

## La causalité, qu'es aquo ?

*The book of Why*, par Judea Pearl et Dana Mackenzie

### 1. Causalité, terra incognita ?

Le rossignol chante avant le lever du soleil mais bâillonner le rossignol n'empêchera pas le soleil de se lever ! On le sait, *corrélation n'est pas causalité*, les statisticiens rappellent avec ténacité ce point, véritable pont-aux-ânes de l'analyse de données.

Fort bien, mais alors quelle part de la science aborde la question de la causalité ? L'affaire est paradoxale : un enfant de cinq ans évoque la causalité (« Je me suis fait mal *parce que* je suis tombé ») mais pas les scientifiques qui se réfugient derrière deux mantras « 1 : *Corrélation n'est pas causalité* » et « 2 : *La science s'intéresse au comment, laissant le pourquoi aux religions* ».

La notion de causalité, pourtant abordée par les philosophes, est-elle-même seulement définie quelque part dans le corpus scientifique ? Quels outils ont été développés pour démontrer l'existence d'une relation causale entre deux événements ? Jusque naguère, les statisticiens en avaient construit un, *presque unique (!)*, souvent connu sous le nom vernaculaire de test en double aveugle.

On tire au sort deux groupes d'individus, à l'un on donne le médicament à évaluer, à l'autre le placebo, ni les « patients » ni les chercheurs ne savent à qui on donne quoi et on compare les résultats. Excellent, mais la méthode n'est pas toujours applicable. Comment par exemple démontrer ainsi l'influence d'un usage du tabac sur une vingtaine d'années ? Pratique et éthique en conviennent : on ne peut évidemment pas ordonner aux membres d'un échantillon test (tiré au sort) de fumer  $n$  cigarettes par jour durant les vingt ans à venir !

Indépendamment même des contre-feux scandaleux allumés par les cigarettiers durant des décennies pour faire avaler la fumée ... et les bobards, les scientifiques ont bataillé des années durant pour trouver une méthodologie permettant de démontrer l'impact du tabac. Comment par exemple éliminer l'hypothèse (évoquée par R. A. Fisher lui-même) qu'un gène accroisse simultanément le désir de fumer et la sensibilité des poumons ?

Cet exemple n'est pas fictif : il a fallu 14 ans (de 1950 à 1964) pour démontrer le lien causal entre tabac et cancer, 14 ans de profondes recherches méthodologiques. Et les outils alors développés ne sont pas aisément transposables à d'autres interrogations.

Certains ont pu penser que l'arrivée des Big data allait bouleverser la situation. L'amoncellement de données observées allait « naturellement » et « automatiquement » dégager les lois physiques sous-jacentes aux phénomènes et les relations de causalités inhérentes. Rien n'est moins sûr : les méthodes employées sont-elles fondamentalement différentes de « simples » ajustements à des lois statistiques – suffisantes, certes, pour décrire les préférences d'un client prospect ou pour construire une machine championne du Jeu de Go mais pas pour étudier les causalités ?

L'affaire finalement est paradoxale. La démarche scientifique aborde presque tous les aspects du réel mais pas (sauf cas très particuliers) les questions de causalité<sup>1</sup> !

## 2. Une nouvelle science de la causalité ?

Judea Pearl, Prix Turing<sup>2</sup> d'informatique en 2011, présente cette question dans un livre récent<sup>3</sup>, *The book of Why* (relativement) facile à lire -- rédigé à la première personne bien que coécrit avec Dana Mackenzie, spécialiste d'écriture scientifique pour le grand public. Il y décrit les avancées de ce qu'il nomme, sans modestie excessive, la « Révolution causale » que lui et ses équipes ont construite depuis plus de vingt ans, suivant les travaux de quelques rares précurseurs injustement oubliés.

Les données, explique J. Pearl, ne suffisent pas à dégager et à démontrer des causalités. Mais si on propose un schéma de causalité (par exemple : « *je fais l'hypothèse que A et B ont une action sur C* ») elles peuvent, démontre-t-il, l'infirmier ou non.

On soulignera que cette hypothèse (A et B ont une action sur C) n'exclut aucunement que X, dont on dispose ou non d'observations, ait aussi une action sur A, B ou C.

En d'autres termes, certaines hypothèses de causalité peuvent être réfutées, ce qui est bien le mécanisme de la démarche scientifique. Et ceci même si on ne dispose pas du vaste ensemble de pré-requis nécessaire à l'élaboration d'un classique test randomisé en double aveugle.

