

Max Kistler
La causalité

in :

Anouk Barberousse, Denis Bonnay et Mikaël Cozic (dir.), *Précis de philosophie des sciences*, Paris, Vuibert, 2011, p. 100-140.

En 1912, Bertrand Russell recommandait aux philosophes d'expurger la causalité de leur répertoire conceptuel. Son argument se fonde sur la prémisse selon laquelle les sciences avancées ne contiennent aucun concept qui correspondrait à notre notion intuitive de causalité. Russell explique en outre que certains aspects de cette notion intuitive interdisent sa réduction à l'aide de notions scientifiques. Or quand il y a un conflit entre une intuition de sens commun et la science, l'attitude naturaliste consiste à résoudre le conflit en donnant raison à la science contre l'intuition. Russell encourage donc les philosophes à écarter le mot « cause » de leur vocabulaire. Le débat lancé par l'article de Russell dure jusqu'à ce jour. Beaucoup de philosophes et de scientifiques considèrent que le recours à la notion de causalité dans l'explication scientifique est légitime et même indispensable.

Le plan de ce chapitre est le suivant. Dans une première section nous analyserons les raisons russelliennes de soutenir qu'il ne peut y avoir aucune analyse du concept de causalité compatible avec la physique du XX^e siècle. Nous verrons que le débat entre les « éliminativistes » qui suivent Russell et les philosophes convaincus que le concept de causalité est indispensable au sens commun et à la science s'articule autour de deux distinctions : d'une part, celle entre le microscopique et le macroscopique ; d'autre part, celle entre les événements concrets et leurs propriétés mesurables. Le débat sur la légitimité du concept de causalité s'avère directement lié à celui de l'existence de lois de la nature en dehors de la physique fondamentale, qui sont des lois admettant des exceptions, ou lois *ceteris paribus*.

Nous parvenons à la conclusion que l'absence de la notion de causalité dans le contenu théorique de la physique fondamentale est compatible avec l'utilité, et même le caractère indispensable de ce concept dans de nombreux contextes : dans les jugements de sens commun, notamment quand il s'agit de planifier nos actions en fonction de leurs conséquences, mais aussi dans les sciences en dehors de la physique fondamentale, c'est-à-dire de la physique macroscopique à la biologie et aux neurosciences, et dans de nombreux projets d'analyse naturaliste de notions philosophiques. La causalité est notamment utilisée dans la réduction philosophique de l'intentionnalité, de la perception, de la connaissance et de l'action.

Une fois justifié le projet d'une analyse philosophique de la causalité, nous examinons les approches majeures actuellement développées et débattues, à savoir l'analyse contrefactuelle, l'analyse probabiliste, l'analyse de la manipulabilité et l'analyse en termes de processus. Nous reviendrons plus loin sur les approches suivantes :

- 1) L'idée directrice de l'analyse contrefactuelle est que c cause e si et seulement si il est vrai que : si c ne s'était pas produit, e ne se serait pas produit¹.
- 2) L'idée directrice de l'analyse probabiliste est que le facteur C exerce une influence causale sur le facteur E si et seulement si l'occurrence d'un événement de type C augmente la probabilité de l'occurrence d'un événement de type E.

¹ Les variables minuscules représentent des événements concrets, alors que les majuscules représentent des propriétés d'objets ou d'événements. Nous justifierons cette distinction par la suite.

3) L'idée directrice de l'analyse de la manipulabilité est que la variable C est dans une relation causale par rapport à la variable E si et seulement si des interventions sur C permettent de modifier la valeur de E.

4) Finalement, l'idée directrice de l'analyse en termes de processus est que l'événement c cause l'événement singulier e si et seulement s'il existe un processus physique de transmission entre c et e, par exemple d'une quantité d'énergie.

Nous reviendrons sur le fait que ces différentes analyses conçoivent différemment les termes de la causalité : selon certaines, ce sont des événements singuliers, alors que pour d'autres, ce sont des propriétés d'événements ou « facteurs » qui peuvent être instanciés par de nombreux événements.

Pour comprendre le débat complexe qui oppose les défenseurs de ces approches, il est indispensable de s'interroger sur le but poursuivi par chacune d'entre elles, et sur les critères utilisés pour juger de leur succès. Pour le dire simplement, on peut concevoir la tâche d'une analyse philosophique de la causalité de deux manières : 1) comme relevant de la pure analyse conceptuelle *a priori*, entièrement détachée de contraintes ayant trait à la réalité de notre monde actuel, tel que les sciences nous le décrivent, ou 2) comme une recherche mi-conceptuelle mi-empirique de « l'essence réelle » d'une relation conçue comme une espèce naturelle de relation. Cette seconde approche conçoit la causalité en analogie à des espèces naturelles comme l'eau, l'or ou les tigres. Le sens commun suppose que ces espèces de substances ou d'animaux possèdent une essence réelle dont la découverte relève de la science empirique. On peut faire l'hypothèse que la causalité a, dans un sens analogue, une essence réelle spécifique à notre monde réel. Mais plutôt que de commencer avec cette réflexion méthodologique, nous l'aborderons après l'exposition du débat sur la conception contrefactuelle : il est plus judicieux d'aborder une question méta-philosophique sur le but, la méthode et les critères d'adéquation d'une analyse après s'être familiarisé avec un échantillon du débat sur lequel elle porte.

1. Russell et l'élimination du concept de causalité

Les arguments de Russell se dirigent en premier lieu contre ce qu'on appelle aujourd'hui la « causalité générique »². On distingue entre des jugements causaux singuliers, comme : « le fait que j'ai frotté cette allumette-ci (que j'ai sous les yeux) est la cause du fait qu'elle se soit enflammée », et des jugements causaux génériques, comme : « de manière générale, le fait de frotter une allumette cause le fait qu'elle prenne feu ». Dans la conception humienne³, la vérité d'un jugement causal singulier dépend de la vérité d'un jugement générique : la vérité de la proposition selon laquelle l'événement singulier *c* cause l'événement singulier *e* présuppose qu'il soit vrai sur le plan générique que des événements du même genre que *c* sont suivis d'événements du même genre que *e*. Autrement dit, il ne peut y avoir causalité entre événements singuliers sans l'existence d'une régularité au niveau des *types* d'événements. Nous verrons plus loin que cette thèse a été remise en question, ce qui a pour effet de dissocier la causalité singulière de la causalité générique. Si l'existence des relations causales singulières ne présuppose pas l'existence de relations causales génériques qu'elles instancient, alors la causalité singulière ne tombe pas sous le coup des arguments de Russell. Cependant, les conceptions contemporaines qui conçoivent la causalité singulière indépendamment d'une relation nomologique au niveau des types d'événements, des facteurs

² Pour des réévaluations récentes des arguments russelliens contre la possibilité de construire un concept scientifiquement respectable de causalité, voir Price et Corry (2007) ; Spurrett et Ross (2007).

³ La conception de la causalité que l'on trouve dans les œuvres de David Hume constitue encore aujourd'hui le point de départ de la réflexion sur la causalité. Cf. Hume (1739-40) et Hume (1777).

ou des propriétés, sont minoritaires. Dans la mesure où les analyses philosophiques de la causalité visent à expliquer et justifier l'usage de concepts causaux en science, le concept générique reste le plus pertinent : on considère généralement que l'on ne peut expliquer scientifiquement pourquoi cette allumette-ci s'est allumée à l'instant t qu'à partir de propositions générales s'appliquant à tous les frottements d'allumettes à n'importe quel instant et en n'importe quel lieu. Une telle explication scientifique fera par exemple appel à la proposition générale selon laquelle l'apport d'énergie sous forme de chaleur produite par un frottement suffisamment intense déclenche la réaction chimique de l'oxydation exothermique de n'importe quel échantillon de trisulfure de tétraphosphore (P_4S_3), qui se trouve être la substance qui recouvre la tête des allumettes ordinaires.

1.1. Le principe de causalité et la répétition des événements

Russell cherche à montrer la vacuité du traditionnel « principe de causalité » selon lequel « les mêmes causes ont toujours les mêmes effets », ou plus précisément : « étant donné un événement e_1 , il y a un événement e_2 et un intervalle de temps t , tel que, toutes les fois que e_1 a lieu, e_2 suit après un intervalle t » (Russell 1912, p. 170). Il s'agit d'une « méta-loi » : on peut l'interpréter comme affirmant que certaines propriétés générales jouent le rôle d'antécédents de relations causales génériques. L'argument que Russell oppose à l'existence de telles relations causales génériques dans les sciences avancées (et donc au principe de causalité), repose sur l'observation que l'on ne peut envisager la récurrence d'un type d'événement qu'à condition de le concevoir de manière 1) vague et 2) étroite, alors que les événements conçus de manière vague ne peuvent pas faire l'objet d'explications scientifiques et que les généralisations portant sur des événements conçus de manière étroite ne sont pas strictement vraies.

1) Les événements qui se répètent sont conçus de manière *vague* : pour reprendre l'exemple de Russell, les lancers de pierres (suivis de bris de vitres) ne se répètent que si l'on les conçoit d'une manière qui fait abstraction des détails microscopiques. Il n'y a pas deux lancers de pierres qui se ressemblent exactement jusqu'au moindre détail microscopique. Le problème est que l'explication scientifique, dans sa forme mûre, requiert la *déduction* de l'*explanandum* à partir d'une description de la situation qui joue le rôle d'*explanans*, jointe à des énoncés nomologiques (voir chapitre 1). Or une telle déduction n'est possible que si l'*explanans* contient une description quantitativement précise, ou « étroite » (*ibid.*, p. 172) de la cause, ainsi que des lois quantitatives de la nature, et l'*explanandum* une description quantitativement précise de l'effet. Cependant, dans la mesure où l'on conçoit les événements de cette manière quantitativement précise - qui rend possible leur explication scientifique - ils ne se répètent pas. Dans la mesure où l'antécédent d'un énoncé conditionnel universel ne s'applique qu'à un seul événement, sa vérité est presque triviale : il est vrai si et seulement si le conséquent est vrai dans l'unique situation où l'antécédent est vrai. Un tel énoncé ne peut pas être utilisé pour expliquer d'autres événements, ce qui est la fonction des lois. Il ne peut pas y avoir de lois strictes qui comportent des prédicats quantitativement précis et qui peuvent être utilisées pour l'explication et la prédiction dans des situations nouvelles : il n'y a de régularités strictes que dans le sens commun et « dans l'enfance d'une science » (*ibid.*, p. 173).

2) Les événements qui se répètent sont conçus de manière *étroite*. La répétition d'un événement requiert qu'on le conçoive de manière localisée, autrement dit comme le contenu d'une région bien délimitée de l'espace-temps. On ne peut envisager qu'il existe de nombreux frottements d'allumettes que dans la mesure où l'on exclut des événements de frottement eux-mêmes les circonstances dans lesquelles ils se produisent. Or dans la mesure où l'on fait abstraction de la personne qui frotte, du temps qu'il fait et d'autres facteurs contextuels, la succession régulière des allumages après des frottements d'allumettes souffre d'exceptions : il

peut y avoir des éléments présents aux alentours du premier événement (le frottement) qui empêchent, en s'interposant, que le second événement (l'allumage) se produise ; autrement dit, la régularité n'existe que *ceteris paribus*, c'est-à-dire toutes choses étant égales par ailleurs. La dialectique est semblable à celle que nous avons établie en ce qui concerne le vague : un événement conçu de manière étroite peut éventuellement donner lieu à une répétition, mais le fait de ne pas prendre en compte les circonstances de l'événement a pour conséquence que la généralité avec laquelle l'événement *c* est suivi d'un autre, *e*, n'est pas parfaite, car les circonstances peuvent interférer et empêcher *e* de se produire. Les généralisations portant sur des événements conçus de manière étroite ne peuvent pas figurer dans des explications scientifiques, car celles-ci nécessitent des énoncés de régularités universelles strictement vrais. « La séquence [...] n'est que probable, tandis que la relation entre cause et effet a été supposée nécessaire » (*ibid.*, p. 176)⁴. De l'autre côté, si et dans la mesure où on exclut le risque d'interférences en concevant les événements de manière plus large, c'est-à-dire en y incluant les environs spatio-temporels de chaque événement, la probabilité qu'ils se répètent diminue. « Aussitôt que nous y incluons l'environnement, la probabilité de la répétition diminue, jusqu'à ce que, à la fin, quand l'environnement tout entier y est inclus, la probabilité de la répétition devienne presque nulle » (*ibid.*, p. 172).

Nous pouvons remarquer que le premier argument ne remet en cause que l'existence de successions d'événements *macroscopiques* conçus avec les concepts du *sens commun* : des événements microscopiques, tels que l'interaction entre un électron et un photon ou la décomposition radioactive d'un noyau d'uranium 238, se répètent même lorsqu'on les conçoit de manière précise. Mais le second argument remet en cause autant l'existence de successions strictes d'événements microscopiques que macroscopiques : si l'on considère un ensemble d'événements-causes microscopiques localisés qui relèvent strictement du même type dans la mesure où on ignore leurs alentours, ils n'ont pas nécessairement les mêmes effets, dans la mesure où ces effets peuvent être influencés par les événements qui se produisent alentour.

La conclusion de Russell vaut donc également pour les événements microscopiques : « Aussitôt que les antécédents ont été donnés assez complètement pour permettre de calculer le conséquent avec quelque exactitude, les antécédents sont devenus si compliqués qu'ils ne reparaitront jamais » (*ibid.*, p. 173). On ne peut trouver d'événements conçus précisément qui se répètent sur le plan macroscopique ; en revanche, les événements microscopiques peuvent se répéter ; cependant, une succession d'événements microscopiques ne se répète que lorsqu'on conçoit les événements de manière localisée, en faisant abstraction des alentours. Par conséquent, le principe de causalité : « même cause, même effet » est, dit Russell, « complètement oiseux », dans la mesure où ce qui permettrait d'obtenir une répétition (« même cause »), à savoir rendre la description vague pour les événements macroscopiques, ou inclure les alentours spatio-temporels pour les événements microscopiques, ou bien empêche qu'on les utilise dans les sciences exactes (pour les premiers) ou bien empêche qu'ils se répètent (pour les seconds).

1.2. Les lois fonctionnelles des sciences mûres

Le second argument de Russell à l'encontre de la possibilité de donner une légitimité scientifique à la notion de cause consiste à montrer que les lois utilisées dans les explications des sciences mûres ne peuvent pas être interprétées comme des lois causales. Les lois utilisées en physique mathématique, comme par exemple en « astronomie de la gravitation » (*ibid.*, p. 167), ont la forme de fonctions (voir également volume 2, chapitre 3) : dans un système de masses soumis à la seule force d'attraction gravitationnelle, il est possible de représenter la

⁴ Russell n'envisage pas la causalité probabiliste, dans la mesure où il considère que la nécessité fait partie de la définition de la causalité.

configuration⁵ du système à un instant donné comme une fonction de cet instant, et de la configuration et des vitesses à un autre instant (ou des configurations à deux autres instants). Une telle fonction « détermine » certes la configuration du système, mais cela ne justifie pas la thèse selon laquelle il s'agit là de détermination *causale*. Il y a deux raisons majeures qui permettent à Russell de dire qu'« il n'y a rien qui puisse être appelé 'cause' et rien qui puisse être appelé 'effet' dans un tel système » (*ibid.*, p. 177). La première est qu'il s'agit d'une détermination purement logique qui est indifférente à la direction du temps : les lois de Newton, jointes à la loi de l'attraction gravitationnelle, permettent de calculer la configuration d'un système de masses à un instant dans le passé, en fonction de la configuration et des vitesses à un instant futur, exactement de la même manière qu'elles permettent de déterminer une situation future à partir d'une situation passée. Etant donné que le concept traditionnel de causalité requiert que la cause précède l'effet⁶, cette détermination fonctionnelle ne peut pas être interprétée comme détermination causale.

La seconde raison concerne les termes des relations : les termes des relations causales sont des événements concrets, alors que les équations fonctionnelles scientifiques mettent en rapport des valeurs de grandeurs mesurables, autrement dit de certaines *propriétés* de ces événements. L'une de ces équations exprime par exemple la loi de la gravitation (ou loi de l'attraction universelle) qui indique la force de l'attraction gravitationnelle entre deux corps massifs en fonction de leurs masses et de leur distance. La loi stipule l'identité de la valeur numérique du produit de l'accélération d'un objet massif et de sa masse, et de la valeur numérique de la force totale qui agit sur lui, identité qui est compatible avec la plus grande diversité des objets accélérés. La connaissance de n'importe quelle loi se heurte au problème de l'induction, dans la mesure où sa portée est en principe infinie. Mais la connaissance d'une loi de ce genre pose un problème supplémentaire : il est difficile de tester une hypothèse qui porte sur une loi exprimant des rapports constants entre les valeurs de certaines grandeurs, dans la mesure où ces grandeurs ne sont pas instanciées seules mais dans des événements concrets dont l'évolution dépend aussi de leurs autres propriétés. On ne peut pas directement tester la loi de la gravitation, pour deux raisons. 1) La première est qu'il n'existe aucun système de deux masses isolées qui ne seraient pas aussi soumises à l'attraction d'autres masses, éventuellement plus éloignées. 2) La seconde est que les objets massifs ont aussi d'autres propriétés qui peuvent être à l'origine d'autres forces. Russell en tire la conclusion que les lois quantitativement exactes des sciences mûres ne sont pas causales parce que leurs termes ne sont pas, comme les causes et les effets, directement accessibles à l'expérience. « Dans toute science, dit-il, nous devons distinguer deux espèces de lois : en premier lieu, celles qui sont vérifiables empiriquement, mais qui ne sont probablement qu'approximatives ; en second lieu, celles qui ne sont pas vérifiables, mais qui peuvent être exactes » (*ibid.*, p. 179). La première espèce correspond aux « lois causales » du sens commun et des sciences qui sont au début de leur évolution, alors que les lois des sciences mûres appartiennent à la seconde espèce : ces dernières ne peuvent pas être interprétées comme causales dans la mesure où leurs termes ne sont pas des événements concrets.

1.3. Les lois *ceteris paribus*

Le problème soulevé par Russell a fait l'objet d'une littérature abondante sur les lois dites *ceteris paribus*⁷. On a observé que l'interprétation de nombreuses lois quantitatives nous place devant un dilemme :

⁵ La configuration d'un système est l'ensemble des positions de ses parties.

