume). En matière de réduction de la pauvreté, ces micro-interventions ont trait au microcrédit, à l'épargne, à la formation à l'entrepreneuriat et aux services d'éducation financière. Une fois de plus, les RCT ne répondent pas à la question des processus de création de richesses régionales ou sectorielles (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume), ni à la question plus large de l'accès aux services de base (Pritchett, chap. 2, ce volume). Dans le domaine de la gouvernance des administrations et des institutions publiques, ces micro-interventions prennent la forme d'inspections ponctuelles, de mesures d'incitation financière, d'audits par des tiers indépendants, de centres d'appels et de retours téléphoniques. Les RCT ne répondent pas à la question de la faible capacité de l'État, des bureaucraties centralisées souffrant d'un manque de confiance, des ressources limitées, des bureaucrates surchargés et des environnements de travail difficiles (Natarajan, entretiens, ce volume).

Contrairement à certaines analyses critiques (voir, par exemple, BERNDT, 2015), les conclusions des RCT ne prônent pas nécessairement la marchandisation des biens privés (ce qui les rapproche davantage de la mouvance humanitaire). Dans le cas des moustiquaires imprégnées d'insecticides et des traitements vermifuges à forte élasticité de prix, les RCT ont précisément plaidé en faveur d'une distribution gratuite, considérée comme plus efficace que de les rendre payantes, ce qui remet en cause la croyance populaire du monde de la santé. Dans le cas du microcrédit, les RCT ont conclu que son impact sur la réduction de la pauvreté reste marginal et que la réduction de la pauvreté nécessite donc d'autres types d'interventions (BANERJEE *et al.*, 2015c). Toujours dans le cas du microcrédit,

les RCT ont montré que les pauvres sont sensibles aux taux d'intérêt, en contredisant là aussi l'idée largement répandue selon laquelle l'accès est plus important que le coût, une croyance populaire parmi les organismes de microfinance et leurs financeurs, qui est invoquée pour légitimer des taux d'intérêt élevés (Morduch, chap. 3, ce volume).

Bien que ces résultats puissent être utiles, les sujets abordés restent limités par rapport à l'ensemble des questions de développement, de pauvreté et d'inégalités. Les conditions requises pour la mise en œuvre des RCT excluent donc un grand nombre de politiques de développement qui mettent en jeu des combinaisons de mécanismes socio-économiques et des boucles de rétroaction (effets d'émulation, d'apprentissage des bénéficiaires, d'amélioration de la qualité des programmes, effets d'équilibre général, etc.). C'est précisément le cas des biens publics (Ravallion, chap. 1, ce volume). Lorsque les interventions concernent des infrastructures et des systèmes de régulation, la manipulation expérimentale est impossible (Spears, Ban, and Cumming, chap. 6, ce volume).

Dans les termes de référence d'une étude commanditée sur le sujet, certains responsables du Department for International Development (DFID) estimaient ainsi le champ d'application des RCT à moins de 5 % des interventions de développement (DFID, 2012). S'il convient de ne pas prendre ce chiffre au pied de la lettre, il ne fait aucun doute que les méthodes expérimentales ne sont pas adaptées pour évaluer l'impact de la grande majorité des politiques de développement. Dans leur papier plus formalisé, PRITCHETT et SANDEFUR (2013b) aboutissent à des conclusions similaires¹⁶. Garchitorena *et al.* (chap. 5, ce volume) soulignent que 97 % des financements de la recherche en santé dans le monde sont consacrés au développement de nouvelles technologies (principalement pharmaceutiques), et que seuls les 3 % restants sont consacrés à la recherche sur la *mise en œuvre*, pourtant essentielle pour comprendre et améliorer les dysfonctionnements des systèmes de santé.

Évaluer l'impact ou tester le comportement ?

Comme le suggère Morduch (chap. 3, ce volume), les RCT visent en fait deux objectifs : mesurer l'impact et explorer « la nature des contrats économiques, des comportements et des institutions ». Il ajoute que ce deuxième type de « RCT exploratoire », moins controversé, est finalement le plus prometteur, puisqu'il représente un gain réel par rapport aux autres méthodes et donc un meilleur potentiel en termes de développement des connaissances.

Avec ce deuxième type de RCT, on ne cherche plus tant à mesurer l'impact des interventions caractéristiques de l'action publique ou de l'aide au développement, mais à expérimenter différents modes d'intervention et à mesurer des résultats en termes de *take up* de l'intervention. Selon Morduch (chap. 3, ce volume), ce

16. « Le champ d'application de l'approche de la "planification avec des preuves rigoureuses" en matière de développement est extrêmement limité » (PRITCHETT et SANDEFUR, 2013b : 1).

type de RCT est une source d'information, voire de « provocation », qui permet de remettre en question certains malentendus en matière d'économie du développement (comme la faible élasticité de la demande de microcrédit par rapport au prix, mentionnée plus haut) et de tester des innovations et la manière dont les personnes y réagissent. Par exemple, il peut permettre de tester différents calendriers de vente d'assurance pour la récolte afin de mieux comprendre les contraintes de temps et de liquidité ; ou de tester le rôle de l'information et de l'assistance dans l'utilisation des téléphones portables par les plus pauvres pour mieux connaître les mécanismes de partage au sein des ménages.

