

# Entretien avec IfBencheikh

## **Pouvez-vous nous présenter le J-PAL ?**

Le J-PAL est un réseau de chercheurs. La notion de réseau est vraiment importante. Il a été créé aux Etats-Unis il y a maintenant 10 ans par trois économistes du développement, dont Esther Duflo et Abhijit Banerjee, et n'était au départ qu'un petit laboratoire du MIT. Pour un certain nombre de raisons, le J-PAL a eu du succès, le réseau s'est étendu et les résultats ont été diffusés. 90 chercheurs, très majoritairement des économistes, sont aujourd'hui affiliés au J-PAL partout dans le monde. En Europe, le J-PAL compte 20 chercheurs.

Ce rassemblement repose sur deux piliers : une forte unicité méthodologie d'une part, les évaluations d'impact randomisées, et l'optique de la lutte contre la pauvreté. Bien entendu, les chercheurs peuvent faire d'autres choses par ailleurs. La pauvreté est entendue au sens très large : par exemple, des études sont menées en Europe sur des politiques de santé, d'éducation, etc. Les études ne se limitent pas aux populations qui sont en dessous du seuil de pauvreté, quelle que soit la définition qu'on en prend. L'intérêt est porté sur l'objet social en général. Il faut savoir que personne n'est salarié du J-PAL : celui-ci n'existe pas juridiquement. Les chercheurs sont rattachés à des laboratoires d'universités. Le J-PAL n'est donc pas une organisation pyramidale mais reste un réseau académique. C'est pourquoi différentes sensibilités cohabitent au sein du réseau et il ne faut pas faire parler le réseau d'une seule voix.

Le J-PAL a levé suffisamment de moyens financiers pour se doter d'une structure administrative d'appui. L'une des particularités de ce réseau est d'avoir pu créer différents bureaux régionaux : Inde, Europe depuis 2008, Amérique Latine, Afrique, Asie du Sud-Est, Amérique du nord. Ces structures qui agissent en amont et en aval des études servent de soutien à la recherche. Leur rôle consiste à préparer les projets, aider au financement, au recrutement, et à la contractualisation. Elles sont aussi en charge de la gestion d'un serveur informatique, de la gestion des données, de la diffusion des résultats, de la communication, et de l'organisation des conférences.

## **Comment les études sont-elles financées ?**

L'argent nécessaire au financement des évaluations est levé au cas par cas la plupart du temps soit par l'institution qui met en œuvre le programme et qui souhaite une évaluation (par exemple pôle emploi), soit par la Banque Mondiale ou ce type d'organisation.

En ce qui concerne le financement de la structure, il s'agit de fonds privés. Le principal mécène est le fils d'Abdul Latif Jameel, et le J-PAL vit en partie sur les intérêts de la dotation importante qu'il a apporté. Il y a également une participation de la Fondation Gates, de la Fondation Hewlett etc. Ces dons permettent de financer les locaux, le personnel, le site internet, et les publications papier.

Au J-PAL, chaque chercheur est absolument indépendant dans la gestion de son agenda de recherche.

Le recrutement de nouveaux membres se fait sur un principe de cooptation : les gens qui font partie du réseau décident et proposent l'adhésion de nouveaux membres. Certains d'entre eux refusent car l'appartenance au J-PAL entraîne quelques contreparties comme le fait d'être disponible pour donner quelques cours. Les formations dispensées par les membres du J-PAL sont de plusieurs types. Chaque affilié peut prendre en charge des enseignements à PSE (l'Ecole d'Economie de Paris), Sciences Po, etc. Il existe également des formations sous forme de stage d'une semaine à destination de professionnels souhaitant se familiariser avec les évaluations d'impact randomisées. Ces formations permettent de s'approprier les notions de base, de comprendre les enjeux et les différentes méthodes existantes. Parmi les professionnels qui y participent, on compte des fonctionnaires ainsi que des employés d'organisations internationales, d'ONG, ou d'entreprises. Sont enfin dispensées des formations académiques, sous forme de séminaires, pour des économistes, économètres, jeunes chercheurs.

## **Quelle est la définition de la pauvreté qu'adopte le J-PAL ?**

Nous ne nous imposons pas de définition restrictive. Habituellement, la définition admise est un seuil par rapport au revenu médian. Cependant, certaines études s'adressent à des publics qui ne sont pas en dessous

du seuil de pauvreté d'après les critères traditionnels mais qui présentent un enjeu social. C'est notamment le cas de l'éducation (par exemple l'insertion des jeunes diplômés).

### **Pouvez-vous nous donner un ordre de grandeur du nombre de personnes dans une expérimentation**

Ce nombre est en fait très variable. En ce qui concerne l'équipe de recherche, il faut généralement compter entre deux et cinq chercheurs, entre un et plusieurs assistants de recherche (avec au moins un à temps plein). L'assistant de recherche est la personne qui va suivre l'évaluation au quotidien, gérer la logistique et la collecte des données, etc. Il est la cheville ouvrière du projet. En Europe, il s'agit la plupart du temps de quelqu'un possédant un master d'économétrie, qui parfois effectue un doctorat. Il faut ajouter ensuite l'ensemble des enquêteurs de terrain. Certains projets internalisent tout, en employant directement les enquêteurs. D'autres sont réalisés avec des prestataires extérieurs. En Europe, c'est à peu près moitié-moitié.

Pour la partie mise en œuvre, le nombre de personnes sur le terrain est très très variable (par exemple une centaine d'agences pôle emploi pour une expérimentation réalisée en France). Enfin, en ce qui concerne le nombre de bénéficiaires, les échantillons peuvent atteindre jusqu'à 20 000 personnes.

### **Quelles sont, selon vous, les raisons du succès des RCTs ?**

C'est une vaste question. Ma réponse est le fruit d'une réflexion personnelle et non institutionnelle. La méthode n'a pas été inventée par le J-PAL, bien qu'on nous reproche parfois injustement de le revendiquer. Le mythe veut que la méthode ait été inventée par un médecin naval anglais au XVIII<sup>e</sup> siècle. Par la suite, plusieurs disciplines ont utilisé des expérimentations aléatoires, notamment l'agriculture, la médecine et les sciences sociales aux USA dans les années 60. Le J-PAL n'est qu'un moment de la méthode, sa volonté n'est pas de planter son drapeau sur la méthode. La méthode a acquis une certaine notoriété et fut appliquée pour l'une des premières fois à l'économie du développement à partir des années 90. Le J-PAL a été créé dans la foulée, dans les années 2000.