⁶ On a récemment tenté de développer un concept de causalité « rétroactive » (*backwards causation*). Cf. Dowe (1996). La causalité simultanée pose des problèmes spécifiques.

⁷ Voir p.ex. le numéro spécial de *Erkenntnis* 57(3) (2002).

1) ou bien on suppose que les lois portent sur des objets ou événements concrets et directement accessibles à l'expérience, et alors il apparaît que ces lois ont des exceptions ou ne valent que *ceteris paribus*,

2) ou bien on suppose qu'elles ne portent pas sur des objets concrets ni sur des événements : cela pose le problème de comprendre comment il est néanmoins possible que de telles lois soient utilisées pour produire des explications et des prédictions scientifiques.

Hempel donne l'exemple suivant : pour tout aimant b en forme de barre, « si b est coupé en deux barres plus courtes et celles-ci sont suspendues à de longs fils fins l'une près de l'autre à la même distance de la Terre, elles s'orienteront le long d'une ligne droite » (Hempel 1988, p. 20). Cette généralisation n'est pas vraie sans exception du comportement de barres aimantées concrètes : dans certaines circonstances, comme lorsqu'il y a un fort courant d'air dans une direction perpendiculaire à l'orientation des barres ou lorsqu'il y a un fort champ magnétique externe, les deux moitiés ne s'alignent pas. De même, si l'on considère que la loi de l'attraction gravitationnelle porte sur des objets massifs concrets, en déterminant leur accélération en fonction de leurs masses et de leurs distances, elle a de nombreuses exceptions⁸ : un objet ayant la masse m_1 qui se trouve à la distance d d'un second objet ayant la masse m_2 , n'est dans la grande majorité des circonstances pas accéléré en direction de ce second objet avec une accélération de $G \frac{m_2}{d^2}$. Il n'est pourtant pas nécessaire d'en conclure,

comme Cartwright (1983), que les lois « mentent »⁹ : il existe plusieurs stratégies de réinterpréter les équations fonctionnelles et autres énoncés nomologiques, de telle sorte qu'ils puissent être vrais même si les objets et événements concrets ne se comportent pas souvent en accord avec ces énoncés nomologiques. Une possibilité est de dire que les lois ne portent que sur des systèmes qui se trouvent dans des situations idéales, ce qui signifie notamment qu'ils sont isolés¹⁰. Pour certaines lois au moins, comme la loi de l'attraction gravitationnelle, cela a pour conséquence qu'elles ne portent sur aucun système réel (puisque aucun système n'est idéal au sens d'être isolé d'influences gravitationnelles extérieures). Même s'il existe, pour certaines lois, des systèmes isolés, cette proposition se trouve confrontée au problème d'expliquer comment une loi qui ne porte pas sur des systèmes réels peut néanmoins être utilisée pour la prédiction et l'explication de faits concernant des systèmes réels.

Une autre possibilité est de dire que les lois ne portent pas sur des systèmes réels, mais sur des modèles abstraits. Smith (2002) propose de résoudre le problème de l'interprétation des lois *ceteris paribus* en distinguant entre lois fondamentales et équations de mouvement. Les lois fondamentales ne s'appliquent pas directement aux systèmes concrets ou réels. La loi de la gravitation universelle détermine la force avec laquelle deux masses s'attirent. Or, cette loi ne peut être directement utilisée pour calculer le mouvement d'objets réels, dans la mesure où aucun corps réel ne soit soumis qu'à la seule force attractive gravitationnelle d'un autre corps. Tout corps réel est attiré par de nombreux autres corps massifs, en plus d'être en général soumis à d'autres forces. Pour Smith, la loi de la gravitation universelle fait partie d'une « recette »¹¹ qui permet de construire un modèle. La dernière étape de l'algorithme conduit à une équation de mouvement qui est spécifique à un système concret ; elle n'a donc pas la généralité requise pour être une loi.¹² Les lois fondamentales correspondent aux lois

⁸ Cartwright (1983), p. 57/8; Hempel (1988), p. 23; Pietroski et Rey, (1995) p. 86; Smith (2002).

⁹ Le titre du livre de Cartwright dit, de manière ambiguë, « *How the Laws of Physics Lie* », ce qui peut signifier « Comment les lois de la physique se trouvent être » mais aussi « Comment les lois de la physique mentent ». Dans son introduction, Cartwright explique qu'elle vise la seconde interprétation : « *laws in physics [...] must be judged false* » (Cartwright 1983, p. 12).

¹⁰ Silverberg (1996); Hüttemann (1998).

¹¹ Smith la nomme « la recette d'Euler (*Euler recipe*) » (Smith 2002, p. 243).

¹² Je critique cette thèse plus loin.

dont Russell dit qu'elles ne sont pas vérifiables mais peuvent être exactes. Parmi ces lois fondamentales, il y a notamment des lois déterminant les différentes forces qui s'exercent sur un objet en fonction de ses propriétés et des propriétés des autres objets représentées dans le modèle A qui contient une spécification partielle des propriétés d'un système concret C que l'on considère. Si C n'évolue pas de la manière prédite par le modèle A, cela indique simplement que A ne représente C que de manière incomplète. Il s'avère dans ce cas nécessaire d'améliorer A en y incluant des objets, propriétés et interactions supplémentaires. Les équations de mouvement calculées (à partir de modèles A) pour représenter l'évolution d'ensembles de systèmes concrets C correspondent aux lois dont Russell dit qu'elles « sont vérifiables empiriquement, mais [...] ne sont probablement qu'approximatives », car rien n'empêche certains systèmes concrets C d'être soumis à l'influence de facteurs non représentés dans A.

Dans le même ordre d'idées, Cummins (2000) a proposé de distinguer entre des « lois générales de la nature » dont le domaine d'application n'est pas limité, et des « lois *in situ* » qui ne s'appliquent qu'à un type particulier de systèmes, par exemple des systèmes planétaires ou des êtres vivants, à cause de la constitution et organisation de ces systèmes. Lorsqu'un tel système, que Cartwright (1999) appelle une « machine nomologique », évolue en accord avec une loi (de système), son évolution peut être jugée causale. A la différence d'une loi de la nature qui s'applique de manière universelle, une loi de système ne s'applique qu'à un type bien particulier d'objet, et elle n'est pas stricte : des exceptions se produisent en fonction des influences perturbatrices qui s'exercent de l'extérieur sur le système.¹³ Par ailleurs, les perturbations peuvent aussi faire l'objet de jugements causaux : cela ressort de l'analyse de Menzies (2004) selon lequel tout énoncé causal présuppose un modèle (composé d'une espèce naturelle et de lois s'appliquant à cette espèce). Un facteur est jugé être une cause s'il fait une différence dans l'évolution du système, sur l'arrière-plan de l'évolution normale du modèle¹⁴. Dans l'un des exemples considérés par Menzies, une personne qui a fumé pendant des années développe un cancer. Intuitivement, le fait que la personne soit née et le fait qu'elle ait des poumons ne sont pas des causes de son cancer quoi qu'ils en soient des conditions nécessaires. Selon Menzies, cette intuition s'explique par le fait que l'identification d'une cause constitue normalement la réponse à une « question contrastée », de la forme : « pourquoi a-t-elle développé un cancer plutôt que non ? » (Menzies, 2004, p. 148/9). On compare l'histoire réelle avec une histoire fictive (ou « contrefactuelle ») dans laquelle la personne ne développe pas de cancer. Les faits d'être née et d'avoir des poumons ne sont pas des causes du cancer parce qu'ils sont également présents dans l'histoire fictive.

L'analyse de Russell montre que les lois ayant la forme d'une dépendance fonctionnelle quantitativement précise, telles qu'on les utilise en physique mathématique, ne peuvent pas être interprétées comme exprimant directement des régularités au sein d'événements observables ; en particulier, elles ne peuvent pas être interprétées comme des généralisations exprimant des successions de causes et d'effets. Mais il s'agit là du problème général de comprendre le rapport entre les lois ou modèles élaborés dans les sciences avancées et leur utilisation pour la prédiction et l'explication de systèmes concrets réels. Comme le montre le débat contemporain sur les lois *ceteris paribus*, cette difficulté n'est pas spécifique à la justification scientifique des jugements causaux. Elle se pose de la même manière s'il s'agit par exemple de déterminer la conformation d'une macromolécule dans

¹³ Cf. Kistler (2006).

¹⁴ L'idée de Menzies selon laquelle une cause est un facteur qui « fait une différence » relativement à un arrière-plan présupposé fixe remonte à Mill et son analyse de la distinction entre causes et conditions. On la retrouve aussi chez Mackie (1974) (l'arrière-plan est le « champ causal (*causal field*) »), dans la nouvelle conception de la causalité comme influence de David Lewis (2000), et dans les conceptions de Hitchcock (1996) et Woodward (2003 ; 2004).

l'espace, à partir d'un modèle de ses composantes et des lois gouvernant les interactions entre ces composantes en fonction de leurs propriétés ; dans la mesure où il s'agit de déterminer cette conformation à un instant donné, à partir des propriétés des composantes *au même instant*, il ne s'agit pas d'un jugement causal. Ce problème pose un défi formidable à la philosophie des sciences, mais il n'est pas spécifique à la justification scientifique des jugements causaux. On peut dire la même chose du problème de l'induction dont Russell constate à juste titre qu'il oppose un obstacle de principe à la connaissance des généralisations causales ; il s'agit là aussi d'un problème général qui s'oppose tout autant à la connaissance des lois non causales.

2. La réduction de la causalité à l'explication déductive-nomologique

Il reste le défi de la justification des caractéristiques spécifiques de la causalité, et en premier lieu de son asymétrie : il est conceptuellement exclu qu'un événement *c* soit à la fois la cause d'un second événement *e*, et son effet. Russell montre qu'aucune asymétrie de ce genre n'existe au niveau des lois fonctionnelles de la physique. Cependant, cela ne montre pas qu'il ne peut pas y avoir de relations asymétriques dans la réalité ; cela pose seulement le problème de trouver l'origine de cette asymétrie ailleurs que dans ces lois.

Le fait que la notion de cause n'apparaisse pas en physique fondamentale ne rend pas illégitime le projet de l'analyse philosophique de cette notion. Les lois de la physique fondamentale et les jugements causaux ne s'appliquent pas aux mêmes objets : les valeurs des variables qui apparaissent dans les premières sont des grandeurs déterminées qui caractérisent certaines *propriétés* de substances ou d'événements, alors que les termes des relations causales sont des événements concrets. Etant donné que les jugements causaux abondent non seulement dans les jugements de sens commun mais aussi dans nombre de projets philosophiques et dans les jugements portant sur le test expérimental des théories scientifiques¹⁵, le projet d'une analyse naturaliste de la causalité a été très activement poursuivi au XX^e siècle, à commencer par Russell lui-même¹⁶.

Il existe plusieurs traditions de recherche visant à une telle analyse. Historiquement, la plus importante est l'analyse déductive-nomologique (DN), héritière de la réduction de la causalité aux régularités et aux lois de la nature. La forme que cette conception de la causalité prend dans la philosophie de l'empirisme logique du XX^e siècle se distingue cependant quelque peu de ses prédécesseurs dans la philosophie classique. Plutôt que de passer, comme Hume, par l'analyse de *l'idée* de causalité à partir de *l'expérience* d'une répétition régulière de certaines successions d'événements, et plutôt que de proposer, comme Galilée, Newton, et beaucoup d'autres, de substituer la notion de loi à la notion de cause, l'analyse DN vise d'abord à analyser *l'explication* causale, telle qu'elle est pratiquée dans les sciences (voir chapitre 1). Selon cette analyse, il est équivalent de dire que C est cause de E et de dire que C figure comme prémisse dans une explication DN de E : l'effet E est l'*explanandum* – ce qu'il s'agit d'expliquer – et occupe le rôle de la conclusion de l'argument, et la cause est le contenu

¹⁵ Cf. Putnam (1984).

¹⁶ « Il y a, dit Russell dès 1914, un sens quelque peu grossier et relâché du mot 'cause' qui peut être préservé. Les uniformités approchées qui conduisent à son emploi préscientifique peuvent s'avérer dans toutes circonstances en dehors de quelques-unes qui sont rares et exceptionnelles, peut-être dans toutes les circonstances qui se produisent réellement. Dans de tels cas, il est convenable de pouvoir parler de l'événement antécédent comme de la 'cause' et de l'événement ultérieur comme de l' 'effet' » (*There is, however, a somewhat rough and loose use of the word 'cause' which may be preserved. The approximate uniformities which lead to its pre-scientific employment may turn out to be true in all but very rare and exceptional circumstances, perhaps in all circumstances that actually occur. In such cases, it is convenient to be able to speak of the antecedent event as the 'cause' and the subsequent event as the 'effect'*) » (Russell 1914/1993, p. 223). Dans (1948/1992, p. 471sq), Russell propose une théorie plus élaborée de la causalité.

de l'une des prémisses qui constituent ensemble l'*explanans* - ce qui explique. Voici comment Carnap défend l'analyse de la causalité en termes d'explication DN : "Que signifie alors l'énoncé : l'événement B est causé par l'événement A? Qu'il existe dans la nature certaines lois dont on peut déduire logiquement l'événement B, à condition de les conjuguer avec la description exhaustive de l'événement A" (1966; trad. p. 189)¹⁷. Pour qu'il s'agisse d'une explication causale, il est essentiel que le lien entre la prémisse qui nomme la cause et la conclusion qui nomme l'effet soit assuré par un ou plusieurs énoncés de lois de la nature. Si E était une conséquence *logique de C seul*, leur lien serait logique ou conceptuel, alors que l'on accepte en général la thèse humienne selon laquelle la relation causale est contingente. Rétrospectivement, la tentative de réduire la causalité à la déductibilité à l'aide de lois apparaît comme une tentative d'*éliminer* la causalité au profit des seules lois. Elle permet certes de garder le mot « causalité » mais l'analyse DN le vide de son contenu : dire que C figure dans une explication *causale* de E ne signifie rien de plus que de dire que C figure dans une explication *scientifique* de E. Si toutes les explications sont causales, le concept de causalité perd son contenu discriminatif.

Si l'analyse DN a été largement abandonnée¹⁸, c'est en premier lieu parce qu'il est apparu qu'il y a des explications scientifiques qui *ne sont pas* causales : cela montre qu'il existe une différence spécifique des explications causales.

De nombreuses explications physiques qui utilisent des dépendances fonctionnelles ne correspondent pas intuitivement à des relations causales : lorsqu'on déduit la conductivité thermique d'un fil métallique de sa conductivité électrique ou inversement (selon la loi de Wiedemann-Franz qui stipule que ces deux grandeurs sont proportionnelles), aucune des deux n'apparaît comme la cause de l'autre. De même, lorsqu'on déduit la température de la pression d'un échantillon de gaz que l'on considère comme parfait (au sens où il tombe dans le domaine de validité de la loi des gaz parfaits selon laquelle la produit de la pression P et du volume V d'un échantillon de gaz parfait égale le produit du volume V qu'il occupe par le nombre n de moles contenues dans l'échantillon et de la constante universelle des gaz, R : $pV=nRT$) et qui est confiné dans un volume fixe, il semble intuitivement clair que la pression n'est pas la cause de la température. Il s'agit de deux grandeurs qui appartiennent au même échantillon individuel en même temps et dont la corrélation systématique s'explique par des processus de niveau moléculaire. La loi des gaz parfaits étant symétrique, on ne peut considérer toutes les explications DN qu'elle permet de construire comme causales, sans enfreindre l'asymétrie de la causalité. Si le fait que P (x,t) (la pression de l'échantillon x de gaz au temps t) est proportionnel à T (x,t) suffisait pour montrer que P (x,t) est cause de T (x,t), alors T (x,t) serait cause de P (x,t) pour la même raison.

3. La conception contrefactuelle

Etant donné le nombre et la diversité des contre-exemples qui s'y opposent, il a paru judicieux à de nombreux philosophes d'abandonner le projet d'analyser la causalité en termes de loi ou d'explication DN. Dans un passage qui marque un tournant dans la réflexion philosophique sur la causalité, David Lewis écrit, en 1973 : « Je n'ai pas de preuve que les analyses en termes de régularités sont irréparables, ni l'espace pour passer en revue les réparations qui ont été tentées. Qu'il suffise de dire que les perspectives paraissent sombres.

¹⁷ Popper aussi identifie l'explication causale avec l'explication D-N tout court : "Donner une *explication causale* d'un événement signifie déduire un énoncé le décrivant en utilisant comme prémisses de la déduction une ou plusieurs *lois universelles* et certains énoncés singuliers." (Popper 1934, trad. p. 57; italiques de Popper).

¹⁸ Je ne développerai pas ici les raisons qui ont conduit à abandonner la conception classique de l'empirisme logique, c'est-à-dire l'assimilation de la causalité à l'explication sous la forme d'un argument déductif-nomologique. Voir le chapitre 4 de Barberousse, Kistler, Ludwig (2000).

Je pense qu'il est temps d'abandonner et d'essayer quelque chose de nouveau. Il ne faut pas chercher loin pour trouver une alternative prometteuse » (Lewis 1973/1980, p. 160). L'idée de l'alternative dont parle Lewis figure dans l'*Enquête sur l'entendement humain*. En effet, Hume y fait suivre sa fameuse définition de la causalité en termes de succession de cette seconde définition : une cause est un objet suivi d'un autre, « de telle sorte que le second objet n'aurait jamais existé sans l'existence du premier » (Hume 1777, p. 76 ; trad. p. 110)¹⁹. Cette seconde définition contient l'idée directrice de ce que l'on appelle désormais l'analyse contrefactuelle de la causalité : la proposition « c est cause de e » signifie que « si c ne s'était pas produit, e ne se serait pas produit non plus ». Cette dernière proposition est couramment représentée par l'expression « $C \rightarrow E$ »²⁰. Cette analyse est censée être *a priori*, au sens où elle vise à découvrir, non pas la nature physique des processus causaux réels, mais quelque chose qui est implicitement connu de tout locuteur compétent du français (ou de toute autre langue contenant un synonyme de « cause »), à savoir la signification du concept exprimé par le prédicat « cause ». Dans la tradition de l'empirisme logique, le recours aux contrefactuels devait apparaître méthodologiquement suspect : il semble nécessaire, pour déterminer la valeur de vérité d'un énoncé contrefactuel, de porter des jugements sur des possibilités et donc non observables²¹. Cependant, l'élaboration d'un formalisme permettant d'interpréter des propositions modales et contrefactuelles en termes de mondes possibles a donné un nouveau souffle à l'analyse contrefactuelle de la causalité. La force de l'approche contrefactuelle repose sur la plausibilité initiale de l'idée qu'une cause « fait une différence », ce qui s'exprime assez directement dans un contrefactuel²².