Si ces objectifs sont utiles et louables (à condition que les critères éthiques et de validité interne soient respectés et que les conclusions soient valables), on peut néanmoins se demander pourquoi les *randomistas* persistent à parler d'impact alors que de nombreuses RCT sont en fait de nature plus « exploratoire » et consistent à comparer différentes modalités d'une seule et même intervention, en se contentant souvent de mesurer les écarts de *take up*. Lorsqu'on analyse le secteur de l'assainissement, la conclusion est la même : les RCT semblent plus à même d'analyser les changements de comportement que de mesurer leur impact en tant que tel (Spears, Ban et Cumming, chap. 6, ce volume).

De fait, la question de l'impact reste souvent sans réponse. Depuis 1992, la plupart des acteurs du secteur de l'aide au développement s'appuient sur cinq critères définis par le Comité d'aide au développement de l'OCDE [Organisation de coopération et de développement économiques] (Development Assistance Committee, 2010), parmi lesquels figure celui de l'impact : « Effets à long terme, positifs et négatifs, primaires et secondaires, induits par une action de développement, directement ou non, intentionnellement ou non. » Les RCT ne peuvent cependant évaluer que l'*impact à court terme de chaînes causales courtes*. Il ne s'agit donc pas à proprement parler d'un impact tel que défini ci-dessus (voir également Picciotto, chap. 9, ce volume). Si l'on prend l'exemple des moustiquaires imprégnées d'insecticide, souvent considérées comme le fleuron des RCT (Ogden, chap. 4, ce volume), les RCT portent généralement sur le *take up* plutôt que sur l'impact, puisque les moustiquaires imprégnées d'insecticide sont considérées comme étant fondamentalement une « bonne chose ». Leurs effets à moyen et long terme sont pourtant controversés, puisque les moustiques ont fini par s'adapter génétiquement et que les systèmes de production locaux ont été détruits (BEISEL, 2015). L'omission des effets à long terme et des effets collatéraux est tout aussi problématique dans le secteur du microcrédit (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume).

Ce type de RCT renvoie en fin de compte à la notion de « marketing social », terme très en vogue dans le monde du développement, qui rejoint tout à fait naturellement la circulation des biens privés, marchands et humanitaires, ainsi que la tendance comportementaliste décrites ci-dessus. Le marketing social consiste à appliquer des outils et des principes commerciaux à la conception, à la mise en œuvre et à l'évaluation de programmes de changement de comportement en vue d'obtenir des avantages individuels et de servir l'intérêt public (FRENCH *et al.*, 2010). Inspirées des sciences du comportement, les techniques

de marketing social comprennent les « *nudges*¹⁷ », mais aussi des méthodes de marketing plus classiques (emballage, prix, identification des canaux et des lieux de distribution les plus appropriés, etc.) Le marketing social a vu le jour dans les années 1970 dans les secteurs sanitaire et social, notamment dans les pays du Sud. Il a aussi été introduit dans des domaines tels que la santé reproductive, la prévention du sida, la réhydratation en cas de diarrhées, l'assainissement, avant de s'étendre pour cibler les changements de comportement dans un grand nombre de secteurs (environnement, agriculture, éducation, gestion financière, consommation, etc.)

Mesurer versus expliquer

Si les RCT peuvent être capables de mesurer et de tester certains impacts et aspects des interventions, ils ne permettent toutefois pas d'analyser leurs *mécanismes*, ni leurs *processus* sous-jacents. Lorsque les causes de la pauvreté sont analysées sous l'angle des manques, comme c'est le cas dans les approches randomisées, la question des processus et des mécanismes n'est pas abordée (SHAFFER, 2015). Pour pallier cette lacune de la théorie probabiliste de la causalité, il faudrait établir un « modèle causal » (CARTWRIGHT, 2010), une théorie cohérente du changement (WOOLCOCK, 2013), une approche structurelle (ACEMOGLU, 2010) et une évaluation contextuelle de l'intervention (RAVALLION, 2009a, et chap. 1, ce volume ; PRITCHETT et SANDEFUR, 2015).

Face à cette critique, les *randomistas* fondent désormais plus souvent leurs résultats sur des théories explicites du changement (Ogden, chap. 4, ce volume), basées en grande partie sur l'économie comportementale. Cette dernière permet de décoder toute la complexité des processus psychologiques et cognitifs, des luttes intérieures des individus, mais aussi de comprendre les multiples pratiques de « comptabilité mentale » (THALER, 2015), ainsi que d'explorer et de tester la manière dont les comportements réagissent aux différentes interventions (voir la section « Évaluer l'impact ou tester le comportement ? »). Cependant, l'économie comportementale ne peut pas saisir la complexité des comportements atypiques, inattendus et « sous-optimaux », ce dernier terme impliquant qu'il faudrait définir ce qui relève de la norme. Il convient ici de distinguer deux niveaux : celui des comportements des individus, qui sortent parfois du cadre comportemental (SERVET, 2018 ; SERVET et TINEL, 2020), et celui des interventions, qui se déroulent rarement comme prévu.

