

Savoir de quoi on parle : pourquoi les preuves ne voyagent pas toujours

Nancy Cartwright*

Article original : Cartwright, N. (2013). Knowing what we are talking about: why evidence doesn't always travel. *Evidence and Policy*, 9(1): 97-112. <https://doi.org/10.1332/174426413X662581> Cet article est republié en français avec l'autorisation de la maison d'édition The Policy Press. Les citations dans le texte sont des traductions libres.

RÉSUMÉ | Quand peut-on considérer qu'un résultat d'étude bien établi, selon lequel une certaine politique, ou un certain programme ou traitement, a produit un effet spécifique dans un contexte d'étude donné (« là-bas »), constitue une preuve que cette politique, ce programme ou ce traitement produira le même effet dans un nouveau contexte (« ici ») ? Cet article souligne l'importance de distinguer fermement « là-bas » de « ici » et répond que nous devons disposer de preuves attestant de deux faits supplémentaires : (a) que la politique peut assumer le même rôle causal de manière étendue (assez étendue pour englober à la fois « là-bas » et « ici ») et (b) qu'un ensemble complet de facteurs de soutien, nécessaires à l'opérationnalisation de la politique « ici », est présent chez certains individus « ici ».

MOTS CLÉS | Efficacité, inférence, validité externe

1 | Introduction

1.1 | Contexte

Imaginons qu'il a été démontré qu'une politique, un programme ou un traitement spécifique est efficace dans le cadre d'une étude de qualité. L'efficacité, comme cela doit être le cas dans toute étude, a été établie auprès d'une population spécifique dans un environnement spécifique. Comment les résultats obtenus dans cet environnement précis, pour cette population précise, peuvent-ils constituer une preuve solide permettant de prédire que cette politique, ce programme ou ce traitement sera efficace si nous décidons de l'appliquer ici, dans un nouvel environnement ? Cette interrogation est au cœur des politiques basées sur les preuves, où il est recommandé aux décideurs et analystes politiques, dans la mesure du possible, de n'adopter que des politiques ayant démontré des résultats positifs dans le cadre d'études rigoureuses destinées à prouver leur efficacité. On peut alors se demander : Quels types d'études sont considérés comme de « bonnes » études pour établir l'efficacité ? Il s'agit d'un sujet largement débattu, un débat auquel je ne participerai pas ici. Mon questionnement émerge après qu'il

* L'affiliation de l'autrice se trouve à la fin de l'article.

a été reconnu que nous disposons d'une étude sérieuse portant sur une population spécifique dans un contexte spécifique, et qui a montré que la politique a engendré l'effet recherché chez au moins certains individus de cette population. Cela constitue un bon point de départ pour deux raisons. Premièrement, de nombreux guides de politique basée sur les preuves recommandent ce point de départ; deuxièmement, il existe à présent plusieurs bases de données qui évaluent les programmes selon ces critères et qui rendent disponibles au public les informations concernant les programmes qui ont été validés. Alors, quand est-ce qu'un résultat positif issu d'une étude sérieuse constitue-t-il une preuve convaincante que la politique produira l'effet souhaité sur certains individus ici, dans notre contexte ?

1.2 | Méthodes

Il existe une abondante littérature concernant la transférabilité des résultats d'études établissant des liens de causalité, portant à la fois sur la nature de cette transférabilité et sur les moyens de la sécuriser. Toutefois, la majeure partie de cette discussion, comme il se doit pour les scientifiques en exercice qui progressent dans leur travail, se situe à un niveau relativement concret. Elle prend place en plein cœur du problème et emploie le jargon courant dans le domaine. Dans ce contexte, il est difficile de trouver des critères clairs pour évaluer les différentes propositions. Mon objectif de recherche est de fournir une assise solide pour ces débats, ancrée dans une compréhension plus fondamentale et, je l'espère, relativement non controversée des questions de causalité et de preuve. Cette assise fondamentale offre un cadre aux catégories de connaissances additionnelles nécessaires pour que les résultats des études démontrant l'efficacité d'un traitement quelque part puissent servir de preuve que le traitement fonctionnera également ici, dans notre nouveau contexte. Cela nous fournira un cadre pour développer des propositions de recherche concrètes qui établiront ce que nous devons savoir pour soutenir les affirmations selon lesquelles les résultats d'une étude peuvent être transposés dans un contexte nouveau spécifique. Donc cela nous dira, bien que de manière très abstraite, ce qu'il en coûte pour une mise en œuvre réussie. Je compte aborder ces questions à travers deux injonctions de philosophes :

- Clarifions ce que nous avançons.
- Déterminons ce qui convertit les résultats d'une étude en preuves.

2 | Que prétendons-nous ?

Les affirmations concernant l'efficacité dans des conditions idéales et contrôlées et les affirmations concernant l'efficacité dans des conditions réelles sont des revendications causales. Depuis longtemps, philosophes et scientifiques se sont interrogés sur ce qui constitue une relation causale plutôt qu'une simple association. Cette préoccupation est légitime, bien que je pense qu'elle puisse être abordée de manière satisfaisante de plusieurs façons différentes.¹ Ce n'est pas l'objet de ma préoccupation ici, et ce que je vais soutenir ne nécessite que des suppositions minimales et largement incontestées sur la causalité.² Le problème que je souhaite soulever est que nous ne portons pas suffisamment d'attention aux différences entre la portée et les types de revendications causales. Dans le contexte des politiques fondées sur des preuves, il est essentiel de distinguer trois types de revendications causales :

- La politique fonctionne quelque part (« là-bas ») : elle produit l'effet ciblé chez certains individus dans certains contextes.
- La politique fonctionne : elle produit l'effet ciblé de manière « étendue ».
- La politique fonctionnera « ici » : elle produira l'effet ciblé chez certains individus dans ce contexte précis.