La contribution de David Lewis à l'analyse contrefactuelle de la causalité fut déterminante pour l'orientation de cette approche : toutes les propositions ultérieures partent de l'analyse de Lewis qui propose de concevoir l'évaluation sémantique des contrefactuels en termes de proximité des mondes possibles. Les termes des relations causales et des contrefactuels sont des événements, que Lewis (1986a) conçoit comme des ensembles de régions spatio-temporelles possibles. S'il s'agit de l'explosion d'une bombe à l'endroit x à l'instant t, toutes les « contreparties » de x à t dans d'autres mondes possibles où une bombe explose font partie de l'événement lewisien. Je dirai indifféremment que l'événement – par exemple l'explosion – « a lieu », et qu'« il est le cas » – par exemple, que la bombe explose.

La stratégie adoptée par Lewis pour donner les conditions de vérité des contrefactuels consiste à comparer différents mondes possibles quant à leur similarité globale par rapport au monde réel – que l'on appellera « actuel » au sens modal. Le point de départ est la thèse selon laquelle l'énoncé contrefactuel « si C était le cas, alors E serait le cas » est vrai dans le monde actuel si et seulement si 1) C n'est vrai dans aucun monde possible ou si 2) un monde où C et E sont vrais est plus proche du monde actuel que tous les mondes où C est vrai mais E, faux. Lorsqu'on pose la question de savoir si c est cause de e, l'on présuppose que c a eu lieu, et que C soit donc vrai dans le monde actuel. Sur la base de cette présupposition, c'est la seconde clause qui détermine la valeur de vérité du contrefactuel.

L'analyse lewisienne de la relation causale en termes de contrefactuels est indirecte ; elle passe par l'intermédiaire de la notion de dépendance causale. Si c et e sont deux

¹⁹ Hume ne poursuit pas cette nouvelle idée, et ne fait même pas remarquer qu'elle n'est pas équivalente à l'analyse de la causalité en termes de régularités.

²⁰ Dans la terminologie de Lewis, une majuscule « C » représente la proposition que l'événement désigné par la minuscule correspondante, « c », a eu lieu.

²¹ J. St. Mill (1843) propose d'analyser un contrefactuel « si A avait eu lieu, alors B aurait eu lieu » en termes de la possibilité d'une déduction de B à partir de A et d'un ensemble d'énoncés auxiliaires S qui doit notamment contenir des énoncés de lois de la nature.

²² Mackie (1974, chap. 2) a enrichi l'analyse contrefactuelle par la distinction entre le « champ causal » d'arrière-plan et le facteur saillant que l'on considère comme cause dans la mesure où il « fait une différence » par rapport à cet arrière-plan.

événements actuels distincts²³, e dépend causalement de c si et seulement s'il est vrai que « si c n'avait pas eu lieu, alors e n'aurait pas eu lieu ». Les événements c et e doivent être distincts au sens où la région spatio-temporelle dans laquelle se produit c ne doit avoir aucune partie commune avec celle où se produit e. Cette restriction permet d'éviter de prendre pour causale la dépendance de certaines propriétés d'un événement par rapport à d'autres de ses propriétés : il est clair que la vérité du contrefactuel « si Jean n'avait pas dit 'bonjour', il n'aurait pas dit 'bonjour' à haute voix » ne repose pas sur l'existence d'une relation causale²⁴.

L'analyse contrefactuelle peut rendre compte à la fois de relations causales déterministes et indéterministes. Dans un monde où il y a des lois indéterministes, e dépend causalement de c (où c et e sont des événements distincts qui ont lieu dans le monde actuel) si et seulement si, si c n'avait pas eu lieu, la probabilité de l'occurrence de e aurait été beaucoup plus petite que sa probabilité réelle (Lewis 1986c, p. 176).

Il existe un certain nombre d'objections à l'adéquation de l'analyse lewisienne de la causalité. On a découvert deux espèces de contre-exemples : il existe des « faux positifs » qui semblent montrer que la dépendance contrefactuelle n'est pas suffisante pour l'existence d'une relation causale, et des « faux négatifs » qui semblent montrer qu'elle n'est pas non plus nécessaire. Nous reviendrons un peu plus loin sur ces contre-exemples et les leçons qu'il convient d'en tirer. Plutôt que de les considérer comme des réfutations de l'approche contrefactuelle, ses partisans les considèrent comme des indications de la manière dont il faut la modifier.

Une première difficulté qui se pose à l'analyse contrefactuelle vient de l'existence d'énoncés contrefactuels « à rebours (*backtracking*) » selon lesquels un événement passé dépend d'un événement présent ou futur. Considérez une vague traversant l'océan. Il semble correct de dire : « si la vague n'avait pas été à x en t, elle n'aurait pas été à x-dx à t-dt », où « x-dx » représente l'endroit où se trouvait le sommet de la vague à l'instant t-dt antérieur à t. En général, de tels énoncés contrefactuels à rebours semblent vrais dans des conditions où un événement donné e ne peut avoir qu'une seule cause à un instant antérieur donné, de sorte que c apparaît comme nécessaire pour e. Considérons une situation où une bombe explose à l'instant t. S'il n'existe, à l'instant antérieur t-dt, qu'un événement qui aurait pu causer cette explosion, disons le déclenchement du détonateur, il semble correct de dire : si la bombe n'avait pas explosé, alors son détonateur n'aurait pas été déclenché. Or, s'il y a des contrefactuels vrais à rebours, la dépendance contrefactuelle n'est pas suffisante (ni à plus forte raison équivalente) à la dépendance causale, car l'événement dans le futur n'est pas (du moins dans des circonstances ordinaires) la cause de l'événement dans le passé qui pourtant dépend contrefactuellement de lui : la vague à (x,t) ne cause pas la vague à (x-dx, t-dt), même si le contrefactuel à rebours semble exprimer le fait que la vague à (x-dx, t-dt) dépend contrefactuellement de la vague à (x, t) ; de même, l'explosion de la bombe ne cause pas le déclenchement du détonateur. En d'autres termes, l'analyse contrefactuelle semble prédire, à tort, que les effets sont causes de leurs causes.

Lewis soutient qu'on peut exclure l'usage des contrefactuels à rebours parce qu'ils ne correspondent pas à notre stratégie « standard » (Lewis 1979/1986, p. 35)²⁵ lorsque nous jugeons de la similarité entre mondes possibles. La justification de cette thèse dépend d'une asymétrie contingente mais réelle de notre monde actuel. Selon Lewis (1979/196, p. 49), un

²³ Pour le cas plus général où c et e sont des événements possibles, il faut à la fois qu'il soit vrai que « si c n'avait pas eu lieu, e n'aurait pas eu lieu » et « si c avait eu lieu, alors e aurait eu lieu ».

²⁴ Cf. Kim (1973), Lewis (1986a).

²⁵ Etant donné que les contrefactuels sont en général vagues et que leur évaluation dépend du contexte, Lewis (1979/1980, p. 32-5) reconnaît qu'il existe des contextes particuliers dans lesquels nous acceptons qu'un contrefactuel à rebours puisse être vrai. Cependant, il soutient qu'il est possible d'exclure ces contextes particuliers de l'évaluation des contrefactuels qu'il faut utiliser dans l'analyse de la dépendance causale.

ensemble de conditions est un « déterminant » d'un événement donné si ces conditions, jointes aux lois de la nature, sont suffisantes pour l'occurrence de l'événement. Parmi les déterminants d'un événement, il y a ses causes, ainsi que les traces qu'il laisse. L'asymétrie du monde actuel est fondée sur le fait que les événements ont en général peu de déterminants antérieurs (causes) mais de très nombreux déterminants ultérieurs (traces). C'est ce que Lewis appelle « l'asymétrie de la surdétermination » (p. 49) : un événement ordinaire n'a en général qu'une seule cause ; c'est un fait contingent caractérisant le monde actuel que la surdétermination d'un événement par plusieurs causes est exceptionnelle. Si l'on considère les ondes qui se propagent à partir d'un point sur la surface d'un lac, il n'y a qu'une cause commune à de très nombreuses perturbations sur la surface de l'eau, alors que l'événement à l'origine de l'onde a de très nombreuses traces : l'origine de l'onde est surdéterminée par ses traces dans l'avenir, alors que ces traces ne sont pas surdéterminées par le passé.

Voici le raisonnement par lequel Lewis justifie sa thèse selon laquelle les contrefactuels à rebours sont sans pertinence pour l'analyse des énoncés causaux. Pour juger si *e* dépend causalement de *c*, il faut, selon l'analyse contrefactuelle, évaluer le contrefactuel « si *c* n'avait pas eu lieu, alors *e* n'aurait pas eu lieu », ce qui nécessite la considération des mondes possibles différents où *c* n'a pas lieu. Ces mondes diffèrent du monde actuel, car dans le monde actuel, *c* et *e* ont lieu. Le monde qui, parmi ceux où *c* n'a pas lieu, détermine la valeur de vérité du contrefactuel en déterminant la valeur de vérité du conséquent *e* est le monde le plus proche du monde actuel. Lewis montre que nous considérons qu'un monde est plus proche du monde actuel

- 1) moins il contient d'infractions aux lois du monde actuel, et
- 2) plus sont étendues les régions où il n'existe aucune divergence à l'égard des états de choses²⁶ singuliers par rapport à notre monde actuel.

Rappelons que les mondes possibles pertinents diffèrent tous du monde actuel par le fait que *c* n'y a pas lieu. Dans le cadre d'événements déterminés par des lois déterministes, cette divergence s'accompagne ou bien d'une vaste divergence des états de choses à l'égard des histoires causales qui conduisent respectivement à *c* (dans le monde actuel) et à non-*c* (dans les mondes possibles considérés), ou bien d'une infraction aux lois, autrement dit, du fait que les mondes possibles considérés n'obéissent pas parfaitement aux lois du monde actuel. Lewis soutient qu'il ressort de l'analyse de notre pratique en matière de jugements contrefactuels que nous considérons comme plus proche les mondes qui divergent du nôtre par rapport à *c* suite à une infraction localisée aux lois de la nature que ceux qui sont indemnes de tels « miracles » mais qui diffèrent du nôtre relativement à une large part de leur histoire factuelle.

À ce point, « l'asymétrie de la surdétermination » intervient dans l'analyse pour assurer que les contrefactuels sont évalués selon l'interprétation « standard », c'est-à-dire de manière à ce que l'avenir dépende contrefactuellement du passé, et non l'inverse. Étant donné l'asymétrie de la surdétermination, les mondes dans lesquels ce miracle a lieu *dans le passé* de non-*c*²⁷ sont plus proches que ceux où le miracle a lieu dans le *futur* de non-*c*. Il faut un miracle plus étendu pour faire « reconverger » les mondes non-*c* vers le monde actuel – et assurer ainsi une correspondance parfaite de l'avenir des deux mondes après le miracle - que

²⁶ Le sens technique des expressions « fait » et « état de choses » remonte au *Tractatus* de Wittgenstein (1921). Selon une interprétation importante, un fait (*Tatsache* en allemand) est ce qui rend vrai un énoncé descriptif : la satisfaction d'un prédicat par un objet. Le concept d'« état de chose » (*Sachverhalt* en allemand) est plus général au sens où il s'applique aussi à ce qui est possible, à ce qui pourrait être le cas. S'il est possible que l'objet *a* satisfait le prédicat *P*, alors « *a* est *P* » désigne un « état de choses ». Si *a* est réellement *P*, alors « *a* est *P* » désigne aussi un fait.

²⁷ C'est-à-dire le passé par rapport au temps où *c* a lieu dans le monde actuel. Un événement *e* dans le monde M_1 apparaît comme un miracle par rapport au monde M_2 si les circonstances dans lesquelles *e* apparaît (dans M_1) ne sont pas conformes aux lois de M_2 . *E* est donc un miracle dans M_1 relativement à M_2 .

pour faire diverger le monde non-c du monde actuel – et assurer ainsi une correspondance parfaite des passés des deux mondes avant le miracle. Lewis tire de ce raisonnement la conclusion générale selon laquelle les mondes possibles pertinents contiennent toujours un miracle situé dans le passé immédiat de l'occurrence de l'antécédent. C'est ce choix « standard » de l'importance relative des critères de similarité entre mondes, qui est sous-jacente à notre pratique courante d'évaluation de contrefactuels, qui, joint à l'asymétrie contingente de notre monde actuel, garantit selon Lewis que tous les contrefactuels à rebours sont faux : considérez « si e n'avait pas eu lieu alors c n'aurait pas eu lieu » où e et c sont, comme toujours dans le contexte de l'évaluation de jugements causaux, des événements qui ont lieu dans le monde réel, et où e a lieu *après* c. Les mondes pertinents sont des mondes où l'antécédent non-e est vrai à cause d'un petit miracle qui a lieu *immédiatement avant* e. Le miracle a donc lieu après c ; par conséquent, c a lieu dans le monde le plus proche où l'antécédent du contrefactuel est vrai ; le conséquent du contrefactuel à rebours y est donc faux, et le contrefactuel est faux lui aussi.

L'argument montrant que les contrefactuels à rebours sont systématiquement faux permet aussi d'éviter ce que Lewis (1986b, p. 170) appelle le « problème des épiphénomènes » : considérons un événement c qui cause deux effets e et f, sans que e soit pour autant cause de f. L'analyse de Lewis semble prédire à tort que e cause f parce qu'il semble y avoir une chaîne de dépendances contrefactuelles entre e et f : si c est la seule cause qui aurait pu, dans les circonstances, causer f, alors f dépend contrefactuellement de c, et si c est suffisant pour e, c semble dépendre contrefactuellement de e : si e n'avait pas eu lieu, c n'aurait pas eu lieu. Or si Lewis a raison d'affirmer que nos critères d'évaluation des contrefactuels garantissent, dans le contexte de l'asymétrie de la détermination, que les contrefactuels à rebours sont toujours faux, alors ce dernier contrefactuel est faux, et il n'y a après tout aucune chaîne de dépendance contrefactuelle entre les deux épiphénomènes e et f.

Plusieurs objections ont été soulevées contre ce raisonnement. Horwich (1987, p. 10) fait valoir que l'asymétrie de la surdétermination n'est connue que de manière scientifique et *a posteriori* ; dans la mesure où il ne s'agit pas d'un aspect de la réalité qui est connu *a priori* par tous les locuteurs compétents, une analyse conceptuelle de la causalité ne peut y faire appel²⁸. Plusieurs auteurs ont remis en question la correction scientifique de la thèse lewisienne (et poppérienne (Popper 1956)) selon laquelle un événement a typiquement peu de déterminants qui le précèdent mais de nombreux déterminants qui lui succèdent, ou en d'autres termes peu de causes et beaucoup de traces. Pour ce qui est des lois déterministes et symétriques de la mécanique classique, cette différence n'est qu'illusoire. Elga (2000) montre que, pour des contrefactuels dont l'antécédent porte sur un événement irréversible au sens thermodynamique (d'augmentation de l'entropie), il est erroné de dire, comme le fait Lewis, que les mondes où cet antécédent est vrai en vertu d'un miracle qui a lieu immédiatement *avant* l'antécédent sont plus proches que les mondes où le miracle a lieu *après* l'antécédent. Elga illustre son propos avec l'exemple de Gretta qui casse, dans le monde actuel W_1 , un œuf dans sa poêle à 8h. Considérons les mondes les plus proches dans lesquels Gretta ne casse pas d'œuf à 8h. Selon Lewis, il ne faut qu'un petit miracle, par exemple dans les processus ayant lieu dans le cerveau de Gretta juste avant 8h, disons à 7h59, pour qu'elle ne casse pas d'œuf, de sorte que le monde miraculeux W_2 correspond exactement aux faits dans l'ensemble de l'histoire avant 8h, et n'en diverge qu'après le miracle. Mais Elga montre qu'il y a des mondes W_3 qui, au contraire, partagent avec le monde actuel l'ensemble du futur à partir d'un moment peu après 8h, disons à partir de 8h05, dans lesquels Gretta ne casse pas d'œuf mais dans lesquels le miracle qui permet la convergence avec le monde actuel n'est pas plus grand que le miracle qui a lieu dans le monde W_2 . Elga nous fait considérer un processus qui

²⁸ Lewis répond à cette objection dans (1979/1986, p. 66).

correspond au processus réel de 8h à 8h05 mais qui se déroule en direction inverse, comme lorsqu'on rembobine un film. L'œuf cassé à 8h05 dans la poêle « décuit (*uncooks*) » et retourne dans la coquille d'œuf. Ce processus est en accord avec les lois physiques mais il est très improbable car il dépend de manière extrêmement sensible des conditions initiales : si on effectue un changement infime dans les conditions de positions et de vitesses des molécules à 8h05, on observera un processus plus banal où l'œuf reste dans la poêle et se met à refroidir. Il suffit donc de faire un petit miracle à 8h05, pour faire en sorte que tout le passé change, y compris l'acte où Gretta casse l'œuf à 8h.

Nous avons vu que Lewis utilise la notion de dépendance causale comme intermédiaire entre la dépendance contrefactuelle et l'existence d'une relation causale : c est cause de e si et seulement si il existe une chaîne finie d'événements intermédiaires e_1, e_2, \dots, e_k , entre c et e, de telle sorte que le deuxième maillon dépend causalement du premier, et en général si, pour tout n, le maillon n dépend causalement du maillon précédent n-1. La dépendance causale est ensuite, comme nous l'avons vu, réduite à la dépendance contrefactuelle.

Cette analyse résout deux difficultés : premièrement elle garantit la transitivité de la relation causale, et deuxièmement elle permet de justifier l'intuition qu'une cause « préemptée » n'est qu'une cause potentielle mais non réelle.