J. Pearl propose même un modèle de diagrammes, une syntaxe spécifique et un mode de raisonnement (le « do-calcul ») permettant de définir rapidement et précisément, *selon l'hypothèse causale envisagée*, les analyses de données que le chercheur devra mener ou les données complémentaires qu'il devra collecter.

Par delà les corrélations, le scientifique peut donc, dans nombre de situations, étudier les causalités. Il peut même aller plus loin et atteindre ce que J. Pearl dénomme « le troisième degré de l'échelle de causalité », celui de l'analyse « contrefactuelle » : « *Que se serait-il passé si ...* » ?

Dans *The book of Why* J. Pearl donne de très nombreux exemples de ces trois niveaux d'analyse : analyse de corrélations, analyse de causalités et analyse contrefactuelle. Il montre aussi comment la nouvelle approche méthodologique qu'il a développée permet d'aborder et de maîtriser certains problèmes ou paradoxes statistiques classiques (dont le redoutable paradoxe de Simpson).

## 3. Un point de passage obligé pour l'IA ?

On connaît la formule : « Je ne comprends pas quelque chose tant que je ne suis pas capable de l'expliquer à un ordinateur – en le programmant ». Judea Pearl insiste sur le paradoxe que l'esprit humain semble fait pour dégager (parfois à tort !) des causalités alors que ce sujet n'est pas même profondément étudié par les scientifiques. Surtout, il affirme que là est l'obstacle majeur actuel au développement de l'intelligence artificielle (IA).

---

<sup>1</sup> Affirmer que « *la pomme tombe parce que la force de gravitation (en  $1/R^2$ ) l'attire* » relève plus du descriptif que du causal.

<sup>2</sup> Le Prix Turing est traditionnellement présenté comme l'équivalent du Prix Nobel pour l'informatique.

<sup>3</sup> Mai 2018 Basic books New York ISBN 978-0-465-09760-9

L'analyse de corrélations, souligne-t-il, peut être simplissime s'il s'agit de construire la droite de régression entre deux paramètres ou devenir atrocement compliquée (et impénétrable par l'esprit humain) si on confie à une machine l'analyse des Big data ... mais fondamentalement rien ne distingue ces deux approches qui décrivent des associations *mais n'abordent en rien les questions de causalité*. Or, la seule véritable percée en matière d'intelligence artificielle ne pourra venir, dit Pearl, que de l'inclusion de certaines formes de raisonnement causal (dont celles proposées dans *The book of Why*) dans la programmation des machines.

Le jury du Prix Turing ne s'y est pas trompé, qui a décerné le Prix 2011 à Judea Pearl « pour des contributions fondamentales à l'intelligence artificielle par le développement de l'analyse probabiliste<sup>4</sup> et du raisonnement causal ».

#### 4. Comme Churchill, il écrit l'Histoire

On demandait à Churchill s'il pensait que l'Histoire serait bienveillante à son égard. « *Je suis optimiste, répondit le vieux lion -- qui s'apprêtait à rédiger les milliers de pages de ses Mémoires -- car l'Histoire, c'est moi qui l'écrirai* ».

De fait il semble que l'un des objets de *The book of Why* soit, pour Pearl, de s'attribuer une place ensoleillée dans l'Histoire des sciences. Pearl jauge et juge l'action de tous ceux qui ont abordé – et surtout refusé d'aborder – l'approche scientifique de la causalité. Peu nombreux, hors de ses équipes et de ses élèves, trouvent grâce à ses yeux. Cette tonalité satisfaite peut agacer le lecteur, elle n'empêche pas que les analyses historiques proposées par Pearl, et notamment celles qui concernent les Pères de la statistique (Galton, Pearson, Fisher) soient passionnantes.

The Book of Why souffre aussi d'un autre défaut qui pourrait perturber le lecteur : il inclut quelques passages ardues au milieu de pages faciles à lire.

Mais ...

**... que ces quelques critiques ne vous rebutent pas : lisez *The book of Why* !**

*Vous découvrirez un domaine (étonnamment) nouveau de la science contemporaine ... et les modes de pensées d'un chercheur de très haut niveau. Oui, *The book of Why* mérite votre pleine attention, plongez-y sans tarder -- sans même attendre une traduction française qui, nous l'espérons, ne se fera pas trop longtemps désirer.*

(Olivier Moch, août 2018)

---

<sup>4</sup> Analyse probabiliste : Judea Pearl a été un des principaux concepteurs et promoteurs des Réseaux bayésiens (modèles graphiques probabilistes permettant de relier les probabilités conditionnelles de nombreuses variables aléatoires). Exemple d'application : j'apprends que la cousine germaine du patient souffrait du diabète et que son neveu a un souffle au cœur, cela me permet-il d'affiner mon diagnostic ?