Mon explication de l'essor de la méthode dans les années 90, est une hypothèse et un ressenti personnel. L'aide au développement s'est cherchée lorsqu'a été perdu le paradigme de la Guerre froide. Celle-ci était jusqu'alors très liée à la volonté de garder les états récipiendaires de l'aide dans le « bon camp ». A la fin des années 80, le tremblement de terre géopolitique que constitue la chute du mur de Berlin touche également l'aide au développement : la pertinence de l'aide devient moins évidente alors qu'il n'y a plus d'ennemi en face et cette mise en question est renforcée par le contexte de crise économique dans les années 90 qui limite l'argent que les pays peuvent allouer à l'aide au développement. Par ailleurs certains pays ayant besoin d'aide se sont développés très vite, comme la Chine. Ces différents ingrédients se sont mis en place et ont été assez propices à ce que de nouvelles idées émergent notamment la question de l'efficacité de l'aide. Ont suivi des discours très pessimistes sur l'aide au développement comme ceux d'Easterly qui ne voient que du gaspillage dans l'aide et pour qui le remède à la pauvreté réside dans l'ouverture au marché : « trade not aid ».

Les RCT, approche à la fois très pragmatique et moins généralisante (pas tout blanc ou tout noir) a séduit : face à la restriction des ressources, il y avait une forte demande de réponses pragmatiques. Ce contexte était évidemment plus propice que les années 50 et le plan Marshall. Il aurait été difficile de demander à ce moment-là : « il faudrait mesurer l'impact de votre plan avant de déverser des millions ». La seule préoccupation était l'arrêt de la vague rouge en RDA.

### **Esther Duflo apparaît comme une figure importante du mouvement. Quelle place possède-t-elle au sein du J-PAL ?**

Il est difficile de pondérer l'importance d'un individu dans un réseau de 90 personnes. L'importance accordée à la personne d'Esther Duflo peut être excessive et elle est la première à le dire. Au sein de l'essor du J-PAL, un certain nombre d'éléments ont fait qu'Esther Duflo est devenue la figure emblématique. Les médias l'ont adoré et ça a eu une répercussion énorme. C'est à la fois juste car elle a fait des travaux remarquables à la fois sur le contenu scientifique et sur sa capacité à diffuser les résultats de manière compréhensible à un public plus large. Elle a fait preuve d'une vraie volonté de ne pas se cantonner au cercle académique uniquement et a d'ailleurs organisé une structure comme le J-PAL qui n'est pas qu'académique. Cela étant, d'autres économistes ont fait des travaux remarquables dans le réseau et l'ultra personnalisation est quelque chose qui

ne nous plaît pas. Lorsqu'Esther Duflo a été nommée au conseil d'Obama, les médias en ont fait une montagne ! Le fonctionnement du système médiatique fait qu'il est difficile de s'opposer à cette ultra personnalisation. On essaie aussi également d'en tirer profit en s'en servant intelligemment pour promouvoir une tendance plus générale et pas uniquement une personne. La critique portant sur la trop forte personnalisation d'Esther Duflo n'est pas juste car elle nous est adressée alors qu'elle n'est pas de notre fait.

### **Cette forte médiatisation a-t-elle un impact sur les financements ?**

Cela n'est pas très clair car les financements ont précédé la médiatisation, en particulier pour le financement de Jamil. Par contre cela a engendré une augmentation d'audience assez forte, y compris dans le milieu institutionnel de l'aide au développement, sur la méthode et ses résultats. La banque mondiale a embrayé, DFID aussi. Il y a eu un phénomène d'essor de la méthode d'évaluation d'impact qui a permis à des gens de financer des projets : c'est un cercle vertueux qui ne tient pas qu'à la personne d'Esther Duflo.

### **Un certain nombre de critiques sont adressées aux RCTs notamment par des personnes comme Heckman, Nubukpon, Deaton: ont-elles entraîné une révision de la méthode par les chercheurs du J-PAL ?**

Je pense que oui. Lorsque les critiques viennent de gens du niveau de Deaton ou d'Agnès Labrousse, qui sont honnêtes intellectuellement et constructifs dans la façon dont ils posent la question et débâtent, nous sommes ravis. C'est stimulant et c'est de cette manière que la science progresse. En dix ans, il y a eu un approfondissement de la réflexion et des améliorations apportées à la méthode. Les critiques extérieures y ont contribué.

Je peux vous donner deux exemples pour lesquels les critiques nous ont fait beaucoup réfléchir. Le premier porte sur l'hétérogénéité de l'impact. Les critiques adressées aux RCTs sur ce plan ont beaucoup nourri la réflexion et toute expérimentation qui commence essaie maintenant d'avoir une capacité d'analyse de l'impact beaucoup plus fine. Il n'existe pas d'obstacles et on peut se donner les moyens de travailler sur la mesure de l'hétérogénéité de l'impact. Ce genre de critiques permet d'avancer : il force à se poser des questions. Le deuxième concerne la critique sur le fait que l'évaluation d'impact apparaît comme une « black box ». Nous sommes très ouverts sur les méthodes mixtes (« mix methodologies ») nous n'avons aucun soucis à essayer de travailler le plus possible avec des gens qui font du qualitatif pour compléter notre analyse quantitative, ou à l'inverse avec des gens qui mènent de grandes analyses multi-pays. Tout savoir qui peut être complémentaire au nôtre est intéressant. C'est donc une critique qu'on a accepté et à laquelle on essaie de répondre, par exemple en faisant des mesures intermédiaires lors de l'évaluation afin d'avoir une traçabilité des output mesurés.

Ensuite, il existe d'autres critiques moins constructives pouvant témoigner d'une certaine jalousie de personnes qui ont peur de perdre des rentes. Certaines personnes jettent un discrédit complet sur le principe même des évaluations d'impact et cela ne nous fait pas avancer...