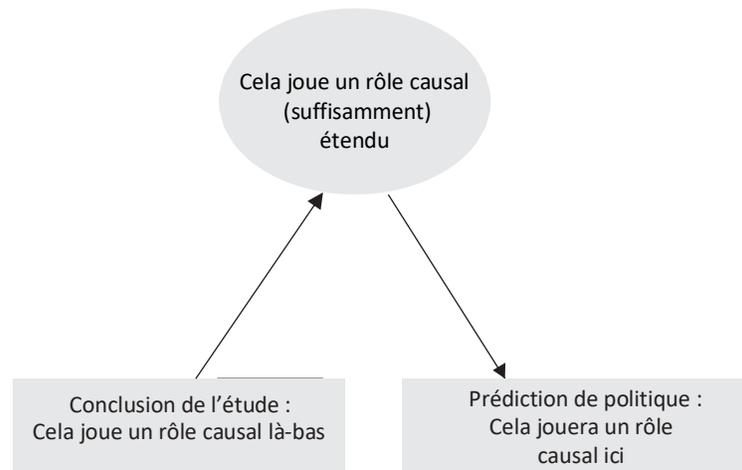
Il est trop courant que ces trois affirmations soient confondues, non seulement dans le discours public mais aussi dans des cercles politiques et académiques sérieux. Prenons, par exemple, le deuxième paragraphe d'un article de deux économistes distingués, Esther Duflo et Michael Kremer, qui prônent l'utilisation d'essais contrôlés randomisés (ECR) en économie du développement. Ils passent sans problème d'une affirmation à l'autre : « Les avantages de connaître *les programmes efficaces...* s'étendent bien au-delà de tout programme ou agence, et des *évaluations d'impact crédibles...* peuvent offrir une *orientation fiable* aux organisations internationales, aux gouvernements, aux donateurs, et... aux ONG [organisations non gouvernementales] au-delà des frontières nationales » (Duflo et Kremer, 2003 : 205, souligné par nous). Dans cette citation, je suppose que :

- les programmes efficaces = cela fonctionne « en général »;
- évaluation de l'impact = cela fonctionne quelque part;
- orientation fiable = cela fonctionnera ici.

Cet amalgame a son importance car des études individuelles peuvent justifier (a). Mais c'est (c) que nous visons. Alors, qu'est-ce qui fait de (a) une preuve pour (c) ? Généralement, la seule réponse disponible est (b). Un rôle causal étendu est le moyen par lequel (a) devient une preuve pour (c). Étendu, jusqu'où ? De manière assez large, au moins, pour inclure à la fois ici et là-bas. La figure 1 illustre comment les résultats de bonnes études peuvent servir de preuves pour prédire qu'une politique sera efficace dans un nouveau lieu spécifique.

3 | Qu'est-ce qui qualifie un résultat d'étude comme preuve en faveur d'une prédiction politique ?

Pour que les preuves se manifestent comme il est illustré dans la figure 1, deux faits immédiats doivent être avérés. Je les examine successivement dans les deux sous-sections suivantes.

Figure 1 | Circulation des preuves

3.1 | Facteurs de soutien

Le premier fait essentiel pour que les preuves circulent de manière adéquate est la présence, ici, des bons facteurs de soutien permettant à la politique d'être efficace dans notre nouveau contexte. Il ne suffit pas que ces facteurs de soutien soient présents ici pour revendiquer les résultats de l'étude comme preuves de l'efficacité de la politique dans le nouveau contexte. Nous avons besoin de solides justifications pour supposer qu'ils le sont.

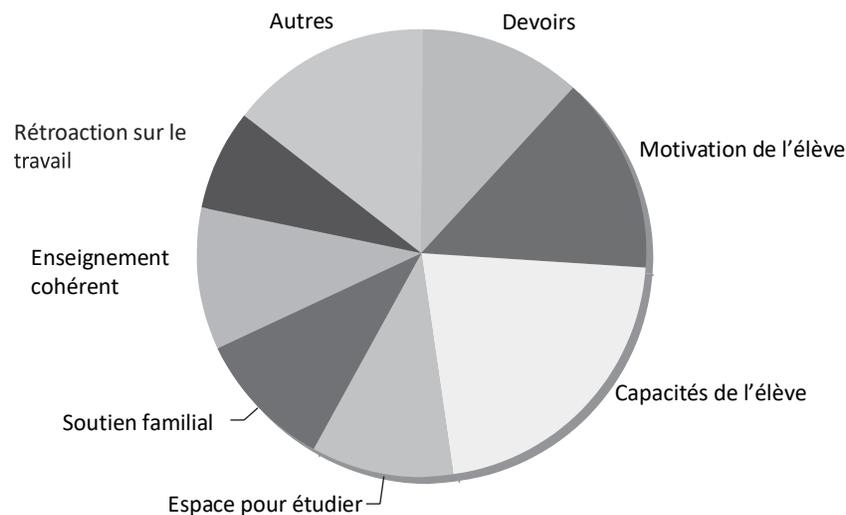
Qu'entend-on par « facteurs de soutien » ? Ce que nous identifions comme cause est rarement suffisante en elle-même pour produire l'effet ciblé. Les causes nécessitent le bon soutien pour agir comme attendu. En Californie, par exemple, la réduction de la taille des classes n'a pas réussi à améliorer les scores ciblés aux examens, malgré de bons résultats d'études sur la réduction de la taille des classes au Tennessee. Cela était dû en grande partie à l'insuffisance de personnel enseignant qualifié en Californie pour le nombre de classes en augmentation. Les petites classes, en elles-mêmes, ne suffisent pas à améliorer les résultats – un personnel enseignant compétent pour ces classes sont un facteur de soutien nécessaire à leur efficacité. Il en va de même pour les devoirs à la maison qui requièrent également de nombreux facteurs de soutien avant d'atteindre les résultats attendus.

Les épidémiologistes représentent parfois une équipe complète constituée d'une « cause saillante » et de ses facteurs de soutien dans des « tartes causales », ce qui peut être très utile pour illustrer l'ensemble des facteurs requis pour produire l'effet. La figure 2 montre un exemple de « tarte causale » pour les devoirs à la maison, basée sur les travaux du spécialiste de l'éducation Harris Cooper (Cooper, 2007). Il est important de noter qu'il peut exister plusieurs ensembles de facteurs pouvant soutenir une politique donnée – plusieurs « tartes » contenant la politique. Par exemple, la pression des pairs parmi les personnes étudiantes peut remplacer le soutien familial dans les écoles où le personnel et les élèves ont des attentes élevées en termes de réussite académique. Les essais contrôlés randomisés (ECR) sont souvent salués comme « l'étalon-or » pour les preuves dans les politiques basées sur des données probantes. Il y a de bonnes raisons à cela. Les ECR sont efficaces pour distinguer les relations causales

des simples associations. Idéalement, ils peuvent démontrer avec certitude que le facteur testé est non seulement associé au résultat ciblé dans la population et le contexte étudiés, mais qu'il en est également la cause chez certains individus dans l'étude. Nous savons que cela ne serait pas possible sans qu'au moins certains membres de la population étudiée ne possèdent un ensemble complet de facteurs de soutien nécessaires à l'action de la cause. La taille de l'effet observé dans l'étude est une moyenne des résultats de toutes les différentes combinaisons de facteurs de soutien présents chez les individus de la population étudiée. Cependant, le design de base d'un ECR ne révèle pas quels sont ces facteurs.

