



HAL
open science

Simulations, explication, compréhension : essai d'analyse critique

Cyrille Imbert

► **To cite this version:**

Cyrille Imbert. Simulations, explication, compréhension : essai d'analyse critique. *Philosophia Scientiae*, 2017, N'allez pas croire! Hommage à Pascal Engel, 21-3, pp.49-109. 10.4000/philosophiascientiae.1302 . hal-03036321

HAL Id: hal-03036321

<https://hal.science/hal-03036321>

Submitted on 21 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le présent document contient le texte de l'article.

Imbert, Cyrille. "Simulations, explication, compréhension : essai d'analyse critique." *Philosophia Scientiæ. Travaux d'histoire et de philosophie des sciences*, no. 21–3 (October 30, 2017): 49–109.

Pour des raisons de droits liés au format éditorial de sa version finale, il ne s'agit pas de la version au format définitif.

Une version pdf de l'article, avec la pagination de référence, peut néanmoins être trouvée sur la page :

<https://journals.openedition.org/philosophiascientiae/1302#text>

<https://doi.org/10.4000/philosophiascientiae.1302>.

Si, dans les années qui viennent, ce lien est amené à changer, le texte dans sa version finale publiée devrait néanmoins pouvoir être trouvé par une recherche sur le site de la revue, qui est publiée en libre accès.

Cyrille Imbert, janvier 2020.

Simulations, explication,
compréhension : essai d'analyse critique

Cyrille Imbert

Laboratoire d'Histoire des Sciences et de
Philosophie, Archives Henri-Poincaré, Université
de Lorraine, CNRS, Nancy (France)

fr

J'analyse dans cet article la valeur explicative que peuvent avoir les simulations numériques. On rencontre en effet souvent l'affirmation selon laquelle les simulations permettent de prédire, de reproduire ou d'imiter des phénomènes, mais guère de les expliquer. Les simulations rendraient aussi possible l'étude du comportement d'un système par la force brute du calcul mais n'apporteraient pas une compréhension réelle de ce système et de son comportement. Dans tous les cas, il semble que, à tort ou à raison, les simulations posent, du point de vue de leur valeur explicative, des problèmes spécifiques qu'il convient de démêler et de décrire précisément. J'essaie dans cet article d'analyser systématiquement ces problèmes en utilisant comme guide les théories existantes de l'explication. J'analyse d'abord le rapport des simulations à la vérité (section 2). J'examine ensuite en quoi les simulations satisfont ou non les exigences de déductivité et de nomicité, qui jouent un rôle central dans le modèle de l'explication de Hempel (section 3). J'étudie dans quelle mesure les simulations sont aptes à véhiculer l'information causale pertinente qu'on attend d'une bonne explication (section 4). Je poursuis en analysant en quoi l'abondance informationnelle et la lourdeur computationnelle des simulations peut sembler problématique par rapport au développement de nos connaissances explicatives et de notre compréhension des phénomènes (section 5). J'analyse enfin en quoi les simulations ont un rôle unificateur comme cela est attendu des bonnes explications (section 6). Au final, cette étude permet de comprendre plus précisément pourquoi les simulations, alors même qu'elles semblent pouvoir satisfaire les conditions que doivent remplir les bonnes explications, semblent spécifiquement problématiques au regard de

l'activité explicative. Je suggère que les raisons sont notamment à chercher dans l'épistémologie de l'activité explicative, dans les attentes méthodologiques envers les bonnes explications et dans l'usage spécifique qu'on fait des simulations pour l'étude des cas difficiles - en plus du fait que les simulations constituent une activité qui n'est plus à taille humaine.

Simulations, explanations,
understanding: a critical analysis essay

In this paper I investigate the potential explanatory value of computer simulations, which are often said to be suitable tools for the prediction, emulation or imitation of phenomena, but not for their explanation. Similarly, they are described as providing brute force (number-crunching) methods that are helpful to investigate the behaviour of physical systems but less so when it comes to understanding them. Be this as it may, wrongly or not simulations seem to be raising specific problems concerning their potential explanatory value. I try to systematically analyze these problems and for this purpose use existing theories of explanations as analytical guidelines. I first investigate how simulations fare concerning truth (section 2) and proceed by analyzing whether they satisfy the deductivity and nomicity requirements, which play a central role in Hempel's model of explanation (section 3). I also study whether simulations are appropriate vehicles for relevant causal information which a good explanation is expected to provide (section 4). I then analyze how much the informational repleteness and the inferential or computational stodginess of computer simulations are problems for the development of explanatory knowledge and our understanding of phenomena (section 5). I finally analyze how much computer simulations can have a unificatory role as good explanations arguably do (section 6). Overall, I try to analyze why computer simulations still raise specific problems concerning explanations while seemingly being able to meet the criteria that explanations should fulfil. I suggest that some of these reasons are to be found in the epistemology of explanatory activity, in methodological expectations towards good explanations and in the specific uses that are made of computer simulations to investigate complex systems—in addition to the fact that

computer simulations are not human-sized activities and that, as such, they may fail to provide first person understanding of phenomena.

49-109

30-10-2019

30-10-2017

1 Introduction

Mon but dans cet article est d'analyser la valeur explicative que peuvent avoir les simulations numériques. On rencontre assez souvent, aussi bien dans la bouche de scientifiques que dans celle de philosophes des sciences, l'affirmation selon laquelle les simulations permettent de prédire, de reproduire ou d'imiter des phénomènes, mais pas de les expliquer¹. Une idée voisine est que les simulations permettent d'étudier numériquement le comportement d'un système, par la force brute du calcul, mais pas d'atteindre une compréhension réelle de ce système et de son comportement – avec l'idée implicite que la compréhension est facilitée par d'autres techniques mathématiques, et notamment par la résolution symbolique analytique d'équations décrivant le système cible. On oppose également parfois les simulations aux théories dans la mesure où ces dernières seraient capables de nous éclairer sur un ensemble de phénomènes qu'elles unifient et expliquent alors que les simulations offriraient seulement un outil, souvent sans assise théorique solide, pour reproduire et prédire les phénomènes, parfois au moyen d'artefacts particuliers et « locaux ». À travers ces reproches, que nous étudierons plus en détails dans ce qui suit, se dessine la thèse selon laquelle les simulations ont peu de valeur explicative et ne nous permettent guère de développer une compréhension authentique des phénomènes qu'elles servent à explorer. Dans tous les cas, il semble que, à tort ou à raison, les simulations posent, du point de vue de leur valeur explicative, des problèmes spécifiques qu'il convient de démêler et de décrire précisément. Quel que soit le verdict de l'analyse sur la valeur explicative des simulations, un critère de réussite d'une telle étude est d'arriver à comprendre pourquoi, même si on finit par conclure que les simulations peuvent avoir une bonne valeur explicative, elles suscitent un tel malaise – afin de réussir ainsi à rendre compte (philosophiquement) des apparences.

¹ Je n'utilise pas ici le terme « expliquer » au sens de Duhem pour lequel la question du caractère explicatif des théories se confond avec celle de leur vérité. Comme nous le verrons, la question de la vérité des ingrédients d'une explication n'est qu'un aspect du problème.

Cette entreprise d'analyse du potentiel explicatif des simulations n'est pas qu'un jeu ou un exercice philosophique. Produire des explications et de la compréhension est un des objectifs majeurs de l'activité scientifique. Il se trouve par ailleurs que la science est en grande partie transformée par l'usage des calculateurs et des simulateurs, qui n'a cessé de progresser depuis leur utilisation dans le projet Manhattan pour l'élaboration de la bombe atomique. Analyser dans quelle mesure l'activité d'explication scientifique est – ou n'est pas – transformée par cette pratique scientifique et comprendre si le malaise mentionné n'est qu'un effet passager lié à la nouveauté d'un type de technique ou le signe de problèmes authentiques est une question légitime.

Pour qui est familier avec les discussions en philosophie sur la notion d'explication scientifique, analyser frontalement la question du pouvoir explicatif des simulations peut sembler une gageure. Depuis les articles séminaux de Hempel [Hempel 1942] & [Hempel & Oppenheim 1948], les discussions sur le sujet n'ont guère cessé. Le lecteur de *Four Decades of Scientific Explanation*, ouvrage dans lequel Wesley Salmon, un des acteurs majeurs de ces discussions, dresse un panorama des différentes contributions à travers quatre décennies, peut parfois avoir le tournis devant la profusion des thèses et des arguments, sachant que les trois décennies qui suivent ne sont pas moins riches en analyses sur ce thème [Salmon 1989]. Pour le philosophe en quête d'unité philosophique, la situation est d'autant plus désespérante – ou stimulante – que la majorité de ces contributions apportent des éclairages authentiques sur cette question... tout en étant souvent incompatibles. Face à cette situation, j'utiliserai certaines de ces théories de l'explication comme des outils théoriques qui conceptualisent et précisent certaines intuitions fortes sur ce sujet, éclairent les différentes dimensions du problème, et permettent ainsi de disséquer le potentiel explicatif d'une pratique comme les simulations. J'essayerai par ce biais de recenser les reproches que suscitent les simulations à propos de leur valeur explicative, d'examiner s'ils sont justifiés et de comprendre si les simulations suscitent ces reproches de façon accidentelle ou non.

En résumé, mon objectif est ici analytique et critique². Je me propose un tel objectif avec la conviction que c'est par une analyse patiente des éléments et des dimensions d'un problème qu'on peut espérer des progrès philosophiques, en atteignant des conclusions reposant le moins possible sur l'utilisation d'hypothèses ou de

² Pascal Engel fut un des premiers philosophes « analytiques » que j'ai rencontrés dans mes études de philosophie et avec lesquels j'ai découvert une façon de pratiquer la philosophie à laquelle je n'avais guère été formé jusqu'alors. Je suis en ce sens heureux de proposer un tel texte pour ces mélanges publiés en son honneur.

suppositions souvent évitables et dont l'accumulation peut fragiliser un résultat, le relativiser ou en affaiblir la généralité.

Deux mises en garde préliminaires doivent finalement être faites. i) Toutes les simulations n'ont pas de visée explicative. On utilise aussi les simulations pour faire des prédictions, explorer ou tester une théorie, un modèle ou une hypothèse, comprendre les effets de mécanismes potentiels, etc. Les conditions requises pour rencontrer le succès dans ces activités sont potentiellement différentes (et peut-être moins exigeantes) que celles pour bien expliquer. De nombreuses simulations n'expliquent donc rien tout en remplissant parfaitement le rôle scientifique qui leur est dévolu (même si la tentation peut exister de prétendre *après coup* qu'une simulation qui joue avec succès un autre rôle (e.g., prédire, contrôler) a aussi un rôle explicatif). ii) Les simulations se déclinent en de nombreux types (e.g., simulations de Monte-Carlo, méthode des différences finies, simulations multi-agents, utilisations d'automates cellulaires, etc.) ayant leurs spécificités. Une analyse bien menée doit essayer de montrer dans quelle mesure les conclusions obtenues sont générales ou bien sont spécifiques à un type particulier de simulations – et naturellement, il convient d'éviter les extrapolations indues à partir de conclusions portant sur des simulations particulières utilisées dans des contextes épistémiques spécifiques (voir [Barberousse & Imbert 2013a] pour une analyse critique des affirmations portant sur les simulations menées à partir d'automates cellulaires).

2 Explication, vérité et simulations

2.1 Explication et vérité

Les théories de l'explication comportent souvent une condition de vérité, qui dit que les différents éléments qui apportent des informations à visée explicative à propos du système dont on essaie d'expliquer le comportement doivent être vrais. (Ces éléments peuvent être de natures diverses : il peut notamment s'agir de lois, de modèles, d'énoncés, de simulations, etc.) Cette condition de vérité est en conformité avec ce que nous attendons usuellement d'une explication. Si je demande l'explication de la mort de John Fitzgerald Kennedy, j'attends qu'on me réponde en indiquant des faits qui sont vrais et non qu'on invente des histoires séduisantes mais imaginaires. Relativement à cette condition de vérité, les simulations ne semblent *dans l'absolu* pas poser plus de problèmes que les autres items scientifiques mentionnés ci-dessus. Une simulation, tout comme une loi, un modèle ou un énoncé peuvent être vrais, approximativement vrais, faux mais en conformité aux phénomènes ou contraires à ce que sont les phénomènes. Tous ces items peuvent en effet être vus comme des

représentations qui peuvent décrire ou présenter plus ou moins correctement ce qu'ils représentent.

2.2 Les simulations s'écartent-elles plus souvent et davantage de la vérité ?

Il est vrai que les simulations s'écartent peut-être *plus souvent et davantage* de la vérité. *De facto*, les simulations servent en général à étudier des phénomènes rencontrés dans des systèmes complexes dont la dynamique, si on devait la simuler exactement, demanderait trop de calculs et qu'on ne sait pas étudier par des méthodes analytiques. Pour cette raison, on n'hésite pas à faire de nombreuses approximations, simplifications ou idéalizations quand cela permet d'économiser du temps de calcul tout en permettant d'étudier correctement le système cible – et potentiellement, ces écarts à la vérité peuvent nuire à la qualité explicative des activités correspondantes, même quand ils ne nuisent pas forcément à leur valeur prédictive.

Néanmoins, la vérité n'est pas *nécessairement* sacrifiée au sein d'une simulation (ou en tout cas, pas davantage que pour toute autre représentation scientifique³). On peut ainsi simuler exactement l'évolution de systèmes simples modélisés par des équations pour lesquels il existe aussi des solutions analytiques qui, d'ordinaire, rendent inutiles les simulations [Humphreys 2004, 108]. C'est d'ailleurs en utilisant de tels cas déjà connus qu'on étalonne les simulations, c'est-à-dire qu'on vérifie qu'elles donnent des résultats fiables. Ces écarts par rapport à la vérité ne caractérisent donc pas intrinsèquement les simulations mais sont plutôt un trait de leurs *usages typiques* – et cette typicité s'explique par le fait que les simulations sont un outil performant qu'on utilise dans les cas « difficiles » et non d'un défaut qui leur serait propre. On ne reproche pas à un bon chirurgien mobilisé dans les cas extrêmes d'avoir plus de patients qui décèdent qu'un collègue ne pratiquant que des opérations de routine.

2.3 Simulations, modèles et vérité : attention aux comparaisons rapides

Les modèles scientifiques s'écartent aussi souvent de la vérité et les débats sur leur caractère explicatif sont également actifs (pour une présentation de ces questions, voir [Frigg & Hartmann 2012]). Il est instructif de comparer et de mettre en perspective les deux discussions.

³ Je mets ici de côté le fait que la discrétisation requise pour effectuer des simulations numériques entraîne un léger écart à la vérité quand la dynamique initiale du système est continue.

Il convient d'abord de noter qu'une simulation n'est jamais que l'exploration d'un modèle computationnel discret M_C , lequel est souvent une version computationnellement traitable d'un modèle initial M_I plus exact. Les simulations ne sont donc que l'exploration d'une sous-classe de modèles et la discussion sur le caractère explicatif des simulations et des modèles doit donc coïncider en partie.

Par ailleurs, *stricto sensu*, d'un point de vue épistémologique, un modèle ne nous *apporte* pas de connaissance explicative. Un modèle est une représentation statique qui peut au plus *contenir* de l'information explicative ou représenter des états de faits qui ont une valeur explicative, mais un modèle ne *montre* pas en quoi l'information qu'il contient, ou les états de faits qu'il représente, permettent de rendre compte d'un comportement – et c'est par un tel processus que nous identifions l'information explicative, et donc le modèle qui la véhicule, comme explicatifs, voire peut-être que nous expliquons. En d'autres termes, même si on considère que ce qui explique, c'est *in fine* l'information explicative ou le modèle qui la véhicule, et non la preuve ou la procédure qui identifie cette information ou ce modèle comme explicatifs, l'identification d'une explication à partir d'un modèle requiert son exploration, et c'est notamment parce que cette exploration peut être problématique que, dans le cas des simulations, on peut être appelé à s'écarter de la vérité. En un mot, quand on parle du pouvoir explicatif des modèles, c'est en écartant par une idéalisation implicite (à juste titre ou non) les aspects liés à l'épistémologie de l'activité explicative, c'est-à-dire ici à l'exploration effective du contenu et du potentiel explicatif des modèles. En revanche, comme une simulation est une activité qui présente *et* explore effectivement des représentations scientifiques rendues propres à cette exploration – au prix souvent d'écarts à la vérité –, quand on discute de la valeur explicative d'une simulation, les contraintes épistémologiques liées à l'exploration sont forcément déjà présentes à travers les contraintes que la représentation explorée (c'est-à-dire le modèle computationnel) remplit déjà. En somme, les simulations sont peut-être en partie considérées comme spécifiquement problématiques d'un point de vue explicatif, parce que précisément, elles « font le travail » et que l'étude de leur valeur explicative se fait à un niveau d'analyse dans lequel on intègre les difficultés effectives des enquêtes qui permettent de nous rendre accessible la connaissance explicative.