1) La dépendance contrefactuelle n'est en général pas transitive : il est facile de trouver des exemples où il est vrai que $A \rightarrow B$ et que $B \rightarrow C$, mais faux que $A \rightarrow C$. La raison en est que l'évaluation d'un contrefactuel dépend des circonstances d'arrière-plan de l'antécédent, et que les arrière-plans des antécédents d'une série de contrefactuels ne sont en général pas identiques. Lorsque l'on réduit la relation causale, non pas à la dépendance causale, mais à l'existence d'une chaîne d'événements contrefactuellement dépendants les uns des autres par paires, le premier et le dernier maillon d'une chaîne causale sont assurés d'être liés comme le sont une cause et un effet, alors que le dernier maillon ne dépend en général pas contrefactuellement du premier. Cependant, cet aspect de l'analyse lewisienne peut aussi donner lieu à une objection : un certain nombre d'auteurs soutiennent qu'il existe des contre-exemples à la transitivité de la causalité. De tels contre-exemples concernent en particulier des jugements dans lesquels une absence, ou un aspect particulier d'un événement, joue le rôle de cause ou d'effet, ou des jugements dans lesquels le lien causal est fondé sur une double prévention²⁹. Dans un exemple proposé par Ehring (1987), Jones met du sel de potasse dans un feu, ce qui provoque un changement de la couleur de la flamme qui devient pourpre. Par la suite, la flamme incendie un morceau de bois à proximité. Il existe une chaîne causale entre l'acte de mettre du sel de potasse dans le feu et l'inflammation du morceau de bois, mais il semble faux de dire que le premier événement a causé le dernier³⁰. On peut défendre la transitivité de la causalité contre certains de ces contre-exemples, en montrant que l'apparence de l'existence d'une chaîne causale est due à une conception trop grossière des termes des relations causales. Si les termes des relations causales ne sont pas les événements concrets directement mais des *faits* portant sur les événements, il n'y a plus de chaîne entre l'acte de jeter le sel dans le feu et l'incendie du morceau de bois : le sel est causalement responsable du fait que la flamme change de couleur ; mais la cause de l'incendie n'est pas le fait que la flamme change de couleur mais le fait qu'elle dégage de la chaleur³¹. On peut aussi défendre la thèse selon laquelle il n'existe pas de relations causales comportant des termes « négatifs », tels que des absences ou omissions : il s'agit souvent, au contraire, d'explications non causales qui peuvent donner l'impression trompeuse de causalité dans la mesure où elles présentent une situation sans processus causal en l'opposant à une situation d'arrière-plan où

²⁹ Voir Bennett (1987), Hall (2004a).

³⁰ On peut trouver d'autres exemples de ce genre dans McDermott (1995), Hall (2000/2004b), Paul (2004).

³¹ Voir Kistler (2001). Paul (2004) propose une analyse semblable où elle défend l'idée selon laquelle les termes de relations causales sont des aspects d'événements.

existe un processus causal³². Les chaînes explicatives impliquant une double prévention n'indiquent donc pas l'existence de chaînes causales. Pour reprendre un exemple que Hitchcock (2001) attribue à Ned Hall³³, un randonneur voit la chute d'un rocher, ce qui l'incite à s'accroupir pour éviter d'être touché. Le fait de ne pas avoir été touché est la cause de la suite de sa randonnée. Il s'agit de double empêchement (*double prevention*) au sens où le geste de s'accroupir empêche le rocher d'empêcher la suite de la randonnée. Il semble faux de dire que la chute du rocher cause la suite de la randonnée alors qu'il semble y avoir une chaîne causale qui va du premier événement au dernier. Mais on peut nier qu'il s'agisse d'une chaîne causale, et ainsi défendre la transitivité de la causalité, en niant que le fait de *ne pas* être touché par le rocher puisse être un effet ou une cause.

2) Le second problème est que l'introduction d'une chaîne d'événements intermédiaires permet de résoudre vient des situations de « préemption ». Lewis parle aussi de « causalité redondante ». De telles situations sont courantes notamment en biologie. L'évolution a souvent fait apparaître, à côté d'un mécanisme important pour la survie, des mécanismes de secours (*backup*) qui prennent la relève en cas de défaillance du mécanisme principal. Les philosophes ont tendance à raisonner sur des cas qui relèvent de l'action humaine : parmi les situations paradigmatiques de préemption qui sont considérées dans la littérature, on trouve celle des deux snipers S_1 et S_2 qui visent la même victime. S_1 décide de tirer (événement a) ; cette décision cause son tir, qui cause la mort de la victime (événement c). S_2 qui voit S_1 tirer ne tire plus et ne cause donc pas c ; cependant, c semble dépendre de la même manière de la décision de tirer de S_2 (événement b) que de la décision de tirer de S_1 (événement a). D'une part, si aucun des deux ne tirait, la victime ne mourrait pas ; d'autre part, il est vrai de chacun de S_1 et S_2 que s'il tirait la victime mourrait. Or, la dépendance contrefactuelle de c, par rapport à la décision de S_2 de tirer, ne semble dans ce cas pas *suffire* pour garantir l'existence d'une relation causale.

L'exigence de l'existence d'une chaîne d'événements intermédiaires permet de surmonter cette difficulté : dans le cas de a, les positions de la balle lors de sa trajectoire de a à c constituent une telle chaîne. Cependant, étant donné que S_2 ne tire pas, il n'existe entre b et c aucun événement intermédiaire dont la mort de la victime dépende contrefactuellement et qui dépende de b. L'analyse de Lewis débouche donc sur le résultat intuitivement correct que b n'est pas cause de la mort de la victime. On parle dans ce cas de « préemption précoce (*early preemption*) », dans la mesure où la chaîne causale potentielle entre b et c est interrompue tôt, c'est-à-dire suffisamment longtemps avant c pour qu'il existe une chaîne d'événements entre a et c à laquelle ne correspond aucune chaîne parallèle entre b et c.

Cependant, les situations de préemption précoce posent un second problème, dans la mesure où elles semblent montrer que la dépendance contrefactuelle n'est pas *nécessaire* pour la causalité. En présence de b, c ne dépend pas plus contrefactuellement de a que de b, car il n'est vrai ni de a ni de b que, s'il ne s'était pas produit, c ne se serait pas produit non plus. Lewis résout ce problème en même temps que celui que posent les situations de « préemption tardive (*late preemption*) ». Elles se caractérisent par l'existence d'une chaîne continue entre un événement b et un événement c, sans que b cause c : Hall (2004a, p. 235) examine la situation où deux enfants (Suzy et Billy) lancent des cailloux sur une bouteille. Suzy lance son caillou un petit peu plus tôt que Billy, de sorte qu'il casse la bouteille (événement c). Mais la trajectoire du caillou lancé par Billy suit celle du caillou de Suzy de près, de sorte qu'il existe non seulement une chaîne d'événements entre le lancer de Suzy et c, mais aussi entre le lancer de Billy et c. Pourtant, dans la mesure où le caillou de Suzy parvient un instant plus tôt à la bouteille que le caillou de Billy, la cause de c est le lancer de Suzy et non le lancer de Billy.

³² Cf. Kistler (1999), Hall (2004a), Kistler (2006).

³³ Selon Hitchcock (2001, p. 276), cet exemple figure dans une version non publiée de Hall (2004a).

Dans « Postscripts to ‘Causation’ », Lewis (1986c) introduit le concept de « quasi-dépendance » qui lui permet de résoudre le problème de la préemption tardive, ainsi que le problème selon lequel la dépendance causale n’est pas nécessaire à la causalité. La présence de l’événement préempté *b*, et dans le cas de la préemption tardive, la présence de toute une chaîne parallèle de *b* à *c*, n’empêche pas l’événement « préemptant » *a* de causer *c*. La raison pour laquelle la présence de la cause redondante *b* ne prive pas *a* de son efficacité par rapport à *c* est le fait que la causalité est une propriété *intrinsèque* du processus localisé entre *a* et *c*. Selon Lewis, chaque événement dans la chaîne entre *a* et *c* est quasi-dépendant de son prédécesseur dans la chaîne parce qu’il ressemble intrinsèquement – c’est-à-dire en prenant en compte uniquement les événements localisés sur la chaîne entre *a* et *c* – à des processus dont les éléments sont pleinement contrefactuellement (et donc causalement) dépendants de leurs prédécesseurs. *a* (c’est-à-dire le lancer de Suzy) est la cause de *c* parce qu’il ressemble intrinsèquement à des lancers possibles que Suzy exécute en l’absence de lancers de Billy. Puisque ces derniers lancers sont caractérisés par la dépendance contrefactuelle, *c* est quasi-dépendant du lancer de Suzy.

Cependant, des cas récalcitrants de préemption sont apparus, dans lesquels il existe une chaîne d’événements intermédiaires qui rendent l’effet *c* « quasi-dépendant » de l’événement préempté *b*. Dans une situation de cette espèce, que son découvreur Schaffer (2000) appelle la « préemption coupante (*trumping*) », un major et un sergent crient des ordres aux soldats. Les deux crient, en même temps, « avancez », et les soldats avancent. Etant donné que les soldats obéissent aux ordres de l’officier supérieur, ils avancent parce que le major le leur ordonne, et non parce que le sergent le leur ordonne. Les décisions des soldats sont quasi-dépendantes aussi bien de l’ordre du sergent que de l’ordre du major. La chaîne qui va de l’un et de l’autre aux décisions des soldats est intrinsèquement semblable aux chaînes qui, en l’absence du second officier, assurent la dépendance contrefactuelle de proche en proche et donc l’existence d’une relation causale. La quasi-dépendance n’est donc pas, après tout, suffisante pour l’existence d’une relation causale.

Cette difficulté a conduit Lewis (2000) à une nouvelle version de la théorie contrefactuelle, en termes d’« influence ». Lewis suggère que pour qu’il soit correct de dire que *c* cause *e*, il n’est pas suffisant que *le fait* qu’*e* se produise soit contrefactuellement dépendant du fait que *c* se produise. Il faut exiger en outre que la manière dont se produit *e* et l’instant auquel *e* se produit soient également contrefactuellement dépendants de la manière dont *c* se produit et de l’instant auquel *c* se produit. La nouvelle analyse utilise le concept d’altération d’un événement : une altération d’un événement *e* donné est un événement actuel ou seulement possible qui diffère légèrement de *e*, soit par ses propriétés, soit par l’instant auquel il se produit. Si un événement *c* influence un autre événement *e*, « il existe une série de dépendances contrefactuelles entre l’occurrence de *e*, la façon dont il se produit, et le moment où il se produit d’une part, et l’occurrence de *c*, la façon dont il se produit, et le moment où il se produit, d’autre part (*a pattern of counterfactual dependence of whether, when and how on whether, when and how*) » (Lewis 2000/2004, p. 91). Plus précisément : « si *c* et *e* sont des événements actuels distincts, disons que *c* influence *e* si et seulement s’il existe une gamme substantielle c_1, c_2, \dots d’altérations de *c* différentes mais peu éloignées les unes des autres (*different not-too-distant alterations*) (y compris l’altération actuelle de *c*) et une gamme substantielle e_1, e_2, \dots d’altérations de *e* dont au moins certaines sont différentes, telles que si c_1 s’était produit, e_1 se serait produit, et si c_2 s’était produit, alors e_2 se serait produit, etc. » (Lewis 2000/2004, p. 91). Comme dans l’analyse originale, le fait que *c* cause *e* est réduit à l’existence d’une chaîne intermédiaire dont chaque maillon influence le maillon suivant.

Une autre objection qui a été soulevée contre l’analyse contrefactuelle concerne le fait qu’elle ne respecte pas la distinction de sens commun entre les causes et les conditions

d'arrière-plan. On peut envisager de rejeter cette distinction, en disant comme Mill qu'elle ne reflète que les intérêts de l'observateur humain, et que, « d'un point de vue philosophique » (*philosophically speaking*), les conditions d'arrière-plan sont des causes au même sens que les facteurs saillants que le sens commun reconnaît comme tels. Cependant, dans la mesure où le but de l'analyse contrefactuelle n'est pas la nature physique de la causalité réelle mais la structure de notre concept naïf de causalité, il semble essentiel que l'analyse respecte cette distinction. Cela est possible en faisant l'hypothèse que les énoncés causaux ordinaires selon lesquels « c cause e » contiennent une comparaison implicite à une situation « normale » d'arrière-plan, ce qui peut être rendu explicite sous la forme : « c au lieu de c* a causé e au lieu de e* ». Par conséquent, l'analyse contrefactuelle appropriée est : « si c* s'était produit à la place de c, alors e* se serait produit à la place de e »³⁴. Cette idée est étroitement liée à l'intuition selon laquelle une cause est ce qui fait une différence par rapport à un effet donné : on compare, quoique souvent implicitement, la situation en présence de la cause en question à la situation, telle qu'elle aurait été si la cause avait été absente. Si l'effet est présent dans la situation où la cause est présente mais absente sinon, on a une raison de penser que la cause est responsable de cette différence. Pour reprendre un exemple de Dretske (1977), le fait que Socrate a bu la ciguë est la cause de sa mort dans la mesure où c'est le facteur qui fait la différence cruciale à l'égard de la mort. Les nombreuses autres caractéristiques de la situation, tels que le fait que Socrate boive la ciguë à l'aube, ne sont pas des causes de sa mort. L'heure ne fait aucune différence à la fatalité de la ciguë.

4. Méthodologie

Les modifications successives de l'analyse contrefactuelle sont motivées par la volonté d'éviter deux sortes de contre-exemples : les « faux positifs » sont des situations dans lesquelles il existe deux événements dont on juge intuitivement qu'ils ne sont pas liés comme une cause à un effet mais dont l'analyse prédit qu'ils le sont. Les « faux négatifs » sont au contraire des situations où un événement c est intuitivement cause d'un autre événement e, alors que l'analyse prédit que ce n'est pas le cas. Il s'agit des deux formes possibles de décalage entre une analyse donnée et l'intuition. Autrement dit, l'accord avec les intuitions du sens commun semble être ici un critère important d'adéquation de l'analyse. Cependant, il n'existe pas de consensus quant au choix d'un tel critère d'adéquation. La diversité des analyses de la causalité dans la littérature philosophique s'explique au moins en partie par l'existence de différentes manières de concevoir le but et la méthode de cette analyse. Une division majeure oppose les analyses *a priori* aux analyses *a posteriori*.

1. Les défenseurs de la conception contrefactuelle visent « l'analyse conceptuelle ». Il s'agit de trouver une analyse de la causalité qui vaut dans tous les mondes possibles. Le concept humain de causalité est considéré comme un objet d'étude indépendant de la nature physique des processus causaux que se déroulent dans le monde. Son analyse en tant que concept peut être menée à bien de manière purement *a priori*, notamment en s'interrogeant « dans son fauteuil » sur ses intuitions spontanées dans un certain nombre de situations fictives. Ces situations peuvent être réalistes : des enfants qui lancent des cailloux sur une bouteille, ou des supérieurs militaires qui donnent des ordres aux soldats. Mais l'analyse *a priori* de notre concept de causalité peut tout aussi bien utiliser des intuitions portant sur des situations irréelles et même physiquement impossibles, telles que des magiciens lançant des sorts. Dans un exemple largement commenté de Schaffer (2004, p. 59), Merlin lance un sort qui transforme un prince en grenouille : de telles interactions causales

³⁴ Cf. Hitchcock (1996a), (1996b), Maslen (2004), Schaffer (2005).

magiques ne sont pas contraintes par la nature physique et peuvent se produire à distance (spatiale et temporelle) sans intermédiaires causaux.

2. Partant de l'analyse du concept du sens commun, la théorie peut imposer des corrections, en vue d'une plus grande cohérence et d'une plus grande systématisme, sans pour autant quitter le cadre des contraintes *a priori*. On découvre par exemple que l'on peut juger intuitivement à la fois que les glaçons (plus exactement la fonte des glaçons) dans un verre d'eau cause le refroidissement de l'eau, et que le refroidissement de l'eau (le fait que l'eau dégage de la chaleur) cause la fonte des glaçons. Or, l'ensemble de ces deux jugements enfreint l'asymétrie, qui est comme nous l'avons vu, une composante essentielle du concept de causalité. La théorie peut donc juger, contre l'intuition première, que l'un des deux jugements est erroné. Cependant, il n'existe dans ce cadre aucune raison de nier l'une plutôt que l'autre.

3. On peut considérer que le but de la réflexion philosophique sur la causalité n'est pas seulement l'analyse conceptuelle du sens commun : la causalité peut être conçue comme une « espèce naturelle » de relation dont l'essence doit faire l'objet d'une découverte *a posteriori*. C'est de cette manière que les théories de la causalité en termes de processus conçoivent leur tâche. Dans une telle perspective, la relation causale dont on cherche « l'essence réelle » n'existe pas dans tous les mondes possibles. Dans ce cadre, on peut chercher une raison scientifique de suivre l'intuition à l'égard de l'un plutôt que de l'autre des deux jugements précédents qui enfreignent l'asymétrie : le jugement selon lequel le refroidissement de l'eau cause la fonte des glaçons correspond au transfert de chaleur, alors qu'aucun processus n'existe en sens inverse³⁵.

Du point de vue du projet de l'analyse conceptuelle, une approche qui se soucie des contraintes que la nature physique impose aux interactions causales réelles paraît souffrir d'un « manque d'ambition » (Collins et al. 2004, p. 14). Pour les approches *a priori*, l'analyse du concept de causalité doit s'appliquer dans tous les mondes possibles, et en particulier à des « mondes où les lois sont très différentes de ce qu'elles sont dans notre monde actuel ». Le fait de se limiter à chercher l'analyse de processus causaux tels qu'il sont dans le monde actuel avec ses lois spécifiques « est non seulement malheureux mais profondément erroné (*misguided*) » (Collins et al., p. 14) car il s'agit selon les défenseurs de l'analyse conceptuelle, de trouver une analyse qui « est non seulement vraie mais nécessairement vraie » (Collins et al., p. 14).

Le défenseur de l'idée selon laquelle la relation causale est une espèce naturelle de relation dont il s'agit de découvrir, à partir de contraintes conceptuelles et empiriques, la nature, peut rétorquer qu'il s'agit là de deux projets distincts quoique liés : la différence entre la recherche sur le concept naïf de causalité et celle sur l'essence de la causalité dans le monde réel est analogue à la différence entre la recherche psychologique sur la « physique naïve » et la recherche physique, ou entre la recherche psychologique sur la « biologie naïve » et la recherche biologique. Les concepts physiques naïfs et les convictions naïves sur les propriétés et l'évolution des objets physiques ne déterminent que très partiellement les concepts et théories de la physique scientifique. De manière analogue, il est envisageable que nos convictions *a priori* sur la causalité ne contraignent que partiellement la théorie de la causalité en tant que relation naturelle objective : cette dernière est également contrainte par la découverte de la nature physique des relations causales réelles.