Pour que l'information selon laquelle la politique a été efficace chez certains membres de la population étudiée soit pertinente pour nos prédictions (c'est-à-dire que la politique fonctionnera ici), il est nécessaire de savoir que nous disposons ici des facteurs de soutien requis et en quantité suffisante pour que la politique soit valable. Cela constitue une information supplémentaire sans laquelle les prédictions positives en matière de politiques ne sont pas bien étayées. Et, en réfléchissant à ces facteurs et à leur présence dans le nouveau contexte, il est crucial de considérer que ce qui est nécessaire dans un contexte pour qu'une politique fonctionne peut différer grandement de ce qui est requis dans un autre contexte.

Figure 2 | Les devoirs et les résultats aux examens



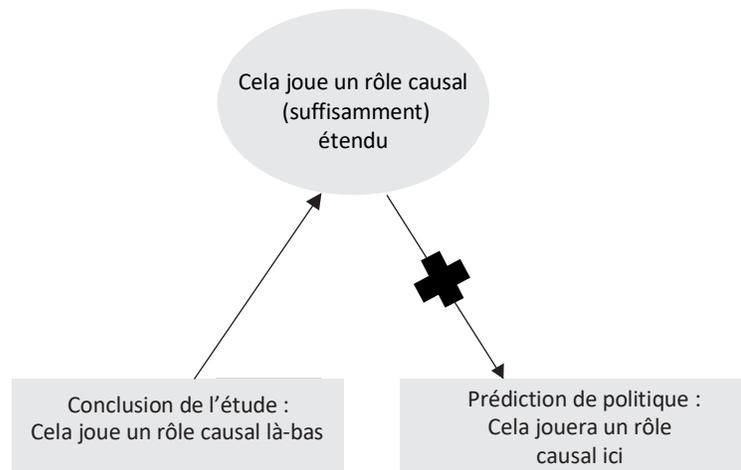
Je suis donc préoccupée par l'usage du terme « étalon-or » pour désigner les preuves dans le cadre des politiques fondées sur des preuves, tel qu'il est souvent appliqué aux résultats des ECR, car cette appellation peut induire un faux sentiment de sécurité. Un ECR peut constituer l'étalon-or pour affirmer que la politique est efficace au sein de la population étudiée dans le contexte de l'étude. Cependant, cela ne représente qu'une partie des preuves nécessaires pour soutenir la prédiction de son efficacité ici, dans notre contexte. À l'instar de l'économiste Edward Leamer, je suis préoccupée par le manque

d'attention accordé aux facteurs de soutien qui permettent à une politique d'être efficace. Dans un symposium sur l'utilité des ECR en économie, il déclare :

En l'absence d'observation des ... [facteurs de soutien], ce qui est estimé est en quelque sorte un effet moyen du traitement... Cela ressemble à l'avocat qui explique qu'étant jeune, il a perdu de nombreuses affaires qu'il aurait dû gagner, mais qu'avec l'âge, il en a gagné beaucoup qu'il aurait dû perdre, de sorte qu'en moyenne, la justice a été rendue... (Leamer, 2010 : 35)

La figure 3 illustre l'importance de se renseigner sur les facteurs de soutien dans le contexte ciblé avant de considérer les bons résultats d'une étude comme preuves que la politique sera efficace dans le nouveau contexte.

Figure 3 | Flux des preuves en l'absence de garantie que les facteurs de soutien soient présents



3.2 | Rôles causaux partagés

Le deuxième fait indispensable pour que les preuves puissent passer d'une population et d'un contexte d'étude à un nouveau contexte est que la politique doit pouvoir jouer le même rôle causal vis-à-vis du résultat dans les deux contextes. Il est crucial, tout comme pour les facteurs de soutien, d'avoir une garantie de cela avant de pouvoir affirmer que l'étude constitue une preuve en faveur de prédictions positives quant à l'efficacité de la politique dans le nouveau contexte. J'ai déjà employé l'expression « rôle causal », partant du principe que le lecteur en comprendrait le sens. Mais qu'est-ce que cela signifie exactement ? Le concept présuppose que les causes ne produisent pas des effets de manière aléatoire mais plutôt en accord avec des principes causaux, du moins dans les cas où nous pouvons espérer un certain succès systématique dans la prédiction. Si un facteur apparaît véritablement comme une cause dans un principe causal pour un effet donné, nous pouvons dire que ce facteur peut jouer un rôle causal par rapport à cet effet dans tout contexte ou population régis par ce principe causal. Cependant, comme discuté dans la sous-section précédente, les facteurs capables de jouer un rôle causal positif dans un contexte donné ne produiront l'effet ciblé que si les facteurs de soutien requis

sont présents, c'est-à-dire les facteurs de soutien nécessaires selon le principe causal qui prévaut dans ce contexte. Par conséquent, si certains individus dans un contexte donné possèdent une combinaison d'autres facteurs conformément au principe qui y prévaut, de sorte qu'augmenter (ou diminuer) la taille de la cause en question augmente (ou diminue) la taille de l'effet pour ces individus, nous dirons que le facteur *joue un rôle causal positif (ou négatif)* dans ce contexte. Selon cette approche, un même facteur peut jouer un rôle à la fois positif et négatif vis-à-vis du même effet dans le même contexte.

Les principes causaux adoptent diverses formes. Considérons une forme linéaire simple, fréquemment utilisée dans les modèles des sciences sociales :

$$z = a + bx + w.$$

Le facteur représenté par x peut exercer un rôle causal dans tout contexte régi par ce principe. Le coefficient b représente en un seul geste l'ensemble complet des facteurs de soutien nécessaires pour que x contribue de manière significative. x joue un rôle positif (ou alternativement négatif) dans tout contexte où les facteurs de soutien, pour au moins certains individus, génèrent une valeur positive (ou alternativement négative) pour b .

Les rôles causaux sont essentiels, car en matière de causalité, l'identique n'est pas toujours équivalent. Un facteur répondant exactement à la même description peut jouer des rôles très différents dans différents contextes. Par exemple, actionner un levier dans mes toilettes déclenche la chasse d'eau, parce qu'il est situé sur le réservoir. Actionner un levier très similaire dans ma cuisine grille mon pain du matin, parce qu'il fait partie de mon grille-pain. Envoyer les pères à des cours de parentalité peut améliorer le bien-être des enfants dans un groupe culturel, car cela améliore leur compréhension de la gestion des comportements de leurs enfants. Dans un autre, cela peut dégrader le bien-être des enfants, car cela est perçu comme une humiliation publique pouvant engendrer une colère qui se répercute sur l'enfant. L'éducation des mères sur la nutrition infantile a eu un rôle causal positif dans l'amélioration de la nutrition de leurs enfants au Tamil Nadu mais pas généralement au Bangladesh, car là-bas, les mères ne sont souvent pas responsables de l'achat ou de la distribution de la nourriture au sein de la famille, où ce sont les hommes qui font les courses et les belles-mères qui décident de la répartition (Banque mondiale, 1995).