Une comparaison raisonnée devrait donc porter sur la valeur explicative des modèles effectivement explorés par simulations et celle des modèles effectivement explorés par d'autres méthodes. Le résultat d'une telle comparaison ne serait par ailleurs nullement évident. Les modèles dont l'exploration est aisée sont peut-être parfois plus explicatifs parce qu'ils correspondent à des cas simples où on peut à la fois représenter exactement les systèmes étudiés et explorer facilement

ces représentations. Néanmoins, dans certains cas, les modèles « simples » qu'on peut explorer sans simulation correspondent à des cas dans lesquels des approximations et des idéalizations extrêmes ont été faites pour rendre le modèle analytiquement explorable (e.g., en économie) – et justifier le caractère explicatif de tels modèles peut être tout aussi compliqué.

En définitive, contrairement à ce que suggère peut-être la grammaire, si les théories, les modèles ou les simulations sont dits « expliquer », c'est en des sens et dans des contextes en partie différents (même si les problèmes correspondants sont proches et s'entrecroisent) et il n'est pas forcément approprié de faire des comparaisons sur ce point.

2.4 Quand toute la vérité n'est pas requise pour expliquer

Le fait d'ôter du modèle ou de la simulation des détails non-pertinents n'est par ailleurs pas forcément une entrave à leur caractère explicatif (je reviendrai sur ce point dans la section 5). En effet, un modèle ou une simulation peuvent ne pas représenter fidèlement un système relativement à certains traits qui sont dans l'absolu physiquement importants pour expliquer d'autres aspects de son comportement mais qui ne sont pas pertinents pour expliquer le fait étudié. Supposons par exemple qu'on étudie des phénomènes de percolation au moyen de simulations. Le comportement des systèmes percolants près du seuil de percolation est décrit par des exposants critiques dont la valeur ne dépend en fait que de quelques traits comme la dimension du système. On dit que de tels systèmes ont des propriétés universelles [Stauffer 1985]. Du coup, si on ne s'intéresse qu'à l'explication de la valeur des exposants critiques d'un système et non à son comportement de détail, une simulation peut être explicative si elle représente correctement le nombre de dimensions du système étudié et ne représente pas fidèlement les autres traits.

2.5 Quand les simulations ne sont que « phénoménologiques »

Un reproche voisin souvent fait aux modèles et aux simulations⁴ est qu'ils permettent de reproduire des phénomènes ou des comportements mais ne permettent pas de les expliquer, parce qu'ils ne sont que « phénoménologiques ». Ce reproche peut se décliner de différentes façons en fonction du sens donné au terme « phénoménologique »,

⁴ J'indique dans ce qui suit différents sens de l'expression « modèle phénoménologique ». Comme une simulation repose toujours sur un modèle, la qualification de « phénoménologique » peut se transmettre aux simulations.

lequel est souvent employé trop vaguement pour permettre d'atteindre des conclusions solides (voir [Barberousse & Imbert 2013b] pour plus de détails).

1. Un modèle phénoménologique peut d'abord être défini comme un modèle qui permet de prédire la phénoménologie du système étudié alors qu'il décrit de façon inexacte les interactions ou mécanismes fondamentaux du système étudié, comme par exemple quand on simule la circulation automobile à partir d'automates cellulaires [Nagel & Schreckenberg 1992]. Pris à la lettre, le reproche est ici que prédire correctement n'est pas expliquer. Beaucoup de théories fausses sur des points fondamentaux (comme la théorie ptoléméenne des corps célestes) permettent de faire des prédictions correctes sans pour autant apporter l'explication de ce que nous observons (pour plus de détails sur ce point, voir la sous-section suivante).
2. Un modèle phénoménologique se définit aussi parfois comme un modèle qui propose des relations entre des variables correspondant à des quantités observables mais qui ne décrit pas les mécanismes sous-jacents du système. Le reproche est alors de ne pas donner une explication profonde dont le contenu soit ancré dans une description des interactions fondamentales (d'ordinaire non-observables) du système étudié. L'explication fournie est donc en partie superficielle.
3. Un modèle phénoménologique peut aussi être défini comme un modèle qu'on n'arrive pas à rattacher à la théorie plus fondamentale du domaine et qu'on établit et/ou justifie de façon en partie indépendante et souvent assez artisanale [McMullin 1968], [Rohrlich 1990]. Ce qui est alors en jeu est moins le contenu de l'explication que sa justification et son rapport aux théories du domaine. Par exemple, il arrive souvent en physique qu'il faille fixer la valeur de paramètres ou de coefficients (par exemple la viscosité d'un fluide), ou même de variables qui ne représentent rien (si, par exemple, on crée un terme artificiel de viscosité pour corriger un autre défaut du modèle), et qu'on ne soit pas capable de déduire ladite valeur à partir de la physique fondamentale. Dans ces cas-là, on peut choisir de fixer la valeur de ces paramètres, coefficients ou variables de façon expérimentale ou par ajustement. Comme la simulation est mal rattachée aux théories fondamentales, on peut dire que l'explication qui est fournie est mal justifiée, et peut-être fausse ou superficielle.

On peut répondre à ces reproches de la façon suivante. D'abord, le fait d'être « phénoménologiques » n'est pas une caractéristique intrinsèque des modèles et des simulations et la caractéristique ne leur est pas non plus spécifique, puisque la caractéristique peut aussi

caractériser une théorie ou toute représentation imagée ou linguistique. En revanche, les simulations sont sans doute *souvent* phénoménologiques en raison des possibilités qu'elles offrent et du type d'usage que ces possibilités favorisent (voir *infra*). Par ailleurs, le caractère purement phénoménologique d'un modèle ou d'une simulation n'obère pas nécessairement sa valeur explicative (même si dans les faits, il est possible que ce soit souvent le cas). Si les aspects phénoménologiques du modèle ou de la simulation ne sont en fait pas explicativement pertinents, car le comportement observé (par exemple une circulation en accordéon) ne dépend que de certaines propriétés fondamentales (par exemple, la conservation du nombre de véhicules, l'application d'un type de règle de freinage – *versus* la règle comportementale exacte, etc.) et si ces quelques propriétés fondamentales sont présentes dans le modèle ou la simulation, alors le caractère explicatif peut éventuellement être conservé malgré ces aspects phénoménologiques. Par ailleurs, être justifié théoriquement est un prédicat qui caractérise comment nous *justifions* une activité scientifique. Il n'est pas inenvisageable qu'une représentation à visée explicative capture correctement de l'information vraie à valeur explicative dont l'utilisation et la valeur explicative soient justifiées de façon non théorique – ce qui n'exclut pas qu'une justification théorique ultérieure puisse être proposée. Dans cette perspective, avoir un support théorique peut être considéré comme un moyen privilégié, mais pas forcément irremplaçable, pour construire, authentifier, et justifier l'explication d'un comportement.

2.6 Simulations et sous-détermination : les simulations comme leurres explicatifs ?

Revenons maintenant au reproche selon lequel les simulations permettent de reproduire la phénoménologie d'un système, mais pas forcément d'expliquer son comportement parce qu'elles s'écartent souvent *fatalement* de ce que devrait être, relativement à des objectifs explicatifs particuliers, une description vraie des mécanismes régissant le système cible (*versus* en présentent une version qui ne soit « que » partiellement, et sans préjudice explicatif, fausse, idéalisée, approchée, mal justifiée théoriquement ou phénoménologiquement).

Cet aspect du problème recoupe la question de savoir si une théorie, un modèle ou une hypothèse qui sauvent les phénomènes sont vrais et potentiellement explicatifs. L'histoire et la philosophie des sciences nous ont appris que la réponse à cette question était souvent négative [Duhem 1981]. En effet, il existe toujours une multitude d'hypothèses théoriques qui permettent de rendre compte d'un ensemble de phénomènes. En d'autres termes, les phénomènes sous-déterminent le choix des théories – et se pose donc la question de savoir quand il est

légitime de croire au contenu non observable de nos théories (ce qui correspond à la version traditionnelle du problème du réalisme scientifique).

Ce problème de la vérité des descriptions théoriques est à la fois i) étroitement lié à la question de savoir ce qui est explicatif et ii) néanmoins en grande partie indépendant de celui de savoir ce qu'est une bonne explication. D'un côté, comme indiqué plus haut, on attend d'une bonne explication qu'elle soit vraie – au moins relativement aux aspects du système qui comptent pour expliquer le phénomène étudié. Cette condition de vérité est la première qui doit être remplie si on souhaite avoir une bonne explication : une explication *complètement et/ou fatalement fausse* ne peut rien expliquer. En même temps, la question de savoir ce qu'est une bonne explication pose une multitude de problèmes distincts. *De facto*, les discussions sur la nature de l'explication scientifique se sont développées indépendamment des discussions sur la vérité des théories scientifiques. Il convient donc de ne pas réduire la question du caractère explicatif d'une activité scientifique à celle de la vérité des représentations utilisées dans cette activité. On comprend néanmoins que, dans les domaines où la question de la vérité des descriptions des mécanismes et des interactions cibles est problématique (e.g., en sciences cognitives, en sciences sociales), cette préoccupation vienne en premier dans les débats sur l'explication et que soit souvent rappelé que simuler n'est pas expliquer, puisqu'une simulation construite sur des hypothèses fausses peut reproduire la phénoménologie du système étudié.

Ce cadre étant fixé, la question de la vérité du matériau explicatif se repose avec acuité dans le cadre de la discussion des simulations. En effet, quand ils jouent avec les hypothèses, les scientifiques cherchent souvent à reproduire des comportements bien établis et leur bricolage conceptuel et théorique est donc orienté. Cela est dans l'absolu problématique en raison de la sous-détermination des hypothèses théoriques par les phénomènes et, comme nous allons le voir, dans le cas des études computationnelles et des simulations, de multiples facteurs aggravent significativement ce risque.

Premièrement, les simulations sont souvent utilisées dans des cas où les données sont rares, incomplètes et parfois médiocres en précision et en fiabilité ([voir Edwards 2010, chap. 10, pour le cas de la difficile production des données climatiques]) – ce qui est justement une motivation supplémentaire pour utiliser les simulations afin de produire des « données » (non empiriques), comme en astrophysique où les expériences sont difficiles ou impossibles à mener [voir sur ce cas Ruphy 2011]. Ce type de situation laisse plus d'incertitude à propos du comportement qui est à expliquer et moins de contraintes dans le paramétrage des simulations (e.g., choix des conditions initiales) et dans le choix des hypothèses.

Deuxièmement, les simulations, et notamment les investigations computationnelles, mettent en jeu de nombreuses variables. Or, plus le nombre de variables augmente, plus il est aisé de reproduire les phénomènes à expliquer. Cette possibilité est utilisée dans le champ de l'apprentissage automatique, par exemple au moyen de réseaux de neurones. Dans de tels cas, l'ordinateur apprend à faire une tâche (ici prédire ou reproduire des observations) à partir d'un corpus de cas, en adaptant la valeur des paramètres d'un grand nombre de neurones artificiels connectés et régis chacun par une fonction mathématique d'un même type simple. Quand le réseau a appris à faire la tâche, on l'utilise pour extrapoler, c'est-à-dire, résoudre des cas nouveaux (ici faire de nouvelles prédictions ou reproduire de nouveaux aspects du comportement). Clairement, dans un tel cas, la capacité du réseau à bien prédire un comportement ne suffit pas pour produire une explication de ce comportement à partir de la structure du réseau et de la valeur des paramètres des neurones artificiels. Même si les simulations mettent en jeu moins de variables indépendantes que les systèmes d'apprentissage computationnels, et même si elles sont scientifiquement plus contraintes, l'ajustement de leurs paramètres est une pratique qui, si elle est à la fois commune et utile, s'accompagne de risques épistémiques forts [Epstein & Forber 2013] et peut compromettre leur valeur explicative.

Pour finir, les simulations sont souvent utilisées dans des contextes épistémiques où les scientifiques n'ont pas le garde-fou d'une théorie générale bien ancrée. De multiples hypothèses locales sont alors parfois explorées au fil des enquêtes sans qu'une théorie couvre et oriente directement le choix de ces hypothèses. Détaillons ce dernier point.

Il peut d'abord exister des zones d'ombres théoriques dans les processus fondamentaux qui gouvernent un système, comme de nouveau en cosmologie avec l'hypothèse de l'existence d'une grande quantité de matière noire et d'énergie sombre [Ruphy 2011]. De plus, même dans les domaines où la dynamique fondamentale du système est connue (comme la dynamique des fluides), puisque cette dernière ne peut souvent pas être simulée dans tous ses détails, il est souvent nécessaire de re-théoriser ou re-modéliser à une autre échelle différents aspects de cette dynamique, comme par exemple les mécanismes de formation des nuages en météorologie. De telles situations laissent de nouveau une latitude théorique.

Dans d'autres domaines, il peut simplement exister de grandes incertitudes sur la nature des mécanismes qui produisent un comportement. Ainsi, dans les sciences cognitives, des théories concurrentes de notre architecture neuronale et cognitive peuvent s'affronter, et même les techniques modernes d'imagerie ne permettent pas de connaître dans le détail les mécanismes de notre cerveau. Du coup, être capable de simuler des aspects du comportement d'un esprit,

notamment en termes de capacités cognitives, est un pré-requis pour affirmer qu'on comprend ce fonctionnement. Néanmoins, cela ne suffit clairement pas, puisque différentes architectures neuronales (ou computationnelles) peuvent accomplir les mêmes tâches. Typiquement, une architecture connexionniste et une architecture computationnelle classique peuvent permettre l'exécution de tâches identiques. Pour que de telles simulations puissent avoir des prétentions explicatives, il faut donc utiliser des stratégies plus raffinées qui contraignent davantage les hypothèses sur l'architecture fonctionnelle du cerveau, par exemple en étudiant le temps de traitement des tâches [Pylyshyn 1984, chap. 5].

De la même façon, dans les sciences sociales, valider empiriquement le contenu d'une simulation [Arnold 2014] est une question compliquée et lui conférer une valeur explicative l'est *a fortiori*, puisque de multiples processus, proches ou non, peuvent produire un comportement global identique. Par exemple, de nombreux mécanismes peuvent produire un phénomène de polarisation des opinions, de consensus ou de coopération entre agents. En l'absence de stratégie spécifique pour donner une confirmation empirique à la simulation (soit par des données, soit indirectement par une théorie ayant une conformation empirique indépendante), il peut être légitime d'être circonspect sur la valeur de ces simulations. On peut du coup comprendre l'existence de soupçons méthodologiques généraux à l'égard des simulations dans de tels champs. Cela n'implique pas pour autant qu'elles doivent être rejetées comme méthodes, comme cela tend à être le cas dans des domaines comme l'économie [Lehtinen & Kuorikoski 2007]. En effet, précautionneusement utilisées, les simulations peuvent jouer un rôle explicatif important dans de tels domaines parce qu'elles offrent de nombreux avantages [Reiss 2011]. En un mot, certaines utilisations inappropriées et l'existence de risques généraux liés à l'usage d'une méthode dans un champ ne doivent pas condamner la méthode même.

On peut noter pour finir que les simulations, parce qu'elles fournissent des moyens d'investigation puissants, sont souvent utilisées pour explorer les effets de mécanismes ou étudier des mondes artificiels et l'émergence de comportements cibles (e.g., des phénomènes de polarisation d'opinions ou de formations de groupes dans le monde social). Dans les domaines où il n'existe pas de descriptions théoriques bien établies qui couvrent le comportement des systèmes étudiés et guident l'exploration, il peut être ensuite tentant de supposer indûment qu'on a ainsi expliqué le comportement correspondant – alors que, dans la plupart des cas, d'autres mécanismes différents peuvent le produire.