Il y a une manière de concilier l'approche conceptuelle *a priori*, adéquate dans tous les mondes possibles, et l'approche de la causalité comme espèce naturelle de processus : selon

³⁵ Rien n'empêche bien entendu de décrire le processus de diffusion de chaleur de manière négative ; au lieu de dire que l'eau transmet de la chaleur sur les glaçons, on peut dire que la présence d'un corps plus froid diminue la chaleur contenue dans l'eau.

ce qui est parfois appelé le « plan de Canberra »³⁶, l'analyse de la causalité procède en deux étapes. La première relève de l'analyse conceptuelle : on détermine les contraintes qu'une relation réelle doit satisfaire pour pouvoir être considérée comme relation causale. La transitivité ou l'asymétrie peuvent compter parmi ces contraintes conceptuelles. Dans une seconde étape, on découvre de manière empirique quelles relations ou quels processus réels satisfont ces contraintes. L'idée est d'appliquer au concept de causalité la stratégie de la réduction fonctionnelle (Jackson 1998, Kim 1998) qui permet de réduire, de manière générale, les concepts de sens commun à des concepts scientifiques. Lors de la première étape conceptuelle, on montre que le concept d'eau, par exemple, est un concept fonctionnel qui s'applique à une substance dans la mesure où elle satisfait à un certain nombre de conditions fonctionnelles : elle est liquide à température ambiante, elle est transparente mais réfracte la lumière avec un indice spécifique, elle gèle à 0°C et s'évapore à 100°C à la pression atmosphérique moyenne du niveau de mer etc. Lors de la seconde étape, on découvre, de manière empirique, que les substances qui satisfont, dans le monde réel, à ces conditions, sont composées essentiellement de molécules d'H₂O.

5. La causalité comme processus

L'analyse contrefactuelle de la causalité était surtout motivée par l'existence de différents types de « faux positifs » : il existe des faits qui peuvent, sur l'arrière-plan des lois de la nature, jouer le rôle de prémisses et de conclusions d'arguments déductifs, sans être liés comme des causes à des effets. Or il s'avère que certaines situations qui réfutent l'analyse déductive-nomologique sont également des faux positifs réfutant l'analyse contrefactuelle. Dans certaines conditions d'arrière-plan, étant donné deux effets e_1 et e_2 d'une cause commune c , e_1 peut servir de prémisse pour prédire et même pour expliquer e_2 , et inversement. Or dans des circonstances appropriées, e_1 et e_2 peuvent aussi être mutuellement contrefactuellement dépendants l'un de l'autre. Il paraît plausible que ce parallèle ne soit pas fortuit : la dépendance nomologique (qui est selon l'analyse DN à la base de tous les jugements causaux vrais) crée une dépendance contrefactuelle, et ce, à la fois dans les cas où la dépendance nomologique correspond à une relation causale et dans les cas où la dépendance nomologique existe sans aucune relation causale. La condition de dépendance contrefactuelle apparaît donc comme trop faible pour garantir l'existence d'un lien causal. Nous avons déjà considéré la controverse autour de la suggestion de Lewis de ne pas considérer la dépendance entre e_1 et e_2 comme pertinente pour la causalité parce qu'elle passe par un événement « intermédiaire » (la cause commune c), à laquelle l'un des effets est lié par un contrefactuel à rebours. De toute manière, cette solution ne s'applique pas aux cas de dépendance contrefactuelle entre des aspects d'un même événement ou d'une même situation : si l'échantillon de gaz g n'avait pas été à la température T (en supposant le volume fixé), il n'aurait pas été à la pression P . Ou encore, si l'énergie cinétique des molécules contenues dans g n'avait pas été E , alors la température de g n'aurait pas été $T=2E/3k_B$. (où k_B représente la constante de Boltzmann). Il est généralement accepté que la relation causale requiert que les termes de la relation soit distincts à la fois spatialement et temporellement. « Il faut que c et e soit des événements distincts – et distincts non seulement au sens d'être non-identiques mais aussi au sens que l'un ne chevauche pas l'autre ni ne l'implique » (Lewis 2000, p. 78)³⁷. La pression et la température du même gaz au même instant ne peuvent pas

³⁶ Cette expression qui fait allusion à l'université nationale australienne (ANU) située à Canberra, a été introduite par O'Leary-Hawthorne et Price (1996) dans le contexte de l'analyse des concepts de vérité, de référence et de croyance. Lewis (2000/2004, p. 76) l'applique à l'analyse du concept de causalité.

³⁷ « C and E must be distinct events – and distinct not only in the sense of nonidentity but also in the sense of nonoverlap and nonimplication. »

être liées comme cause et effet parce qu'il n'y a entre ces instances de propriétés aucune distance spatiale ni temporelle. Le même constat s'impose pour la température et l'énergie cinétique moyenne des molécules. Les situations de préemption montrent que la dépendance contrefactuelle n'est pas non plus nécessaire : en présence d'une cause parallèle redondante, une cause peut causer un événement sans que celui soit contrefactuellement dépendant de sa cause.

Il est possible d'éviter les deux problèmes en analysant la causalité en termes d'un processus local qui s'étend entre deux événements localisés dans l'espace et le temps. Il existe plusieurs versions de ces analyses en termes de processus. L'une de leurs sources historiques est l'analyse de la causalité en termes de « lignes causales » développée par Russell (1948), à partir de la notion physique de « ligne de monde » (ou « ligne d'univers ») (en allemand *Weltlinie*, en anglais *world line*). On parvient au concept de ligne de monde à partir de celui de la trajectoire d'un objet à travers le temps. La trajectoire elliptique de la Terre autour du Soleil est une courbe elliptique fermée dans l'espace à trois dimensions. Lorsqu'on ajoute la représentation du temps sous forme d'une quatrième dimension, suivant en cela l'unification des dimensions spatiales avec la dimension temporelles accomplie par la théorie de la relativité, on parvient à la ligne de monde de la Terre qui est une courbe ouverte dans un espace à quatre dimensions.

Une ligne causale est une ligne de monde qui satisfait une condition supplémentaire : le long de la ligne, il existe des qualités ou structures constantes ou qui changent de manière continue mais non de manière soudaine et avec une grande magnitude : « Tout le long d'une ligne de monde donnée, il peut y avoir constance de qualité, constance de structure, ou changement graduel dans l'un des deux, mais aucun changement soudain de grandeur considérable » (Russell 1948, p. 477). Cette exigence est censée garantir que la causalité nous permet d'acquérir des connaissances. Pour Russell, comme pour Hume, la seule manière dont nous pouvons justifier des croyances dont la portée dépasse les données sensorielles immédiates, c'est de s'appuyer sur la causalité. La perception d'une table ne me donne des connaissances sur la table, en non seulement sur mon impression sensorielle de la table, parce que cette impression sensorielle est liée par une chaîne causale à la table, plus précisément à des événements d'interaction entre la lumière du Soleil et la surface de la table. Russell définit la ligne causale par rapport à son pouvoir de justifier nos inférences à ce qui se passe à quelque distance de nous : « Une 'ligne causale', comme je souhaite définir le terme, est une série temporelle d'événements qui sont liés entre eux de telle sorte que, étant donné certains d'entre eux, on peut inférer quelque chose sur les autres, quoi qu'il arrive ailleurs. » (Russell 1948, 477). Il est clair que toute inférence de cette sorte est inductive, et donc faillible. Russell fait notamment remarquer que l'inférence des effets à partir d'une cause donnée est plus fiable que l'inférence à rebours, qui infère la cause à partir d'un effet donné. La raison est qu'un même type d'événement peut souvent avoir des causes différentes. Or, les inférences qui nous procurent des connaissances sur le monde à l'extérieur de nos organes sensoriels sont précisément de cette seconde sorte plus fragile.

Russell définit les lignes causales comme étant les lignes de mondes dont la continuité qualitative peut nous servir de justification inductive pur élargir nos connaissances au-delà de nos perceptions. Le fait que les lignes causales soient définies en fonction d'une exigence épistémique est ce qui les rend inadéquates à un concept métaphysique de causalité : fonder la causalité sur les lignes causales russelliennes, ce serait commettre l'erreur de rendre l'existence des liens causaux dépendante de l'existence de processus d'inférence humaine. La faillibilité des inférences fondées sur la continuité des lignes causales montre qu'une telle ligne ne peut pas être plus qu'un symptôme faillible de l'existence d'un processus causal réel ; mais le fait d'être une ligne causale ne peut pas être une condition nécessaire est suffisante de l'existence d'un processus causal réel. Ce n'est pas une condition suffisante : la continuité de

structure ou de qualité caractérise également certains pseudo-processus (Salmon 1984). Ce sont des lignes de monde qui donnent à l'observateur humain l'apparence trompeuse d'un processus causal. Leur continuité qualitative les qualifie de lignes causales russelliennes, mais il ne s'agit pas réellement de processus causaux. L'exemple de Salmon est la tâche lumineuse projetée sur la surface intérieure d'un cylindre par un phare qui tourne au centre du cylindre. La ligne de monde caractérisée par la série des lieux sur lesquels apparaît un point lumineux est une ligne causale sans être un processus causal. Le parcours de la tâche lumineuse se caractérise par une parfaite continuité qualitative. Pourtant, il ne s'agit pas d'un processus causal dans la mesure où les tâches lumineuses n'exercent aucune influence causale les unes sur les autres : le point lumineux qui apparaît à x en t ne cause pas le point immédiatement adjacent, car chacun de ces points résulte d'un processus causal qui a son origine dans le phare.

Partant de Russell et du critère de la « transmission d'une marque »³⁸ de Reichenbach (1956), Salmon (1984) propose d'analyser le concept de processus causal ainsi : c'est un processus qui 1) a une structure ou des qualités qui sont permanents ou ne changent que de manière continue et 2) est capable de transmettre une marque, autrement dit, une modification locale de structure. Le point sur la paroi du phare n'est par exemple pas un processus causal parce que, si l'on modifie sa couleur en interposant un filtre rouge entre le phare et la paroi en un point, cette modification ne se propage pas aux points lumineux suivants.

Le critère de la transmission de marques soulève plusieurs difficultés³⁹ : il semble qu'il existe des processus causaux qui se caractérisent par des changements qualitatifs importants et très rapides, par exemple lorsque de nombreuses particules différentes se succèdent au cours d'une « cascade » de décomposition radioactive. Cela contredit l'exigence de continuité de structure. Dès qu'une ligne de monde connaît des changements rapides relativement à l'échelle humaine, de telle sorte que son observation ne donne pas à un observateur humain ordinaire l'impression de la constance qualitative ou du changement continu, il ne s'agit pas d'une ligne causale russellienne. Il ne s'agit pas non plus d'un processus causal, tel qu'il est défini par Salmon. Salmon part du concept russellien de ligne causale, ce qui requiert l'existence d'une structure préservée le long de la ligne, pour lui ajouter l'exigence supplémentaire de la transmission. « Un processus donné, fut-il causal ou pseudo, a un certain degré d'uniformité – nous pouvons dire, de manière quelque peu relâchée, qu'il exhibe une certaine structure. La différence entre un processus causal et un pseudo-processus est que le processus transmet sa propre structure alors que le pseudo-processus ne le fait pas. » (Salmon 1984, p. 144) Une ligne de monde sujette à des changements qualitatifs rapides et importants, toujours relativement à l'échelle temporelle de l'observateur humain ordinaire, ne satisfait même pas aux conditions que Salmon impose aux processus : « les processus peuvent être identifiés comme paths spatio-temporels qui affichent (*exhibit*) de la continuité et un certain degré de constance de marque » (Salmon 1990, p. 298). A plus forte raison, elle ne peut pas correspondre à un processus causal. D'un autre côté, il semble qu'il existe des pseudo-processus capables de transmettre des marques. Kitcher (1989, p. 463) mentionne les marques dérivées : lorsque le passager d'une voiture tend un drapeau par la fenêtre, l'ombre projetée par la voiture sur un mur porte la marque du drapeau. Par ailleurs, l'analyse des notions de marque et d'interaction causale semble circulaire : une marque est une modification de structure qui est introduite dans un processus par une interaction causale, mais une interaction est causale lorsqu'elle provoque l'introduction d'une marque.

Une tradition qui remonte au 19^e siècle⁴⁰ identifie les processus causaux aux processus de transmission d'énergie ou de quantité de mouvement (Aronson 1971, Fair 1979), ou, plus

³⁸ Je traduis par « marque » l'allemand *Merkmal* et l'anglais *mark*.

³⁹ Ces difficultés ont conduit Salmon (1994) à l'abandonner.

⁴⁰ Cf. Krajewski (1982).

généralement, d'une quantité d'une grandeur conservée (Salmon 1994, Kistler 1998, 1999). Cette approche est motivée par une intuition « mécaniste » selon laquelle l'influence causale ne se propage que par contact et à une vitesse finie. Cette intuition se fait jour lorsqu'on contemple certaines situations qui posent problèmes aux théories qui analysent la causalité en termes de régularités nomologiques ou de dépendance contrefactuelle : il y a succession régulière entre les chutes de baromètre et les orages ; les orages dépendent aussi contrefactuellement des chutes de baromètre : si le baromètre n'avait pas chuté il n'y aurait pas eu d'orage. Or il semble intuitivement clair que la raison pour laquelle le baromètre n'est néanmoins pas la cause de l'orage est que le baromètre n'intervient absolument pas dans le mécanisme de la genèse de l'orage. Certains auteurs contestent la possibilité qu'une quantité, par exemple l'énergie, puisse être transmise au sens strict : la raison est qu'il manque aux quantités particulières d'énergie l'individualité qui donnerait un sens à l'idée qu'il s'agit de la même quantité à travers le temps (Dieks 1986). Par conséquent, la version la plus élaborée de la théorie de processus en termes de grandeurs conservées (Dowe 1992 ; 2000) n'a pas recours au concept de transmission, mais reprend l'idée russellienne de « manifestation » continue d'une grandeur conservée. Par « manifestation » continue d'une propriété par une ligne de monde, Dowe entend simplement que cette propriété caractérise tous les points de la ligne, sans que cela présuppose l'idée d'une transmission. Cela la rend vulnérable à l'objection selon laquelle certains pseudo-processus manifestent des grandeurs conservées, sans pour autant être causaux⁴¹. Nous avons déjà considéré l'exemple de la tâche lumineuse qui parcourt la paroi interne d'un cylindre. Le parcours de cette tâche constitue une ligne de monde parfaitement homogène : dans les conditions stipulées de cette expérience de pensée, la tâche lumineuse contient, ou « manifeste », à chaque instant exactement la même énergie ; chaque instant est qualitativement parfaitement semblable à chaque autre instant. Pourtant, la situation est telle que la ligne de monde constituée par le parcours de la tâche n'est pas un processus causal. Seul le processus de propagation des rayons lumineux du phare vers la paroi est causale.

Les théories qui analysent la causalité en termes de transmission ou de manifestation continue de grandeurs conservées évitent les problèmes mentionnés plus haut, notamment celui posé par les effets de causes communes et celui des processus préemptés. Le fait que deux événements soient effets d'une cause commune n'en fait pas des termes d'une relation causale, dans la mesure où il n'y a aucun processus de transmission entre eux. Aussi, le fait qu'un processus P_1 soit accompagné d'un doublon redondant (préempté) P_2 n'empêche pas P_1 de transmettre une quantité de grandeur conservée. Reprenons la situation où deux snipers tirent sur la même victime dont ils sont séparés par la même distance. Imaginons que le sniper S_1 tire un petit instant plus tôt que le sniper S_2 , de sorte que ce soit la balle tirée par S_1 qui tue la victime. Dans ce cas, le tir de S_2 (l'événement b) n'est pas une cause de sa mort (l'événement c). Ni l'analyse probabiliste ni l'analyse contrefactuelle ne peut rendre compte de l'intuition selon laquelle ce qui détermine si le tir de S_1 (l'événement a) est cause de la mort ou non doit être localisé au processus entre a et c. Aussi bien l'analyse contrefactuelle que l'analyse probabiliste font dépendre l'existence d'une relation causale entre a et c de facteurs qui sont localisés ailleurs qu'entre a et c. Dans la situation où le tir du sniper S_1 a lieu dans une situation où le sniper S_2 tire lui aussi, il n'existe pas de dépendance contrefactuelle entre a et c : étant donné le tir de S_2 , il n'est pas vrai que, si S_1 n'avait pas tiré, la victime ne serait pas morte. Or l'une de nos intuitions semble indiquer que l'existence d'une relation causale entre a et c ne peut dépendre que des processus qui ont lieu entre a et c, et qu'elle ne peut donc pas dépendre d'événements et processus qui n'interfèrent pas avec les processus entre a et c. Le même problème se pose à la théorie probabiliste : le tir de S_2 (l'événement b)

⁴¹ Cf. Salmon (1994, p. 308), Kistler (1998), (1999).

augmente la probabilité de la mort de la victime (l'événement c). Pour expliquer que b n'est néanmoins pas cause de c, la théorie probabiliste doit faire appel des circonstances qui ne concernent pas les processus qui se produisent entre b et c, à savoir à l'événement a. En revanche, l'analyse qui postule que la causalité est fondée sur la transmission tient compte de cette intuition de localité selon laquelle l'existence d'une relation causale entre a et c ne dépend que des processus allant de a à c. Si a transmet quelque chose, disons de l'énergie, sur c, alors a est cause de c, peu importe s'il existe par ailleurs d'autres événements, comme b, qui ont également un impact causal sur c.

Néanmoins, les défenseurs de cette analyse sont minoritaires parce qu'elle rencontre un certain nombre de problèmes importants.

1. Nous avons déjà évoqué l'objection selon laquelle l'analyse fondée sur la transmission souffre d'un certain manque d'ambition, en ne visant que la causalité dans le monde actuel, et non le concept tel qu'il s'applique à tous les mondes possibles. Ce n'est qu'une objection que dans la mesure où on suppose que l'analyse conceptuelle est le seul but légitime ou suffisamment ambitieux de l'analyse philosophique de la causalité.