Une fois énoncé, ce point semble évident et familier. Pourtant, il n'occupe généralement pas le devant de la scène, comme il le devrait, dans les discussions sur le moment (ou la possibilité) d'utiliser un résultat d'étude sur l'effet d'une politique comme preuve pour prédire l'effet dans un nouveau contexte. Sans informations confirmant que la politique peut jouer le même rôle dans le nouveau contexte que dans le contexte de l'étude, nous ne pouvons pas déterminer la pertinence du résultat de l'étude. Si la politique n'assume pas le même rôle causal dans les deux contextes, les résultats de l'étude n'ont aucune incidence sur la véracité des prédictions concernant ses effets dans le nouveau contexte.

3.3 | Pas de manuels pour trouver ce dont nous avons besoin

Comme mentionné précédemment, pour que les informations relatives au résultat de l'étude soient pertinentes pour les prédictions dans le nouveau contexte, nous devons connaître les facteurs de soutien

nécessaires dans notre nouveau contexte pour que la cause soit efficace ici, ainsi que s'assurer qu'au moins certains individus dans notre contexte disposent de l'ensemble complet des facteurs de soutien requis ici. Il a été ajouté que nous devons également confirmer que la politique joue le même rôle causal ici, dans le nouveau contexte. Sans preuve de ces trois affirmations, nous n'avons aucune base pour considérer le résultat de l'étude comme preuve de ce que la politique produira ici, dans notre nouveau contexte.

Où trouver ces preuves supplémentaires ? Nous avons besoin d'informations sur les rôles causaux que les politiques reconnues efficaces quelque part peuvent jouer ailleurs, sur leurs facteurs de soutien dans chaque nouveau contexte où nous envisageons de mettre en oeuvre la politique, ainsi que sur la présence des facteurs de soutien nécessaires dans ces nouveaux contextes. Ces différentes catégories d'informations nécessitent un grand nombre d'outils, de théories et de méthodes scientifiques issus de disciplines variées. Il n'existe aucun système pour déterminer précisément ce qui est nécessaire ni comment en garantir la qualité.

Il y a des cours de méthodologie et des manuels qui enseignent comment réaliser des méthodes d'enquête spécifiques, allant des entretiens structurés aux ECR. Cependant, il n'existe aucun manuel pour indiquer comment définir des concepts significatifs applicables de manière générale, ni comment identifier les facteurs de soutien nécessaires dans des contextes donnés, ni quels rôles causaux une politique spécifiquement décrite peut jouer dans différents contextes ou, à l'inverse, étant donné un facteur décrit de manière abstraite connu pour jouer un rôle causal de manière générale, comment déterminer ce que cela implique dans un nouveau contexte. Par exemple, il est assez largement admis que l'humiliation publique des parents « défailants » tend à augmenter les cas de maltraitance des enfants, mais qu'est-ce qui est considéré comme une humiliation publique dans cette culture ? Dans le même sens, qui incarne la description « responsable de l'approvisionnement et de la distribution alimentaire » afin que nous puissions offrir des conseils nutritionnels aux personnes appropriées, comme l'analyse post hoc le suggère pour le Tamil Nadu et non pour le Bangladesh ? Le fait qu'il n'existe pas de manuel sur la façon de procéder n'implique pas, bien sûr, que nous ne disposons pas de bonnes méthodes au cas par cas. Néanmoins, pour ceux qui aimeraient des règles générales pour déterminer ce qui constitue une preuve solide de l'efficacité, c'est une conclusion décourageante.

4 | Une objection à mes affirmations sur les rôles causaux, une réponse et une ironie méthodologique

On pourrait contester mes affirmations sur la nécessité d'informations concernant les rôles causaux partagés, en arguant que mes exemples, destinés à illustrer à quel point le rôle causal joué par un même facteur peut varier, ne décrivent pas suffisamment les causes. Je concède que les interprétations changent avec la description. En effet, si j'appuie sur le levier de mes toilettes ou sur celui au sol de ma voiture sous le volant, cela ne grille pas mon pain, contrairement à ce qui se passe lorsque j'appuie sur le levier semblable de mon grille-pain. Avec ces descriptions revisitées, nous ne traitons finalement pas de la « même » cause. Si nous qualifions les cours non simplement de cours de parentalité mais de cours de parentalité qui constituent (ou non) une humiliation publique, alors nous ne sommes plus confrontés à des effets différents issus des « mêmes » causes.

Cela est vrai, mais cela ne dissipe pas mes inquiétudes. Nous testons toujours une relation causale sous une certaine description de la cause et de l'effet. Dans les études scientifiques empiriques établissant des liens causaux, il est crucial de décrire aussi précisément que possible les causes et les effets pour garantir que nous savons exactement ce que nous testons. Supposons que les résultats soient positifs. Ce que nous apprenons, c'est que cette cause (celle précisément décrite dans le protocole de l'étude) provoque l'effet ciblé dans le contexte de l'étude. Ce que l'étude ne nous apprend pas, c'est que cette cause, celle que nous avons décrite et testée, peut jouer le même rôle causal de manière universelle, étendue ou presque nulle part ailleurs. Cette information doit provenir d'une source externe à l'étude.

C'est là l'ironie de la méthodologie. Une procédure scientifique rigoureuse pour tester les affirmations causales exige des descriptions précises et concrètes des variables de traitement et de résultat. Cependant, les rôles causaux largement partagés impliquent souvent des concepts bien plus généraux et abstraits. La concrétisation/opérationnalisation de ces concepts abstraits peut varier considérablement d'un cas à l'autre. La poignée du réservoir des toilettes et celle de mon grille-pain, la balançoire qui soulève mes enfants dans le parc, l'accélérateur au sol de ma voiture et la fourchette avec laquelle je mange sont tous des exemples concrets de leviers, bien qu'ils puissent différer lorsqu'ils sont décrits de manière plus détaillée. La personne qui distribue la nourriture au Tamil Nadu est, de façon plus précise, une mère; au Bangladesh, c'est la belle-mère. Et l'humiliation publique revêt clairement des formes très différentes selon les contextes culturels. Pour assurer la rigueur et la validité, nous sommes contraints de tester des affirmations causales impliquant des descriptions très spécifiques des causes et des effets. Mais ces descriptions ne correspondent probablement pas à celles selon lesquelles ces mêmes causes, qui opèrent dans l'étude, peuvent jouer un rôle causal étendu.