En résumé, il existe de multiples raisons pour lesquelles les simulations numériques font partie d'un environnement épistémique qui favorise la production d'explications indues. Les praticiens des

simulations doivent donc spécifiquement prendre garde à ne pas succomber aux biais de confirmation usuels (qu'on rencontre également chez les scientifiques), et qui, dans le cas considéré, sont d'autant plus dangereux qu'il est « facile » d'obtenir les résultats souhaités.

2.7 Typologie des écarts à la vérité, épistémologie de l'explication et simulations

Les considérations précédentes peuvent nous amener à distinguer au moins trois types de violations de la condition générale de vérité.

1. Dans certains cas, le modèle ou la simulation ne sont pas littéralement vrais mais ce n'est nullement gênant d'un point de vue explicatif (voire, comme nous le verrons, cela est même avantageux). En d'autres termes la représentation est en partie fausse mais l'explication reste complètement correcte. C'est par exemple le cas si le modèle représente correctement toutes les propriétés qui sont responsables de l'apparition du phénomène observé et s'écarte de la vérité pour les autres propriétés sans que cela n'altère la représentation des aspects des processus étudiés qui produisent le comportement cible.
2. Dans d'autres cas, le modèle ou la simulation ne sont qu'approximativement vrais ou sont idéalisés et ils permettent d'expliquer approximativement ou en partie leur comportement. En d'autres termes, la qualité de l'explication est partiellement dégradée et on peut parler d'explication approchée.
3. Dans d'autres cas, le modèle ou la simulation représentent de façon partiellement fausse, approchée, idéalisée ou complètement fausse le système étudié et cela altère complètement la qualité de l'explication.

Savoir dans quels cas de figure on se trouve est crucial pour évaluer la qualité des explications qu'on produit, sachant qu'une simulation peut combiner plusieurs de ces types d'écart à la vérité. Le problème est que l'impact exact des écarts à la vérité des représentations scientifiques sur leur valeur explicative n'est que rarement maîtrisé. C'est une tâche scientifique propre et difficile d'évaluer les effets de ces écarts sur le comportement d'une simulation et la façon dont il est produit. En d'autres termes, il est en général difficile de savoir avec précision dans lequel des cas ci-dessus on se trouve. Au final, les écarts à la vérité au sein des simulations à visée explicative posent deux problèmes. Ils posent d'abord un problème du point de vue de la logique de l'explication au sens où il convient de caractériser par des critères précis quels écarts sont fatals ou partiellement dommageables à

la valeur des explications. Ils posent ensuite un problème épistémologique au sens où, même si on arrive à philosophiquement trouver de tels critères logiques, reconnaître dans quel cas on se trouve est d'ordinaire difficile⁵. En conséquence, nous sommes souvent *de facto* dans l'incertitude relativement à la valeur explicative des représentations (resp. simulations) partiellement fausses, même dans les cas où nos doutes légitimes ne sont en fait pas fondés. C'est ce qu'on peut appeler le problème de l'opacité partielle du caractère dommageable (resp. non-dommageable) de la fausseté au sein des représentations (resp. simulations) à visée explicative. Cette opacité est un problème pour les scientifiques qui veulent expliquer, mais aussi pour les philosophes qui ont besoin de réfléchir à partir d'exemples clairs à ce qu'est l'activité explicative et le potentiel explicatif des simulations. Au final, être méthodologiquement par défaut dans le soupçon relativement à la valeur explicative des simulations eu égard à leurs écarts par rapport à la vérité est une posture épistémique compréhensible – même si, dans de nombreux cas, ces soupçons sont en fait non fondés (sans que nous soyons toujours en mesure de le montrer).

Dans ce qui suit, je ne traiterai plus de ce problème et considérerai donc implicitement des cas où la satisfaction de la condition de vérité n'est pas problématique. Comme nous allons le voir, la question de la valeur explicative des simulations pose de nombreuses autres questions.

3 Le modèle déductif-nomologique (DN) de l'explication et les simulations

Je discute maintenant le caractère explicatif des simulations par rapport à des aspects du modèle explicatif de Hempel. Ce modèle a relancé les débats sur la question de la nature de l'explication scientifique au milieu du XXe siècle et, même s'il a été critiqué, il reste très utilisé et commenté aujourd'hui encore pour discuter de la question de l'explication dans les différentes sciences.

Le modèle déductif-nomologique [Hempel & Oppenheim 1948, réimprimé *in* Hempel 1965] indique qu'une explication est constituée de deux éléments : un *explanandum*, c'est-à-dire un énoncé qui décrit le phénomène à expliquer, et un *explanans*, c'est-à-dire un ensemble d'énoncés qui sont invoqués pour rendre compte du phénomène [Hempel 1965, 247]. Plusieurs conditions sont requises pour que l'*explanans* explique avec succès l'*explanandum*. i) L'*explanandum*

⁵ De même, il n'est nullement gênant qu'une pratique soit souvent peu fiable si on est capable d'identifier précisément quand elle exactement elle l'est. Il n'est pas non plus préjudiciable que nos bibliothèques comportent beaucoup de livres médiocres si on est capable d'identifier quels livres il convient de lire.

doit être une conséquence logique de l'explanans. ii) Les énoncés qui constituent l'explanans doivent être vrais. iii) Enfin, les énoncés composant l'explanans doivent comporter au moins une loi de la nature, d'où l'appellation de modèle « nomologique ». Une explication doit donc se présenter sous la forme suivante :

$L_1, \dots,$	(Enoncés nomologiques)
L_r	
$C_1, \dots,$	(Enoncés particuliers : conditions initiales, conditions
C_k	aux limites, etc.)
E	(Description des phénomènes empiriques à expliquer)

La barre signifie que l'ensemble des propositions constituent un argument, et celui-ci doit être valide⁶. Un tel modèle est censé s'appliquer à l'explication de généralités et de phénomènes particuliers. Hempel donne comme exemple d'explication la dérivation valide de la position de Mars à un moment futur (*explanandum*) à partir des lois de la mécanique et des positions, des vitesses et des masses du Soleil et de Mars.

Ce modèle déductif-nomologique est complété par d'autres schémas d'arguments explicatifs destinés à expliquer les lois statistiques ou les événements ayant lieu seulement avec une certaine probabilité. Comme cette partie de la théorie de Hempel est notoirement problématique et n'est pas importante pour la présente discussion, je ne m'étends pas dessus ici.

3.1 Déductivité, simulations et épistémologie de l'activité explicative

J'analyse maintenant les simulations à l'aune de la condition i) sur le caractère déductif de l'explication. Je discute d'abord de la façon dont les simulations satisfont cette condition de déductivité puis analyse les conséquences qu'il faut en tirer à propos des qualités explicatives des simulations.

Dans une simulation, on étudie d'ordinaire l'évolution temporelle d'un système en tant qu'elle obéit à une dynamique dont on possède une description mathématique. On peut par exemple simuler l'évolution des planètes (resp. d'un fluide), dont la dynamique est (approximativement) décrite par la deuxième loi de Newton (resp. les équations de Navier-Stokes). Simuler le système étudié, c'est trouver

⁶ Un argument est valide quand le fait d'accepter ses prémisses nous contraint logiquement à aussi accepter sa conclusion. On distingue la validité sémantique de la validité syntaxique. Un argument est sémantiquement valide quand sa conclusion est forcément vraie si ses prémisses sont vraies, ce qui est ici noté avec une barre, e.g., $A, B/C$. Un argument est syntaxiquement valide quand on peut utiliser des règles syntaxiques bien définies pour dériver sa conclusion à partir de l'ensemble de ses prémisses, ce qui sera ici noté par \vdash , par exemple $A, B \vdash C$.

un algorithme, dérivé des lois décrivant la dynamique du système, qui permet de calculer mécaniquement les états successifs de ce système (en tant qu'ils instancient des descriptions d'un type donné). En d'autres termes, une simulation vise à produire une déduction logique syntaxique des états successifs du système et peut par là même satisfaire la condition i). (Pour être exhaustif, il faudrait analyser comment le modèle DN doit être étendu⁷ pour être appliqué aux processus probabilistes ou stochastiques puis montrer comment les simulations, en tant qu'activités qui étudient de tels processus, se conforment à cette extension du modèle DN.)

Précisons maintenant la façon dont les simulations, *en tant que processus physiques*, se conforment à cette exigence de déductivité (voir [Barberousse, Franceschelli *et al.* 2009], [Beisbart 2012], [Arnold 2013], [Durán 2013] pour plus de détails). Dans une simulation, à tout état physique de l'ordinateur correspond (dans la limite des conditions de bon fonctionnement) un état computationnel *EC*. Les règles de transition (déterministes ou indéterministes) décrivent comment on peut passer d'un état computationnel à un autre. L'ordinateur, s'il est bien construit et programmé, passe mécaniquement par sa propre dynamique physique d'état computationnel en état computationnel selon ces règles de transition. Cette succession peut être interprétée comme une déduction syntaxique valide $EC_1 \vdash EC_2 \vdash \dots \vdash EC_n$ à partir de l'état computationnel initial en utilisant les règles de transition comme règles de déduction syntaxique. À ce stade, les états computationnels, qu'on symbolise dans les ordinateurs par des suites de 0 et de 1, ne signifient rien. Ce calcul syntaxique peut ensuite être interprété au moyen de règles sémantiques comme une déduction portant sur des objets mathématiques (par exemple des nombres ou des vecteurs codés en binaires), laquelle nous renseigne sur l'évolution d'un système dynamique discret, par application d'un opérateur d'évolution (correspondant au modèle computationnel exact). Cette évolution peut enfin être considérée comme une représentation mathématique qui est informative sur une trajectoire du système dans l'espace des états du modèle qui décrit ce système. Une simulation permet donc, si toutes les étapes inférentielles décrites ci-dessus sont effectuées correctement, de faire une déduction sémantiquement valide entre des propositions décrivant les états C_1 d'un système physique, en enracinant cette déduction dans des calculs logiques syntaxiquement valides.

Examinons maintenant les conséquences de cet état de fait pour les activités explicatives menées au moyen de simulations. Tout d'abord, le fait que les simulations puissent être vues comme des procédures

⁷ Je songe ici à des extensions qui ne seraient pas identiques à celles proposées par Hempel pour traiter des inférences probabilistes et des raisonnements inductifs.

syntactiquement valides signifie que, en tant qu'activités déductives, elles se conforment aux idéaux les plus contraignants développés au XXe siècle à propos de la notion de preuve. En effet, la déduction faite par une simulation est purement formelle : elle consiste à avancer dans le processus inférentiel selon des règles syntaxiques qui peuvent être appliquées mécaniquement, c'est-à-dire sans aucun recours à l'intuition. Quand on peut mécaniser une procédure, plus rien de psychologique ne peut entrer dedans et tout est contenu *explicitement* dans les règles syntaxiques de déduction. Toutes les déductions ne sont pas de ce type. Les déductions d'Euclide nécessitent par exemple un recours à l'intuition du mathématicien pour pouvoir être accomplies. Cette mécanisation du processus déductif a pour conséquence qu'une cause possible d'erreur inférentielle est supprimée dans les simulations. Cela ne signifie pas en revanche que les simulations proposent des déductions infaillibles. Une simulation doit être codée par un être humain qui peut faire des erreurs. Vérifier la validité d'un programme informatique est en général un problème difficile [Oberkampff & Roy 2010]. Et des difficultés peuvent par ailleurs surgir du fait de la réutilisation des codes numériques [Lenhard 2016], du travail des scientifiques en équipe, de l'utilisation simultanée de différents codes interconnectés, de la qualité des nombres aléatoires utilisés et plus généralement du fait que les simulations sont utilisées pour étudier des systèmes complexes et sont donc elles-mêmes des objets logiques extrêmement élaborés. Nonobstant ces difficultés réelles, il n'en reste pas moins que les simulations sont des outils performants qui étendent le domaine de la science déductive et donc potentiellement le domaine de ce que nous pouvons expliquer déductivement.

Un autre avantage de la mécanisation du processus déductif est que cela contraint les scientifiques à expliciter intégralement ce qui joue un rôle déductif dans les modèles et donc à faire qu'aucune hypothèse explicative cachée n'intervienne dans le processus déductif. En effet, quand il faut implémenter un modèle, tous les éléments effectivement nécessaires pour la déduction doivent être explicitement intégrés au code (ce qui est souvent l'occasion d'identifier certaines hypothèses explicatives cachées), sinon la machine ne peut pas calculer. De ce point de vue, en tant que déductions explicatives, les simulations nous contraignent à formuler des explications *complètes* au sens où le matériau explicatif utilisé doit être suffisant pour pouvoir déduire l'explanandum – alors que, comme le note Hempel dans une section sur l'incomplétude des explications [Hempel 1965, 415-425], beaucoup d'explications sont souvent elliptiques (*id est*, elles laissent tacites certains aspects connus) et même incomplètes (c'est-à-dire informativement insuffisantes, même quand on fournit toutes les informations dont on dispose).

Deux points supplémentaires méritent d'être encore mentionnés. *Expliquer comment*⁸ l'événement explanandum survient étant donné les faits explanans n'est pas rendu manifeste par un argument explicatif ayant la forme d'un couple (prémisses, conclusion). De ce point de vue, une simulation explicative fait plus qu'indiquer un contenu informatif à valeur explicative dont on peut déduire l'explanandum, car elle produit inférentiellement le passage entre ce contenu explicatif et l'explanandum. En cela, comme souvent une preuve en mathématique, elle peut permettre de développer la compréhension de la façon dont l'explanans permet d'inférer l'explanandum et des processus précis qui mènent à l'occurrence de l'événement à expliquer (voir section 5.2).

Enfin, même si on considère comme Hempel qu'une explication n'est qu'un ensemble d'énoncés, les simulations à visée explicative sont dignes d'intérêt du point de vue de l'épistémologie de l'explication. En effet, elles fournissent non seulement l'explanans (ici, l'état initial couplé au modèle computationnel) qui mène potentiellement à l'explanandum (l'état final ou un de ses aspects), mais en plus, elles insèrent cet explanans, en tant qu'entité linguistique, au sein de l'activité explicative inférentielle qui permet de l'identifier comme tel. Une explication présentée sous forme de couple (prémisses/conclusion) n'atteste pas par elle-même sa validité en tant qu'argument. Si on sait par ailleurs que l'argument est valide, l'explication ne présente au mieux que le *résultat* d'une enquête explicative, mais n'indique guère comment l'enquête s'est déroulée, comment l'explication a été produite et nous est devenue accessible et comment cette production a été authentifiée comme explicativement correcte. Les simulations à visée explicative correspondent donc à une unité d'analyse qui est plus informative à propos de l'épistémologie de l'activité explicative (et notamment à propos des processus permettant de la produire et de la vérifier). L'étude de la complexité computationnelle de la simulation rend notamment possible d'évaluer les ressources en mémoire et en calcul que la production et l'authentification de l'explication ont nécessitées [Imbert 2008].

3.2 Lois, causes, régularités nomologiques et simulations

J'examine maintenant les simulations par rapport à la condition iii) de Hempel qui dit qu'une explication doit contenir une loi. De nombreux philosophes des sciences considèrent que la plupart des explications scientifiques doivent mentionner des informations à propos des causes des phénomènes à expliquer. Hempel pense, lui, que cette information causale peut être décrite en termes de régularités

⁸ Cette notion est distincte de celle de « *how-possibly explanation* » proposée par Dray [Dray 1957, chap. 6 : « Explaining Why and Explaining How »].

nomologiques. Caractériser le sous-ensemble des régularités qui peut jouer le rôle que Hempel souhaite leur donner est un des obstacles principaux de son modèle. Sans entrer dans ce débat, je souhaite ici simplement montrer que, par rapport à ce point précis, les simulations semblent ne pas poser de problème spécifique, voire peuvent fournir d'excellentes explications. Je discute à cet effet la façon dont Hempel répond au contre-exemple suivant proposé par Michael Scriven [Scriven 1962].

(G) : L'impact de mon genou sur le bureau a été la cause de la chute de l'encrier.