En ce qui concerne le comportement des individus, ces derniers ne peuvent être réduits à de simples populations cibles, puisqu'ils sont des êtres sociaux et pluriels. Une personne ne se limite pas à sa capacité de refuser ou d'accepter, ni d'ailleurs à ses « biais » cognitifs ou sociaux. La rationalité et les motivations d'une population locale donnée sont le fruit d'une construction : elles découlent

17. En économie comportementale, un « *nudge* » (« coup de pouce » en français) désigne un dispositif de petite taille et peu coûteux qui vise à influencer le comportement des personnes de manière prévisible, sans qu'il ne s'agisse d'une obligation ni d'une interdiction formelle.

des normes et des réalités sociales et politiques et en sont le reflet. Elles renvoient à des formes préexistantes d'interdépendance, d'équilibre des pouvoirs et de structures sociopolitiques, mais aussi à des désirs et des aspirations. Si la dimension sociale est si imprévisible, elle ne doit pas pour autant être considérée comme un obstacle et une contrainte à éliminer à coups de « *nudges* ». Les populations locales ont parfois de bonnes raisons d'agir comme elles le font, surtout lorsque l'environnement global n'évolue pas. En effet, elles ont leurs propres conceptions et représentations du monde (et leurs propres théories du changement), ainsi que leurs propres connaissances et savoir-faire en matière de soins, de maladie et de bien-être, de propreté et de saleté, de finances, de pauvreté et de richesse, etc. Si certaines de ces représentations sont source de discrimination, il n'en demeure pas moins qu'elles façonnent les comportements. Elles reflètent par ailleurs des visions du monde particulières, qui ne sont pas nécessairement moins « optimales » que celles des chercheurs (voir l'exemple du microcrédit avec Bédécarrats, Guérin et Roubaud, chap. 7, ce volume).

Les interventions sont également complexes, puisqu'elles combinent plusieurs niveaux et acteurs. Les réalités locales façonnent, encadrent, contraignent et influencent les interventions (MOSSE, 2004 ; OLIVIER DE SARDAN, 1995). C'est du moins le cas pour les trois secteurs représentés dans cet ouvrage. Dans le domaine de la santé mondiale, par exemple, « l'une des questions fondamentales [...] consiste à savoir pourquoi les technologies connues – celles qui ont fait leurs preuves dans certains contextes – ne parviennent pas systématiquement à toucher les personnes auxquelles elles sont destinées » (Garchitorea *et al.*, chap. 5, ce volume). Pour y répondre, il faut forcément s'intéresser au fonctionnement des « systèmes » : que ce soit les systèmes de santé locaux, les systèmes d'organisation, les particularités des interactions entre les populations « cibles » et les prestataires de soins, etc. De même, l'assainissement et le microcrédit ne revêtent pas le même sens selon les personnes et englobent d'innombrables réalités, méthodes et formes de mise en œuvre. Cette diversité limite considérablement les possibilités de généralisation des RCT (Spears, Ban et Cumming, chap. 6 ; Bédécarrats, Roubaud et Guérin, chap. 7, ce volume). Ces trois secteurs ne sont sans doute pas les seuls à présenter une telle complexité et diversité.

Pourquoi une controverse scientifique est-elle nécessaire et pourquoi n'a-t-elle pas eu lieu ?

Comme nous l'avons vu, les RCT sont loin de faire l'unanimité et suscitent de nombreux débats et critiques. Alors qu'elles auraient dû déclencher une controverse scientifique, vitale pour le progrès scientifique et le débat démocratique, celle-ci n'a pas (encore) eu lieu. Comment l'expliquer ? Sans prétendre couvrir

le sujet de manière exhaustive, et compte tenu de son importance, il nous a semblé utile d'esquisser quelques pistes d'analyse empruntées au champ de l'épistémologie des sciences.

Contrairement à une vision naïve de la science, il faut garder à l'esprit que le progrès scientifique n'est pas toujours un processus rationnel et linéaire, où les méthodes et les résultats les plus efficaces et les plus utiles l'emportent systématiquement sur les autres, et où les connaissances attestées font l'objet d'un consensus. Les connaissances scientifiques sont également le fruit de l'histoire, de la société et de la politique, forgées par des succès et des échecs, des cycles, des débats et des désaccords, lesquels débouchent parfois sur des *controverses*, c'est-à-dire des divergences entre plusieurs parties exposées et débattues sur la scène publique.

Les controverses scientifiques ne doivent donc pas être vues d'un mauvais œil, comme étant la manifestation d'erreurs de raisonnement (où le « vrai » l'emportera finalement sur le « faux ») ou d'une ingérence indésirable de la politique ou d'intérêts autres que le progrès des connaissances (un domaine censé être exempt de toute subjectivité). Les controverses sont inhérentes à la production collective de connaissances, et permettent souvent l'émergence de progrès scientifiques majeurs. Tous les domaines scientifiques sont marqués par de grandes controverses parfois violentes (mémoire de l'eau, organismes génétiquement modifiés [OGM], « scandale de l'Eldorado », ondes gravitationnelles), mais parfois aussi étouffées dans l'œuf (CALLON, 2006a).