Ce problème est aggravé par le fait que dans la plupart des cas où les politiques produisent les résultats escomptés, cela se fait via plusieurs étapes intermédiaires. Souvent, comme dans le cas de l'actionnement de la chasse d'eau, pour chaque étape reliant la cause initiale (appuyer sur le levier) à l'effet final (l'actionnement de la chasse), il existe une description de la cause à cette étape selon laquelle cette cause joue un rôle causal très général – appuyer sur le levier du réservoir pour soulever une tige à l'autre extrémité du levier n'est qu'un exemple de l'utilisation d'un levier pour soulever un objet. Souvent, la séquence implique une série de causes qui, une à une et décrites de manière abstraite, jouent le même rôle causal de manière générale que dans la séquence considérée. Mais ce sont les arrangements très spécifiques (la construction des toilettes ou les normes familiales et culturelles) qui permettent la succession de chaque étape. De ce fait, la cause initiale ne joue pas un rôle général vis-à-vis de l'effet final. Pousser sur un levier qui ressemble à celui de mes toilettes, même si nous le décrivons comme « actionner une extrémité d'un levier », ne joue aucun rôle causal général dans l'actionnement de la chasse. Pousser sur un levier ne joue un rôle causal dans l'actionnement de la chasse que dans une gamme très restreinte de contextes où le levier est attaché à une tige dans des toilettes correctement conçues.

5 | Cumuler les essais contrôlés randomisés (ECR)

Découvrir quels facteurs de soutien sont nécessaires et dans quelle mesure les rôles causaux se généralisent est une tâche complexe et sans fin précise. Il serait idéal de pouvoir éviter cela, surtout si

nous pouvions limiter nos problèmes à l'amélioration d'un ensemble défini de méthodes que nous pourrions lister et surveiller, comme celles présentées dans les hiérarchies de preuves typiques des guides sur les politiques basées sur les preuves. Les ECR sont habituellement la méthode de choix pour ceux qui prônent cette stratégie. Il existe trois propositions courantes sur comment utiliser les ECR à cette fin. Cependant, aucune d'elles n'est suffisante. La première proposition est de réaliser de plus en plus d'études – les ECR étant le design le plus recommandé – pour établir l'efficacité dans différents contextes d'étude. La seconde est de réaliser une étude rigoureuse ici, dans le contexte même où nous souhaitons faire des prédictions, où, encore une fois, un ECR est le design d'étude privilégié. La troisième est d'effectuer une méta-analyse statistique de différentes études de qualité, particulièrement des ECR. Je vais brièvement aborder chacune de ces propositions.

5.1 | Que peut-on conclure de multiples observations du même résultat ?

Les ECR sont souvent défendus par ceux qui se méfient de la théorie. Ils considèrent nos prétentions à la connaissance théorique comme trop incertaines, ils ne veulent tout simplement pas s'y fier. Ils résistent donc à des affirmations comme la mienne, selon laquelle il est nécessaire d'utiliser un large éventail de méthodes et de connaissances de base, y compris une bonne dose de théorie, pour justifier l'utilisation du succès d'une politique dans une étude comme preuve de son succès dans un nouveau contexte. Leur proposition alternative est de multiplier les ECR, avec autant de variations de circonstances que possible. Je conviens que davantage d'ECR, surtout dans une variété de circonstances, peuvent renforcer la justification d'une prédiction d'efficacité. Ils le font en soutenant l'hypothèse que la politique peut jouer un rôle causal positif ici, dans notre nouveau contexte. Comment ? C'est là que les choses se compliquent. L'argument pourrait reposer sur une simple induction énumérative : le cygne 1 est blanc, le cygne 2 est blanc..., donc tous les cygnes sont blancs; x peut jouer un rôle causal positif dans la situation 1, x peut jouer un rôle causal positif dans la situation 2..., donc x peut jouer un rôle causal positif partout.

Quelle est la solidité de cet argument ? Pour une bonne induction, nous avons besoin non seulement d'une base inductive large et variée (de nombreux cygnes, de nombreux endroits, de nombreux ECR réalisés dans différentes populations et différents contextes) ; nous avons également besoin de raisons de croire que les observations sont extrapolables, et d'une explication de l'étendue de l'extrapolation. Une caractéristique est extrapolable à une variété d'exemples s'il y a quelque chose dans chaque exemple de la population qui garantit qu'ils doivent toutes être identiques par rapport à cette caractéristique. La charge d'un électron est extrapolable partout (une bonne expérience suffit pour généraliser à tous les électrons); la couleur des oiseaux l'est parfois; la causalité, c'est plus complexe. Comme il est mentionné dans la section précédente, de nombreux liens causaux dépendent d'interactions intimes et complexes entre les facteurs présents, de sorte qu'aucun rôle spécial pour le facteur d'intérêt ne peut être isolé et projeté dans de nouvelles situations. Par conséquent, on ne peut pas supposer que les connexions causales sont extrapolables – que ce qui se passe dans un contexte est ce qui se passera toujours. Cela a déjà été bien illustré. Appuyer sur un levier peut vous permettre d'accélérer, mais seulement si c'est la bonne pédale au sol d'une voiture, pas si elle est attachée à un grille-pain.

Parfois, l'induction par simple énumération est une bonne manière de raisonner. Mais très souvent, ce n'est pas le cas. L'induction sur les liens causaux établis dans les ECR donnera des conclusions correctes seulement si, en rapportant les résultats des ECR, nous avons identifié une cause qui peut jouer le même rôle de manière générale, ou du moins suffisamment générale pour atteindre les nouveaux contextes envisagés. Pour cela, deux types d'informations sont nécessaires. Premièrement, cela nécessite de trouver la bonne description de la paire cause-effet, une qui se maintiendra dans d'autres contextes : quelle est la caractéristique de la cause qui lui permet de produire cet effet ? Deuxièmement, cela nécessite d'identifier les circonstances sous-jacentes qui permettent à cette relation de cause à effet entre les causes et les effets ainsi caractérisés de se maintenir : quelles sont les circonstances qui rendent possible la connexion causale entre le fait d'appuyer sur un levier et d'accélérer ou entre l'éducation des mères à la nutrition et une meilleure nutrition des enfants ?

Il existe essentiellement deux approches ici. Nous pouvons tenter le coup. Nous pouvons utiliser des descriptions qui semblent justes et varier les circonstances de la manière que nous jugeons pertinente, puis espérer le meilleur. Ou nous pouvons essayer de comprendre de manière systématique et fondée ce qui permet à la cause de produire son effet et quels arrangements de circonstances sous-jacentes rendent cette connexion causale possible. Trop souvent, nous optons pour la première option, parfois parce que nous le devons. Nous devons prendre des décisions sur les actions à entreprendre dans de nouveaux contextes avant de disposer du type d'informations nécessaires pour justifier la confiance dans nos prédictions. C'est la condition humaine. Mais c'est une réalité à laquelle nous devons faire face, et non pas détourner le regard. Si nous manquons d'informations soutenant une induction, nous n'avons pas de solides raisons de supposer que la conclusion de cette induction est vraie. Cela peut être notre meilleur pari, mais cela ne le rend pas pour autant un bon pari. Nous pouvons avoir les résultats de certaines bonnes études. Mais cela ne constitue pas une preuve solide de notre conclusion si nous n'avons pas également de bonnes raisons de soutenir les hypothèses supplémentaires sur les rôles causaux nécessaires pour rendre ces résultats d'étude pertinents de manière générale, ou au moins pour les nouveaux contextes envisagés.