Scriven indique que dans cette explication de la chute de l'encrier ne figure aucune loi ou généralisation puisqu'est seulement indiquée la succession d'une cause et d'un effet particulier ; de plus, selon Scriven, il n'est pas possible de remplacer la mention de la cause par la mention d'une loi sans substantiellement modifier cette explication.

La réponse de Hempel est que l'explication *G* est insatisfaisante à bien des égards [Hempel 1965, 360 sq.]. Elle correspond à ce que Hempel décrit dès 1942 comme une *esquisse d'explication* [Hempel 1965, 238, 424]. Une esquisse d'explication est une explication qui est caractérisée par différents degrés d'incomplétude ou d'imperfection relativement aux qualités qu'on peut attendre d'une bonne explication. Elle indique de façon approximative des informations qui semblent explicativement pertinentes, mais son contenu descriptif (à la fois à propos de l'explanandum et de son explanans) reste vague, sa justification est incomplète, la transition entre circonstances antécédentes et fait particulier à expliquer est mal étayée et, du même coup, le degré de certitude associé à cette explication est faible – et de fait, dans certains cas, un coup de genou dans la table n'occasionnera pas de chute de l'encrier. Ce qui manque à cette explication, c'est justement de l'information nomologique permettant d'indiquer précisément, et avec plus de certitude, dans quelles conditions un choc entraîne une chute de l'encrier et une tache. Et Hempel de remarquer « qu'il ne semble y avoir aucune raison de douter de la possibilité d'ajouter ou d'établir un ensemble de lois de plus en plus étendu qui permettrait de fournir une explication de plus en plus précise et détaillée du phénomène considéré⁹ ».

Dans le cas présent, Hempel renvoie aux lois du domaine physique correspondant quand il indique que des lois élémentaires de la dynamique des fluides¹⁰ (et, peut-on ajouter, des lois sur les transferts

⁹ « And there seems to be no reason to doubt the possibility of adducing or establishing a gradually expanding set of laws which would afford an increasingly accurate and detailed explanation of the phenomenon at hand» [Hempel 1965, 362].

¹⁰ « Then some elementary laws in the mechanics of fluids might well provide adequate nomological support for the explanatory statement» [Hempel 1965, 361].

de quantité de mouvement lors d'un choc et sur la transmission des ondes dans les matériaux et à des interfaces) peuvent être utilisées pour établir que l'encrier ouvert a versé suite à un choc [Hempel 1965, 361]. Pour cela, on doit aussi disposer d'une description précise des conditions initiales (les rapports entre la force de l'impact, la masse de la table et de l'objet qui la heurte, la position des différents corps, la rugosité de la table, etc.). En résumé, l'explication initiale, *si elle doit satisfaire les qualités d'une bonne explication scientifique*, doit en effet être substantiellement modifiée. *In fine*, en complétant l'esquisse explicative proposée par Scriven, on doit pouvoir arriver à une explication ayant une forme déductive-nomologique proche de celle-ci :

- Ensemble de lois L_1 : Loi de Newton (pour le choc) ; équations de Navier-Stokes (décrivant les mouvements des fluides), lois sur la propagation des ondes, e.g., loi de d'Alembert,
$$\Delta\phi - \frac{1}{c^2}\partial_t^2\phi = 0$$
, etc.
- Ensemble de conditions particulières C_1

Chute de l'encrier puis existence d'une tache, etc.

Hempel est néanmoins conscient de la difficulté potentielle qu'il y a à changer une esquisse explicative en une explication en bonne et due forme en précisant rigoureusement l'explanandum et en incluant dans l'explanans toute l'information nomologique et particulière nécessaire pour déduire l'explanandum ainsi précisé. De plus, plus l'étude d'un processus est complexe, moins il est d'ordinaire couvert par des régularités nomologiques simples et moins il est probable que le problème mathématique correspondant à son étude puisse se résoudre par des techniques mathématiques simples. En conséquence, aller au-delà de la production d'une esquisse explicative, ce que Hempel considère comme tout à fait souhaitable, passe en général par l'utilisation de simulations numériques, comme à chaque fois qu'un traitement mathématique simple n'est pas possible.

Encore faut-il être conscient que le perfectionnement de l'explication par le recours à une simulation numérique rencontre également des limites. Passer d'une esquisse d'explication à une explication complète nécessite d'abord de préciser rigoureusement ce qu'est l'explanandum ; il convient ensuite d'enrichir les lois¹¹ et la

¹¹ « But we should note also that an expansion of the set of supporting laws will normally call for a corresponding expansion of the set of antecedent circumstances which have to be taken into account » [Hempel 1965, 362].

description des circonstances particulières qui composent l'explanans jusqu'au point où il est possible de déduire l'explanandum. Néanmoins, dans le cas présent, si on demande non pas l'explication de l'existence d'une tache, mais celle d'une tache d'une taille et d'une forme précise, alors « on ne connaît aucune loi qui permette « une telle inférence¹² ». Même si Hempel ne détaille pas ce point, on peut noter que l'impossibilité en question a de multiples origines. Pour arriver à expliquer la taille et la forme, il faudrait être capable d'expliquer comment et où se forment les gouttelettes éjectées, le détail de leurs trajectoires et de leurs chutes sur le tapis. Cela demanderait de connaître précisément le détail des conditions initiales, de simuler les mouvements du fluide à un niveau mésoscopique, voire au niveau microscopique des molécules. Même si dans l'absolu on possède la connaissance des lois qui couvrent de tels phénomènes, il nous est impossible de connaître de telles conditions initiales et, dans tous les cas, la complexité de la simulation correspondante serait rédhibitoire.

En définitive, les simulations semblent un outil adapté pour essayer de transformer, *autant que cela est possible*, les esquisses d'explication en explications nomologiques complètes et par là même d'étendre les frontières du domaine des phénomènes que nous sommes *en pratique* en mesure d'expliquer [Humphreys 2004], [Imbert 2008, chap. 6].

3.3 Les simulations comme narrations « déductives-nomologiques »

Un aspect qui rapproche les simulations de l'explication proposée par Scriven est qu'elles proposent une description séquentielle d'états particuliers qui se causent les uns après les autres. Néanmoins, dans la perspective hempélienne, ni la particularité extrême d'une explication, ni la séquentialité de son contenu ne sont des points gênants. Le rôle des lois n'est pas forcément de fournir une régularité générale qui lie *explicitement* et directement les conditions antécédentes à l'explanandum en un seul énoncé (comme dans une explication de la forme « si *A* alors *B* (*loi*), *A* (*conditions particulières*)/*B* »), mais de connecter nomologiquement ces événements.

De ce point de vue, si les différentes étapes d'une narration séquentielle sont reliées nomologiquement, celle-ci peut avoir valeur explicative¹³. Hempel nomme de telles explications « génétiques ».

¹² « If, by contrast, the *q*-statement is taken to specify, not only that the ink spilled out, but also that it produced a stain of specified size and shape on the rug, then, to be sure, no laws are known that would permit the inference from the *p*-statement (in any plausible construal) to this *q*-statement» [Hempel 1965, 361].

¹³ « An explanation may well be put into the form of a sequential narrative, but it will explain only if it at least tacitly presupposes certain nomic connections between the different stages cited» [Hempel 1965, 362].

Hempel revient sur la notion d'explication génétique quand il discute les explications en histoire qui présentent l'événement à expliquer comme « l'étape finale d'une séquence qui se développe » [Hempel 1965, 447]. Pour Hempel, de telles explications sont satisfaisantes et permettent de développer des attentes justifiées, etc., *pour autant que les étapes qui les composent soient reliées nomologiquement*¹⁴. Ainsi, c'est parce que l'histoire ne se résume pas à une accumulation de faits mais repose comme les autres sciences sur un tel tissu nomologique que Hempel peut proposer, dès le départ (en 1942), son modèle pour l'analyse des explications en histoire.

De même, une simulation chronologique de processus, par exemple de l'évolution de l'univers, *pour autant que son modèle computationnel est dérivé de lois*, peut aussi être considérée comme une explication génétique et être acceptable d'un point de vue hempélien.

3.4 Conclusion : les simulations et la *logique* de l'explication

En résumé, si on en reste aux conditions énoncées par Hempel dans sa théorie déductive-nomologique de l'explication, les simulations peuvent clairement fournir des explications satisfaisantes. Leur utilisation appropriée semble même faciliter la production de telles explications dans les cas complexes.

Il semble donc raisonnable de s'attendre à ce que les simulations puissent aussi apporter les bénéfices épistémiques que les explications doivent selon Hempel fournir. Un de ces effets est notamment la compréhension qui provient du fait qu'une explication montre comment, étant donné les circonstances particulières et les lois décrivant la situation étudiée, on devait s'*attendre* à l'occurrence de l'explanandum¹⁵. Une simulation, parce qu'elle externalise dans un ordinateur le processus computationnel explicatif, ne nous permet néanmoins pas de développer immédiatement, et en première personne, de la certitude ou des attentes directes relativement à sa conclusion –

¹⁴ « Undeniably, a genetic account of this kind can enhance our understanding of a historical phenomenon. But its explanatory role seems to me basically nomological in character. For the successive stages singled out for consideration surely must be qualified for their function by more than the fact that they form a temporal sequence and that they all precede the final stage, which is to be explained [...]. In a genetic explanation each stage must be shown to "lead to" the next, and thus to be linked to its successor by virtue of some general principles which make the occurrence of the latter at least reasonably probable, given the former» [Hempel 1965, 448–449].

¹⁵ « By pointing this out, the argument shows that, given the particular circumstances and the laws in question, the occurrence of the phenomenon *was to be expected*; and it is in this sense that the explanation enables us to *understand why* the phenomenon occurred» [Hempel 1965, 337, italiques dans le texte original].

nous reviendrons sur cet aspect dans la section 5.2. Néanmoins, pour en rester à un point de vue purement hempélien, on doit ici préciser que l'attente dont il est question doit être comprise *en un sens purement logique*¹⁶. Dans cette perspective, les simulations permettent bien de développer des attentes fortes, et donc de la compréhension au sens hempélien, dans la mesure où elles sont bien justifiées et qu'on croit qu'elles sont exécutées proprement, mêmes si ces attentes ne sont possibles qu'au prix d'une certaine dépendance épistémique [Hardwig 1991] envers la simulation¹⁷ et les personnes qui garantissent son bon fonctionnement informatique et son utilisation scientifique appropriée.

Malheureusement, comme noté par Salmon [Salmon 1989, 120], la notion d'attente élevée¹⁸ demeure insatisfaisante dans le présent contexte, ce qui est illustré par Railton ainsi : si on explique la désintégration radioactive d'un atome particulier, on peut être amené à conclure de façon déductive sur la base des lois pertinentes que cette désintégration avait une probabilité très faible d'advenir et que l'attente correspondante devait donc être très faible [Railton 1980].

Hempel pense par ailleurs que les analyses en termes de causalité peuvent être reconstruites en termes de régularités nomologiques et sa théorie nécessite donc de pouvoir utiliser une notion suffisamment satisfaisante de ce qu'est une loi, ce qui reste un sujet controversé (voir par exemple [Van Fraassen 1994] pour un examen critique des positions). Il est de plus apparu au fil des discussions que spécifier quelles régularités nomologiques peuvent être employées dans une bonne explication restait une question problématique. Par exemple, la théorie de Hempel semble vulnérable à l'utilisation de régularités qui suscitent des problèmes de pertinence explicative. Ainsi, pour reprendre un exemple de Kyburg, si je dis « tous les échantillons de sel qu'on ensorcelle se dissolvent dans l'eau, or cet échantillon a été ensorcelé, donc il doit se dissoudre », je propose une explication du fait que l'échantillon va se dissoudre qui est conforme au modèle de Hempel mais qui n'est pas satisfaisante, car l'ensorcellement n'est pas un fait explicatif ici pertinent (même si la régularité correspondante est

¹⁶ « The two kinds of explanation by covering laws have this feature in common : they explain an event by showing that, in view of certain particular circumstances and general laws, its occurrence was to be expected (*in a purely logical sense*) [...]» [Hempel 2001, 299, italiques ajoutées].

¹⁷ La notion de dépendance épistémique caractérise pour Hardwig une relation entre deux personnes. L'extension de cette notion, ou l'utilisation d'une notion voisine, pour décrire notre relation à certains objets comme les simulations est une question qui commence à être discutée.

¹⁸ Il conviendrait de distinguer ici entre la force de l'attente (ou plutôt de la croyance) et l'objet de cette attente ou de cette croyance, qui peut être décrit en termes de probabilité d'arriver d'un événement. Si j'ai une croyance forte à propos de la faible (resp. forte) probabilité d'arriver d'un événement, je dois avoir une attente faible (resp. forte) relativement à l'occurrence de cet événement.

avérée) [Kyburg 1965]. Par ailleurs, comme nous le verrons dans les sections 4.4, 5.1 et 6.3, dans certains cas, la qualité de certaines explications utilisant des lois scientifiques authentiques peut également prêter à caution. Ce type de problèmes, associés à d'autres difficultés de la théorie de Hempel, ont poussé les théoriciens de l'explication à chercher d'autres caractérisations de cette notion, notamment en termes de causalité.

4 Les simulations comme vecteur approprié d'un texte causal idéal ?

Alors que Hempel essaye dans sa théorie d'éliminer la notion de cause, trop suspecte à ses yeux, au profit de la notion de loi, Salmon en fait une notion centrale de sa théorie causale de l'explication. La notion de cause doit notamment servir à éliminer les pseudo-explications dans lesquelles des régularités fortuites, apparentes ou dérivées, sont utilisées indûment pour expliquer un fait¹⁹. À partir de là, une explication vise à montrer « comment “les événements” s'insèrent dans le réseau causal “du monde²⁰” » [Salmon 1984, 276].

Je présente d'abord la théorie causale de Salmon puis l'idéal explicatif des théoriciens causaux de l'explication en discutant à chaque fois dans quelle mesure les simulations peuvent, si on suit cette théorie de l'explication, être explicativement satisfaisantes. Je conclus sur les problèmes que pose cette théorie et les simulations relativement à la condition de pertinence explicative.

4.1 Théories causales de l'explication, simulations et représentation des processus causaux

Si on adopte une théorie causale de l'explication, il convient de préciser ce qu'est cette information causale, qui doit être la matière des explications. Selon cette théorie, les relations explicativement pertinentes sont les relations ou processus causaux, et une telle caractérisation doit permettre de résoudre des cas comme celui du sel ensorcelé. Il existe plusieurs théories causales de l'explication comme celles de Salmon, celle de Dowe [Dowe2000] ou les théories centrées sur la notion de mécanisme [Machamer, Darden *et al.* 2000], [Tabery 2004]. Je présente à des fins d'illustration celle de Salmon, qui, dans ce

¹⁹ Salmon avait auparavant développé une autre théorie de l'explication (la théorie de la pertinence statistique) qui ne s'est pas avérée satisfaisante sur ce point.

²⁰ « We explain events by showing how they fit into the causal nexus » [Salmon 1984, 276].

cadre, est la plus connue, mais les arguments que je donne à propos des simulations vaudraient en général aussi pour ces théories.

Salmon définit un processus causal comme un processus physique ayant la propriété de pouvoir transmettre des marques [*marks*]. Ainsi, une voiture en mouvement est un processus causal, car une raie sur la voiture sera transmise d'un point spatio-temporel à un autre. En revanche, l'ombre de la voiture est un pseudo-processus car, si on altère l'ombre en changeant la source lumineuse, par exemple avec un cache, la modification produite sur l'ombre ne se transmettra pas. Enfin Salmon définit une interaction causale comme l'intersection de deux processus causaux qui modifie la structure de chacun des processus. Des objections faites à ce premier modèle de l'explication [Dowe 1992, 1995], [Kitcher 1989], amènent par la suite Salmon à définir un processus causal comme la ligne du monde²¹ [*world line*] d'un objet qui transmet une quantité non nulle d'une quantité invariante (comme l'énergie, la charge, le moment angulaire, etc.) à chaque moment de son histoire (chaque point spatio-temporel de la trajectoire) [Salmon 1994, 257].

À partir de là, Salmon définit l'explication d'un événement *E* comme l'ensemble des processus causaux et des interactions causales qui mènent à *E* (c'est la partie étiologique de l'explication), ainsi que les processus causaux et les interactions qui constituent l'événement lui-même (c'est la partie constitutive de l'explication). Une explication montre ainsi comment un événement « s'insère dans un réseau causal » (voir la citation complète ci-dessus²²).