### **Etes-vous témoins d'une différence de perception de votre action par la population selon l'endroit où est menée l'expérimentation ?**

L'idée d'expérimentation est très admise aux EU où des expérimentations sont menées depuis déjà 40 ans. Dans certains pays en Europe, il n'est au contraire pas possible d'utiliser le terme même d'« expérimentation ». En Allemagne par exemple, si vous parlez d'expérimentation, les gens pensent tout de suite au nazisme, à l'eugénisme. Il y a donc des sensibilités diverses dont il faut tenir compte lorsque l'on mène une évaluation.

Sur le rapport au tirage au sort, les réactions sont assez c'est variable. Dans certains pays en développement, le principe d'accéder ou non à un programme de manière aléatoire est perçu comme une grande avancée égalitariste et démocratique (comme l'était la démocratie athénienne !). Cela est préféré aux avantages qui se distribuent au copinage ou lorsque l'on est le cousin du chef etc. Le fait même de tirer au sort, qui est souvent une question éthique qui ressort, est donc très variable selon les contextes. Cela est accueilli très favorablement dans certains endroits ou au contraire moins facilement comme en France même si c'est aujourd'hui possible. En France, l'idée qu'il faut que soit respectée l'égalité de tous devant la loi est globalement admise : si quelqu'un est éligible à un programme, tout le monde doit être éligible. Pour pouvoir expérimenter il a ainsi

fallu changer la constitution, comme Martin Hirsch l'explique : en 2005 il a fallu introduire dans la constitution française la possibilité de faire des traitements différenciés dans le cadre d'une expérimentation sociale. Il est en effet nécessaire d'exclure des gens d'un traitement que d'autres reçoivent si l'on souhaite expérimenter, et ce qu'elle que soit la méthode.

C'est donc une question très sensible. Il est alors important de connaître les contextes et d'être toujours très explicatif, d'autant plus que nous sommes très dépendants de nos partenaires et des gens qui mettent en œuvre les programmes. S'ils ne sont pas d'accord avec l'évaluation d'un point de vue éthique, tout s'arrête.

### **Avez-vous des exemples d'expérience qui n'ont pas été mises en œuvre à cause d'un refus éthique ?**

L'arrêt d'une évaluation résulte rarement d'un problème éthique. Cette question de l'éthique apparaît tout le temps quand on discute avec les gens, on parle d'éthique 50% du temps. Dans la réalité des projets, c'est rarement cette question qui achoppe. En effet, c'est un point qui surgit vite et l'on s'interdit absolument de se diriger vers des domaines dans lesquels exclure des gens est manifestement criminel : comme le fait de randomiser l'aide d'urgence ou de laisser mourir des gens pour mesurer un impact. Un deuxième filtre est assuré par les comités éthiques en particulier aux Etats-Unis où ils sont très institutionnalisés : on parle d'« institutional review board ». Tous les projets J-PAL sont soumis à un comité comme ceux-ci. Enfin, les partenaires constituent un filtre très fort car ils n'hésiteront pas à dire si l'évaluation ne leur semble pas acceptable.

Ce qui bloque souvent les projets en pratique c'est de ne pas avoir réussi à convaincre l'ensemble de la chaîne de commandement. On surestime souvent le degré de décentralisation des structures (de type associatifs ou dans le secteur privé etc...) : même si on est arrivé à se mettre d'accord avec le dirigeant de la structure, celui-ci peut être la seule convaincue et les gens sur le terrain peuvent ne voir l'évaluation que comme des ennuis. Il faut savoir qu'une expérimentation aléatoire est très contraignante. Très souvent, au niveau de la mise en œuvre sur le terrain, des gens peuvent soit refuser ouvertement de se plier à l'évaluation, soit faire de la résistance passive car cela complique leur travail. Ils peuvent également le ressentir comme de l'évaluation/sanction et craindre des reproches futurs. C'est beaucoup plus souvent lorsque l'on n'a pas réussi à convaincre l'intégralité de la structure que pour des considérations éthiques qu'un projet peut capoter.

### **Est-il vraiment possible de faire une randomisation rigoureuse sur le terrain ?**

Il faut veiller à s'assurer que le plan de randomisation est respecté. Par exemple dans le cadre de l'évaluation d'une offre de micro crédit, il faut s'assurer que l'agent de micro crédit, lorsqu'il a fini de vendre son produit dans les villages du groupe test, n'aille faire la même chose dans le groupe témoin parce qu'il peut avoir intérêt à développer son activité par exemple. Il est important d'avoir des points de contrôle même si on ne peut pas mettre quelqu'un derrière chaque personne sur le terrain : si vous commencez à mélanger les groupes, vos résultats n'ont plus de valeur.

### **Et au niveau même des villages ? Est-ce que les décideurs locaux tentent de privilégier le choix de certains individus au risque de biaiser la randomisation ?**

On a effectivement souvent des demandes : par exemple, lorsqu'on randomise des établissements scolaires en France, il y a toujours un inspecteur d'académie pour demander à ce que tel ou tel établissement soit dans l'échantillon test pour des raisons personnelles de préférence. Il faut être très ferme là-dessus : on ne peut pas changer la randomisation en cours de route et si la randomisation n'est pas respectée l'expérience doit s'arrêter. Cependant, d'habitude, si vous avez fait le bon travail d'information au préalable, ce problème est normalement réglé au moment du tirage au sort. Tout le monde en a alors accepté le principe. Il y a évidemment toujours des mauvais perdants qui a posteriori cherchent à changer de groupe mais cela reste quand même assez rare.