5.2 | Des ECR locaux pour des prédictions locales

Certes, la meilleure preuve que la politique fonctionnera ici est un ECR sur cette politique mené localement, du moins selon la majorité des partisans des ECR. Ainsi, la seconde stratégie pour pallier notre manque de connaissances sur les rôles causaux et les facteurs de soutien consiste à réaliser un bon ECR localement. Je concède que cela constituerait une bonne preuve – ne nous disputons pas sur le terme « meilleure ». *Cela serait idéal*, si cela était réalisable. D'abord, il y a le problème déjà mentionné : sans informations extérieures à l'étude – émanant notamment d'une théorie solide – nous ne savons jamais ce qui constitue « exactement la même politique ». Ensuite, un ECR local, c'est-à-dire réalisé sur la population même à laquelle nous nous intéressons et au même moment, ou sur un échantillon représentatif de cette population, est pratiquement impossible. Et à la fois la population exacte et le moment peuvent être déterminants. Un échantillon n'est presque jamais représentatif. Être représentatif signifie être régi par les mêmes principes causaux et présenter la même distribution de probabilité sur les facteurs causaux pertinents. Et le temps ne peut certainement pas être négligé. Les causes sont-elles les mêmes aujourd'hui qu'au moment où l'étude a été réalisée ? C'est une question particulièrement critique pour les programmes socioéconomiques, comme l'ont souligné des

économistes allant de J.S. Mill à l'éminent économètre britannique David Hendry, ainsi que Robert Lucas et d'autres membres de l'« École de Chicago » en économie. Les régularités passées sont souvent de mauvais indicateurs pour l'avenir en économie, car l'agencement des causes change fréquemment et de manière imprévisible. Bien sûr, la population expérimentale pourrait être suffisamment représentative et les causes suffisamment stables. Mais cela ne peut pas être présumé. C'est une hypothèse qui nécessite des preuves solides avant que nous puissions avoir une confiance élevée dans nos prédictions concernant l'efficacité de la politique ici.

5.3 | Méta-analyse

La méta-analyse désigne un ensemble de techniques statistiques destinées à amalgamer les résultats de différentes études menées sur diverses populations dans différents contextes. De nombreux problèmes se posent. Rarement les différentes études ont testé exactement la même relation causale – avec le même protocole pour le traitement dans toutes les études et les mêmes procédures de définition et de mesure de l'effet, elles n'ont pas toutes assigné les personnes participantes de manière aléatoire ou masqué les informations de manière également satisfaisante, ni géré les abandons de la même façon, etc. Supposant que tous ces problèmes puissent être suffisamment gérés, la méta-analyse est surtout efficace pour fournir une taille d'échantillon plus grande, rendant ainsi les résultats positifs moins susceptibles d'être le fruit du hasard. Cela signifie que si nous trouvons un effet positif dans la population amalgamée, nous pouvons être plus confiants que pour n'importe quelle population contributive. Mais ce dont nous pouvons être sûrs, c'est que la cause à l'étude a produit l'effet chez certains membres d'au moins certaines des populations contributrices. Cela ne nous aide pas à déterminer quels sont les facteurs de soutien ou ce qui caractérise les populations dans lesquelles la relation causale se maintient.

Nous pouvons bien entendu procéder à une analyse de sous-groupes à ce stade, tout comme nous le ferions pour chaque étude séparément, en cherchant à voir comment les résultats des différentes études se comparent dans l'espoir de déterminer ce qui caractérise les populations où la relation causale se vérifie, ou chercher à identifier les individus ayant obtenu des résultats positifs afin de déterminer quels sont les facteurs de soutien (tout en gardant à l'esprit que dans différentes populations, dans différents contextes, les facteurs de soutien peuvent varier). Cependant, sans une solide assise théorique ou des études supplémentaires spécifiquement conçues pour tester ces hypothèses, toute conclusion restera spéculative, peut-être utile, mais spéculative quand même.

6 | Pourquoi ne pas parler de « validité externe » ?

Vous vous demandez peut-être si tout ce dont j'ai parlé jusqu'à présent ne concerne pas simplement le problème bien connu de la *validité externe*. En partie oui, en partie non. Dans la mesure où la réponse est affirmative, il est crucial de ne pas laisser la familiarité engendrer le mépris. Pourquoi « non » ? Selon la vision traditionnelle, une étude possède une *validité externe* lorsque le « même traitement » produit le « même résultat » ailleurs qu'il ne l'a fait dans l'étude. Le conseil classique est que l'on peut s'attendre à une validité externe si le contexte ciblé ressemble au contexte de l'étude. Mais quelle est l'efficacité de ce conseil pour passer de « Cela a fonctionné là-bas » à « Cela fonctionnera ici » ? Ma réponse : c'est une chance si cela vous mène quelque part.

Premièrement, ce conseil est vague, étonnamment vague étant donné la spécificité de nos directives pour évaluer les études elles-mêmes.

Deuxièmement, beaucoup repose sur la bonne interprétation de ce que signifie « même traitement ». Utiliser le même traitement est acceptable, à condition d'avoir identifié la bonne description du traitement, comme je l'ai souligné auparavant. Et la bonne description est celle qui joue le même rôle causal ici que dans l'étude.

Nous ne devrions pas non plus nous attendre au « même résultat », sauf si nous entendons par-là quelque chose de très général, et certainement pas si le résultat doit être l'effet du traitement : la différence dans la moyenne des résultats entre le groupe de traitement et le groupe témoin. Sauf cas de chance exceptionnelle, l'effet du traitement sera le même dans l'étude et dans un nouveau contexte seulement si la politique peut jouer le même rôle causal dans les deux contextes, si les facteurs de soutien sont les mêmes et si la distribution de leurs valeurs est identique. Cela représente un véritable défi.

Finalement, viser le même effet de traitement est inefficace. L'effet du traitement est moyen sur l'ensemble des arrangements pour les facteurs de soutien. Certains de ces arrangements permettent à la politique de faire une grande contribution, d'autres une petite contribution. Et pour d'autres, la politique peut même s'avérer contre-productive. Vous ne devriez pas viser le même mélange de ces arrangements dans votre contexte que dans l'étude. Au contraire, vous devriez rechercher un bon mélange, c'est-à-dire un mélange qui se concentre sur les arrangements permettant à la politique d'être le plus bénéfique pour vous.

7 | Et la fidélité ?