Si pour expliquer, il convient de fournir une telle information causale, on voit mal pourquoi les simulations numériques devraient avoir une valeur explicative faible, ou plus faible que les modèles, expériences de pensée, arguments et autres vecteurs potentiels du discours scientifique explicatif. Une simulation numérique peut bien sûr échouer à représenter correctement les processus causaux qui sont effectivement à l'œuvre dans un système mais il en va de même pour tous ces autres vecteurs. De la même façon, quelle que soit la « bonne » notion d'interaction et de processus causal à laquelle une telle théorie causale finisse par aboutir, les simulations numériques, parce qu'elles fournissent des outils logiquement riches, au sens où elles peuvent effectuer tout calcul algorithmique ou reproduire tout langage produit par un autre outil de représentation symbolique, doivent pouvoir représenter ces processus causaux – du moins pas plus mal que tout autre type de représentation symbolique à visée explicative.

²¹ Une ligne du monde [*world line*] est la collection de points d'un diagramme de l'espace-temps (ou de Minkowski) qui représente l'histoire d'un objet.

²² Pour plus de détails sur la dernière version de ce modèle voir [Salmon 1994, 1997], [Hitchcock 1995].

Dans le cadre de la théorie causale de l'explication de Salmon, les simulations numériques peuvent même avoir dans certains cas une valeur explicative optimale. Une simulation numérique peut permettre de retracer le détail des processus et interactions causaux qui mènent à un événement. Si on s'attache à définir les processus causaux en fonction de quantités invariantes (voir *supra*), le verdict est *a fortiori* le même. Ainsi, l'équation de Navier-Stokes est une équation de continuité décrivant le transport et la conservation de quantités invariantes comme l'énergie, la masse ou la quantité de mouvement. Elle ne peut en général être résolue autrement que par des simulations, qui permettent de suivre le transport et la diffusion de ces quantités. On peut ajouter que les simulations permettent de visualiser aisément, par de riches interfaces visuelles, l'ensemble des processus et des interactions qui mènent au fait à expliquer, et donc de les rendre épistémiquement accessibles, par exemple en montrant comment une dépression naît et se développe en passant dans des zones humides et chaudes. En un mot, les simulations sont un outil naturel pour produire et présenter les explications telles que les décrivent les théories causales de l'explication comme celles de Salmon, surtout dans le cas de processus complexes.

On peut arriver à la même conclusion relativement à toute théorie selon laquelle expliquer nécessite de décrire séquentiellement et précisément l'évolution de processus causaux décrits à une même échelle, évolution qui est en général complexe et qui ne peut souvent se faire (au mieux) que par des simulations.

4.2 Texte explicatif idéal, information explicative et simulations

Dans la perspective causaliste défendue par Salmon, peut être incluse dans l'explication d'un événement toute information qui nous renseigne sur la façon dont cet événement s'insère dans le tissu causal qui mène à cet événement, et plus on fait de lumière sur ce dernier, plus complète est l'explication. Cela amène Railton, pour préciser la conception de l'explication sous-jacente, à proposer la distinction entre les notions d'*information explicative* et de *texte explicatif idéal* qui contient toute l'information causale et nomologique pertinente relativement à un événement étudié – distinction que Salmon juge extrêmement féconde [Salmon 1989, 160] :

Le texte idéal pour l'explication du résultat d'un processus causal ressemblerait à quelque chose comme ceci : une série interconnectée de comptes rendus reposant sur des lois de tous les nœuds et liens composant le réseau causal et aboutissant dans l'explanandum, livrée avec une description complètement détaillée des mécanismes causaux impliqués et des dérivations théoriques de toutes les lois de couverture

impliquées. Ce compte rendu causal pleinement développé s'étendrait, via diverses relations de réduction et de survenance, à tous les niveaux d'analyse ; en d'autres termes, le texte idéal serait clos par toutes les relations de dépendance causale, de réduction et de survenance. Il constituerait l'histoire complète expliquant pourquoi l'explanandum a eu lieu, relativement à une théorie correcte de toutes les dépendances nomologiques du monde. [Railton 1981, 241, ma traduction]

Si on suit cette perspective – et en supposant que la notion d'information causale soit précisée (voir *supra*), le rôle quasi inévitable des simulations pour produire et présenter de larges fragments de ce texte explicatif idéal est de nouveau patent. Si on se contente de présenter dans une explication certains aspects parcellaires de ce texte, une explication brève et facile à traiter mathématiquement est peut-être envisageable. Néanmoins, si on souhaite recenser l'information explicative plus systématiquement, alors la description du système doit se faire plus riche et plus précise et, *in fine*, seule une simulation permet en général de traiter mathématiquement cette représentation explicative et d'en dérouler le contenu pour retracer le déroulement de tous les événements ayant mené au fait étudié, en indiquant leurs connexions. Cela ne signifie pas pour autant qu'une simulation peut présenter l'intégralité de ce texte explicatif idéal. Ainsi, même les simulations « directes », qui simulent les fluides de l'échelle à laquelle les processus de diffusion rendent les fluides homogènes à l'échelle des plus grands tourbillons [Pope 2000], malgré leur précision extrême et leur complexité d'ordinaire rédhibitoire, ne permettent pas forcément de suivre le détail des comportements turbulents, ne s'étendent pas à tous « les niveaux d'analyse » ni ne présentent explicitement les relations éventuelles de « réduction et de survenance » présentes dans les fluides.

En résumé, dans cette optique, une simulation est un vecteur approprié pour produire et présenter des explications qui indiquent, *autant que nous pouvons le faire*, une grande partie de ce texte explicatif idéal, même si d'autres méthodes peuvent être appropriées pour éclairer d'autres aspects de ce texte (e.g., la physique statistique pour les relations entre échelles).

4.3 Les simulations : l'oubli des intérêts épistémiques et de la pragmatique explicative ?

Railton ne prétend pas pour autant que pour expliquer, il soit forcément nécessaire de produire un tel texte explicatif idéal. Dans les contextes d'enquêtes ou de requêtes explicatives usuels, il suffit « d'avoir la capacité (en principe) d'en produire des segments arbitraires » [Railton 1981, 247]. En effet, quand on recherche une explication, nous ne sommes d'ordinaire intéressés que par certains de

ces segments, et cet intérêt varie de façon contextuelle. Une requête explicative ne vise donc en général qu'à renseigner certains de ces segments. De la même façon, nous sommes potentiellement capables d'expliquer le mouvement de n'importe quel corps céleste connu au sein de notre système solaire, même si nous ne sommes pas en mesure de déterminer la trajectoire exacte de tous les astéroïdes ou météorites qui y gravitent. Du coup, nous nous contentons d'étudier surtout les principales planètes et les astéroïdes géo-croiseurs et, en cas de risque de collision, nous raffinons à l'extrême l'étude de la trajectoire de ces derniers. De ce point de vue, notre capacité explicative ou prédictive est prise en défaut seulement quand nous ne sommes pas capables de prédire ou expliquer *ce qui suscite spécifiquement notre intérêt*.

Ainsi, il convient de faire une place importante à la pragmatique pour comprendre comment des explications d'un même phénomène peuvent varier dans leur forme et leur contenu et néanmoins satisfaire pleinement nos requêtes explicatives. Il s'agit d'une tradition qui a été explorée par des auteurs comme Bromberger [Bromberger 1966], Achinstein [Achinstein 1983], ou Van Fraassen [Van Fraassen 1980] – indépendamment des tenants d'un modèle causal de l'explication. Dans une telle perspective, une explication doit être vue comme une réponse à une question d'un sujet épistémique. La qualité des réponses explicatives doit alors aussi être évaluée en fonction des attentes du requérant et du type d'information qu'il attend. Une réponse explicative peut ainsi être considérée comme satisfaisante même si elle met entre parenthèses des éléments que, par exemple, des hempéliens jugeraient centraux, comme la dimension nomologique de l'explication. L'intérêt du requérant détermine le segment du texte explicatif idéal sur lequel porte l'enquête, mais également le type approprié de réponse. En effet, ce segment du texte explicatif idéal suscite l'intérêt en tant qu'il est d'un certain type. Le requérant a en général à l'esprit une *classe de contrastes* (pour reprendre une notion de la théorie pragmatique de l'explication de Van Fraassen [Van Fraassen 1980]) décrivant ces réponses possibles et il souhaite savoir quel élément de cette classe apparaît dans le texte explicatif. Par exemple, quand on demande pourquoi l'encrier est tombé, on peut vouloir savoir si cela est suite à un coup de genou, un geste maladroit ou au passage d'un chat, toutes causes qui peuvent occasionner la chute d'un encrier.

De nouveau, cette perspective a le mérite d'expliquer en quoi une explication peut à la fois être pleinement satisfaisante pour un individu ayant ses intérêts propres, avoir une dimension objective (on ne remplit pas n'importe comment tel ou tel aspect du texte explicatif) et être accessible à des créatures finies – remplir tel ou tel segment est beaucoup moins exigeant pour des êtres naturels limités que d'essayer de produire tout le texte explicatif idéal correspondant à un événement.

De plus, l'extrême diversité des explications que nous fournissons en différentes circonstances pour un même événement et qui, dans leur contexte, semblent appropriées, même si elles ne présentent qu'un aspect des mécanismes menant à un phénomène, devient également compréhensible. De ce point de vue, chercher à ordonner toutes ces explications sur une seule échelle décrivant la valeur explicative intrinsèque de chaque réponse est une perspective en partie inadéquate. La meilleure explication n'est pas celle qui apporte le plus d'information causalement pertinente, mais celle qui apporte juste celle qui est attendue. Néanmoins, la question de la valeur explicative n'est pas non plus dissoute car un requérant peut se tromper sur la classe de contraste pertinente et la réponse qui est faite peut également être erronée. On peut ainsi distinguer la *pertinence explicative*, qui correspond à une notion objective relativement au contenu du texte explicatif idéal, de la *saillance*, qui est une affaire d'intérêts particuliers [Salmon 1989, 161-162, 185].

Cette perspective pragmatique permet de discuter de façon plus aiguë la façon dont les simulations contribuent efficacement ou non à l'activité explicative. Une simulation peut apporter beaucoup d'information causalement pertinente. Néanmoins, donner une masse d'informations à propos du texte explicatif idéal à quelqu'un qui est intéressé par un seul aspect de ce texte, et donc chercher à rendre la réponse explicative aussi riche que possible est une erreur – même si la réponse ne contient que des éléments authentiquement explicatifs. Si, pour répondre à une demande explicative sur un aspect d'un processus causal menant à un événement, on fournit un disque dur contenant la simulation d'un ensemble énorme de processus menant à cet événement, le requérant peut ne même pas être capable d'extraire l'aspect du texte explicatif idéal qui l'intéresse (d'où un préjudice épistémique lié à cette erreur pragmatique). Une simulation peut donc être une mauvaise réponse explicative même quand elle ne contient que des éléments causalement pertinents car elle noie la réponse attendue dans un flot informatif encombrant – et viole donc une exigence de « pertinence » pragmatique eu égard à l'objet spécifique de la requête, qui est bien plus spécifique que ce qui est explicativement pertinent pour expliquer l'événement étudié.

De la même façon, une attente pragmatique implicite peut être qu'une réponse ne dépasse pas, pour être comprise, un certain niveau cognitif (e.g., si on s'adresse à un enfant) ou soit adaptée à un certain format de transmission (e.g., une page, une heure de présentation). De ce point de vue, une simulation peut de nouveau être une réponse appropriée du point de vue de l'objet précis d'une requête explicative (« expliquez-nous au mieux les processus causaux menant à cet

ouragan »), mais violer une maxime²³ disant qu'une explication doit être adaptée aux circonstances contextuelles dans lesquelles elle doit être proposée. De la même façon, souvent dans une conférence, voire dans un article scientifique, on ne fournit pas la preuve d'un théorème mais une version de cette preuve, ou un résumé de cette preuve adapté au public et à la durée de l'exposé. De ce point de vue, de nouveau, fournir comme représentation explicative une simulation est un acte communicationnel en général inadapté aux échanges explicatifs entre créatures humaines.

En résumé, même si l'activité explicative a sa logique objective, il s'agit également d'une activité humaine. Les simulations, contrairement à la réponse d'un bon pédagogue, ne sont pas des processus humano-centrés dans lesquels les caractéristiques de celui qui émet une requête explicative sont prises en compte et, de ce point de vue, elles peuvent peiner à satisfaire nos attentes explicatives. Elles peuvent donc à la fois fournir des représentations explicatives appropriées pour livrer une description rigoureuse des processus causaux menant à un événement tout en étant inappropriées par rapport à la majorité des contextes humains dans lesquelles sont échangées des explications.

4.4 Simulations et processus causaux : le même aveuglement dans les détails ?

Le reproche adressé dans le paragraphe précédent aux simulations est de ne pas être, en tant qu'actes de communication, pertinentes relativement aux circonstances et au contexte de l'échange explicatif correspondant – tout en permettant d'exposer une partie importante du texte explicatif idéal et du contenu explicatif causalement pertinent. Je souhaite maintenant montrer que les simulations – et le détail des processus causaux – peuvent dans certains cas apporter trop d'informations du point de vue de la logique de la pertinence explicative même.

Prenons l'exemple d'un système complexe comme un fluide ou un gaz. L'idéal explicatif prôné par la théorie causale de l'explication est la description de tous les processus et interactions causaux qui mènent à l'occurrence d'un fait à expliquer. Supposons qu'on souhaite expliquer le fait que tel système isolé atteigne après un laps de temps assez grand une température uniforme $T=300K$, alors qu'il avait initialement une température inhomogène. Une première façon d'expliquer cela est de simuler l'évolution du système, c'est-à-dire de suivre pas à pas le détail causal de son évolution jusqu'à l'équilibre.

²³ J'adopte volontairement une terminologie proche de celle utilisée par Grice [Grice 1975] pour décrire ces règles pragmatiques d'un échange ou d'une conversation à visée explicative.

Néanmoins, dans de tels cas, on peut considérer qu'il est plus explicatif de montrer à partir de la physique statistique que cette température est la température d'équilibre du système et que les systèmes tendent vers l'équilibre. Il ne semble pas alors que la connaissance du mouvement de toutes les molécules soit requise pour l'explication et que, bien au contraire, une telle information n'est pas pertinente et nous noierait dans des détails explicativement inutiles. Mais c'est précisément ce qu'on peut reprocher aussi aux simulations, à savoir de potentiellement nous submerger sous une information potentiellement inutile sans extraire de façon sélective ce qui permet, à moins de frais et plus directement, d'expliquer l'occurrence du fait étudié. L'esprit de cette critique faite à la théorie causale et aux simulations est qu'une bonne explication doit écarter tous les détails non pertinents, y compris le détail des mécanismes causaux si cela s'avère nécessaire. En rester au détail causal d'une situation particulière, c'est potentiellement être aveugle à des explications plus générales et meilleures. En un mot, la référence à des relations causales ne permet pas de régler le problème de la pertinence explicative [Hitchcock 1995], [Batterman 2002], [Imbert 2008, 2011, 2013]. J'analyse ce point plus en détail dans la section qui suit en rapport avec les simulations.

5 Information explicative et ressources inférentielles : des simulations trop replètes et trop coûteuses ?

Je discute maintenant spécifiquement de la question des ressources utilisées dans l'activité explicative. Si on souhaitait être extrêmement précis, il conviendrait de distinguer les ressources utilisées pour *produire* ou *identifier* une explication et celles utilisées pour la *présenter* et la *vérifier*, puisque, en général, ces quantités sont distinctes [Imbert 2008, § 6.2]. Ainsi, en général, trouver la solution d'une équation est beaucoup plus difficile que la vérifier. Afin de ne pas compliquer outre mesure la discussion, je ne rentre pas en détail dans ces distinctions, qui sont par ailleurs presque complètement absentes de la littérature philosophique (anglophone) sur la question (voir en revanche [Zwirn 2006] pour une prise en compte de cette dimension).