2. Mais on peut la soupçonner d'un manque d'ambition d'un autre genre : la théorie de transmission semble ne s'appliquer qu'aux processus causaux physiques. En ce sens, elle semble inadéquate en tant qu'analyse de jugements causaux ordinaires qui ne visent pas des processus physiques mais des évolutions et interactions qui sont déterminées par des propriétés non-physiques des objets, par exemple par leurs propriétés chimiques ou psychologiques. Le fait qu'on sonne à la porte réveille Pierre. Le bruit de la sonnette semble bien être la cause du réveil, mais il peut sembler peu pertinent de considérer le processus causal sous-jacent sous l'angle de la transmission d'énergie⁴². En effet, l'application de l'analyse à des jugements causaux de sens commun présuppose que tous les causes et tous les effets sont physiques. Plusieurs solutions ont été proposées pour articuler la théorie de la transmission aux objets des jugements causaux courants. Le jugement causal selon lequel la sonnette a réveillé Pierre ne porte pas directement sur une transmission d'énergie. La dépendance du réveil par rapport à la propagation des ondes sonores, leur transduction en signaux nerveux et la transmission de ces derniers au cortex auditif de Pierre fait l'objet de différentes sciences « spéciales », notamment l'acoustique, la psychophysique, la physiologie et la neurophysiologie. Dans un cadre physicaliste, on suppose que tous les faits surviennent⁴³ sur l'ensemble des faits physiques. On peut alors défendre la thèse selon laquelle le déclenchement du réveil par le bruit de la porte survient sur un processus physique de transmission. Dans une perspective réductionniste, il paraît envisageable que les propriétés dont le jugement affirme la dépendance causale sont des formes spécifiques que prennent les grandeurs conservées. On parvient à une conception en deux volets : deux conditions rendent ensemble vrai le jugement selon lequel le fait que c (l'activation de la sonnette à l'instant t) est F (fait du bruit) est causalement responsable du fait que e (Pierre à l'instant immédiatement après t) est G (se réveille). Il faut qu'il y ait 1) un processus de transmission de la cause c à l'effet e et 2) une loi de la nature qui exprime la dépendance de G par rapport à F (Kistler 1999). Pour juger que la sonnette a réveillé Pierre, il faut qu'il y ait une loi « in situ » selon laquelle, dans des circonstances ordinaires et sauf exception, les sonnettes

⁴² Cf. Collins et al. (2004), p. 14.

⁴³ En première approximation, on dit qu'un premier ensemble de propriétés M « survient » sur un second ensemble de propriétés P si et seulement si il est impossible que deux objets diffèrent à l'égard de l'une des propriétés M, sans différer à l'égard de l'une des propriétés P. Selon le physicalisme, les propriétés mentales surviennent sur les propriétés physiques. Cela implique notamment qu'il est impossible qu'une personne change mentalement sans changer physiquement, et qu'il est impossible qu'il y ait une réplique r d'une personne p qui ne diffère physiquement pas du tout de p, mais qui en diffère pourtant mentalement. Il existe plusieurs concepts de survenance qui se distinguent notamment par l'interprétation que l'on donne de la nécessité (ou impossibilité) qui figure dans la définition. Cf. Kim (1990) et l'introduction de Savellos et Yalcin (1995).

réveillent les gens endormis, ou du moins augmentent la probabilité avec laquelle ils se réveillent. Une autre approche consiste à articuler la condition de transmission à une condition contrefactuelle : selon Menzies (2004), 1) le fait de « faire une différence » et 2) l'existence d'un processus sont nécessaires et ensemble suffisants pour la causalité. La transmission garantit l'existence d'un processus entre *c* et *e* (condition (2) de Menzies). Le fait que *c* est *F* fait une différence à l'égard du fait que *e* soit *G*, dans la mesure où, si *c* n'avait pas été *F* (si la sonnette n'avait pas été activée), *e* n'aurait pas été *G* (Pierre ne se serait pas réveillé) (condition (1) de Menzies).

3. Le concept courant de transmission étant causal, cette analyse semble vouée à une certaine circularité. On peut éviter cette objection par une redéfinition du concept de transmission : étant donné deux endroits spatio-temporels distincts *x* et *y*, une quantité *A* est transmise entre *x* et *y* si et seulement si *A* est présent à la fois en *x* et *y*.

4. Si on conçoit la transmission de cette manière, la causalité n'est plus asymétrique. On peut résoudre ce problème en s'appuyant sur le fait que l'origine de l'asymétrie des relations causales est une caractéristique physique du monde réel : il y a notamment, dans notre région de l'univers, d'une pléthore de processus irréversibles, tous dirigés dans la même direction, comme cela est garanti par la seconde loi de la thermodynamique. Un tel fondement physique de l'asymétrie de la causalité peut également servir de fondement de la direction du temps (Reichenbach 1956 ; Lewis 1979/1986 ; Hausman 1998 ; Savitt 2006).

5. Les processus de transmission sont omniprésents. Deux événements quelconques qui sont suffisamment proches (spatialement et temporellement) sont en général liés par des transmissions de photons. La théorie de la transmission semble donc condamnée à provoquer une inflation de jugements de causalité. On peut répondre à cette critique en faisant valoir premièrement que ces jugements causaux pléthoriques sont vrais mais manquent de pertinence communicationnelle. Deuxièmement, on peut faire valoir que les sciences spéciales sélectionnent les processus causaux pertinents, à partir des propriétés de l'effet indiquées dans l'*explanandum*. S'il s'agit d'indiquer la cause du réveil de Pierre, le processus causal pertinent conduit à la modification de la propriété physiologique et psychologique qu'est l'état d'éveil.

6. Selon un argument récent (Curiel 2000, Lam 2005), l'énergie n'est pas conservée localement dans le cadre de la relativité générale et ne peut donc pas être « transmise » localement.

7. Il semble qu'un problème moins technique condamne de toute façon la théorie de transmission : aussi bien dans le sens commun qu'en sciences, de nombreux jugements causaux comportent des faits négatifs comme causes ou comme effets, comme c'est le cas dans les jugements de causalité par omission ou prévention. Si je tue une plante en omettant de l'arroser, il semble que j'aie causé sa mort sans lui avoir rien transmis⁴⁴. Si je préviens au contraire la mort de la plante en l'arrosant, l'événement de la mort n'a pas lieu et ne peut donc faire l'objet d'aucune transmission. Schaffer (2000a) montre que dans de très nombreux jugements de sens commun, aucune transmission ne semble avoir lieu, notamment dans les jugements de double prévention où quelqu'un empêche un empêchement. Schaffer (2006) donne l'exemple du terroriste qui empêche le surveillant d'une tour de contrôle d'aéroport d'empêcher la collision d'un avion.

Les jugements causaux où la cause ou/et l'effet est un fait négatif sont incompatibles avec trois propriétés intuitives de la causalité notées par (Hall 2000) : il s'agit d'un processus local (où la cause est liée à l'effet par l'intermédiaire d'une série d'événements intermédiaires), intrinsèque (il ne dépend pas de ce qui a lieu ailleurs), et transitif. Si *a* peut

⁴⁴ L'exemple est de Beebe (2004). Plus précisément, je ne transmets rien de pertinent, quoiqu'il y ait sans doute d'innombrables processus sans pertinence, comme par exemple la transmission de photons.

causer b par omission, prévention, ou double prévention, alors certaines relations causales ne respectent ni la localité, ni le caractère intrinsèque ni la transitivité. On peut en tirer trois conséquences incompatibles.

1. Les omissions ne sont pas des instances de causalité mais nous apparaissent comme telles, par exemple parce que nous confondons l'explication non-causale ou l'attribution de responsabilité morale avec le jugement causal. (Dowe 1999 ; 2000, Armstrong 2004 ; Beebe 2004 ; Kistler 2006)
2. Ce sont de vrais jugements causaux, et la localité, le caractère intrinsèque et la transitivité ne sont pas des conditions nécessaires à la causalité (Schaffer 2000 ; 2004).
3. Il y a deux concepts de causalité ou deux aspects du concept de causalité : l'un correspond à la dépendance contrefactuelle (ou à l'augmentation de probabilité, ou à l'influence nomologique), l'autre correspond à l'existence d'un processus de transmission. Selon Hall (2000), il y a deux concepts de causalité indépendants.

6. L'analyse probabiliste

Lorsqu'il s'agit de découvrir, dans des situations complexes, des lois en général et des lois causales en particulier, il semble y avoir deux types de stratégies. La recherche de corrélations statistiques qui s'expriment dans les probabilités conditionnelles trouvées au sein d'un ensemble de données, et la manipulation contrôlée. Chacune de ces méthodes correspond aussi à une stratégie d'analyse du concept de causalité : la première a donné lieu à l'analyse probabiliste de la causalité qui fait l'objet de la présente section; la seconde a donné lieu à l'analyse de la causalité en termes de manipulation : nous l'examinerons dans la section suivante.

Dans les situations complexes auxquelles sont confrontées l'économie, la sociologie, l'épidémiologie ou la météorologie, les lois et les liens causaux ne se manifestent pas par des conjonctions constantes : tous les fumeurs ne développent pas de cancer des poumons. En macroéconomie, la relation dite de Phillips pose une dépendance entre le taux de chômage et la variation des salaires ; elle implique notamment que plus le chômage est élevé, plus la croissance des salaires est faible, et que si au contraire le taux de chômage baisse, les salaires et indirectement l'inflation ont tendance à augmenter ; or il s'avère qu'un fort chômage peut coexister, pendant des périodes assez longues, avec une forte inflation.

Dans la perspective d'une amélioration de l'analyse de la causalité en termes de régularités, l'analyse probabiliste propose d'associer la causalité à l'influence d'un facteur sur un autre facteur, une influence qui n'est pas nécessairement universelle mais statistiquement significative. L'hypothèse fondamentale est qu'il existe une influence causale d'un premier facteur A sur un second facteur B si et seulement si la probabilité de B étant donné A est plus grande que la probabilité de B en l'absence de A.

(AP « Augmentation de probabilité ») [[revoir l'écriture](#)] A cause B si et seulement si $P(B|A) > P(B|\text{non-A})$

Il faut distinguer deux sortes de motivations encourageant à passer d'une analyse de la causalité en termes de régularités universelles à une analyse en termes d'augmentation de probabilité. La première est que les influences nomiques et causales sont, dans une situation complexe, souvent masquées par d'autres influences et ne se manifestent donc pas directement et de façon pure dans une régularité universelle, comme c'est le cas des exemples qu'on vient de donner. La seconde est l'hypothèse selon laquelle il existe des lois intrinsèquement statistiques, en ce sens que, même dans une situation dépourvue de toute interférence, la loi ne donne lieu qu'à une augmentation de la probabilité de l'effet en présence de la cause. La question de savoir s'il existe de telles lois en dehors de la mécanique quantique est sujette à controverse, mais la capacité de l'analyse probabiliste à de tenir

compte de lois de ce genre lui procure un avantage sur l'analyse en termes de régularités universelles.

Avant de considérer le développement de cette hypothèse fondamentale, il convient de faire deux remarques. La première est que l'analyse probabiliste rapproche l'ontologie de l'épistémologie : la relation causale est identifiée à ce qui permet de découvrir l'existence d'influences causales dans des situations complexes, à savoir une inégalité entre probabilités conditionnelles. La seconde est que l'approche probabiliste ne concerne pas – en tout cas pas directement – les relations causales entre événements particuliers, mais des relations d'influence causale entre « facteurs », autrement dit entre propriétés ou types d'événements. Le formalisme de cette approche présuppose que les termes de la relation causale peuvent être soumis aux opérations de la logique propositionnelle, telles que la négation et la conjonction. Cela revient à adopter la conception des termes de la relation causale comme des faits (Vendler 1967a, 1967b, Bennett 1988, Mellor 1995) plutôt que comme des événements particuliers (Davidson 1967).

La condition (AP) se heurte à deux obstacles que rencontrent également les analyses nomologique et contrefactuelle :

1) La condition d'augmentation de probabilité est symétrique : si A et B sont statistiquement corrélés positivement, de sorte que $P(A|B) > P(A|\text{non-B})$, alors on a aussi $P(B|A) > P(B|\text{non-A})$.

2) Les effets de causes communes sont généralement statistiquement corrélés sans être cause l'un de l'autre. Si le fait de fumer (F) augmente à la fois la probabilité d'avoir un cancer des poumons (C) et la probabilité d'avoir un infarctus (I), C et I sont aussi, *ceteris paribus*, corrélés positivement entre eux. Le fait que l'analyse probabiliste dispose, avec la condition de l'absence de « facteurs écran »⁴⁵, d'une solution à ce dernier problème, est sans doute l'une des raisons de son succès. Si A et B sont statistiquement corrélés positivement, alors on appelle un facteur C un « facteur écran » par rapport à A et B, si la corrélation positive entre A et B disparaît si les probabilités sont calculées en tenant fixe la présence ou l'absence de C. En formules, $P(B|A) > P(B|\text{non-A})$, mais $P(B|A \text{ et } C) = P(B|\text{non-A et } C)$ et $P(B|A \text{ et non-C}) = P(B|\text{non-A et non-C})$.

A l'aide du concept de facteur écran, on peut construire l'analyse suivante : le facteur A, instancié à l'instant t, est cause du facteur B, instancié en même temps ou plus tard, si et seulement si deux conditions sont remplies :

1) $P(B|A) > P(B|\text{non-A})$

2) Il n'y a aucun facteur C, instancié à t ou plus tôt, qui fait écran à la corrélation entre A et B.

Cette condition permet de résoudre le problème posé, dans le cas des corrélations dues aux causes communes, par le fait qu'une corrélation positive n'est pas en général *suffisante* pour l'existence d'une relation causale. Mais il y a également des circonstances dans lesquelles elle n'est pas *nécessaire* ; ce sont des situations où la présence d'un facteur A qui est causalement corrélé positivement à un facteur B *diminue* la probabilité de B. Si les fumeurs (F) font plus de sport (S) que les non-fumeurs, de sorte que F est positivement corrélé à S, il est possible que l'effet bénéfique de S, qui fait diminuer le risque de contracter une maladie cardio-vasculaire (M), surcompense l'effet négatif de F qui augmente le risque de M. Il y a donc des situations où une cause, F, diminue la probabilité de son effet, M :

$P(M|F) < P(M|\text{non-F})$.

Il existe une solution à ce problème, dont différentes versions ont été proposées par Cartwright (1979 p. 423) et Skyrms (1980). Dans la version de Cartwright, A cause B si et seulement si la probabilité de B est plus grande en présence de A qu'en son absence, dans

⁴⁵ Ce concept a été introduit par Reichenbach (1956).

tous les ensembles qui sont homogènes à l'égard de toutes les causes de B qui ne sont pas des effets de A.

A cause B si et seulement si $P(B|A \text{ et } C_i) > P(B|C_i)$ pour tout C_i , où les C_i sont les causes de B qui ne sont pas causées par A.

L'ensemble des facteurs qui causent B mais ne sont pas causés par A est appelé une « situation de test » : le fait de tenir fixes les facteurs dans cet ensemble permet en effet de tester l'hypothèse selon laquelle A est cause de B. Par exemple, cette condition permet de justifier le jugement intuitif selon lequel F est bien une cause de M : à l'intérieur d'un ensemble de personnes qui ont tous le même niveau de pratique sportive (S), la probabilité de M est plus grande en présence de F qu'en son absence.

Cependant, il convient de noter qu'analyser le fait que A cause B en termes d'augmentation de probabilité dans des situations de test change la nature du projet de l'analyse probabiliste. Premièrement, dans la forme proposée par Cartwright et Skyrms, l'analyse ne peut plus servir de base à une réduction du concept de causalité : en effet, l'*analysans* contient de manière essentielle le concept de cause. Pour pouvoir déterminer si A est cause de B, il faut déjà connaître les autres causes de B, plus précisément l'ensemble de tous les facteurs qui causent B de manière indépendante de A.

Deuxièmement, le fait que l'on doive mesurer les probabilités conditionnelles dans un ensemble homogène à l'égard de tous les facteurs qui peuvent influencer la probabilité de B indépendamment de celle de A est incompatible avec l'une des motivations principales de l'approche probabiliste : il s'agissait de détecter des influences causales dans des situations de corrélation imparfaite, où la présence de facteurs interférents empêche l'apparition universelle de l'effet avec la cause. Or dans la mesure où on fait abstraction des lois indéterministes, dans une situation où toutes les causes indépendantes de B sont tenues fixes, si A est cause de B, alors $P(B|A)=1$. Les probabilités inférieures à 1 ne semblent en effet mesurer que l'effet net des facteurs inconnus différents de A qui influencent B négativement ou positivement.

Nous avons déjà mentionné un autre problème important pour l'analyse probabiliste : la corrélation statistique est symétrique, de sorte que si la probabilité de B est plus grande en présence de A qu'en son absence, la probabilité de A est aussi plus grande en présence de B qu'en son absence. Plusieurs propositions ont été émises pour associer une condition supplémentaire à la condition d'augmentation de la probabilité, destinées à distinguer la cause de l'effet. On peut simplement stipuler que le facteur instancié plus tôt dans le temps est la cause, et celui instancié plus tard, l'effet. Cela s'accorde pourtant mal avec le cadre d'une théorie qui a pour objet des relations causales générales entre facteurs, plutôt que des relations particulières entre leurs instances. Par ailleurs, une telle stipulation empêche de manière *a priori* d'envisager la possibilité de processus causaux dirigés à l'inverse de la direction du temps ; et finalement, elle rend impossible la réduction de la direction du temps lui-même à la causalité. L'une des approches traditionnelles pour expliquer l'asymétrie du temps consiste à faire l'hypothèse qu'elle est dérivée de l'asymétrie de la causalité : on explique le fait que t_2 soit objectivement plus tard que t_1 par le fait qu'un événement qui se produit à t_1 peut causer un événement qui se produit à t_2 , mais que l'inverse n'est pas vrai. Selon une autre hypothèse, l'asymétrie de la causalité et celle du temps dérivent toutes les deux d'une asymétrie plus fondamentale. Selon certains, il s'agit de l'asymétrie intrinsèque des processus irréversibles, notamment des processus d'évolution des systèmes dont l'entropie augmente. Il a aussi été suggéré que certains processus microphysiques intrinsèquement asymétriques, en particulier la désintégration des mésons-K, ou « kaons », peuvent fonder l'asymétrie de la causalité⁴⁶. Aucune de ces hypothèses n'est compatible avec l'idée que l'asymétrie de la causalité se réduit à celle du temps.