Lorsqu'une politique a été démontrée efficace quelque part par de bonnes études, ou mieux, dans plusieurs endroits, on vous incite parfois à mettre en œuvre la politique exactement comme elle a été réalisée dans les situations d'étude. C'est le principe de *fidélité*. Par exemple, les directives actuelles de la Greater London Authority (GLA, 2010) pour l'adoption de politiques sociales à Londres soulignent que si vous souhaitez utiliser une intervention efficace ailleurs, vous devriez l'exécuter exactement comme elle a été réalisée là où elle a réussi. Certains fournisseurs commerciaux de programmes d'éducation et autres politiques insistent pour que la mise en œuvre soit exactement conforme à leurs directives. La raison est évidente. Ils ne peuvent être tenus responsables d'un échec si les instructions ne sont pas suivies. Le gâteau ne lèvera pas si vous ne suivez pas la recette à la lettre.

Il y aura des cas où l'analyse d'un programme échoué mettra effectivement en évidence le non-respect des instructions comme problème. L'erreur en Californie a été de penser que le succès dans le Tennessee était en soi une bonne preuve de succès potentiel en Californie. Un des principaux problèmes était le manque de personnel enseignant qualifié et de salles de classe, alors que ces aspects avaient été bien gérés dans le Tennessee. Ainsi, une mise en œuvre fidèle de la politique telle qu'elle avait été menée dans le Tennessee aurait très bien pu fonctionner, et le non-respect des instructions en était le problème.

Le cadre théorique que j'ai présenté aide à clarifier les questions de fidélité. Pour que la politique fonctionne ici, elle doit jouer un rôle causal positif et les facteurs de soutien nécessaires à son fonctionnement doivent être présents. Le conseil de « rester fidèle » au protocole de l'étude suppose donc que la politique joue le même rôle causal ici que dans l'étude, que la politique intègre des facteurs de soutien qui seront également nécessaires dans les deux endroits et que tout facteur de soutien supplémentaire nécessaire ici, qui n'est pas intégré dans la politique, sera présent. Si l'on a de bonnes raisons de croire tout cela, alors rester fidèle est un très bon conseil. Sinon, en modifiant le programme, vous pourriez éliminer des facteurs de soutien nécessaires ou changer le rôle causal que le programme peut jouer. Vous pourriez même transformer un programme bénéfique en fléau. L'exemple du programme nutritionnel au Bangladesh montre que la fidélité n'est peut-être pas la meilleure approche. Même si au Bangladesh, on avait fait exactement comme au Tamil Nadu, cela aurait échoué car les mères n'ont pas le même rôle causal en matière de nutrition dans les deux endroits.

Si l'appel à la fidélité ne signifie pas « faire exactement la même chose de la même manière » mais plutôt agir conformément au principe de haut niveau illustré par le succès précédent, alors c'est acceptable, à condition que les deux contextes relèvent des mêmes principes causaux. Mais il faut alors beaucoup travailler pour identifier ce principe de haut niveau, vérifier s'il s'applique suffisamment au nouveau contexte et comprendre ce qui incarne les concepts de haut niveau dans ce nouveau contexte. Et là encore, il n'y a pas de manuels pour cela. Cela relève de la véritable découverte scientifique.

8 | Résultats et implications

Des études de qualité peuvent prouver qu'une politique sera efficace dans notre contexte. Elles le peuvent si :

1. la politique peut jouer le même rôle causal de manière étendue (au moins assez étendue pour englober à la fois ici et là-bas);
2. tous les facteurs de soutien nécessaires au fonctionnement de la politique dans notre contexte sont présents.

Pour être justifiés dans la prédiction que la politique sera efficace ici, nous avons besoin de bonnes raisons pour soutenir à la fois les points (a) et (b).

Cette description de ce qu'il faut pour qu'une étude fournisse des preuves qu'une politique sera efficace ici est fondée théoriquement : d'abord, sur l'idée fondamentale que les causes agissent en accord avec les principes de causalité, là où une certaine prévisibilité systématique est attendue, ensuite, sur une explication de ce qui constitue un principe de causalité – en se concentrant particulièrement sur les facteurs de soutien nécessaires selon le principe pour que la variable de politique produise ses effets. Cette description est cependant très abstraite. Et j'ai souligné qu'il n'existe pas de manuel pour « concrétiser » cela, pour acquérir les types d'informations requises en (a) et (b).

Cela ne signifie pas qu'il n'existe pas de stratégies. Et le fait d'être clair, bien que de manière abstraite, sur ce que nous devons établir, offre un bon point de départ pour développer ces stratégies et une bonne base pour les évaluer. Nous avons vu précédemment que s'appuyer principalement sur les ECR fournira rarement une stratégie suffisamment bonne. Cela découle de la logique du cadre de base. Qu'en est-il des méthodes plus positives ? Le cadre ne les dictera pas, mais il montrera si elles répondent aux bonnes questions. Examiner les propositions actuelles pour voir si elles aident ou comment elles aident à répondre aux questions nécessaires pour que les résultats d'une étude puissent être transposés dépasse largement mon objectif ici. Mais laissez-moi évoquer de manière succincte quelques exemples où la connexion peut être assez directe.

Premièrement, les effets de la plupart des politiques ne sont attendus qu'un certain temps après leur mise en œuvre. Si la politique doit jouer son rôle causal attendu par rapport à l'effet final, il doit y avoir un processus causal continu entre les deux. Si l'une des étapes de ce processus causal est rompue, il est peu probable que la politique atteigne l'effet ciblé. Il est donc important de détailler étape par étape comment ce processus est censé se produire dans le nouveau contexte ici, pour voir si la politique peut jouer le rôle causal attendu. Le suivi des processus est également extrêmement utile pour identifier les facteurs de soutien : ce qui doit être mis en place à chaque étape pour qu'elle mène à la suivante comme requis. Ce type de suivi des processus est l'une des actions qui pourraient être entreprises sous l'étiquette de « modèle logique » pour la politique. C'est aussi souvent ce qui est fait dans l'analyse de séries temporelles, comme en économétrie, et dans la modélisation des réseaux bayésiens causaux.

Modéliser les structures sociales sous-jacentes qui rendent certains processus causaux possibles et d'autres impossibles est une autre stratégie qui gagne en attention. Dans toutes les sciences, de la biologie neuronale à la recherche sur la dépendance aux drogues, les philosophes de la science étudient comment modéliser les « mécanismes » ou « machines nomologiques » sous-jacents qui donnent lieu à des processus causaux reproductibles. Plus proche de la pratique dans les politiques basées sur des preuves, un travail similaire est central à ce qu'on appelle l'« évaluation réaliste ». Le nouveau livre de Pawson (2013, à paraître) – *The Science of Evaluation: A Realist Manifesto* – est particulièrement efficace pour utiliser cette méthode afin de suggérer des manières de gérer des interventions complexes. Mais on ne peut attendre d'aucune intervention (ou aspect d'une intervention), qu'elle soit complexe ou simple qu'elle produise les mêmes résultats, ou le même sous-ensemble de résultats, dans un nouveau contexte comme celui d'une étude, à moins que les bonnes informations sur les rôles causaux et les facteurs de soutien ne soient en place.