Une controverse peut être définie comme un différend opposant deux camps et prenant à témoin un public composé de pairs scientifiques ou bien plus large (LEMIEUX, 2007). Si les divergences s'avèrent parfois virulentes, les participants sont néanmoins tenus de respecter les conventions du monde académique, notamment le principe d'égalité entre les participants, l'importance du raisonnement logique, la maîtrise de l'agressivité, et le respect du principe de la dignité des intervenants. Ces conventions restent cependant floues, et lorsqu'un adversaire est accusé d'abuser d'une position dominante ou de manquer de civisme, c'est souvent un moyen de renverser le rapport de force ou de disqualifier son rival.

Comme dans de nombreux domaines de la sociologie, la manière d'aborder les controverses diffère selon les écoles. Alors que certaines privilégient la logique et les preuves (RAYNAUD, 2018), d'autres se concentrent sur les croyances, les conventions sociales ou les rapports de force qui influent sur le contenu des arguments et l'arbitrage entre les logiques rivales (AKRICH *et al.*, 2013). Mais, d'après les études scientifiques, quelle que soit l'école, le fait de contester les processus reflète une réalité sociale et historique, et met en lumière les rapports de force, les positions institutionnelles et les réseaux sociaux. Les controverses font avancer nos sociétés en modifiant les rapports de force, en redistribuant le prestige et les ressources, et en créant de nouvelles normes qui limiteront les actions et les positions futures (LEMIEUX, 2007).

Pour en revenir à notre question – pourquoi la controverse n'a-t-elle pas eu lieu ? –, les interprétations conceptuelles développées par CALLON (2006a ;

2006b) fournissent quelques éléments de réponse. Premièrement, ce qui justifie ou non qu'une controverse ait lieu fait toujours l'objet d'accords négociés dans le cadre des processus de contestation. Dans notre cas, la communauté professionnelle des économistes du développement privilégie ce qu'elle estime relever de la recherche fondamentale, notamment la pureté statistique des essais randomisés et la maîtrise des biais d'identification des causes. Cet aspect prime ici sur des considérations, jugées secondaires par la communauté professionnelle, qui relèvent davantage du domaine appliqué et qui impliquent la reconnaissance des différentes « astuces du métier », des tactiques et des mises au point nécessaires à la mise en pratique de la méthode, mais aussi la reconnaissance du rôle des expérimentateurs et des sujets de l'essai, ainsi que des groupes de participants qu'il implique (KABEER, 2019 ; BÉDÉCARRATS *et al.*, 2021). Cela nous ramène aux différences épistémologiques évoquées plus haut. La naissance d'une controverse implique donc la mise en place de forums suffisamment structurés pour que des discussions approfondies puissent avoir lieu. Sans ces espaces de débat, il est difficile de savoir qui parle et dans quel contexte, les mêmes acteurs pouvant défendre une argumentation partielle dans certains forums et, sans jamais se rétracter, faire des déclarations beaucoup plus équilibrées et prudentes dans les forums d'experts.

Éviter la controverse tout en écoutant et en s'adaptant

L'absence de dialogue public n'empêche pas les *randomistas* d'adapter leurs méthodes et pratiques (Ogden, chap. 4, ce volume), même si les réponses varient selon les groupes de chercheurs. Certains mettent leurs données à disposition, encourageant ainsi leur reproduction, pendant que d'autres reconnaissent le bien-fondé du pluralisme méthodologique et combinent les RCT avec d'autres méthodes. Alors que certains se concentrent en détail sur les mécanismes et les processus d'impact et utilisent des théories spécifiques (basées principalement sur l'économie comportementale), d'autres s'attaquent à la question de la validité externe et multiplient les études de cas dans différents contextes (le numéro spécial sur le microcrédit édité par BANERJEE *et al.* [2015c] en est un exemple typique ; Bédécarrats, Guérin et Roubaud, chap. 7, ce volume), ou réanalysent *ex post* un certain nombre de RCT (MEAGER, 2019). D'autres encore s'intéressent sérieusement à la question du « penser petit » et se concentrent sur les programmes à grande échelle et les politiques nationales. Face à la faible influence sur les politiques publiques (Pritchett, chap. 2, ce volume), certains *randomistas* créent des organes dédiés, ou deviennent parfois eux-mêmes décideurs.

Reste à voir dans quelle mesure la *mise en œuvre* de cette nouvelle génération de RCT dans l'économie du développement peut faire face aux contingences du terrain et évaluer réellement des interventions plus complexes. Au risque de nous répéter, il convient de souligner cette obsession pour le protocole, jugé plus important que sa faisabilité et ses enjeux éthiques, et qui constitue l'un des points cruciaux du débat. Or, plus les programmes et les politiques

étudiés sont compliqués, plus il y a de risques que le protocole initial soit ajusté et que des compromis soient faits. Il ne s'agit pas seulement de revoir la technique, mais de se défaire d'une position épistémologique scientifique telle que définie ci-dessus.

La controverse peut-elle vraiment être évitée compte tenu des effets d'éviction ?

Si la controverse est indispensable, c'est aussi parce que la prétendue hiérarchie des méthodes produit des effets d'éviction, tant sur le plan des méthodes elles-mêmes (les autres étant discréditées) que sur celui du financement et des types d'interventions, avec par conséquent une dimension performative : le succès des RCT transforme le domaine du développement.