⁴⁶ Ces processus de décomposition "violente" la symétrie à l'égard de l'inversion temporelle ("T"). Cf. Dowe (1992a, p. 189).

Il a aussi été envisagé que la différence entre cause et effet n'est qu'un effet de perspective de l'observateur ou de l'agent humain, à laquelle ne correspond aucune différence objective sur le plan des liens entre les facteurs eux-mêmes⁴⁷.

La proposition la plus influente rendant compte de l'asymétrie de la causalité à partir de conditions probabilistes est due à Reichenbach (1956) qui a suggéré d'utiliser les causes communes pour déterminer la direction de la causalité (et du temps). Si A et B sont corrélés positivement et si C est un facteur écran, de telle sorte que la corrélation entre A et B disparaît à la fois en présence et en l'absence de C, et de telle sorte que la présence de C augmente à la fois la probabilité de A et celle de B, alors le triplet ACB est appelé une « fourche conjonctive ». Si le facteur C est instancié dans le passé de A et de B, et s'il n'existe aucun facteur D qui satisfait les mêmes conditions que C mais qui est instancié dans le futur, alors ACB constituent une fourche ouverte en direction du futur (et C est une cause commune des deux effets A et B) ; si le seul facteur D qui satisfait ces conditions est instancié dans le futur par rapport à A et B, alors ADB constituent une fourche ouverte vers le passé ; si enfin il existe à la fois un facteur C dans le passé et un facteur D dans le futur qui satisfont les conditions indiqués, ACBD constituent une fourche fermée. Reichenbach suggère d'identifier la direction de la cause à l'effet (qui est aussi la direction du temps) à la direction dans laquelle les fourches ouvertes prédominent.

Parmi les nombreuses tentatives d'amélioration de l'analyse de la causalité qui passent par la synthèse d'éléments conceptuels de différentes approches, il faut mentionner ici l'analyse en termes de contrefactuels probabilistes. Cette théorie, proposée par D. Lewis (1986c) et récemment élaborée par Noordhof (1999, 2004), propose d'analyser la relation causale entre événements particuliers ainsi : A cause B s'il existe une série d'événements intermédiaires C_1, C_2, \dots, C_n , tels que C_1 dépend causalement de A, C_2 dépend causalement de C_1, \dots et B dépend causalement de C_n . La dépendance causale est analysée en termes d'une condition contrefactuelle portant sur des probabilités : B dépend causalement de A si A et B ont eu lieu, et si la probabilité de B, au moment de A, était significativement plus grande que la probabilité qu'aurait eue B si A n'avait pas eu lieu.

7. L'analyse en termes d'équations structurelles

L'une des contributions les plus importantes de ces dernières années est l'analyse philosophique des modèles de la causalité qui ont été élaborés en intelligence artificielle pour représenter les interactions causales dans des situations complexes, notamment en économie. Dans la version élaborée par Pearl (2000), cette analyse en termes de « manipulation » ou « d'intervention » apparaît comme une variante de l'analyse contrefactuelle⁴⁸. Comme l'approche probabiliste, l'analyse en termes de manipulabilité part de l'épistémologie : elle essaye d'extraire le concept de causalité de l'analyse de la structure logique de la recherche scientifique des causes, telle qu'elle se fait notamment en sciences sociales. En psychologie autant qu'en sociologie et en économie, on cherche à extraire de l'analyse de probabilités conditionnelles certaines informations sur les facteurs qui influencent d'autres facteurs. Les graphes orientés sont un outil récent qui a été développé afin de représenter les relations d'influence causale entre des variables qui correspondent aux propriétés mesurables dans un système interactif. Ce formalisme reflète la méthode de découverte des causes qui est utilisée dans les sciences expérimentales : on intervient de l'extérieur sur une variable pour observer d'éventuels changements dans d'autres variables, et en tirer des conclusions sur l'existence

⁴⁷ Fair (1979), Price (1992) et Menzies et Price (1993).

⁴⁸ Une autre version a été élaborée par Spirtes, Glymour et Scheines (2000). Woodward (2003) a élaboré sur le plan philosophique leur travail et celui de Pearl (2000). Keil (2000, 2005) propose une analyse originale de la causalité en termes de manipulation sans avoir recours à l'appareil technique des équations structurelles.

d'influences causales. La formalisation de cette stratégie de recherche à l'aide d'équations structurelles ou de graphes orientés a inspiré la conception philosophique de la causalité en termes de manipulabilité.

Ces recherches renouent avec l'idée traditionnelle selon laquelle une cause C d'un effet E est une action qui permettrait à un agent humain d'obtenir E s'il décidait de faire C⁴⁹. Les deux difficultés majeures de cette analyse sont sa circularité et l'anthropocentrisme implicite de la thèse selon laquelle seule peut être cause un événement qui peut en principe être le résultat d'une action. En effet, von Wright (1971) juge que le constat selon lequel l'intervention humaine dans l'expérimentation est indispensable à l'analyse de notre *connaissance* des relations causales n'autorise pas la conclusion selon laquelle l'action humaine est essentielle à l'*ontologie* de la causalité. Nous verrons que l'analyse manipulationniste ou interventionniste toute récente semble pouvoir éviter un tel anthropocentrisme. Par ailleurs, il n'est pas possible de construire une analyse non-circulaire de la causalité qui soit fondée sur la notion d'intervention, dans la mesure où une intervention est un processus causal. Par conséquent, les versions récentes de la théorie de la manipulabilité n'ont pas pour prétention d'analyser la notion de causalité, mais d'étudier la logique du raisonnement causal à partir de son lien avec les interventions expérimentales.

Nous ne pouvons pas donner ici plus qu'une idée très sommaire de la démarche sous-jacente à l'analyse de la causalité en termes d'équations structurelles. On représente la structure causale d'un système complexe à l'aide d'un modèle qui consiste en un ensemble de variables V et un ensemble d'équations structurelles qui expriment des relations fonctionnelles entre ces variables. Reprenons l'analyse (due à Menzies 2001) de l'exemple simple des enfants qui essayent de casser une bouteille en lançant des cailloux. Nous l'avons rencontré plus haut comme exemple d'une situation de préemption : le lancer de T ne casse pas la bouteille alors qu'il l'aurait cassée si S n'avait pas lancé son caillou un instant plus tôt, de sorte qu'il a cassé la bouteille avant que le caillou de T puisse le faire. Cette situation peut être représentée à l'aide des variables suivantes. Cet exemple ne requiert que des variables à deux valeurs, mais il est bien entendu possible d'utiliser des variables continues.

- $TL = 1$ si T lance un caillou, sinon 0;
- $SL = 1$ si S lance un caillou, sinon 0;
- $TA = 1$ si le caillou de T atteint la bouteille, sinon 0;
- $SA = 1$ si le caillou de S atteint la bouteille, sinon 0;
- $BC = 1$ si la bouteille casse, sinon 0.

A chaque variable est associée une équation structurelle. Une variable est appelée « exogène » si sa valeur est déterminée par des facteurs extérieurs au système causal dont on construit le modèle. Ainsi, les équations $TL=1$ et $SL=1$ stipulent la valeur des variables exogènes TL et SL. En revanche, la valeur d'une variable endogène est déterminée en fonction d'autres variables à l'intérieur du système. Par exemple, on peut poser $SA=SL$, ce qui signifie que le caillou lancé par S atteint la bouteille si S lance un caillou ($SL=1$ et $SA=1$) et qu'il ne l'atteint pas si S ne lance pas de caillou ($SL=0$ et $SA=0$). La préemption du processus provoqué par T s'exprime dans l'équation pour TA : $TA=TL \ \& \ \text{non-SA}$. Le caillou lancé par T n'atteint la bouteille que si celui lancé par S ne l'atteint pas. Enfin, $BC = SA$ ou TA . La bouteille casse si le caillou de S ou celui de T atteint la bouteille.

⁴⁹ Cf. Gasking (1955), Menzies et Price (1993).

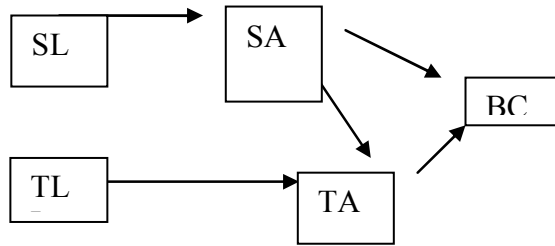


Figure 1

On peut représenter le contenu des équations structurelles dans ce qu'on appelle un graphe structuré. Le graphe de la figure 1 représente les équations du système de S, T et la bouteille : chaque variable correspond à un nœud dans le graphe. Une flèche allant de X à Y représente le fait que la valeur de Y dépend de la valeur de X ; dans ce cas, X est appelé un « parent » de Y. Un chemin orienté (*directed path*) est un ensemble de flèches qui connecte X à Y. Chaque flèche et chaque équation représentent un ensemble de propositions conditionnelles contrefactuelles. Dans ce modèle, il est possible de déterminer la valeur de vérité de contrefactuels plus complexes qui ne correspondent pas directement à une flèche. Disons que nous voulons déterminer ce qui se serait passé si le caillou lancé par S n'avait pas atteint la bouteille. On pose alors que la variable correspondant à l'antécédent prend la valeur qu'elle a dans la situation contrefactuelle : on pose $SA=0$. Une telle « intervention atomique » (Pearl 2000, p. 70) correspond à ce que Lewis appelle un « miracle » : on ne considère pas le passé qui aurait pu conduire à l'antécédent, mais on pose l'antécédent (SA), tout en laissant son passé tel qu'il est dans la réalité. Dans la représentation graphique, on efface toutes les flèches conduisant à la variable SA, de sorte à la transformer en variable exogène. Dans l'interprétation manipulationniste de ce formalisme, cela correspond à une intervention expérimentale localisée sur la variable SA, qui vient de l'extérieur du système et qui est directe au sens où elle ne passe pas par une intervention sur des facteurs qui influencent SA à l'intérieur du système. Comme le concept du miracle de Lewis, cela garantit qu'aucun contrefactuel « à rebours » ne peut être vrai. En changeant la variable X, on laisse inchangé le passé de X, autrement dit, les valeurs de toutes les variables représentées à gauche par rapport à X. A partir des équations correspondant aux flèches partant de X, on peut déterminer les valeurs que prennent les variables à droite de X, dans la situation correspondant à la nouvelle valeur.

Pearl (2000, p. 70) définit l'effet causal de X sur Y comme $P(y/do(x))$, c'est-à-dire la distribution de probabilité des différentes valeurs de Y, étant donné qu'une intervention (« do ») a fixé x comme valeur de la variable X. Cela a pour conséquence que l'on considère que l'ensemble des facteurs autres que X qui influencent également sur Y font partie de l'impact de X sur Y. Pour éviter cela, Woodward (2003) impose des contraintes supplémentaires sur l'intervention I qui permet de déterminer si X cause Y. 1) I doit être la seule cause de X, au sens que toutes les autres influences sur X doivent être considérées comme coupées. 2) Il ne faut pas qu'I cause Y par un chemin qui ne passe pas par X, comme c'est le cas dans la situation suivante : I est l'administration d'un comprimé placebo ; X est l'action du placebo sur le corps après son ingestion ; Y est la guérison. Par définition d'un placebo, lorsque I est efficace en changeant la valeur de Y, son efficacité ne passe pas par l'intermédiaire X de changements dans le corps provoqués par l'absorption du comprimé. Dans une telle situation, le fait qu'I influence Y ne signifie pas que X cause Y. 3) I ne doit pas

être l'effet d'une cause qui influence Y par un chemin qui ne passe pas par X. Si, pour déterminer si l'indication X d'un baromètre cause la tempête Y, mon intervention I sur X dépend elle-même de ma connaissance de la pression de l'air, il est possible que je trouve que Y dépend en effet des valeurs que j'impose à X, alors que X ne cause bien entendu pas Y. 4) On maintient fixes les valeurs de toute cause éventuelle de Y qui ne se situerait pas sur un chemin qui mène de I à X et à Y.

Dans ce contexte, Woodward définit l'effet causal de X sur Y en termes de la différence des valeurs de Y correspondant à la différence entre deux valeurs x et x* de la variable X sur laquelle on intervient par I.

(CT) (« cause totale ») $Y_{do(x), Bi} - Y_{do(x^*), Bi}$

où « $Y_{do(x), Bi}$ » représente la valeur de la variable Y étant donné qu'une intervention a imposé la valeur x à la variable X, dans les circonstances B_i .

Si la relation entre X et Y est déterministe, X est une cause de Y s'il existe des paires de valeurs x et x* tels que (CT) est différent de zéro ; si la relation est indéterministe, X est une cause de Y s'il existe des paires de valeurs x et x* tels que la probabilité de Y est différente pour les deux valeurs de X.

Tout en partageant avec la conception contrefactuelle l'idée de considérer une situation qui partage avec la réalité le passé mais diffère à partir de l'événement décrit dans l'antécédent, l'analyse en termes d'équations structurelles donne un résultat intuitivement correct dans le cas de préemption que nous avons pris pour exemple.

Dans la figure 1, TA est la seule variable intermédiaire entre SL et BC qui n'est pas sur la route SL – SA – BC. Pour évaluer si SL cause BC, il faut donc tenir fixe la valeur actuelle de TA=0. Si on change la valeur de SL pour poser SL=0, la valeur de BC n'est plus la valeur actuelle mais devient BC=0. SL est donc bien la cause de BC.

Le problème pour l'analyse de Lewis était que BC ne dépend pas contrefactuellement de sa cause SL, car BC=1 même si SL=0. L'analyse interventionniste évite cette difficulté en fixant les valeurs de toutes les variables qui ne sont pas situées sur le chemin qui connecte la cause hypothétique à son effet hypothétique. Or dans la situation réelle TA=0. De cette façon, on obtient le résultat intuitivement correct que si SL=0 (si S ne lance pas), alors SA=0 et BC=0 (la bouteille ne casse pas).

Sur le plan de l'interprétation, l'analyse en termes d'équations structurelles permet d'interpréter l'antécédent non plus comme décrivant un fait réel dans un autre monde possible, mais comme décrivant une situation qui correspond à une intervention expérimentale. Cette différence d'interprétation s'accompagne de différences formelles : l'analyse de Lewis rend la causalité transitive, alors qu'elle ne l'est pas nécessairement dans l'approche des équations structurelles⁵⁰. Dans l'analyse de Lewis, l'évaluation d'un contrefactuel requiert de maintenir fixes tous les événements dans le passé de l'événement décrit dans l'antécédent, alors que l'approche des équations structurelles requiert de maintenir fixes les valeurs de toutes les variables qui ne sont pas situées sur un chemin reliant cause et effets hypothétiques.

Le cadre des équations structurelles permet d'introduire différentes notions causales qui correspondent à différents usages de la notion intuitive de causalité. L'analyse des différences entre ces notions montre la fécondité de cette approche même si elle ne peut pas prétendre déboucher sur une analyse non circulaire. Il arrive par exemple qu'une variable X influence une autre variable Y de deux manières indépendantes qui s'annulent. Pour reprendre un exemple célèbre de Hesslow (1976), le fait de prendre une pilule contraceptive X a une influence positive sur la probabilité de subir une thrombose Y. Mais X a aussi une influence négative sur la probabilité d'être enceinte Z, où Z est un autre facteur augmentant le risque de

⁵⁰ Cf. Hitchcock (2001).

thrombose. Il est possible que l'influence positive directe de X sur Y est exactement compensée par l'influence négative de X sur Y par l'intermédiaire de Z, de sorte que l'influence nette de X sur Y soit nulle. Il y a à la fois un sens intuitif auquel il semble correct de dire que le fait de prendre la pilule cause un risque accru de thrombose et un sens auquel il semble tout aussi correct de dire que le fait de prendre la pilule ne cause pas de tel risque.

On peut faire disparaître le paradoxe en distinguant deux notions de causalité⁵¹. Dans la situation envisagée par Hesslow, X n'est pas une « cause totale » de Y, au sens de la condition (CT) introduite plus haut. En revanche, X est une « cause directe » de Y si l'on définit une « cause contributive » ainsi :

(CD) X est une cause contributive de Y si et seulement si la valeur de Y change suite à un changement de la valeur de X, les valeurs de toutes les autres variables en dehors de X et Y étant tenues fixes, y compris de celles qui sont sur un chemin entre X et Y.

En effet, si l'on fixe la valeur de Z dans l'exemple, on trouve qu'une intervention sur X modifie la valeur de Y, de sorte que la prise de la pilule est bien une cause directe du risque accru de thrombose même si elle n'en est pas une cause totale.

Les versions récentes de la théorie de la manipulabilité évitent l'objection d'anthropocentrisme en caractérisant la notion d'intervention de telle sorte qu'elle ne requiert pas d'intervention de la part d'un agent humain. Un événement naturel sans aucun lien avec une action intentionnelle peut en effet modifier un système à la manière d'une intervention. Il peut alors servir d'« expérience naturelle » qui permet de tirer des conclusions causales. La neuropsychologie élabore ainsi des raisonnements causaux sur la contribution de certaines parties précises du cerveau à partir de l'observation des changements Y qui accompagnent la modification accidentelle X de ces parties du cerveau.

Les versions plus anciennes de la théorie de la manipulabilité qui font dépendre le jugement « X cause Y » de la possibilité d'agir sur X rencontrent le problème de rendre compte de jugements causaux portant sur des événements qui sont en principe soustraits à la sphère de l'influence de l'action humaine : les éruptions volcaniques et les explosions de supernovae sont ainsi des causes sans qu'elles puissent être modifiées par des actions humaines physiquement possibles. Dès lors que la notion d'intervention est définie de manière indépendante de l'action humaine, ce problème ne se pose plus. Cependant, il peut y avoir des relations causales dans lesquelles même une intervention au sens de la nouvelle théorie semble impossible. Pour évaluer le jugement selon lequel l'attraction gravitationnelle de la Lune est la cause des marées, il faut examiner les conséquences d'une intervention sur la position ou sur la masse de la Lune. On peut douter de la possibilité physique d'une « intervention » sur la Lune en ce sens, c'est-à-dire d'une modification de la position ou de la masse de la Lune qui n'influence pas elle-même les marées, par des chemins qui ne passent pas par la Lune.