### **Il faut quand même que les écoles, villages aient au préalable accepté de participer au tirage au sort. Cela ne nuit-il pas à la représentativité de l'échantillon par rapport à la population ?**

Le fait que les gens acceptent ou non la randomisation peut fausser la représentativité de l'échantillon par rapport à l'intégralité de la population. Si, par exemple, les écoles qui n'ont pas accepté de participer à la

randomisation sont très différentes de l'ensemble des autres écoles, cela limite votre capacité à généraliser les résultats. C'est cependant quelque chose que l'on peut contrôler, et cela relève de la question de la validité externe. Il est important de savoir que la validité interne des RCTs est très forte. C'est une méthode qui permet d'être vraiment assez sûr de l'impact que l'on mesure. Il n'existe pas de biais de sélection comme dans les autres méthodes. Par contre, la validité externe n'est pas, par principe, plus forte que pour une autre méthode : on sort là du domaine de la certitude contrairement à ce qui concerne la validité interne. On est sûr de notre résultat dans un contexte particulier mais on n'est pas sûr qu'il reste valable de manière générale dans d'autres contextes, à d'autres endroits, à d'autres moments. La validité externe comprend toujours une part de théorie : il faut, pour pouvoir généraliser un résultat, essayer d'analyser les différents facteurs qui permettent d'établir des similitudes etc. De toute façon, chercher à généraliser nécessite de faire un certain nombre d'hypothèses : il s'agit un pari en quelque sorte. Cela ne veut pas dire que l'on n'a pas de raison rationnelle de faire ce pari. Seulement, on sort du domaine de la certitude absolue. Cette question de la validité externe se pose quelle que soit la méthode, pour les RCTs comme pour les autres. Cependant, dans le cas des RCTs, on est très sûr des résultats dans l'expérimentation. Pour les autres méthodes, « avant/après », « matching », « diff and diff », on n'est jamais complètement sûr de la validité interne et on doit ainsi prendre d'autant plus de pincettes pour la validité externe. C'est pour répondre à la volonté de généraliser un résultat qu'il est important d'ouvrir la « boîte noire » et de comprendre les mécanismes en œuvre ainsi que les explications du comportement des acteurs, et non juste de mesurer l'impact. Ces questions-là permettent d'évaluer de la manière la moins incorrect possible la possibilité de généraliser : il est important de passer par la théorie.

Il est possible d'ouvrir un peu cette boîte noire lors d'une évaluation randomisée en essayant d'avoir des outcome intermédiaires et un suivi avec des étapes. Pour ouvrir la boîte noire il existe d'autres méthodes complémentaires, comme les approches qualitatives. La différence entre le quantitatif que l'on fait et le qualitatif c'est que nous faisons un questionnaire fermé à 10 000 enfants. C'est détaillé mais cela reste de la grosse machine comparé à un enquêteur qualitatif qui passe une journée entière dans un foyer à discuter avec les parents etc. On n'obtient pas du tout les mêmes résultats. Notre avantage est d'obtenir quelque chose de beaucoup plus quantifiable. Il est donc possible de commencer à ouvrir la boîte noire avec du quantitatif et de compléter avec d'autres méthodes qualitatives. Plus vous avez appris de choses et compris quels étaient les mécanismes en jeu grâce à différents types de méthodes, plus vous pouvez mener une réflexion pertinente sur la capacité à généraliser les résultats obtenus. L'un de nos chantiers est de nous efforcer à articuler les résultats précis que l'on a avec des approches théoriques plus larges. Notre approche est assez proche de l'économie comportementale. La mesure d'impact rigoureuse que l'on fait doit venir nourrir les modèles d'économie comportementale pour essayer de bien comprendre comment les gens se comportent et c'est cela qui permettra ensuite de concevoir les programmes les plus pertinents.

### **Comment les approches quantitatives et qualitatives s'articulent-elles concrètement ?**

Cela est souvent simultané. Je ne vous cache pas qu'il n'est jamais très facile de faire de l'interdisciplinaire parce qu'il y a des « chapelles ». Les disciplines ne sont pas forcément très habituées à travailler entre elles mais on essaie autant que possible de casser les barrières. Par exemple en science de l'éducation, on travaille à la fois avec des économistes chevronnés comme Marc Gurgand qui vont mettre en place le protocole quantitatif mais également avec des spécialistes des sciences de l'éducation qui apportent une compréhension complémentaire des enjeux pédagogiques. Cela est également possible avec des anthropologues ou des sociologues. On a notamment travaillé avec le CREDOC.

Il est aussi possible d'internaliser le qualitatif : rien n'empêche un économiste quantitatif de faire lui-même du travail qualitatif comme des entretiens plus détaillés etc. Un quantitatif n'est pas dans l'incapacité de faire du qualitatif et réciproquement. Cette internalisation peut être faite par l'équipe de recherche ou en partenariat avec d'autres spécialistes.

### **Les résultats des travaux, publiés par exemple sur le site duJ-PAL, ont-ils vocation à être utilisés comme base de données pour l'économie comportementale ?**

Nous les mettons dans le domaine public, ils sont publiés dans les revues académiques qui sont souvent des revues économiques. Le fait d'utiliser un site web permet de viser un plus large public, qui reste quand même principalement les décideurs politiques, les organisations internationales. Tous les canaux pour toucher d'autres disciplines sont les bienvenus. D'ailleurs, quand on organise des formations, des gens parfois même

réticents au début s'intéressent à la méthode, dont des anthropologues. Cela se passe généralement bien. Je suis sûr qu'il y a des potentiels de complémentarité qui sont importants. La question qui est souvent difficile est celle des publications académiques : il n'est pas facile de faire un partenariat qui permette à la fois de publier votre article dans l'American Journal of Economics et dans une revue d'anthropologie. Il faut alors réussir à monter des projets communs ou chacun peut tirer des résultats et publier un papier selon les critères de sa discipline. C'est pour ça que c'est plus souvent des partenariats d'équipe qu'une seule et même équipe unie rassemblant plusieurs disciplines. On ne peut pas écrire le même article dans les deux disciplines or les chercheurs sont de plus en plus jugés sur leurs publications.

### **Les travaux des chercheurs affiliés au J-PAL se rapprochent-ils de l'économie comportementale ?**

Ils entrent en résonance avec les questionnements de l'économie comportementale. Ce sont avant tout des économètres qui réalisent des évaluations d'impact. Mais comme nous réalisons des évaluations d'impact avec l'ambition de comprendre pourquoi les impacts sont obtenus, forcément des questions d'économie comportementale émergent. Il s'agit d'essayer d'affiner la compréhension de la manière dont les gens prennent leur décision. N'étant pas économiste, je ne peux pas trop entrer dans les détails.