Par ressources inférentielles, je désigne les ressources au sens logique et informatique (comme l'espace et le temps de calcul) associées à la mise en œuvre des activités explicatives. Par ressources informationnelles, j'entends les informations à propos du système physique qui sont utilisées dans l'explication, comme le contenu des prémisses dans le cadre d'un argument déductif. Informations explicatives sur le système et ressources nécessaires à une activité explicative sont assurément des aspects distincts de la logique et de l'épistémologie de l'explication. Je les associe néanmoins ici car,

comme nous allons le voir, il existe des analogies fortes entre le traitement de ces deux questions. Pour le dire brièvement, toutes choses égales par ailleurs, moins on a besoin de ressources informationnelles et inférentielles pour bien expliquer un phénomène, plus l'explication peut être produite par un agent qui en sait peu et a peu de capacités déductives, plus elle est facile à appréhender une fois produite et plus elle est épistémiquement utile pour son utilisateur – même si, en deçà d'un certain seuil, peu facile à identifier, une explication qui est construite avec trop peu d'information sur le système étudié et au moyen de trop peu de ressources inférentielles risque d'être une mauvaise explication – et les choses ne peuvent plus alors être « égales ».

5.1 Pertinence et explication : rien de trop dans des simulations informationnellement saturées ?

J'ai mentionné à plusieurs reprises le problème de la pertinence explicative. Il s'agit d'un problème central à la fois d'un point de vue philosophique et scientifique et il est particulièrement aigu dans le cas des simulations. Pour comprendre pourquoi il en est ainsi, je commence par une présentation générale du problème sans laquelle la spécificité des simulations ne peut guère être comprise.

5.1.1 Le problème de la pertinence explicative : présentation générale

Le problème de la pertinence explicative est soulevé initialement dans le cadre de la discussion du modèle DN de Hempel. Comme indiqué plus haut, Kyburg note que, selon ce modèle, la régularité (non-scientifique) « tous les échantillons de sel ensorcelés se dissolvent dans l'eau » permet de déduire et donc expliquer la dissolution d'un morceau de sel [Kyburg 1965], ce qui est gênant. Les discussions philosophiques sur cette question pourraient laisser penser que la question de la non-pertinence explicative est un problème purement philosophique qui se rencontre seulement dans le cadre d'exemples introduisant des prédicats non scientifiques et visant à illustrer la difficulté à trouver une théorie philosophique de l'explication satisfaisante. Il n'en est en fait rien : le problème de la pertinence explicative se retrouve même à propos de prédicats scientifiques authentiques et touche aussi des aspects de lois scientifiques respectables. Par exemple, les lois de Newton couvrent les mouvements des planètes ; néanmoins, pour expliquer la loi des aires, une grande partie du contenu de la deuxième loi de Newton (ou principe fondamental de la dynamique) est non pertinent (voir *infra*, section 6.3 et [Imbert 2011] pour plus de détails sur cet exemple). Ainsi, le problème de la pertinence scientifique est-il avant tout un problème

scientifique : quand on cherche à expliquer un phénomène, le problème est bien sûr de trouver une description théorique *scientifique* qui couvre le domaine correspondant (*problème général de la pertinence scientifique*), mais aussi de déterminer ce qui, *au sein de cette* description théorique scientifique, est pertinent pour un phénomène particulier (*problème de la pertinence intra-scientifique*) [Imbert 2011].

Il s'agit de plus d'un problème propre à l'activité explicative : ajouter un élément non pertinent dans les prémisses d'un argument valide n'altère pas sa validité ; en revanche, la présence d'un élément non pertinent dans le matériau explicatif d'une explication semble gêner celle-ci. C'est ce que note Wesley Salmon avec acuité quand il demande « Pourquoi l'inclusion d'énoncés non pertinents est-elle inoffensive pour les arguments mais fatale pour les explications ? » [Salmon 1989, 102, ma traduction]. Même si, comme nous allons le montrer, le jugement de Salmon est sans doute trop sévère, il pointe le caractère réellement nocif des éléments non pertinents au sein d'une explication.

5.1.2 Pertinence explicative, vœu de pauvreté informationnelle et extension des connaissances explicatives

Présenter l'explication d'un phénomène, c'est essayer indiquer les éléments d'un processus physique qui sont cruciaux à son occurrence et en l'absence desquels il peut arriver que l'explanandum ne soit pas instancié ou ait une probabilité différente de l'être. C'est en ce sens que de tels événements font une différence relativement à l'occurrence (ou la non-occurrence) de l'explanandum. Inclure dans une explication des éléments en fait non pertinents, c'est (implicitement) affirmer que ces éléments sont pertinents et font une différence. C'est donc faire implicitement une affirmation fautive à propos de l'élément (non pertinent) concerné, laquelle a comme corollaire des affirmations fautes à propos de ce qui pourrait arriver dans des situations où cet élément non pertinent n'est pas instancié, à savoir que l'explanandum ne serait pas instancié ou le serait avec une probabilité différente. En d'autres termes, quand on inclut à son insu des éléments non pertinents dans une explication, on est *privé de la connaissance* que, en l'absence de cet élément, l'explanandum est quand même instancié ou a la même probabilité de l'être.

C'est pour cela que vouloir expliquer et produire de la connaissance explicative, c'est vouloir éliminer tous les éléments non pertinents d'une explication. Il serait en général paradoxal de dire « voici l'explication de X , mais la majeure partie de cette explication n'est pas explicativement pertinente relativement à X », au sens où les éléments cités ne contribuent en rien à son occurrence ou ne font aucune

différence pour celle-ci. Quand on propose sérieusement une explication, c'est donc implicitement en affirmant que, à *sa connaissance*, tous les éléments en son sein sont pertinents. Cette norme explicative, consubstantielle à l'activité de la recherche de bonnes explications peut être rapprochée de la maxime conversationnelle de quantité [Grice 1975] qui dit les choses suivantes : 1°) Faites une contribution à la conversation qui apporte autant d'information qu'il est requis. 2°) Faites une contribution à la conversation qui n'apporte pas plus d'information qu'il est nécessaire. Comme indiqué plus haut, appliquer l'équivalent de cette maxime de quantité à une « conversation explicative » est indissociable de l'objectif de recherche des connaissances que véhicule une bonne explication. Ce n'est donc pas un aspect pragmatique extérieur à la logique même de l'explication, dont on aurait seulement besoin pour comprendre certains aspects supplémentaires de l'activité explicative, comme la transmission ou la réception des explications. Autrement dit, toute personne qui recherche une explication ou en présente une afin de développer ou de transmettre des connaissances explicatives s'engage *ipso facto* à purger cette dernière de toutes les informations sur la situation étudiée qu'elle sait être non-pertinentes (sauf si les circonstances rendent cette « purge » impossible ou non souhaitable pour d'autres raisons). Chercher une explication, c'est ainsi faire vœu – dans la mesure du possible – de pauvreté informationnelle dans la description du système dont on explique le comportement.

5.1.3 La pertinence explicative : une vertu graduelle

Faut-il pour autant dire comme Salmon que les éléments non pertinents sont fatals aux explications ? Une explication qui comporte des éléments non pertinents peut néanmoins véhiculer de l'information explicative authentique. Dans le cas d'une explication déductive, où les prémisses sont suffisantes pour déduire l'explanandum, la personne qui propose l'explication indique que l'explanandum sera instancié dans tous les cas où les prémisses sont vraies. En cela, elle ne se trompe pas et, d'un point de vue extensionnel, elle possède une partie de la connaissance qu'il est possible d'avoir à propos du cas considéré, à savoir la connaissance d'une partie des cas où l'explanandum est instancié en vertu de faits identiques. Ainsi, l'ajout d'éléments non pertinents réduit la taille de l'ensemble dans lequel les prémisses sont vraies et donc notre connaissance explicative. Il altère l'explication mais ne détruit pas toutes les connaissances qu'elle véhicule à propos de ces cas où l'explanandum est instancié. Une explication comportant des éléments non pertinents peut donc encore avoir de la valeur explicative. Il s'agit simplement d'une explication qui n'est pas aussi

bonne et informative qu'elle pourrait l'être. De ce point de vue, l'affirmation de Salmon devrait être reformulée comme suit : « Pourquoi l'inclusion d'énoncés non pertinents est-elle inoffensive pour les arguments mais partiellement préjudiciable aux explications et aux connaissances explicatives qu'elles véhiculent ? »

Quand on produit une explication, l'objectif est donc de supprimer si possible toute l'information non pertinente du matériau explicatif afin d'obtenir une explication « minimale », dont tout le contenu est indispensable, du point de vue de la pertinence explicative. Comment définir cette minimalité est une question difficile : supprimer toutes les prémisses inessentiels d'une explication déductive ne suffit pas, puisque les prémisses peuvent comporter des prédicats, et plus généralement de l'information, qui les rendent *en partie* non pertinentes, mais qui ne sont pas aisément identifiables [Schurz 1991, *passim*, en particulier 421], [Imbert 2008, § 4.3.3.1]. C'est typiquement le cas quand certains aspects d'une loi ou d'un principe sont non pertinents. Le problème est similaire quand on analyse la minimalité en termes de modèles. Retirer des éléments explicativement non pertinents amène clairement à produire des modèles plus abstraits [Strevens 2008, § 3.4-3.5], ou des descriptions de modèles plus pauvres [Imbert 2008, § 4.3.5.2], mais comment définir un modèle minimal n'est pas une question philosophiquement facile puisque, trivialement, le modèle minimal le plus pertinent est le modèle défini au moyen de l'explanandum !

5.1.4 Épistémologie de la pertinence, opacité explicative et méthodologie par provision

Dans tous les cas, en présence d'un argument déductif valide, trouver un autre argument valide qui donne exactement la même conclusion *C* mais qui repose sur des prémisses moins riches, dont des éléments non pertinents ont été supprimés, est une tâche scientifique ardue – et naturellement, cela est impossible quand tout le contenu des prémisses est essentiel à la dérivation. Établir qu'un élément *X* est intégralement essentiel à une dérivation est également une tâche difficile, puisque cela nécessite de montrer que la conclusion *ne peut plus être dérivée* si on supprime *X* ou une partie de son contenu²⁴. En définitive, établir que tout le matériau explicatif d'une explication est complètement pertinent est extrêmement délicat, puisque cela nécessite de montrer que, *quel que soit la partie ou l'aspect du contenu du*

²⁴ On peut néanmoins remplacer *X* par un élément incompatible et montrer qu'une conclusion *C'* incompatible avec *C* peut être dérivée. Mais même dans ce cas, certains aspects de *X* (une partie de son contenu logique) peuvent être non pertinents. Sans la 2^e loi de Newton, on ne peut déduire la loi des aires mais une partie de son contenu est non pertinent relativement à cette loi.

matériau explicatif qui est supprimé, la dérivation est *impossible* (c'est-à-dire de prouver un énoncé correspondant à une quantification universelle à propos de résultats d'impossibilité). Pour cette raison, en général, il est difficile d'être certain qu'une explication repose exclusivement sur du matériau explicatif pertinent, ce qu'on peut nommer le *manque de transparence ou l'opacité de la satisfaction de la condition de pertinence*.

Ainsi, contrairement au cas des prédictions et des déductions, qui sont des tâches qu'on peut avec certitude achever (d'un point de vue déductif, en établissant leur validité), il n'y a en général pas d'indication claire qui indique que la tâche explicative pour un phénomène donné, relativement au critère de pertinence, est achevée.

Par ailleurs, nous savons qu'il est parfois possible de faire reposer les explications sur des prémisses peu nombreuses en quantité ou faibles en contenu informatif – et les explications, construites intégralement par des esprits humains, sont en un sens de ce type. L'existence de telles explications à taille humaine nourrit ensuite l'espoir qu'il soit possible d'expliquer tel ou tel autre phénomène par une explication à taille humaine, qui utilise elle aussi un matériau explicatif peu abondant. Inversement, il est aisément possible d'ajouter *ad libitum* des informations non pertinentes dans une explication et d'ainsi produire des explications suspectes. Sur la base de ces deux états de fait clairement attestés, il n'est peut-être pas illégitime d'adopter par défaut la posture méthodologique suivante :

Maxime par provision relative à la pertinence explicative : « Plus une explication (resp. un argument) repose sur des contenus informatifs abondants, plus nous sommes en droit de suspecter qu'elle (resp. il) contient des contenus non pertinents relativement à l'explanandum (resp. relativement à la conclusion de l'argument). »

Cette maxime ne fait que retraduire notre attente à propos des explications et notre connaissance que, en l'absence de preuve du contraire, une explication *peut* être indûment replète, surtout si elle repose sur des descriptions riches. Cette maxime explicite et reflète peut-être aussi les normes disciplinaires implicites qu'on trouve dans les différents champs disciplinaires et qui sont souvent en défaveur des simulations. Elle ne dit en revanche rien de définitif à propos du caractère réellement adéquat de telles explications. Et je ne m'engage nullement ici sur sa fiabilité en tant que maxime devant guider notre pratique au sein des différents champs (voir infra pour la discussion de ce point).

5.1.5 Simulations, pertinence, opacité et soupçon de surabondance informationnelle

Nous sommes maintenant en mesure d'analyser précisément le problème que les simulations posent relativement au problème de la pertinence explicative.

Dans des cas comme celui de la chute de l'encrier (voir la section 3.2), si on souhaite expliquer de façon précise les détails de la chute et de la formation de la tache, il semble quasiment inévitable de devoir simuler le détail des processus physiques qui se sont déroulés et de fournir une multitude d'informations sur les circonstances de la chute. En revanche, dans de nombreux cas, on peut soupçonner que les simulations numériques, qui correspondent en général à des descriptions replètes, représentent trop précisément le système étudié et son évolution et ne sélectionnent pas les faits réellement explicatifs.

Prenons de nouveau l'exemple de la percolation. La mer, lorsque son niveau descend lentement, laisse apparaître entre les continents des îles de plus en plus nombreuses qui finissent, au seuil de percolation, par se rejoindre et par permettre de passer d'un continent à l'autre par voie terrestre. Comme indiqué plus haut, des études montrent que les propriétés statistiques du système sont décrites à ce seuil par des exposants dits « critiques ». Si on simule le retrait de la mer et qu'on calcule au fur et à mesure les quantités appropriées, on peut obtenir dans des cas particuliers la valeur de ces exposants critiques. Néanmoins, on n'aura pas pour autant correctement expliqué le fait que de tels systèmes soient décrits de façon générale au seuil de percolation par les mêmes exposants critiques. La simulation peut en effet nous faire croire que le détail de la dynamique de chaque système a une valeur explicative pour rendre compte de la valeur particulière de ces exposants. Il n'en est en fait rien car, comme le montrent les preuves analytiques existantes [Stauffer 1985], quelques traits spécifiques du système, comme sa dimension, suffisent à expliquer la valeur (générale) de ces exposants. Ce que montre la simulation, c'est donc seulement que la description du système constituée par la condition initiale, sa géométrie, etc., et la dynamique de retrait est *suffisante* pour dériver le fait explanandum. Et tant qu'une preuve générale n'est pas fournie, il est difficile de déterminer ce qui, dans le modèle computationnel utilisé, est en fait non pertinent relativement à la valeur des exposants et jusqu'à quel point cette valeur est générale (voir aussi [Barenblatt 1996] et, pour une analyse philosophique, [Batterman 2002]). En un mot, parce qu'elles fournissent en général des

explications potentielles replètes²⁵, les simulations tombent sous le coup de la maxime énoncée plus haut et suscitent le soupçon.

5.1.6 Le nouveau problème de la simplicité (explicative) informationnelle

On se trouve *in fine* dans la situation épistémique suivante²⁶.

1. Il est vraisemblable que la majorité des phénomènes physiques, pour peu qu'on souhaite expliquer leurs détails, nécessite irréductiblement de partir de descriptions riches des systèmes dans lesquels ils surviennent (chaque détail physique a de multiples effets !) et dont seule une simulation informationnellement précise peut nous indiquer comment ils évoluent.
2. Il est scientifiquement légitime de vouloir développer *autant que possible* des explications reposant sur des descriptions pauvres (afin de développer notre connaissance explicative, voir *supra*) et donc d'essayer les descriptions sur lesquelles reposent nos explications – ce qui est possible dans de nombreux cas.
3. Nous ne sommes que rarement en mesure d'affirmer de façon pleinement justifiée qu'une explication repose sur une description explicativement minimale du système étudié – et le soupçon peut donc porter sur toute explication.
4. Le caractère informationnellement (et inférentiellement, voir *infra*) pesant des simulations est sans doute une motivation supplémentaire (mais extérieure) pour essayer de se passer autant que possible de ces méthodes – et donc d'être peut-être plus exigeant avec elles relativement à la condition de pertinence.