Pour un autre exemple, considérez le conseil de rechercher des éléments communs dans les programmes qui ont prouvé être efficaces quelque part, ce qui est maintenant un sujet de débat central pour les interventions sociales et psychologiques. Rechercher des éléments communs peut être un bon pas vers la découverte de concepts généraux capables de jouer le même rôle causal de manière étendue.

Je termine par deux figures qui illustrent graphiquement les conclusions essentielles découlant du cadre fondamental présenté ici. La figure 4 illustre les types de soutien probant requis pour une justification solide.

La figure 5 montre à quel point notre base de preuves est instable lorsque nous tentons de nous appuyer sur les ECR de référence sans fournir le soutien approprié pour les hypothèses nécessaires concernant les rôles causaux et les facteurs de soutien.

Notes

- ¹ Pour une discussion approfondie de ce sujet, voir Cartwright (2007) ainsi que les références aux auteurs mentionnés dans cet ouvrage.
- ² Une formulation explicite de ces hypothèses est disponible dans Cartwright et Hardie (2012), ouvrage qui constitue la source principale du matériel de cet article et où l'on peut trouver une explication plus détaillée des théories de la preuve et de la causalité décrites ici.
- ³ Ils sont ainsi nommés parce que les processus causaux répétables qu'ils génèrent sont décrits dans des principes causaux ou des « lois » causales.

Figure 4 | Ce qui est nécessaire pour une prédiction d'efficacité bien étayée

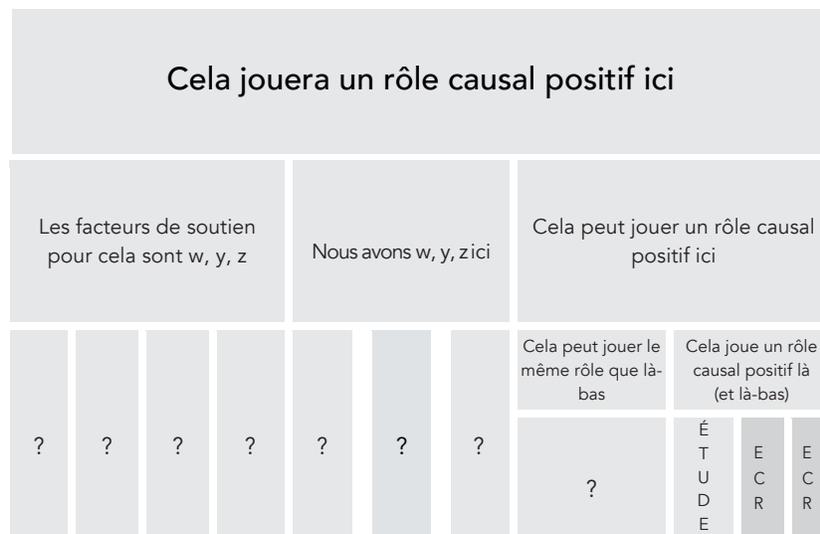
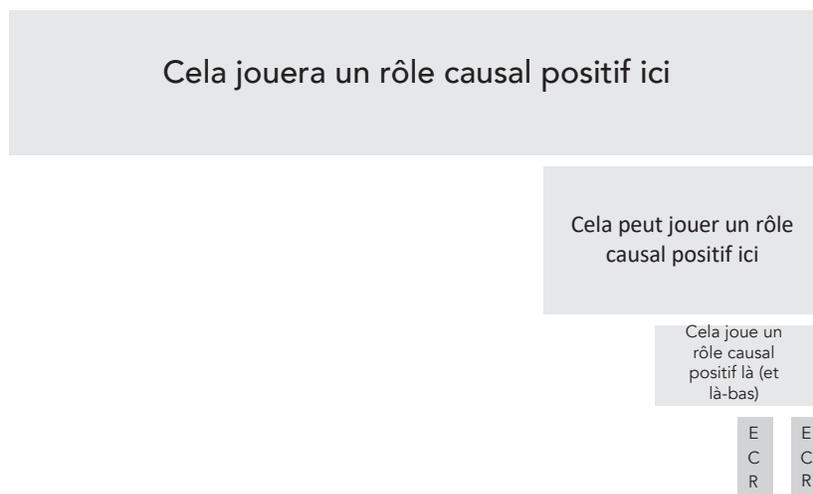


Figure 5 | Pourquoi les résultats des ECR seuls ne suffisent pas à des prédictions d'efficacité bien étayées



AFFILIATION DE L'AUTRICE

Nancy Cartwright Ph.D.

Département de philosophie | Université de Durham, Royaume-Uni

TRADUCTION

L'article a été traduit en suivant ces étapes : 1) Utilisation de la traduction automatisée avec le logiciel DeepL Pro; 2) Amélioration de la traduction automatisée à l'aide d'un modèle GPT développé par Mathieu Ouimet; 3) Révision humaine par Mathieu Ouimet, Julie Desnoyers et Aurélie Hot.

RÉFÉRENCES

Cartwright, N. (2007) *Hunting causes and using them: Approaches in philosophy and economics*, Cambridge: Cambridge University Press.

Cartwright, N. et Hardie, J. (2012) *Evidence-based policy: A practical guide to doing it better*, New York, NY: Oxford University Press.

Cooper, H. (2007) *The battle over homework: Common ground for administrators, teachers, and parents* (3rd edition), Thousand Oaks, CA: Corwin Press.

Duflo, E. et Kremer, M. (2003) Use of randomization in the evaluation of development effectiveness. Dans G. Pitman, O. Feinstein and G. Ingram (éditeurs) *Evaluating development effectiveness*, New Brunswick, NJ: Transaction Publishers, 205-32.

GLA (2010) *Standards of Evidence the Greater London Authority*, Oracle Project, London, UK: Greater London Authority.

Leamer, E. (2010) Tantalus on the road to asymptopia, *Journal of Economic Perspectives*, 24 (2): 31-46.

Pawson, R. (2013: forthcoming) *The science of evaluation: A realist manifesto*, London: Sage Publications.

World Bank (1995) *Tamil Nadu and child nutrition: A new assessment*, Washington, DC: World Bank.

CITATION SUGGÉRÉE

Cartwright, N. (2024). Savoir de quoi on parle : pourquoi les preuves ne voyagent pas toujours.
Revue sur le transfert et l'utilisation des connaissances, 8(3).
<https://doi.org/10.18166/tuc.2024.8.3.47>



ISSN | 2369-8896

www.revue-tuc.ca



Cet article est mis à disposition selon les termes de la Licence Creative Commons
Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International