Pour illustrer cela, nous allons examiner deux exemples sur la question du financement. En Inde, une étude permettant d'évaluer véritablement l'impact de l'assainissement sur la mortalité infantile (l'indicateur le plus approprié, mais que les RCT ne permettent pas de mesurer statistiquement) coûterait environ 90 millions de dollars (sous certaines conditions ; Spears, Ban, et Cumming, chap. 6, ce volume). Une RCT classique coûte entre 500 000 et 1 500 000 dollars¹⁸, et chaque RCT fait généralement l'objet d'une seule publication de recherche. Est-ce bien raisonnable, lorsqu'on sait que cette même somme permettrait de financer le système d'enquête statistique auprès des ménages d'un pays pauvre, et de donner lieu à une multitude d'études issues de ces données non expérimentales ? C'est l'une des questions cruciales que pose Patnaik (entretien, ce volume).

Le domaine de la santé illustre particulièrement bien les effets performatifs des RCT. Sans en être la raison première, les RCT ont contribué à l'essor des approches verticales dans le domaine de la santé (projets en silos). Axées sur le traitement individuel de maladies spécifiques, elles ont pris le pas sur les approches horizontales qui, elles, visent à développer des systèmes de santé complexes et intégrés (Garchitorena *et al.*, chap. 5, ce volume). D'autres études mettent en évidence les effets performatifs (et problématiques) du recours croissant aux RCT (ADAMS, 2016 ; BIEHL *et al.*, 2014) : programmes non randomisables délaissés, modification des programmes pour les rendre plus facilement randomisables, priorité donnée à l'évaluation plutôt qu'à l'intervention elle-même (notamment en modifiant le travail du personnel de terrain : ADAMS, 2016). Les perturbations causées par les RCT et affectant la qualité des interventions ont été documentées dans d'autres domaines, tels que le microcrédit (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume) et la micro-assurance (QUENTIN et GUÉRIN, 2013).

18. Il n'existe à notre connaissance aucune estimation précise du coût des RCT, mais PAMIÉS-SUMNER (2015) donne quelques approximations. Voir également Ravallion, chap. 1, ce volume.

Quelles sont les alternatives de recherche ?

Notre objectif n'est pas de rejeter les RCT, car elles constituent une méthode prometteuse... parmi d'autres. Cependant, elles doivent être menées dans les règles de l'art, prendre au sérieux la question de la faisabilité et des enjeux éthiques en s'alignant sur les bonnes pratiques du monde médical, et se conjuguer à d'autres méthodes. Si les RCT conviennent à certaines politiques précisément définies, d'autres méthodes peuvent et doivent être utilisées, comme le montrent plusieurs chapitres de ce livre. Pour les projets que les RCT peuvent (en partie) couvrir, il convient de combiner les méthodes.

Une alternative à l'étalon-or consiste à adopter une approche pragmatique en définissant les questions de recherche et les outils méthodologiques requis au cas par cas en fonction des connaissances préalables disponibles, de la conception de l'intervention et des particularités des contextes, en liaison avec les différentes parties prenantes, qu'il s'agisse des opérateurs de terrain, des bailleurs de fonds, des gouvernements ou des populations locales largement négligées.

Ces méthodes alternatives s'appuient également sur une série de méthodes fondées sur l'interdisciplinarité et la reconnaissance des différentes manières de produire des preuves, tant quantitatives que qualitatives. Ces approches ne visent pas à fixer des lois universelles, mais à expliquer les liens de causalité spécifiques à un moment et à un lieu particuliers. Si le recours aux méthodes mixtes est souvent préconisé, que ce soit par les chercheurs¹⁹ ou les institutions (voir Rioux, entretiens, ce volume)²⁰, on remarque qu'il y a un décalage avec la pratique, où cela se fait finalement peu. Alors que certains *randomistas* reconnaissent publiquement la légitimité des méthodes alternatives (Ogden, chap. 4, ce volume), ils ne tiennent souvent pas compte des résultats des méthodes non randomisées, ce qui semble contredire leur apparente ouverture d'esprit (Bédécarrats, Guérin et Roubaud, chap. 7, ce volume).

Dans le domaine de la santé mondiale, les interventions sont d'une telle complexité que la randomisation est souvent impossible et laisse donc place à des méthodes non expérimentales et quasi expérimentales, plus appropriées. Comme le montrent Garchitorena *et al.* (chap. 5, ce volume), il existe de nombreux exemples de méthodes alternatives et complémentaires aux RCT, même si ces dernières restent utiles pour certaines interventions spécifiques. Ces méthodes alternatives présentent quelques particularités : elles reposent sur la théorie de la complexité (avec un système de santé global, plutôt que des

19. Voir, par exemple, les deux ouvrages mentionnés dans l'introduction des éditeurs (COHEN et EASTERLY, 2010 ; TEELE, 2014), dont la plupart des chapitres et des déclarations liminaires insistent sur la nécessité de recourir à des méthodes mixtes. Voir également CAMFIELD et DUVENDACK (2014).