En conséquence, nous pouvons nous trouver dans une situation où nous soupçonnons régulièrement que nos explications, qui reposent sur des simulations, ne sont pas satisfaisantes alors que, potentiellement, elles le sont dans la majorité des cas et que l'explication des phénomènes correspondants – par exemple de la naissance d'une

²⁵ Toute simulation ne repose pas nécessairement sur des prémisses ou un modèle complexe dont la description et l'état soient informationnellement riches. Typiquement, la règle dynamique d'un automate cellulaire peut être simplissime et s'appliquer à des conditions initiales également très simples. Néanmoins, en pratique, les simulations reposent souvent sur des modèles dynamiques complexes qui sont alimentés par des conditions initiales et aux limites compliquées qu'il est long de décrire.

²⁶ Un parallèle peut être fait avec le cas des nombres aléatoires. La majorité des nombres sont aléatoires, mais il nous est très difficile de produire ou de reconnaître un nombre aléatoire. Du coup, si on nous présente un nombre (sans aucune information sur sa source) doit-on supposer qu'il est aléatoire?

tornade à tel endroit précis – nécessite d’entrer dans le détail des processus qui les produisent. Ces soupçons reposent sur une maxime qui a une certaine légitimité même si une application systématique de cette maxime, qui nous pousserait à rejeter ou ignorer les explications correspondantes comme étant encore trop superficielles, pourrait nous amener à écarter souvent des explications en fait informationnellement optimales. En d’autres termes, il est vraisemblable que la maxime ci-dessus, malgré sa légitimité, correspond dans le cas général à une pratique ayant une fiabilité faible (même si, dans certains sous-ensembles de cas, peut-être caractérisables, elle peut avoir une haute fiabilité). On peut donc reformuler le présent problème sous la forme de la question suivante :

Quand est-il légitime de considérer qu’une explication, parce qu’elle repose sur un matériau explicatif riche (var. complexe, très précis, peu simple) est superficielle (et doit être traitée comme telle) ?

Ce problème a les caractéristiques suivantes. i) Ce qui est en jeu est l’accès à la connaissance explicative, puisqu’une explication reposant sur des informations non pertinentes nous prive de connaissances (voir *infra*), mais que rejeter une explication comme trop superficielle (alors qu’elle est peut-être optimale) nous dessert également épistémiquement. ii) Ce problème se pose effectivement dans la pratique scientifique face à des recherches reposant sur des simulations : face à une proposition d’explication qui repose sur une simulation, un évaluateur doit-il rejeter l’article au motif que l’explication *semble* superficielle car elle repose sur beaucoup d’information²⁷ ? iii) Le problème est analogue, mais distinct, du problème usuel de la fiabilité des extrapolations faites à propos de la vérité des hypothèses à partir de leur plus ou moins grande simplicité, et de la nature des maximes méthodologiques (non infaillibles) à appliquer pour guider au mieux notre pratique. Il n’est en l’occurrence pas question de faire des suppositions sur la valeur de vérité des hypothèses qui couvrent un domaine mais sur la quantité d’information explicative qui est irréductiblement requise pour bien expliquer un phénomène dans le cadre d’une théorie ou d’un ensemble d’hypothèses.

En résumé, on peut conclure que les simulations à visée explicative sont une pratique qui, relativement au problème de la pertinence explicative, suscitent presque inévitablement le soupçon, voire la critique (à titre méthodologique), même s’il est possible que dans la

²⁷ Une pratique (par ailleurs également légitime) consiste à demander aux chercheurs d’établir la robustesse (par rapport à des changements de conditions initiales, de valeurs de paramètres, de modèles, etc.) des résultats obtenus par simulation afin de garantir la solidité de leurs résultats. Mais si un résultat est robuste par de multiples changements, il est plus vraisemblable qu’il puisse être établi de façon plus générale, sans simulation.

majorité des cas (lesquels sont hélas difficiles à identifier) ces soupçons ou ces critiques leur soient adressés de façon erronée.

5.2 Ressources inférentielles et compréhension : le problème de l'opacité des simulations

Je passe maintenant à la question des ressources inférentielles ou computationnelles requises par l'activité explicative, lesquelles, dans le cas des simulations numériques, sont un aspect particulièrement crucial puisque, si la quantité de ressources requises est trop importante, l'activité explicative ne peut avoir lieu. Je traite cette question après celle de la question des ressources informationnelles, car, comme nous allons le voir, elles posent des problèmes analogues, et des conclusions proches peuvent être défendues dans les deux cas. Néanmoins, d'un point de vue explicatif, les enjeux sont différents, puisque dans le premier cas, il s'agit d'identifier le cœur d'une explication, à savoir les informations explicatives qu'elle doit contenir, alors que, dans le cas des ressources inférentielles, il s'agit « simplement » de discuter ce qu'est le coût de la production ou de l'authentification d'une explication, si on suppose qu'on possède déjà (mais sans le savoir) le bon matériau explicatif.

Dans le présent contexte, la question des ressources inférentielles utilisées recoupe de façon importante celle de la compréhension scientifique. Pour cette raison, même si l'analyse de la compréhension scientifique reste une question en partie distincte, et qui a ses propres spécificités, je discute conjointement ces deux questions²⁸.

Il convient par ailleurs de noter que le développement des calculateurs a parallèlement bouleversé les mathématiques et suscité d'intenses discussions sur la nature et le rôle des preuves et sur la notion d'explication mathématique. Une étude *comparative* approfondie mériterait d'être menée. Pour une introduction à ces questions, voir notamment [De Millo, Lipton *et al.* 1979], [Asperti, Geuvers *et al.* 2009] et [Imbert 2008, 34.4-34.5].

5.2.1 Explication et compréhension : être continuiste et en rester aux aspects « logiques » ?

Un des rares points sur lesquels sont d'accord la plupart des théories de l'explication est qu'une bonne explication doit permettre de développer notre compréhension – même si elles ne s'accordent pas sur le contenu de cette notion. Selon Hempel, une explication produit cet effet en montrant qu'on pouvait s'attendre à ce que l'événement explanandum survienne. Selon la théorie causale de l'explication, la

²⁸ Voir de nouveau [Zwirn 2006] pour un traitement précoce de cette question.

compréhension est produite par l'exposé des mécanismes internes du monde et de leur articulation [Salmon 1984, 133], [Salmon 1989, 134], [Salmon 1997, 90]. Selon la théorie unificationniste de Kitcher (voir la section sec6.2), la compréhension est produite en présentant une dérivation qui s'insère dans le stock de dérivations explicatives qui permet d'unifier le mieux notre représentation du monde.

Puisque la plus grande partie de notre science se fait désormais au moyen d'ordinateurs et de simulations numériques, sauf à supposer que l'usage des ordinateurs ait transformé l'activité scientifique en éradiquant une de ses dimensions – à savoir la production de bonnes explications –, on devrait en conclure que nous expliquons *et donc que nous comprenons* de plus en plus au moyen d'ordinateurs et de simulations numériques. Cela est sans doute en partie le cas. Si on définit la compréhension objective comme le fait de *pouvoir objectivement s'attendre* à un événement (pour reprendre la définition de Hempel) sur la base de bonnes justifications scientifiques, ou de connaître objectivement les causes d'un phénomène, en l'occurrence sur la base d'une simulation, alors *de facto*, les simulations développent bien notre compréhension objective. Néanmoins, une telle position « continuiste²⁹ », même si elle est partiellement correcte, semble manquer au moins une des dimensions de la notion de compréhension.

Avant l'avènement de la science computationnelle, les avancées explicatives sont le produit d'esprits humains, éventuellement aidés de supports élémentaires. En conséquence, les explications produites sont nécessairement à taille humaine au sens où il est possible à d'autres esprits compétents et informés de jouir des bénéfices épistémiques légitimes que ces explications recèlent (ce qu'on peut appeler leur *profitabilité épistémique*). La situation est différente dans le cas des simulations. L'activité de production de l'explication, et notamment de connexion de l'explanans et de l'explanandum, est externalisée dans l'ordinateur et ne fait donc plus partie de la cognition propre de l'explicateur. Une des conséquences de cette externalisation est que les simulations sont en général épistémiquement opaques [Humphreys 2004, 147-151], au sens où il n'est pas possible à un esprit humain de parcourir le détail de l'activité inférentielle ou computationnelle relative à l'activité explicative considérée, et encore moins d'avoir à l'esprit tout ce processus inférentiel – puisque les calculs ou les inférences en jeu portent sur des objets que notre esprit ne peut embrasser (par exemple des tableaux de grosse taille) et qu'un nombre extrêmement important de pas de calcul est nécessaire. Il est néanmoins important de signaler que *chaque* opération de calcul logique et

²⁹ Pour une discussion de la question du côté révolutionnaire (ou non) de la philosophie des simulations, voir [Frigg & Reiss 2009], [Humphreys 2009], [Imbert 2017].

mathématique du processus reste dans l'absolu inspectable et vérifiable. En conséquence, la vérifiabilité et la fiabilité du processus producteur de connaissance peuvent être maintenues³⁰. Néanmoins, il n'est pas possible à un esprit humain de parcourir toute la déduction *in extenso*³¹ et encore moins de l'embrasser *uno intuitu*, pour reprendre une expression de Descartes [Descartes1684]. En résumé, l'*opacité épistémique globale* de ce type d'activité est compatible avec la *transparence locale* du processus.

5.2.2 Conséquences de l'externalisation de l'activité explicative dans le cas des simulations

La conséquence de l'opacité épistémique décrite ci-dessus est que les bénéfices épistémiques authentiques et les effets psychologiques³² qui peuvent résulter pour un esprit du fait d'être directement et intégralement le porteur de l'activité de production de l'explication et de maîtriser celle-ci sont potentiellement absents dans le cas des simulations à visée explicative – et les définitions « continuistes » (voir *supra*) de la notion de compréhension, malgré leurs réelles qualités, sont inaptes à caractériser cette différence. De fait, suite à une simulation, l'esprit prend connaissance de ce qu'il doit croire sur la base des raisons qu'il a de croire que la simulation s'est déroulée correctement et que le modèle qu'elle explore peut fournir à propos de l'objet d'étude des informations satisfaisantes. Dans cette perspective, la connaissance explicative, en tant que relation identifiée entre des faits (ou des énoncés) constituant l'explanans et un fait (ou un énoncé), explanandum est toujours atteignable, et de façon justifiée. Néanmoins, puisque l'activité de production et d'authentification de l'explication est externalisée, les aspects liés à l'existence de ces opérations dans l'esprit de l'explicateur disparaissent. C'est notamment le cas des effets psychologiques de l'activité explicative comme l'effet « eureka » et de tout ce qui ressort de la « cognition chaude » (la cognition influencée par les émotions). De ce point de vue, l'externalisation computationnelle de l'activité permet de réduire l'influence de biais éventuels (et d'éviter les erreurs de raisonnement). Néanmoins, elle a sans doute aussi pour conséquence de rendre plus délicat le développement d'une « intuition scientifique » ou d'une aptitude à sentir les résultats [*a feel for results*] et de toute capacité cognitive

³⁰ Par ailleurs, les informaticiens s'arrangent pour garder la possibilité de pouvoir, dans les cas où ils en ont besoin, vérifier (automatiquement) un calcul par ordinateur bit par bit.

³¹ De même, on peut visualiser n'importe quel entier naturel entre 1 et 10^{50} mais on ne peut les parcourir ni les avoir tous à l'esprit.

³² Je ne me prononce pas dans ce qui suit sur le rôle et la valeur épistémiques de ces effets psychologiques dans l'activité scientifique.

permettant de court-circuiter de façon partiellement fiable les procédures longues et fiables (capacités qui sont sans doute proches de ce que Feynman nomme la compréhension, voir *infra*).

Dans tous les cas, la jouissance des bénéfices de l'explication n'est plus immédiate pour l'utilisateur et elle requiert une seconde étude ou analyse de la simulation, dans le but de rendre accessible ou saillant certains aspects de son contenu. Typiquement, quand un argument explicatif n'est plus porté par un esprit humain, en plus de la perte de transparence et de certitude décrite plus haut, le détail de la façon dont ses prémisses se combinent pour mener à la conclusion devient opaque ; le scientifique n'est plus en mesure de développer *par lui-même* des attentes légitimes par rapport à des situations semblables ou à des situations contrefactuelles dans lesquelles on adapterait l'argument explicatif, ni même toujours de reconnaître par lui-même si une situation tombe sous le coup de l'explication ; et les gains épistémiques que seules une vue d'ensemble et une maîtrise de l'argument explicatif peuvent fournir sont perdus.

Un des aspects de cette profitabilité épistémique des explications a été spécifiquement thématiqué sous le nom de *compréhension* par Feynman [Feynman, Leighton *et al.* 1963] et elle est particulièrement pertinente ici pour éclairer en quoi l'activité explicative portée par des simulations peut être insatisfaisante. Feynman, citant Dirac, propose l'analyse suivante de la notion de compréhension :

« Je comprends le sens d'une équation si j'ai un moyen de me représenter les caractéristiques de sa solution sans réellement résoudre cette équation. » En conséquence, si nous avons un moyen de savoir ce qui devrait se passer dans des circonstances données sans avoir à réellement résoudre les équations, alors nous « comprenons » les équations, en tant qu'elles sont appliquées à ces circonstances³³. [Feynman, Leighton *et al.* 1963, vol. 2, 2-1, ma traduction]

Dans cette perspective, la compréhension (également parfois appelée « *intelligibilité* » dans cette tradition³⁴) suppose donc une *capacité*, à savoir celle d'anticiper des caractéristiques, typiquement qualitatives, du comportement d'un système, capacité qu'il doit être possible de mettre en œuvre avec une certaine immédiateté afin d'avoir un certain contrôle mental du comportement étudié. Et naturellement, pour que cette capacité puisse être mise en œuvre, il faut notamment

³³ « What it means really to understand an equation – that is, in more than in a strictly mathematical sense – was described by Dirac. He said : “I understand what an equation means if I have a way of figuring out the characteristics of its solution without actually solving it.” So if we have a way of knowing what should happen in given circumstances without actually solving the equations, then we “understand” the equations, as applied to those circumstances».

³⁴ Par exemple, de Regt & Dieks [De Regt & Dieks 2005] utilisent une telle notion pour caractériser la notion de théorie intelligible.

que son coût en termes de ressources soit suffisamment bas pour qu'elle puisse s'exercer, ce qui, pour les simulations, n'est que rarement le cas. En ce sens, les simulations sont *par elles-mêmes* inintelligibles (sauf si bien sûr on possède en parallèle un détour cognitif permettant de court-circuiter la simulation).

Le même reproche d'inintelligibilité peut être fait à certaines preuves mathématiques qui sont parfois très longues et compliquées et que même les membres de la communauté scientifique correspondante ne peuvent pas toujours suivre (pensons à la preuve du grand théorème de Fermat par Andrew Wiles). La différence est néanmoins que dans de telles preuves, il est souvent possible d'exposer les grandes étapes de la preuve en énonçant les principaux résultats intermédiaires, la façon dont ils sont liés entre eux, et le principe de leur démonstration (de fait Wiles a exposé ses résultats à travers trois conférences progressives). Cela n'est pas forcément possible dans le cadre des simulations. En effet, ces dernières semblent souvent *in-résumables*³⁵, car elles consistent en des calculs compliqués et répétitifs, pour lesquels on ne peut formuler de résumés qui demeurent épistémiquement informatifs et qui permettent peut-être *in fine* de récupérer une certaine maîtrise épistémique (partielle) des résultats.

5.2.3 Le nouveau problème de la simplicité (explicative) inférentielle

De même que nous nous sommes demandé plus haut s'il fallait reprocher aux simulations d'être informationnellement replètes, nous pouvons nous demander ici s'il est légitime de reprocher à ces simulations (et aux études computationnelles) d'être inférentiellement coûteuses (voir *supra*) et de ne pas permettre la compréhension en première personne. Ce qui est ici en cause est donc le caractère potentiellement blâmable (d'un point de vue épistémique) de la lourdeur du processus inférentiel. Deux cas de figures sont à considérer.