### **Les évaluations d'impact sont-elles des études pilotes, réalisées ex ante, autrement dit avant que le programme ne soit mis en place ?**

Nous ne réalisons pas d'étude ex ante. Le terme ex ante désigne en effet le fait d'essayer de prévoir ce qui va arriver si l'on met un programme en place. Cette prédiction repose sur de la modélisation et consiste à faire une projection en partant de ce que l'on sait déjà. Avant une réforme fiscale, vous essayez de modéliser l'impact de cette réforme sur les contribuables. Ce n'est pas du tout ce que l'on fait. Nos évaluations d'impact sont ex ante dans le sens qu'il n'y a d'évaluation randomisée possible que s'il on arrive au début du processus. On ne fait pas du ex post, qui consisterait à évaluer des programmes mis en place des années auparavant.

Il faut également se mettre d'accord sur le mot « pilote ». On fait de l'expérimentation, c'est à dire qu'on fait un test à l'échelle la plus petite possible qui nous permette de mesurer des effets. Ce n'est pas par plaisir qu'on a des grands échantillons : on en a besoin pour obtenir des résultats probants et rigoureux. Mais l'idée est de tester à la plus petite échelle possible une politique nouvelle. Le terme « pilote » désigne souvent ce qui vient avant : on fait un pilote à l'expérience qui consiste à tester les questionnaires, observer la réaction des gens face à ce qui est proposé. Mais vous pouvez appeler l'expérience le pilote à la grande politique qui viendra ensuite. Il faut juste faire attention à ce terme.

### **L'expérience intervient toujours avant que le programme ne soit généralisé...**

Tout à fait. Il s'agit d'expérimentation. Une politique nationale par définition ne pourrait pas être évaluée comme on fait car on n'a pas de groupe de comparaison. On fait de l'expérience sur un petit groupe que l'on compare avec un groupe similaire.

### **Le but est quand même que le programme évalué devienne une politique nationale...**

Si l'impact est positif, oui. La situation idéale est de réaliser l'expérience, de mesurer un impact positif et de généraliser. Ou bien vous n'avez pas d'impact et c'est abandonné. Mais souvent, cela ne se passe pas ainsi. Dans la réalité des faits, il y a des problèmes de calendrier : la décision politique ne correspond pas au rythme du travail de recherche. Vous pouvez avoir des résultats bien après que la décision a été prise. Cela nous est arrivé assez souvent. Vous pouvez être confrontés à d'autres facteurs qui font que le résultat de recherche pure ne va pas peser assez lourd face à d'autres considérations. Cela nous est arrivé à de nombreuses reprises que le choix effectué ne soit pas celui qu'on aurait préconisé. Mais ce n'est pas grave. Il faut qu'on reste dans la position humble du chercheur. Nous ne sommes pas les éminences grises qui tirent les fils des marionnettes politiques. C'est une vision qu'on n'a pas, qu'on nous prête parfois et qui est absurde. On est là pour produire

du savoir, pour essayer de mieux comprendre comment les choses se passent, les impacts obtenus et on fournit les résultats au débat public.

Je prends un exemple. On a mené l'évaluation de l'internat d'excellence de Sourduin. On a mis en évidence un impact bon aux niveaux cognitifs et comportementaux pour les élèves. La décision de modifier le dispositif a été prise en considérant que cela coûtait trop cher par rapport à ce qui était obtenu comme résultat. Ce n'est pas au chercheur de se positionner là-dessus. On apporte du savoir mais c'est ensuite aux hommes politiques de faire leur boulot et de se poser les questions : est-ce le bon moment ? Est-ce trop cher ? Il ne faut pas mélanger les rôles.

L'objectif est de généraliser, mais c'est avant tout de construire du savoir, de le répandre et de faire qu'il soit utilisé à bon escient. La logique n'est pas juste de continuer ou non un programme, mais d'apprendre grâce à ce programme. D'ailleurs la continuation du programme ne va pas dépendre que de l'article publié. Il y a d'autres éléments qui entrent en ligne de compte. Par contre, si on a réussi à apprendre quelque chose, on espère que cela aura un poids dans la culture générale du décideur public et que cela finira par sédimenter, ressortir. On est jeune encore : on est dans une phase de construction du savoir. Par exemple, sur le programme microfinance, on a publié un papier qui rassemble tous les résultats d'une centaine d'évaluations qui ont été faites. On est encore dans cette phase où on essaye d'additionner le plus possible de résultats pour dégager des leçons qu'on offre ensuite au débat public.

#### **Lorsque la décision politique intervient avant la fin de l'expérience, met-elle un terme à celle-ci ?**

Pas nécessairement. Si cela brise la randomisation, oui. Après la prise de décision, le temps de mise en œuvre fait qu'on a souvent fini la phase de collecte. On a l'impression d'être le bec dans l'eau mais en fait non, car on est un peu plus que des évaluateurs qui disent si le programme est efficace ou non. On a aussi une ambition de participation à la construction d'un savoir. Et cet aspect est un peu déconnecté des aléas de la décision publique imminente.

#### **Arrive-t-il que des décideurs politiques utilisent de manière malhonnête le travail des chercheurs pour justifier une politique, sans respecter toutes les subtilités des conclusions ?**

Oui forcément. Les hommes politiques ont leur jeu et leurs règles du jeu. Il y a un cas amusant que l'on cite souvent : Progreso au Mexique. Ce programme de *conditional cash transfers* a été lancé et les résultats ont très satisfaisants. Puis, le gouvernement a changé, le nouveau président a clamé que c'était un scandale de payer des gens pour qu'ils envoient leurs enfants à l'école ou qu'ils les soignent. Il voulait tout arrêter. Ensuite, il s'est rendu compte de l'impact et des résultats du programme et a réalisé qu'il ne pouvait y mettre un terme. Du coup, le nom a été modifié, devenant Oportunidades. Cela leur a permis de garder l'idée mais en plantant le drapeau. C'est aussi le jeu des démocraties que les gens ont besoin de se faire réélire et d'afficher des résultats. On ne va pas se plaindre qu'il y ait des élections.