20. Voir Picciotto, chap. 9, ce volume, à propos du monde de l'évaluation dans le développement. Voir aussi PAMIÉS-SUMNER (2015) pour l'AFD, et les travaux du Centre of Excellence for Development Impact and Learning (CEDIL) (WHITE et MASSET, 2018) pour le DFID.

composantes fragmentées), combinent des méthodes et des échelles d'analyse, puisent autant que possible dans les systèmes statistiques nationaux et s'intéressent non seulement à l'impact, mais aussi à l'efficacité (en introduisant dans l'analyse des réalisations [*outputs*] et des résultats [*outcomes*], mais aussi des moyens [*inputs*] et des processus).

Outre les exemples cités dans le livre, nous soulignons également la nécessité de mener des méta-analyses et des répliques. Elles commencent à faire leur apparition en économie du développement, mais sont encore trop rares (CAMFIELD et DUVENDACK, 2014). Ces répliques peuvent également être qualitatives et consistent à revisiter un terrain, comme cela a été fait au Maroc et au Bangladesh (KABEER, 2019 ; MORVANT-ROUX *et al.*, 2014). Les méthodes qualitatives (entretiens semi-structurés, groupes de discussion, observation participante, ethnographie, études de cas, monographies, etc.) peuvent servir plusieurs objectifs : contextualiser les interventions, élaborer des hypothèses originales, identifier des phénomènes nouveaux et inattendus et analyser les interventions dans leur ensemble, en étudiant la complexité des liens de causalité et les nombreuses interactions dynamiques et contradictoires entre différentes entités de manière spécifique au lieu. Lorsque l'on est confronté à des chaînes causales complexes, ce qui est le cas de nombreuses interventions, les méthodes qualitatives sont souvent la seule façon de traiter réellement l'épineuse question de la causalité (WHITE et MASSET, 2018). Pourtant, elles sont souvent critiquées (à tort) pour leur incapacité à « démontrer » les résultats, et sont aussi utilisées de manière superficielle et non rigoureuse. Labrousse (chap. 8, ce volume) illustre cette dérive en évoquant le *storytelling* – un type de récit destiné à étayer un argument, mais qui n'a aucun pouvoir de démonstration – que certains *randomistas* utilisent à mauvais escient pour interpréter des résultats quantitatifs. La seule norme valable est finalement celle qui fait « bon usage des bonnes preuves » (Spears, Ban, and Cumming, chap. 6, ce volume).

Pour résumer et conclure cette introduction, nous pensons que certains principes clés devraient guider la recherche pour le développement, non pas en tant qu'alternatives aux RCT, mais en leur accordant une place à proportion congruente. Ces principes ne sont probablement pas révolutionnaires... mais leur mise en œuvre serait déjà un grand pas pour les sciences humaines et sociales. Premièrement, et pour passer du général au spécifique, la recherche doit être guidée par les grandes questions à traiter plutôt que par les méthodes pour lesquelles il faut trouver des applications. Pour paraphraser une citation célèbre : « Ne vous demandez pas ce que vous pouvez faire avec une RCT, mais demandez-vous plutôt ce qu'une RCT peut faire pour votre recherche ! » Deuxièmement, nous devons dépasser l'obsession de l'impact causal²¹, qui domine la communauté des économistes du développement depuis la fameuse révolution de la crédibilité (ANGRIST et PISCHKE, 2010). Les analyses de données non expérimentales, les descriptions fouillées (*thick description*), les récits analytiques (*analytical narra-*

21. RUHM (2019) qualifie cette obsession de « police de l'identification », qu'il suggère de « mettre aux fers ».

tives) sont autant d'autres questions et d'approches de recherche au moins aussi importantes pour faire progresser les connaissances, surtout si l'on considère que la pauvreté n'est pas seulement un problème de privation, mais aussi et surtout le résultat de rapports sociaux et de pouvoir. Troisièmement, en ce qui concerne les approches quantitatives, il est indispensable de rééquilibrer les efforts de recherche pour englober d'autres composantes de la chaîne analytique : les gains éventuels en termes d'attribution causale (en théorie, puisque différents chapitres de ce livre montrent que rien n'est garanti en pratique dans ce domaine) et le surinvestissement dans ce domaine ont relégué au second plan d'autres aspects tout aussi importants. Il y a tout d'abord la question de la qualité des données, trop souvent sacrifiée par manque d'intérêt et de compétence, ce que certains des plus grands *randomistas* commencent à reconnaître (DILLON *et al.*, 2020). Ensuite, il y a la multiplication des répliques qui s'attaquent de front au diagnostic méticuleux des données, et à leur inclusion dans les critères d'évaluation des revues académiques des pairs. Dans le même temps, il convient d'accorder une plus grande attention à la question des plans d'échantillonnage, car, s'ils sont complexes, leurs conséquences sont trop souvent négligées. Ces négligences entraînent la sous-estimation de la variance des estimateurs et la prise en compte d'impacts jugés statistiquement significatifs, alors qu'ils ne le sont pas (GIBSON, 2019) pendant que d'autres le sont, mais que les RCT, pas toujours performantes, ne parviennent pas à identifier. Quatrièmement, il est temps de mettre réellement en pratique deux recommandations approuvées par tous, mais qui demeurent pour l'instant des paroles en l'air sans aucun résultat tangible : une prise en compte réelle des questions éthiques et la combinaison de méthodes qualitatives et quantitatives²². Le plaidoyer pour les méthodes mixtes et autres approches (MMA) est un vrai sport de combat (*Mixed Martial Sports*²³) ! Enfin et surtout, il est temps de reconnaître une fois pour toutes que les expérimentations randomisées ne sont pas l'étalon-or de l'évaluation. L'orgueil démesuré qu'affiche une partie du mouvement pro-RCT pousse la recherche dans une impasse. C'est pourquoi modérer cette démesure s'avère essentiel et ne peut qu'être bénéfique. Il est également crucial de tirer les leçons du passé et de tenir compte des recherches antérieures, au moins à deux égards : les faiblesses des RCT (Heckman, chap. 12, ce volume) et les résultats des méthodes autres que les RCT. Autrement, toutes les tentatives évoquées jusqu'ici visant à écouter les critiques et à s'adapter sont vouées à l'échec, ou, pour paraphraser la célèbre expression de LAMPEDUSA (1960) dans son roman, *Le Guépard*, « pour que tout reste comme avant, il faut que tout change ».