Un autre problème peut être posé par des jugements causaux portant sur des variables pour lesquelles il semble dénué de sens de concevoir une intervention qui change leurs valeurs. Selon certaines conceptions de la causalité, on donne une explication causale du fait qu'un oiseau donné est noir en indiquant qu'il s'agit d'un corbeau. Pour un oiseau, le fait d'être un corbeau augmente par exemple la probabilité d'être noir. L'approche DN donne le même résultat. Cependant, il est douteux qu'on puisse attribuer un sens à l'idée de manipuler (ou d'intervenir sur) les valeurs de la variable correspondant à l'espèce naturelle à laquelle appartient un oiseau donné. L'énoncé suivant ne semble pas avoir de valeur de vérité déterminée : « si une intervention sur ce cygne (intentionnelle ou « naturelle ») le transformait en corbeau, il deviendrait noir ».

⁵¹ Cf. Woodward (2003, p. 50sq.).

8. Conclusion

La recherche philosophique sur la notion de causalité constitue un champ riche et complexe. Plusieurs approches fondées sur des notions et prémisses très différentes sont poursuivies depuis l'abandon de l'analyse déductive-nomologique. Chacune peut se prévaloir de certains succès au sens où elle rend compte d'intuitions ou de prétendus « faits » concernant la causalité qui font échouer des conceptions concurrentes. Mais chacune semble aussi se heurter à des contre-exemples. Une partie de la perplexité que peut ressentir un observateur de ce débat peut être surmontée lorsque l'on constate que le but poursuivi n'est pas toujours le même. Si la plupart des analyses philosophiques ont pour but l'analyse conceptuelle *a priori*, d'autres comme les théories en termes de processus naturels ou la théorie en termes de manipulabilité ont des ambitions plus modestes. Il existe un certain nombre d'efforts « œcuméniques » pour rendre compte de la multitude des approches poursuivies à partir de l'idée selon laquelle différentes conceptions sont adéquates à différents domaines : on peut considérer l'hypothèse selon laquelle l'analyse probabiliste correspond aux jugements causaux effectués en économie ou dans d'autres sciences sociales, alors que la théorie en termes de processus et de quantités conservées rend compte de la causalité physique. La conception contrefactuelle peut apparaître comme la plus adéquate pour rendre compte des raisonnements causaux que nous effectuons dans notre vie quotidienne. Outre un tel « régionalisme », on peut aussi envisager des formes de pluralisme ou de relativisme selon lesquels il existe plusieurs concepts causaux. Une chose peut causer une autre relative à l'un, sans la causer relativement à l'autre. Au sens contrefactuel, le caillou de S ne cause pas le bris de la bouteille dans la mesure où elle aurait été de toute façon cassée, étant donné la présence d'une seconde cause potentielle ou préemptée (ou *back-up*). Mais au sens de la conception du processus physique, c'est bien le caillou de S qui casse la bouteille. Des formes plus ambitieuses d'œcuménisme visent une théorie unifiée capable de rendre justice à toutes les situations, en utilisant des ressources conceptuelles puisées dans différentes théories ; tel est par exemple le cas des théories contrefactuelles probabilistes (Noordhof 1999), ou de la théorie selon laquelle la causalité se caractérise par l'augmentation de la probabilité d'un processus (Schaffer 2001). La conception de la réduction fonctionnelle pourrait servir de cadre à la recherche d'une telle position synthétique. Selon cette approche, la causalité est un concept dont les conditions d'application sont en partie *a priori* et en partie *a posteriori*. Historiquement, Armstrong (1968) et Lewis (1972) ont été les premiers à suggérer un modèle à deux étapes pour rendre compte du rapport entre esprit et cerveau. Depuis, cette stratégie d'analyse réductive a été appliquée à nombre d'autres concepts, dont la causalité. Dans une première étape de pure analyse conceptuelle *a priori*, on découvre pour un concept donné, ce que pourrait appeler son « profil fonctionnel » : on décrit les contraintes auxquelles doivent obéir les objets auxquels le concept s'applique. Pour reprendre l'un des exemples classiques de la réduction fonctionnelle en philosophie de l'esprit, la douleur est l'état d'un sujet A qui est causé par un dommage au corps de A et qui cause des états mentaux et comportements caractéristiques, tel que le désir que la douleur cesse, et des actions visant à interrompre ou à diminuer le processus à l'origine du dégât. Cette première étape de l'analyse peut être effectuée sans aucune recherche empirique et correspond à la découverte des conditions *a priori* d'application du concept. C'est dans une seconde étape que l'on découvre l'état naturel qui possède, dans notre monde réel, le profil fonctionnel en question. En ce qui concerne les concepts cognitifs comme celui de douleur, il est envisageable que l'on découvre que différents états naturels occupent le rôle fonctionnel dans différents systèmes cognitifs, appartenant par exemple à différentes espèces animales. Du coup, il y aurait bien un concept général de douleur, mais ce concept s'appliquerait à des espèces d'états différents dans différentes espèces animales.

Lorsqu'on transpose cette stratégie à l'analyse de la causalité, il est possible que l'on découvre que ce sont des relations différentes qui jouent, dans différents domaines, le rôle du concept de causalité. On aboutirait à une conception pluraliste qui autoriserait de juger qu'en épidémiologie et en économie, l'augmentation de la probabilité joue le rôle du concept de causalité, alors que dans le domaine de l'explication des actions des individus, le rôle est occupé par la dépendance contrefactuelle, dans le domaine de la biologie, par un mécanisme et dans le domaine de la physique, par une transmission. Il y aurait à la fois un concept général de causalité qui correspond aux contraintes *a priori*, notamment de séparation dans l'espace et dans le temps et d'asymétrie, et des concepts « régionaux », spécifiques aux différents domaines d'explication.

Références

- Aronson J.J. (1971), The Legacy of Hume's Analysis of Causation, *Studies in the History and Philosophy of Science* 2, p. 135-165.
- Armstrong, David (1968), *A Materialist Theory of Mind*, éd. révisée, Londres, Blackwell, 1993.
- Bennett, Jonathan (1987), Event Causation: the Counterfactual Analysis, *Philosophical Perspectives* 1, p. 367-368, repr. in Sosa et Tooley (1993), p. 217-233.
- Bennett, Jonathan (1988), *Events and Their Names*, Hackett, Indianapolis/Cambridge.
- Carnap, Rudolf (1966), *Philosophical Foundations of Physics*, New York, Basic Books; trad. *Les fondements philosophiques de la physique*, Paris, A. Colin, 1973.
- Cartwright, Nancy (1979), Causal laws and effective strategies, *Noûs* 13, p. 419-427; repr. In *How the Laws of Nature Lie*, Clarendon Press, Oxford, 1983.
- Cartwright, Nancy (1999), *The Dappled World, A Study of the Boundaries of Science*, Cambridge, Cambridge University Press.
- John Collins, Ned Hall, and L. A. Paul (éds.), *Causation and Counterfactuals*. Cambridge (Massachusetts), MIT Press, 2004.
- Cummins, R. (2000), How Does it Work? Vs. What Are the Laws? Two Conceptions of Psychological Explanation, in F. Keil & R. Wilson (eds.), *Explanation and Cognition*, Cambridge, MA: MIT Press, p. 117-45.
- Curiel, Erik (2000), The Constraints General Relativity Places on Physicalist Accounts of Causality, *Theoria* (San Sebastian) 15, p. 33-58.
- Davidson D. (1967), Causal Relations, in Davidson D. (1980), *Essays on Actions and Events*, Oxford, Clarendon Press 1980; trad. de P. Engel: *Actions et événements*, Paris, P.U.F., 1993.
- Dieks D. (1986), Physics and the Direction of Causation, *Erkenntnis* 25, p. 85-110.
- Dowe, Phil (1992), Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory, *Philosophy of Science* 59, p. 195-216.
- Dowe, Phil (1992a), Process Causality and Asymmetry, *Erkenntnis* 37, p. 179-196.
- Dowe, Phil (2000), *Physical Causation*, Cambridge, Cambridge University Press, 2000.
- Dretske, Fred (1977), Referring to Events, in P. French, T. Uehling Jr., and H. Wettstein (eds.), *Midwest Studies in Philosophy II*, Minneapolis, University of Minnesota Press, p. 90-99.
- Elga, Adam (2000), Statistical Mechanics and the Asymmetry of Counterfactual Dependence, *Philosophy of Science*, Supp. Vol. 68, PSA 2000, p. 313-324.
- Fair D. (1979), Causation and the Flow of Energy, *Erkenntnis* 14, p. 219-250.
- Ned Hall (2004a), Two Concepts of Causation, in Collins et al. (2004), p. 225-276.

- Hall, Ned (2000/2004b), Causation and the Price of Transitivity, in Collins et al. (2004), p. 181-204.
- Hausman, Daniel (1998), *Causal Asymmetries*, Cambridge, Cambridge University Press.
- Hesslow, Germund (1976), Two Notes on the Probabilistic Approach to Causality, *Philosophy of Science* 43, p. 290-2.
- Hitchcock, Christopher (1996a), The Role of Contrast in Causal and Explanatory Claims, *Synthese* 107, p. 395-419.
- Hitchcock, Christopher (1996b), Farewell to Binary Causation, *Canadian Journal of Philosophy* 26, p. 335-364.
- Hitchcock, Christopher (2001), The Intransitivity of Causation Revealed in Equations and Graphs, *Journal of Philosophy* 98, p. 273-99.
- Hitchcock, Christopher (2007), Prevention, Preemption, and the Principle of Sufficient Reason, *Philosophical Review* 116, p. 495-532.
- Hume, David (1739-40), *Treatise of Human Nature*, L.A. Selby-Bigge et P.H. Nidditch (eds.), Oxford, Clarendon Press, 1978; trad. en 3. vols., par P. Baranger et P. Saltel, *Traité de la nature humaine*, Paris, Flammarion, 1993-1995.
- Hume, David (1777), *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, L.A. Selby-Bigge et P.H. Nidditch (eds.), Oxford, Clarendon Press, 1975; trad. par M. Malherbe, in : D. Hume, *Essais et traités sur plusieurs sujets. Enquête sur l'entendement humain, Dissertation sur les passions*, Paris, Vrin, 2008.
- Keil, Geert (2000), *Handeln und Verursachen*, Frankfurt a.M., Vittorio Klostermann.
- Keil, Geert (2005), How the *Ceteris Paribus* Laws of Physics Lie, in: Jan Faye et al. (eds.), *Nature's Principles*, Dordrecht, Kluwer.
- Kim, Jaegwon (1973), Causes and Counterfactuals, *Journal of Philosophy* 70, p. 570-2.
- Kim, Jaegwon (1990), Concepts of Supervenience, repr. In Kim, Jaegwon, *Supervenience and Mind*, Cambridge, Cambridge University Press, 1993; trad. in *Survenance et esprit*, vol. 2, Editions Ithaque, à paraître.
- Kistler, Max (1998), Reducing Causality to Transmission, *Erkenntnis* 48 (1998), p. 1-24.
- Kistler, Max (1999), *Causalité et lois de la nature*, Paris, Vrin, Collection Mathesis, 1999 ; Traduction anglaise, *Causation and Laws of Nature*, Londres, Routledge, 2006.
- Kistler, Max (2001), Causation as transference and responsibility", in Wolfgang Spohn, Marion Ledwig & Michael Esfeld (eds.), *Current Issues in Causation*, Paderborn, Mentis, p. 115-133.
- Kistler, Max (2006), La causalité comme transfert et dépendance nomique, *Philosophie* 89, p. 53-77.
- Krajewski W. (1982), Four Conceptions of Causation, in: W. Krajewski ed., *Polish Essays in the Philosophy of the Natural Sciences*, Reidel, Dordrecht, 1982.
- Lam, Vincent (2005), Causation and space-time, *History and philosophy of the life sciences* 27, p. 465-478.
- Lewis, David (1972), Psychophysical and Theoretical Identifications, *Australasian Journal of Philosophy* 50, p. 249-258, repr. in : David Chalmers (éd.), *Philosophy of Mind: Classical and Contemporary Readings*, New York, Oxford University Press, 2002, p. 88-94.
- Lewis, David (1979/1986), Counterfactual Dependence and Time's Arrow, with Postscripts, in *Philosophical Papers, vol. II*, New York, Oxford University Press, p. 32-66.
- Lewis, David (1986a), Events, in *Philosophical Papers, vol. II*, New York, Oxford University Press, p. 241-269.
- Lewis, David (1986b), Causation, in *Philosophical Papers, vol. II*, New York, Oxford University Press, p. 159-172.
- Lewis, David (1986c), Postscripts to "Causation", in *Philosophical Papers, vol. II*, New York, Oxford University Press, p. 172-213.

- Lewis, David (2000), Causation as Influence, in Collins, Hall and Paul eds. (2004), p. 75-106.
- Mackie J.L. (1974), *The Cement of the Universe*, Oxford, Clarendon Press.
- McDermott, Michael (1995), Redundant Causation, *British Journal for the Philosophy of Science* 46: 523-544.
- O'Leary, John et Price, Huw (1996), How to Stand Up for Non-Cognitivists, *Australasian Journal of Philosophy* 74, p. 275-292.
- Paul, L.A (2004), Aspect Causation, in Collins et al. (2004), p. 205-224.
- H. Price et R. Corry (éds) (2007), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, Oxford, Clarendon Press.
- Maslen, Cei (2004), Causes, Contrasts, and the Nontransitivity of Causation, in Collins et al. (2004), p. 341-357.
- Menzies, Peter (2001), Counterfactual Theories of Causation, in *Stanford Encyclopedia of Philosophy*, revised 2008.
- Menzies, Peter (2004), Difference-making in Context, in Collins, Hall and Paul eds.
- Noordhof, Paul (1999), Probabilistic Causation, Preemption and Counterfactuals, *Mind* 108, p. 95-125.
- Noordhof, Paul (2004), Prospects for a counterfactual theory of causation, in Phil Dowe et Paul Noordhof (eds.), *Cause and Chance*, Routledge, 2004, p. 188-201.
- Pearl, Judea (2000), *Causality. Models, Reasoning, and Inference*, Cambridge, Cambridge University Press.
- Popper Karl R. (1934), *Logik der Forschung*, 3ème éd. augmentée, Tübingen, J.C.B. Mohr, 1969, trad. par N. Thyssen-Rutten et P. Devaux: *La logique de la découverte scientifique*, Paris, Payot, 1973.
- Popper, Karl R. (1956), *The Arrow of Time*, Nature 1977, p. 538.
- Putnam, Hilary (1984), Is the Causal Structure of the Physical Itself Something Physical ?, repr. in H. Putnam, *Realism with a Human Face*, ed. by J. Conant, Cambridge, MA, Harvard University Press, 1990 ; trad. par C. Tiercelin : *Le réalisme à visage humain*, Paris, Ed. du Seuil, 1994.
- Reichenbach H. (1956), *The Direction of Time*, Berkeley, Univ. of California Press, 1991.
- Russell, Bertrand (1912), On the Notion of Cause, *Proceedings of the Aristotelian Society*, 13 (1912-13) et *Scientia (Bologna)*, 13 (1913); repr. in *Mysticism and Logic* (1917), repr. Londres, Routledge, 2004 et *The Collected Papers of Bertrand Russell, vol. 6 : Logical and Philosophical Papers 1909-13*, John G. Slater (éd.), Londres et New York, Routledge, 1992, p. 193-210. Trad. Sur la notion de cause, *Philosophie* 89 (2006), p. 3-20, et in : B. Russell, *Mysticism et logique*, Vrin, 2007, p. 167-187. La pagination indiquée correspond à cette édition.
- Russell, B. (1914), *Our Knowledge of the External World*, Londres, Routledge, 1993.
- Russell, B. (1948), *Human Knowledge, Its Scopes and Limits*, Londres, Routledge, 1992.
- Salmon, Wesley (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.
- Salmon, Wesley (1994), Causality Without Counterfactuals, *Philosophy of Science* 61, p. 297-312.
- Savellos Elias E. et Yalçın, Ümit D. (eds.), *Supervenience: New Essays*, Cambridge, Cambridge University Press, 1995.
- Savitt, Steven (ed.) (2006), *Studies in History and Philosophy of Modern Physics* 37, no. 3 : « The arrows of time », p. 393-576.
- Schaffer, Jonathan (2000), Trumping Preemption, *Journal of Philosophy* 97, p. 165-181, repr. in Collins et al. (2004), p. 59-73.
- Schaffer, Jonathan (2000a), Causation by Disconnection, *Philosophy of Science* 67, p. 285-300.

- Schaffer, Jonathan (2001), Causes as Probability-Raisers of Processes, *Journal of Philosophy* 98, p. 75-92.
- Schaffer, Jonathan (2005), Contrastive Causation, *Philosophical Review* 114, p. 297-328.
- Schaffer, Jonathan (2006), Le trou noir de la causalité, *Philosophie*, no. 89 (2006).
- Smith, Sheldon (2002), Violated Laws, *Ceteris Paribus* Clauses and Capacities, *Synthese* 130, p. 235-264.
- Spirtes, Peter, Glymour, Clark and Scheines, Richard (2000), *Causation, Prediction and Search*, Second edition, Cambridge (Mass.), MIT Press.
- D. Spurrett et D. Ross (2007), Notions of Cause : Russell's Thesis Revisited, *British Journal for the Philosophy of Science* 58, p. 45-76
- Vendler Z. (1967a), Causal Relations, *Journal of Philosophy* 64, p. 704-713
- Vendler Z. (1967b), Facts and Events, in: *Linguistics and Philosophy*, Ithaca, N.Y., Cornell University Press
- Wittgenstein, Ludwig (1921), *Tractatus logico-philosophicus*, trad. G.G. Granger, Paris, Gallimard, 2001.
- von Wright G.H. (1971), *Explanation and Understanding*, Ithaca, N.Y., Cornell University Press.
- James Woodward (2003), *Making Things Happen: a Theory of Causal Explanation*, Oxford, Oxford University Press, 2003.
- James Woodward (2004), Counterfactuals and Causal Explanation, *International Studies in the Philosophy of Science* 18 (2004), p. 41-72.