Il peut aussi y avoir détournement des conclusions. Il y a eu un cas où le débat public a été difficile à mener correctement, et qui a donné lieu à la rédaction d'un article de la part de nos chercheurs. Nous avons mené une expérimentation sur l'impact du CV anonyme. Il ne s'agissait aucun cas de mesurer la discrimination à l'embauche, puisque l'expérimentation concernait des entreprises volontaires. La moitié d'entre elles recevait des CV anonymes de Pôle emploi, c'est à dire ne comportant ni les noms ni les adresses. L'autre moitié continuait à recevoir des CV habituels. Le fait que les entreprises soient volontaires biaisait à l'évidence les choses. Cela ne permettait pas une mesure de la discrimination puisque que l'on pouvait assez légitimement penser que les entreprises discriminantes n'allaient pas accepter de participer.

Mais ensuite, un certain nombre de personnes se sont saisi des résultats pour conclure à l'absence de discrimination à l'embauche en France. Cela était très malhonnête intellectuellement et ne correspondait absolument pas ce que l'on avait dit. La seule chose qu'on avait montrée était que pour les entreprises a priori

ouvertes sur la question, l'anonymisation avait plutôt des effets qui égalisaient les chances de recrutement entre femmes et hommes. Avec des CV non anonymes, il y avait une tendance que les hommes recrutent surtout des hommes et les femmes recrutent surtout des femmes. Le fait de rendre anonyme change le profil des gens qui accèdent à l'entretien, et le fait d'accéder à l'entretien change la possibilité de recrutement. C'était une vraie leçon. Par contre pour ce qui était de l'origine étrangère et du lieu d'habitation, cela avait plutôt un effet contreproductif. Il y avait moins de recrutement de gens avec des patronymes étrangers et originaires de certains quartiers difficiles. Ce qui semble en être la raison, c'est qu'au sein des entreprises qui jouaient le jeu et qui n'étaient donc a priori pas discriminantes, il y avait une reconnaissance que les candidats venant de quartiers difficiles étaient plus méritants à niveau égal. Si vous lisez un CV anonyme avec des trous, vous allez considérer que son propriétaire n'a rien fait pendant une période. Si vous savez qu'il vient d'un quartier extrêmement dur où très peu de gens s'en sortent, vous vous dites qu'il a galéré mais qu'il est vraiment méritant et vous décidez de lui donner une chance. Il ne s'agit pas de discrimination positive au sens d'aider des jeunes en difficulté, mais de la reconnaissance des qualités liées à la difficulté du parcours d'un candidat. Ces résultats sont intéressants mais subtils. Et vous ne pouvez pas juste vous contenter de balancer dans la presse que les CV anonymes n'ont pas d'impact et que cela montre bien qu'il n'y a pas de discrimination à l'embauche en France, ce qui est par ailleurs contredit par toutes les études réalisées sur la question de la discrimination. Nous ne faisons pas une mesure de la discrimination, mais une mesure de l'impact du CV anonyme dans un pool d'entreprises dont on peut penser qu'elles n'étaient pas discriminantes à la base. C'est très différent. Si vous avez une page dans le Monde pour l'expliquer, vous y arrivez à peu près. Si vous avez 10 secondes à la radio, c'est déjà beaucoup plus dur.

Ceci témoigne de la difficulté des chercheurs qui essaient de se confronter au monde réel, et de ne pas rester dans publication académique. Il est beaucoup plus facile de rester entre pairs, de faire un très bel article sur le CV anonyme qui est lu par 50 personnes dans le monde qui trouvent cela absolument génial. Mais vous ne vous confrontez pas à la difficulté du débat public qui est forcément simplificateur, et souvent malhonnête sur des questions sensibles.

Avec la façon dont le J-PAL fonctionne, pour reprendre la terminologie de Weber, le savant est dans le monde. On n'est pas dans une bulle isolée à nous préoccuper que des questions qui intéressent le monde académique. Notre slogan est « Traduire la recherche en action ». On a eu la volonté de ne pas se limiter à être un réseau académique, mais d'être en plus une structure d'appui qui va aussi diffuser les résultats et donc qui va être en prise avec le débat public où qu'il soit. L'implication dans le débat public est indéniable. Par contre on ne fait pas de politique. Je pense qu'il faut quand même bien faire la distinction entre le rôle du chercheur et le rôle du politique : à nouveau, on n'est pas là pour prendre les décisions. Il est vrai qu'il existe des chercheurs qui prennent des responsabilités politiques. Vous ne pouvez pas empêcher un chercheur de travailler dans un ministère, de rentrer au gouvernement ou d'avoir des convictions politiques comme n'importe quel citoyen. Mais il y a quand même une séparation entre les deux fonctions. On est là pour donner des contributions au débat public, mais pas pour prendre des décisions à la place des hommes politiques ni même des administrations : il ne faut pas se tromper de rôle.

**Vous avez parlé de Martin Hirsch. Le J-PAL a-t-il participé aux expérimentations sur RSA en France ?**

Non, on a refusé. Parce que la randomisation ne pouvait être effectuée correctement. Martin Hirsch est quelqu'un de très convaincu des bienfaits de l'expérimentation. Par-delà le RSA, il a beaucoup aidé l'idée de l'expérimentation en France en créant le Fonds d'Expérimentation pour la Jeunesse, qui a permis de financer des projets. Le RSA, ça n'a pas pu se faire. Ils ont fait ce qu'ils ont pu mais le J-PAL n'a pas été associé. Le problème méthodologique faisait qu'il n'était pas possible de faire rentrer cela dans le cadre d'un travail J-PAL.