La consécration du prix Nobel d'économie conduira-t-elle les *randomistas* à mieux apprécier les avantages des différentes méthodes ou, au contraire, vont-ils

22. Comme le soulignent van der Meulen Rodgers *et al.* (2020) dans leur éditorial du numéro spécial du *World Development* sur les RCT, les contributeurs sont très nombreux à réclamer une triangulation, un pluralisme et une collaboration (tant au sein de la communauté scientifique qu'entre les universités, les bailleurs de fonds et la société civile).

23. Il s'agit d'un clin d'œil au documentaire réalisé par Pierre Carles en 2001 sur les travaux du sociologue Pierre Bourdieu, intitulé *La sociologie est un sport de combat*.

en profiter pour consolider leur position déjà quasi hégémonique ? Seul l'avenir le dira²⁴, mais nous tenons à souligner que la fin de l'étalon-or et de la quête de l'« incontestable », caractéristiques de la supériorité revendiquée par les *randomistas*, appelle une rupture épistémologique, mais aussi l'avènement de cette controverse que nous souhaitons tant. En se basant sur l'examen de controverses autour du changement climatique, LATOUR (2012) préconise de construire des espaces de débat ainsi que des méthodes pour discuter et débattre des différentes formes de connaissances scientifiques (dans toute leur diversité), et non scientifiques, en veillant à ce que les fondements idéologiques et politiques de ces multiples formes de savoirs ne soient ni rejetés ni occultés, mais explicités et débattus (EGIL, 2015). Nous pensons que ce projet, aussi ambitieux soit-il, est absolument nécessaire sur les plans scientifique et démocratique, si l'on veut véritablement améliorer les politiques de développement.

D'ailleurs depuis le début de l'année 2020, la pandémie de Covid-19, dont le monde n'est toujours pas sorti deux ans plus tard à l'heure d'écrire ces lignes, est venue éclairer d'un regard neuf le débat global autour des RCT, apportant son lot d'éléments nouveaux pour alimenter la plus que jamais nécessaire controverse scientifique sur le sujet. Les malheurs du monde offrent une extraordinaire « expérience naturelle » pour en apprécier la véritable contribution. Alors que la pandémie constitue le choc de pauvreté le plus brutal jamais connu à l'échelle mondiale en temps de paix, et que les RCT ont été primées pour leur contribution majeure à sa réduction, quel a été leur apport pour lutter contre cette catastrophe ? De plus, la pandémie de Covid-19, dans sa double dimension sanitaire et économique, ouvre une voie royale pour apprécier les mérites respectifs des RCT dans le domaine du développement et des essais cliniques (par exemple sur les vaccins). Par ailleurs et loin des fureurs du monde, l'attribution en octobre 2021 du prix Nobel d'économie à Joshua Angrist, Guido Imbens et David Card semble apporter de l'eau au moulin à l'obsession expérimentale, même si leur approche de l'inférence causale apparaît moins étriquée que celle des *randomistas*, en redonnant leur lettre de noblesse aux « expériences naturelles ». Il n'est pas question de tirer les enseignements de ce méga-épisode, en laissant la question ouverte, offerte à la sagacité du lecteur, et peut-être l'objet d'un second tome à venir, mais simplement d'annoncer qu'il ne fait que conforter les conclusions de cet ouvrage, tout en l'éclairant d'un jour nouveau.

24. Si l'on se base sur notre propre expérience, nous avons bien des raisons d'être pessimistes. En effet, il est de plus en plus difficile de publier des articles critiques dans les grandes revues universitaires et de trouver des interlocuteurs pour débattre de la contribution effective des RCT dans le domaine des politiques politiques. Si certains sont tentés d'émettre des opinions critiques, ou nuancées, vis-à-vis de la puissante et fameuse nouvelle doxa, ils se gardent bien de le faire, par autocensure ou pour des raisons institutionnelles.

Grandes lignes de l'ouvrage

Le livre est structuré comme suit. La première série de chapitres offre un aperçu des RCT dans le domaine du développement (à quels types de questions permettent-elles de répondre ou pas ?) ainsi que divers points de vue sur le potentiel de cette méthode (**Partie 1. Que peuvent les RCT ?** : chap. 1-4). La deuxième série de chapitres (**Partie 2. Perspectives sectorielles**) porte sur les analyses sectorielles dans le domaine de la santé (chap. 5), de l'assainissement (chap. 6) et du microcrédit (chap. 7), et pose les questions suivantes : quels enseignements avons-nous tiré des RCT dans chaque secteur et quelle est la contribution des autres méthodes ? La troisième série de chapitres (**Partie 3. Économies politiques**) propose des pistes de réflexion en matière d'économie politique, à la fois en ce qui concerne spécifiquement la rhétorique des *randomistas* (chap. 8) et plus généralement en inscrivant les RCT dans le champ et l'histoire de l'évaluation des politiques de développement (chap. 9). La quatrième et dernière série de chapitres (**Partie 4. [De quelques] Pistes de réflexions [ciblées] : éthique et méthode**) approfondit les propositions d'amélioration évoquées dans les chapitres consacrés aux secteurs, en mettant l'accent sur des aspects spécifiques, à commencer par la question de l'éthique, dont nous avons déjà souligné l'importance et la nécessité (chap. 10), puis en explorant les améliorations statistiques, l'utilisation des *a priori* (*priors* ; chap. 11) et la non-adhésion comme source d'information (chap. 12) ; sachant que cette partie fournit quelques éléments pour mieux faire dans le champ des RCT, et non un véritable dépassement, dont de nombreuses pistes sont explorées au fil des chapitres, y compris dans cette introduction. En guise d'épilogue, James Heckman propose une nouvelle relecture de son article phare de 1992, à la lumière de la nouvelle vague des RCT dans le domaine du développement. Il démontre que la plupart de ses conclusions d'alors, qui portaient sur la première génération des expérimentations randomisées dans le domaine de la politique sociale aux États-Unis (le « Premier Réveil » selon ses termes), sont toujours valables. Il appelle la nouvelle génération d'économistes à réfléchir et à tirer les leçons du passé. Enfin, l'ouvrage se conclut par une série de trois entretiens avec des décideurs politiques de premier rang, en laissant le point de vue de la recherche pour adopter une perspective de politique publique. Ces entretiens posent la question de l'utilisation, de l'utilité et des réponses apportées par les RCT pour prendre des décisions dans le monde réel. Le premier est un entretien croisé avec les présidents-directeurs généraux de nos institutions respectives, spécialisées dans le domaine du développement : l'aide pour l'Agence française de développement (AFD) (Rioux) et la recherche pour l'Institut de recherche pour le développement (IRD) (Moatti) ; il propose la vision d'un pays du Nord (la France). Les deux autres entretiens font intervenir des décideurs, confrontés quotidiennement à l'élaboration et au suivi des politiques économiques en Inde, terrain d'application privilégié des RCT dans les pays du Sud (Natarajan, haut fonctionnaire indien ; et Patnaik, ancien conseiller économique principal du gouvernement indien). En amont, Angus Deaton revisite, dans son introduction,

avec ses « onze variations », ses propres réflexions, à la lumière des contributions de cet ouvrage.

Remerciements

Cette introduction, dont les opinions sont de la seule responsabilité des auteurs, a été largement nourrie par les discussions lors de l'atelier qui a eu lieu à Paris le 17 mars 2019 et auquel ont participé la plupart des auteurs, ainsi que par les commentaires reçus d'Agnès Labrousse et d'Ariane Szafarz, que nous remercions chaleureusement.

Expérimentations aléatoires dans le champ du développement

Une perspective critique

Éditeurs scientifiques

Florent BÉDÉCARRATS, Isabelle GUÉRIN et François ROUBAUD

IRD Éditions

INSTITUT DE RECHERCHE POUR LE DÉVELOPPEMENT

Collection Synthèses

Marseille, 2022

Ce texte a fait l'objet d'une première édition en anglais :

BÉDÉCARRATS F., GUÉRIN I., ROUBAUD F. (eds), 2020, Randomized Control Trials in the Field of Development : a Critical Perspective, Oxford, Oxford University Press, 448 p.

<https://global.oup.com/academic/product/randomized-control-trials-in-the-field-of-development-9780198865360>

Photo 1^{re} et 4^e de couverture

© Ermell – Paul Klee, *Klippen am Meer*, 1931, Musée Lenbachhaus, Munich.

https://upload.wikimedia.org/wikipedia/commons/6/69/Klee_Klippen_am_Meer_1290074.jpg

Coordination éditoriale

IRD/Romain Costa

Préparation éditoriale

Marie-Laure Portal (11)

Mise en page

Desk (53)

Maquette de couverture

IRD/Michelle Saint-Léger

Maquette intérieure

IRD/Pierre Lopez



© IRD, 2022

Certains droits réservés. Il s'agit d'une publication en open access, disponible en ligne et distribuée sous les termes de l'attribution Creative commons, non commerciale, non modifiable 4.0 licence internationale (CC BY-NC-ND 4.0), dont une copie est disponible à cette adresse :

<http://creativecommons.org/licences/by-nc-nd/4.0/>

ISBN papier : 978-2-7099-2947-9

ISSN : 2431-7128

ISBN PDF : 978-2-7099-2948-6

ISBN epub : 978-2-7099-2949-3