**Vous arrive-t-il d'être tenu à un certain secret par vos partenaires quand vous faites des expérimentations?**



Qu'on ne parle pas trop d'une expérimentation lorsqu'elle est en cours, c'est assez légitime. Il n'y a pas de raison de le faire de notre part. Ensuite, notre règle quasiment intangible est qu'on publie tout, que les résultats soient positifs ou négatifs. A tel point qu'on a lancé avec l'American Assembly un registre des expériences randomisées. En médecine, avant d'entamer des essais cliniques, les chercheurs s'enregistrent de manière ouverte et transparente, déclarent ce qu'ils comptent faire, leurs hypothèses, les outcomes mesurés, le plan d'analyse. Il n'est alors pas possible de taire des résultats qui ne vont pas dans le sens prévu à l'avance, ou de faire du datamining. Cela n'existait pas en sciences sociales. On est en train d'enregistrer le plus de projets possibles. Tous les projets J Pal devront être enregistrés et on pourra nous demander des comptes sur tout. On essaye de se tenir à cela. Je vous ne cache pas que ce n'est pas facile car se posent des questions de propriété intellectuelle des résultats. Souvent, quand c'est la même institution qui paye à la fois le programme et l'évaluation, il peut y avoir l'idée que c'est eux qui sont propriétaires des résultats et qu'on les publie s'ils en ont envie. Mais on essaye de bien définir les règles dès le début : on est de nature universitaire, l'expérimentation donnera lieu à des publications académiques. Autant on peut les laisser gérer la communication comme ils le souhaitent sur le rapport final autant pour la publication académique, on fera ce qu'on voudra, y compris en faire mention sur notre site web. Je pense pouvoir dire que l'on se tient assez bien à cette règle.

### **Pensez-vous qu'il y a assez d'expériences aléatoires conduites en économie du développement ?**

Je ne suis pas sûr qu'il faille se poser la question comme cela. L'objectif n'est pas de les multiplier de façon absurde. Il faut réfléchir aux secteurs dans lesquels on en a besoin. Typiquement, ces expérimentations répondent à des questions précises et on essaye de partir des questions plutôt que d'être à la remorque des projets sur le terrain. On lance des initiatives qui consistent à essayer de rassembler de l'argent sur une thématique et ensuite à faire ce que l'on appelle un *match making conference* rassemblant tous les acteurs intéressés par le même sujet : praticiens, chercheurs... Ils réfléchissent ensemble aux domaines pour lesquels on manque d'information et sur lesquels il faudrait travailler.

Il vaut mieux faire quelques bonnes expérimentations que d'en faire 50 mauvaises, parce que soit techniquement faibles, soit manquant de pertinence. Il faut être sélectif et bien cibler. Ce sont des gros projets qui coûtent de l'argent et prennent du temps. Il faut donc essayer d'être le plus pertinent possible. C'est un domaine dans lequel on a appris et on a plus aujourd'hui la capacité d'être sélectif au bon sens du terme.

Dans le champ académique, certaines personnes, qui ne font pas d'expérimentation aléatoire et qui voient que dans les publications il y a beaucoup d'articles portant sur ces expérimentations, boudent un peu en prétendant que désormais, pour être publié, il faut faire des expérimentations. Ceci est quand même très exagéré parce que le pourcentage d'articles reposant sur les expérimentations reste minoritaire. On ne peut pas dire qu'il y ait une invasion du champ de la publication en économie par les expérimentations et même en terme d'évaluation, si l'on compare le nombre de programmes qui existent et le nombre de programmes évalués...L'expérimentation est juste en train de prendre une place légitime, la place qui devrait être la sienne dans le champ de la réflexion en économie. Il ne s'agit pas que cela exclue tout le reste, que ce soit la seule méthodologie à employer.

### **Justement, que peut-on dire des expérimentations aléatoires par rapport aux autres méthodes d'expérimentation qui préexistaient ?**

C'est une bonne question. La question du coût dépend de ce à quoi l'on compare les expérimentations. Si vous comparez à un programme pour lequel il n'y a pas de mesure d'impact, c'est sur cela coûte plus cher de faire de l'expérimentation. Si vous comparez à d'autres méthodes rigoureuses, ce n'est pas clair du tout. Ce qui coûte cher, c'est la collecte de données. Prenons le matching par exemple. Le matching consiste à essayer de constituer des paires : une fois la population cible déterminée, il s'agit d'identifier un groupe témoin qui soit comparable, et de le constituer individu par individu. Lors de l'expérimentation sur l'internat de Sourduin,

parmi tous les éligibles, on a tiré au sort ceux qui entraînent et ceux qui n'entraînent pas. Si l'on avait procédé par matching, on aurait fait entrer tout le monde, et ensuite on aurait essayé de trouver, pour chaque élève, un élève qui lui ressemble le plus possible sur des critères observables comme l'âge, les notes, le revenu des parents, et qui ne soit pas à l'internat d'excellence, afin de constituer le groupe témoin. Je peux vous dire qu'il est moins coûteux de randomiser. La question du coût est très relative. Est-ce moins cher que d'autres méthodes ? Dès qu'on arrive à des méthodes rigoureuses, cela se vaut souvent car vous avez besoin à peu près du même nombre de données.

Ensuite, il y a la question du coût de l'évaluation par rapport au coût du programme. Si l'évaluation en elle-même coûte deux fois plus cher que le programme, il faut vraiment être sûr que le savoir qu'elle va produire sera utilisable plus largement. Il ne faut pas se lancer dans des expérimentations trop spécifiques, sur une population trop limitée. Vous ne pouvez rien apprendre et vous allez dépenser autant en évaluation qu'en action : cela n'a pas de sens.

La question du coût ne doit pas être considérée en absolu. Des personnes vont dire que telle évaluation a coûté 500000 euros. Par rapport à plein de choses, c'est très cher et par rapport à plein d'autres, c'est peu. Il faut toujours comparer le coût des évaluations avec l'utilité que vous en retirez, la possibilité de généralisation, et si vous êtes convaincu que cela vaut le coup de faire une évaluation d'impact rigoureuse, il faut comparer avec le coût des différentes méthodes. C'est cette approche là qu'il faut avoir.

**Qui est moteur d'une expérimentation ? Est-ce que ce sont les institutions, ou plutôt les chercheurs, qui lancent des idées nouvelles ?**

Cela dépend. Le modèle vers lequel on veut aboutir, c'est que cela soit un travail commun : la rencontre de chercheurs intéressés par une thématique et d'un partenaire de terrain faisant émerger des projets qui dès le début sont conçus comme étant communs. Dans la réalité des faits, le plus souvent, ce sont des partenaires de terrain qui viennent nous voir avec des projets. Ils ont envie ou besoin d'évaluer- il y a à la fois des interlocuteurs qui veulent réellement savoir l'impact de ce qu'ils font, mais aussi pas mal d'interlocuteurs qui sont dépendants de bailleurs de fonds leur demandant d'évaluer l'impact de ce qu'ils font. Il peut y avoir plein de choses intéressantes qui naissent comme cela. Mais si l'on pouvait réussir à être dans le partenariat conceptuel dès le départ, cela serait plus fructueux, on l'espère. C'est vers ce modèle là qu'on essaye de tendre. Pour l'instant, les cas dans lesquels le chercheur a son idée et trouve un partenaire de terrain pour la mettre en œuvre sont assez rares.

**Par qui sont financés les groupes de réflexion et les initiatives que vous décriviez précédemment ?**

On a un certain nombre de programmes thématiques : éducation, micro finance...Au sein de chacun de ces programmes, on va essayer de lever de l'argent pour faire ce que une initiative. Des donateurs privés financent certaines initiatives. On peut aussi récolter de l'argent public. Souvent, l'argent public est un peu plus dur à décrocher car il est assez rare qu'on nous fasse un chèque en blanc. Quand on rencontre une institution, on évoque des pistes sur lesquelles on veut travailler et lancer des projets. Il faut que les gens aient suffisamment confiance en vous pour vous donner de l'argent. Les bailleurs institutionnels vont financer un programme précis, et ont leur propre agenda. Pour le financement des initiatives, c'est plus souvent de l'argent privé qu'on réussit à lever. Si on en arrive là, c'est qu'on commence à construire sur une notoriété et une confiance. Cela ne veut pas dire qu'il n'y a aucun contrôle sur l'argent que l'on nous donne.

L'idée de récolter des enveloppes qu'on peut utiliser de manière un peu collégiale avec acteurs de terrain et chercheurs est plus facile à vendre à des fondations privées qu'à des bailleurs institutionnels de manière générale.

**Avez-vous déjà participé à une expérimentation par assignation aléatoire ?**

Non, je suis administrateur. En amont, c'est souvent la directrice du bureau et moi qui rencontrons ceux qui demandent des évaluations d'impact de ce qu'ils font. Avant de les envoyer voir nos chercheurs affiliés, on discute avec eux, on explique et on conseille. Une fois que c'est lancé, ce n'est plus mon boulot. Je continue à suivre par intérêt. Dans ces bureaux, on est trois dans la partie administrative et tous les autres travaillent sur des projets de recherche donc évidemment on discute entre nous. Mais je ne suis pas directement sur un projet.

**Le chercheur doit se préoccuper des besoins des partenaires de terrain, se confronter aux enjeux politiques locaux... On peut dire qu'ils sont très ancrés dans la réalité, n'est-ce pas ?**

On est dans la recherche appliquée. Je ne dis pas ça en suggérant qu'il ne faut pas de recherche fondamentale. Il faut qu'il y ait des gens dans leur labo et qui se posent des questions dont on ne voit a priori pas le lien avec la réalité concrète. Si on n'avait fait que de la recherche appliquée depuis le début de l'humanité, on ne serait pas très loin. Les deux ont leur pertinence et leur légitimité. Des gens comme nous, qui se prétendent ancrés dans l'action, il faut qu'on soit au contact et dans la réflexion partagée avec les gens qui sont sur le terrain. Parce qu'ils ont plein de choses à nous apprendre... Les chercheurs qui font de l'évaluation randomisée disent que c'est une des choses qu'ils préfèrent, car ils sont forcés d'être dans l'échange. C'est pénible aussi, car les jeux institutionnels sont plus complexes, et il faut convaincre tous les partenaires. Le recteur de l'académie est d'accord et dans un établissement, on vous jette des pierres. Cela complique les choses mais c'est porteur d'une richesse incroyable. Cela fait vraiment partie de la beauté de l'évaluation.

**La recherche fondamentale en économie est-elle utile?**

C'est une bonne question. Je ne peux pas vous répondre. Les économistes savent beaucoup mieux que moi.

**Le point de départ des expérimentations aléatoires était quand même de se débarrasser des théories qui ne donnent pas de résultat concret et d'aller voir sur le terrain...**

Je ne suis pas très compétent pour répondre à cette question. Nous clairement, on est dans une démarche pragmatique et de terrain. Est-ce pour autant que cela disqualifie toutes les autres démarches ? Pas nécessairement. On peut estimer qu'il y a un manque à combler sans considérer que ce que font les autres n'a pas d'intérêt. Je ne suis pas assez au fait des subtilités de la théorie économique.

**Quels sont les objectifs du J-PAL ?**

Je vais vous faire une réponse un peu bateau, un peu « cérémonie des césars ».

L'objectif, c'est d'essayer d'accumuler suffisamment de connaissance pour faire une différence concrète sur un certain nombre de politiques publiques. Il y a deux mois, j'étais aux dix ans du J-PAL : il y avait une petite cérémonie au MIT. Le sentiment dominant qui ressortait des interventions des gens qui ont accompagné le J-PAL pendant les dix ans de son existence, c'est que le réseau est encore très jeune. On a vraiment cette conscience que c'est le début d'une histoire, qu'il y a plein de choses qui ont été débroussaillées, et qu'on commence à accumuler des résultats qu'on peut commencer à organiser. Mon souhait est qu'il y ait une sédimentation fructueuse. Tout ce qui est fait, qu'on en tienne compte, qu'on construise à partir de cela. Qu'il y ait aussi de nouvelles personnes qui arrivent dans le réseau et qui apportent de nouvelles choses. On est encore très clairement dans une phase d'effervescence, et il faut garder cette créativité. Cela, ce sont les aspects internes.

En externe, plus on pourra avoir des belles réussites de programmes dont on a montré qu'ils ont de la valeur et qui sont généralisés, plus on aura des Deworming et des Progresas, mieux ce sera pour nous et surtout pour les gens qui en bénéficient. On se détournera de choses qui ne fonctionnent pas trop et qui sont de fausses bonnes idées. Ce sont les objectifs à plus long terme.