1. *Cas de la balourdise inférentielle*. Il peut se faire que la lourdeur inférentielle de l'activité explicative soit en fait évitable. De fait, les scientifiques prouvent souvent une première fois des résultats et, par la suite, trouvent des chemins inférentiels plus simples, directs et élégants pour établir ces mêmes résultats. La lourdeur inférentielle du processus explicatif – typiquement, le recours à une simulation numérique – est alors le signe d'un défaut épistémique de l'explicateur qui échoue à trouver des preuves moins lourdes par manque de clairvoyance sur son

³⁵ Une solution peut parfois être de développer un nouveau cadre conceptuel pour étudier les mêmes résultats. Voir [Humphreys 2008] pour plus de détails.

objet d'étude ou sur les façons de l'explorer. La lourdeur inférentielle est donc ici le signe du caractère perfectible de l'activité explicative et peut être reprochée à ceux qui la mettent en œuvre.

2. *Cas de la lourdeur inférentielle irréductible.* Il peut également se faire que la lourdeur inférentielle soit irréductible et qu'il n'existe pas de chemin plus rapide pour produire la même activité explicative. Dans ce cas, la lourdeur inférentielle n'est pas imputable à l'explicateur, mais serait une caractéristique intrinsèque de l'activité explicative particulière considérée. Une étude plus précise de ce point nécessiterait de discuter rigoureusement comment on caractérise cette lourdeur inférentielle puis d'établir en quoi elle peut être intrinsèque ou irréductible. Je laisse de côté ici ces deux questions difficiles (voir néanmoins [Imbert 2008] et [Zwirn & Delahaye 2013] pour des avancées dans cette direction).

L'existence avérée de cas de balourdise inférentielle, et le fait que la science produite exclusivement par des humains nous ait habitués à nous attendre à ce que les explications fournissent de la compréhension accessible en première personne, légitiment en partie le fait que, d'un point de vue méthodologique, on demande aux scientifiques d'essayer de fournir des explications (et des preuves) qui soient le plus frugales possibles dans leurs utilisations de ressources inférentielles. De même, en présence de preuves ou d'explications inférentiellement lourdes, il est peut-être légitime de *souçonner par défaut* qu'on se trouve en présence de cas de balourdise inférentielle. Ces postures épistémiques peuvent être synthétisées dans la maxime suivante :

Maxime par provision relative à la lourdeur inférentielle (ou computationnelle) explicative : « Plus une explication (resp. un argument), pour être produite (resp. authentifié, présenté, etc.), repose sur un processus inférentiel (ou computationnel) lourd, plus nous sommes en droit de suspecter qu'un processus moins lourd peut jouer le même rôle sans perte épistémique. »

Néanmoins, on voit mal pourquoi toute explication de tout phénomène naturel devrait toujours être assez simple pour pouvoir être cognitivement accessible à cette entité surgie dans l'histoire qu'est un esprit humain non assisté d'un ordinateur. Les études sur la complexité des modèles mathématiques semblent de fait indiquer le contraire. Par exemple, résoudre le modèle d'Ising en trois dimensions est un problème *NP*-complet : cela signifie que la complexité du calcul pour trouver l'état du système à l'équilibre nécessite un temps de calcul croissant de façon exponentielle avec le nombre d'éléments du réseau (ici des spins) et qu'on n'a guère d'espoir que ce problème puisse un jour avoir une solution plus rapide. De même, la simulation d'un gaz

sur un réseau est un problème *P*-complet. Cette fois-ci la croissance du coût en calcul, même en multipliant le nombre de processeurs travaillant en parallèle, est polynomiale, et elle nécessite donc aussi un ordinateur (pour une introduction aux questions de complexité appliquées à la physique, voir [Mertens 2002]). Ces exemples semblent montrer que l'explication la plus courte peut nécessiter une quantité de calcul telle qu'aucun esprit humain ne peut produire, authentifier ou parcourir l'explication correspondante. En d'autres termes, le type de compréhension à laquelle la science nous a longtemps habitués en nous fournissant des démonstrations ou des prédictions analytiques cognitivement accessibles semble ne pas pouvoir être systématiquement atteint – sans pour autant que les acteurs scientifiques impliqués soient épistémiquement blâmables. De ce point de vue, il semble légitime de conclure que l'activité explicative à mesure humaine correspond à un sous-ensemble restreint de ce qu'il est envisageable d'étudier et d'expliquer scientifiquement – même si des activités explicatives à taille humaine fécondes et informatives peuvent encore être rencontrées.

De nouveau, il serait important d'être en mesure de pouvoir identifier ou reconnaître les cas dans lesquels on est dans la situation (i) (lourdeur inférentielle explicative induite, faute épistémique) et ceux dans lesquels on est dans la situation (ii) (lourdeur inférentielle explicative inévitable, absence de faute épistémique) et d'avoir une idée de la fréquence avec laquelle ces cas se rencontrent.

Néanmoins, d'un point de vue épistémologique, déterminer si la lourdeur inférentielle ou computationnelle est évitable ou irréductible, et donc si le processus inférentiel est optimal ou perfectible, semble être difficile dans de nombreux cas (outre le fait que cela pose des questions conceptuelles ardues). Dans les faits, nous ne savons en général guère ce qu'il en est. En un mot, la complexité minimale (inférentielle ou computationnelle) de la tâche explicative est une information en général inaccessible³⁶. De nouveau, il pourrait même se faire que la majorité des propriétés physiques aient de façon inévitable des explications inférentiellement ou computationnellement lourdes, et que nous ayons néanmoins de grandes difficultés à reconnaître les cas correspondants (voir de nouveau la note 26 pour le parallèle avec le cas des nombres aléatoires).

Cela signifie que, eu égard à la question générale que nous posons à propos du caractère blâmable ou non de la lourdeur computationnelle (ou inférentielle) des explications, et donc de leur améliorabilité, la situation est forcément inconfortable. D'un côté, il est légitime, en

³⁶ De nouveau, je n'essaie pas du tout de rentrer dans la difficile question de la potentielle définition de cette complexité « minimale », ni même de discuter quel sens cette notion peut avoir.

présence d'une procédure explicative inférentiellement lourde, d'être toujours soupçonneux sur son caractère explicatif et de supposer, au moins à titre méthodologique, qu'il faut essayer de trouver des explications inférentiellement ou computationnellement plus simples. D'un autre côté, dans de nombreux cas, *même si nous ne savons pas forcément lesquels*, cette lourdeur est sans doute inévitable. Une application stricte de la maxime, qui amènerait à écarter les explications computationnellement coûteuses (ce qui n'est pas fait dans un champ comme la physique) risquerait donc de geler nos progrès explicatifs. La maxime énoncée plus haut justifie donc une posture épistémique mais n'indique pas ce qu'il convient de faire. On peut de nouveau reformuler le problème sous la forme de la question suivante :

Quand est-il légitime de considérer qu'une explication, parce qu'elle repose sur une activité inférentiellement ou computationnellement lourde, est superficielle (et doit être traitée comme telle) ?

Ce problème a les caractéristiques suivantes. i) Ce qui est en jeu n'est pas ici l'accès à la connaissance explicative (voir section 5.1), mais à la profitabilité épistémique d'une explication (voir *supra*). ii) Ce problème se pose de nouveau dans la pratique scientifique : face à une proposition d'explication qui repose sur une simulation, un évaluateur doit-il rejeter dédaigneusement l'article au motif que l'explication est peu profitable épistémiquement et développe peu notre compréhension ? Il s'agit d'une question de norme disciplinaire et, *de facto*, certaines disciplines, comme l'économie, n'acceptent guère les résultats reposant sur des simulations [Lehtinen & Kuorikoski 2007].

Dans tous les cas, l'exposé de la situation permet de nouveau de comprendre pourquoi les simulations à visée explicative sont une pratique qui suscite presque inévitablement le soupçon, la critique, voire le rejet, même si, dans la majorité des cas (lesquels sont hélas difficiles à identifier), il est possible que cela soit de façon erronée. « L'ère du soupçon » à propos de l'optimalité inférentielle ou computationnelle des simulations risque donc de perdurer.

5.3 Par delà les soupçons : l'importance de la suffisance explicative ou *at least, we can !*

D'un point de vue explicatif, la situation n'est néanmoins pas désespérée. L'opacité des explications eu égard à la façon dont elles satisfont le critère de pertinence informationnelle ou de minimalité inférentielle est clairement préjudiciable. Néanmoins, même dans de tels cas, une partie authentique des bénéfices d'une explication (et notamment la possibilité de développer des attentes objectives, voir les sections 3.4 et 5.2.1) est *déjà* accessible. Une explication qui repose sur des prémisses en partie non pertinentes nous apprend néanmoins que le

type d'événement décrit par l'explanandum est réalisé dans les cas où toutes ces prémisses sont vraies et nous pouvons (éventuellement aidés de calculateurs) contrôler en partie l'occurrence du phénomène correspondant. Il se trouve seulement que nous découvrirons peut-être incidemment que l'explanandum est aussi réalisé dans des situations où certaines prémisses (en fait non-pertinentes) sont fausses. De même une explication qui a un coût établi K permet *déjà* à des agents d'avoir une certaine maîtrise épistémique s'ils ont accès à de telles ressources, quitte à s'aider d'un calculateur, faute de mieux – même s'il y aurait en fait parfois moyen de rendre la production de l'explication et ces bénéfices cognitivement accessibles.

De ce point de vue, contrairement à ce que Hempel appelle une esquisse explicative, qui est une explication scientifiquement inachevée, au statut épistémique incertain et dont les bénéfices sont peu assurés et encore douteux, une explication informationnellement ou inférentiellement trop lourde peut être considérée comme *déjà suffisante*, car elle nous apporte *déjà correctement* une partie non altérée des bénéfices explicatifs visés – même si nous ne sommes pas toujours en mesure d'atteindre un état de félicité épistémique totale dans lequel nous avons des explications optimales (c'est-à-dire qui fournissent au coût le plus bas possible toute la connaissance explicative possible relativement à un explanandum) et dans lequel nous savons également que nous possédons de telles explications (l'état dans lequel nous possédons des explications optimales mais ne le savons pas n'étant qu'un demi-état de félicité épistémique puisqu'il peut déboucher potentiellement sur une recherche vaine d'améliorations).

6 Explication, unification et simulations

Je conclus cette analyse en discutant les liens entre explication, unification et simulations. On attend en général d'une bonne explication qu'elle contribue à unifier nos représentations des phénomènes physiques. Ainsi, l'explication par Newton des mouvements des corps réunit sous un même schème explicatif général la science du monde sublunaire et celle du monde supralunaire (lesquelles, dans la physique aristotélicienne, étaient traitées séparément), le mouvement linéaire et le mouvement circulaire, mais également des phénomènes ou des domaines aussi variés que la statique, les marées et, jusqu'à un certain point, via la physique statistique, le mouvement des fluides, le comportement des gaz et l'étude calorimétrique des corps. La question des liens entre unification et explication est néanmoins compliquée, car il existe différents types d'unités ou d'unifications possibles, à des étapes ou des niveaux différents de l'activité explicative, lesquels mettent en œuvre des énoncés généraux aux statuts et aux rôles différents. De plus, des

divergences d'appréciation peuvent provenir des différences de statut qu'on donne à la notion d'unification dans l'analyse de ce qu'est une bonne explication.

J'essaie dans cette section de distinguer certains de ces types d'unités et discute en quoi les simulations peuvent ou non y contribuer. Un enjeu de cette discussion est de montrer que ces unités ne se réduisent pas les unes aux autres et que, selon la perspective qui est prise, les simulations sont plus ou moins un facteur d'unification.

6.1 L'unité par le matériau scientifique explicatif : avec une grande puissance inférentielle vient un grand pouvoir d'unification

De façon traditionnelle, on considère qu'unité et unification sont apportées par les théories ou les descriptions théoriques que proposent les scientifiques. Les phénomènes sont divers et multiples. Les scientifiques, par leurs théories, montrent que des principes, lois ou mécanismes plus fondamentaux permettent d'en rendre compte en identifiant de l'unité dans ce qui est multiple, et ils rendent ainsi manifeste qu'une description complète des phénomènes, pris séparément, est extrêmement redondante. De plus, ils montrent également qu'en supposant l'existence d'entités non-observées (e.g., les atomes, les électrons, les gluons), on peut unifier encore davantage notre représentation scientifique de l'ensemble des phénomènes en réduisant la quantité d'information véhiculée par les énoncés de base et ainsi identifier encore plus de redondance dans notre description du monde.

Une telle unité caractérise avant tout le *matériau scientifique*, dont les scientifiques se servent pour construire des explications. Dans cette perspective, la qualité d'une explication s'évalue donc à un double niveau. i) Elle dépend de critères « locaux » dont la satisfaction met en jeu des propriétés spécifiques de l'explication proposée. ii) Elle dépend de la qualité du matériau explicatif utilisé, dont l'évaluation est globale – le pouvoir d'unification explicative de ce matériau (défini notamment par la taille de l'ensemble des explications satisfaisantes qu'on peut construire à partir de ce matériau) étant une des dimensions possibles de cette évaluation globale. C'est à une séparation des questions de ce type que procèdent Hempel & Oppenheim quand ils discutent des critères que doit satisfaire une explication et réservent une dernière section distincte à la question du pouvoir systématique des théories [Hempel & Oppenheim 1948]³⁷.

³⁷ En même temps, dans le cadre de leur théorie de l'explication, le critère de vérité rend peut-être en partie inutile le recours à un critère d'évaluation du pouvoir systématique des théories utilisées. Il y a ici une tension.

Comme indiqué plus haut, dans l'absolu, les simulations correspondent à un type d'outil flexible et neutre par rapport au type de contenu théorique ou empirique qu'elles permettent de véhiculer et d'explorer [Barberousse & Imbert 2013b]. De ce point de vue, en tant qu'outils qui étendent nos ressources et nos capacités scientifiques, elles permettent d'accroître la quantité d'explications que nous produisons ou que nous pouvons produire à partir des matériaux théoriques dont nous disposons. Elles ont donc une contribution positive à l'activité explicative en permettant de concrétiser le potentiel explicatif de nos représentations théoriques.

6.2 L'unification par des schèmes explicatifs comme but primaire de l'activité explicative (Kitcher)

On peut arriver à des conclusions similaires quand l'unification est considérée cette fois comme le but primaire de l'activité explicative, comme dans la théorie de Kitcher [Kitcher 1989]. La thèse inédite de ce dernier est que les questions de l'unification, de l'explication, et celle de savoir ce que sont les éléments de base de notre représentation du monde – théories, modèles, croyances ou énoncés fondamentaux – doivent se résoudre d'un même coup à travers la recherche de schèmes explicatifs ayant globalement un fort pouvoir unifiant.

Ainsi, selon Kitcher, la science améliore notre compréhension de la nature en nous montrant comment rendre compte des différents phénomènes naturels de la façon la plus économique et la plus unifiée possible, en partant d'un nombre de croyances de base toujours plus réduit et en utilisant autant que possible les mêmes types de dérivation [Kitcher 1989, 423]. Pour lui comme pour Hempel, une explication est un argument – mais pour Kitcher, une explication est une dérivation et non simplement un argument sémantiquement valide (c'est-à-dire un couple prémisses/conclusion). Une dérivation est explicative si elle fait partie de l'ensemble des dérivations qui permettent d'unifier le mieux l'ensemble K de nos croyances sur le monde. Plus un type d'argument a un domaine d'application large, qui nous permet de rendre compte de nombreux phénomènes, plus il a de chances de figurer dans l'ensemble de dérivations explicatives. Par exemple, un type de dérivation qui permet de rendre compte de la position de toutes les planètes et du mouvement des marées a un fort pouvoir unificateur. Kitcher propose de minimiser le nombre de formes d'arguments et non le nombre d'arguments parce qu'en cherchant à unifier nos croyances de la façon la plus économique possible on ne veut pas compter plusieurs fois des arguments certes différents mais en fait très proches. Par exemple, on ne veut pas compter comme deux arguments distincts les dérivations de la trajectoire de Vénus et celle de la Terre, car elles sont très similaires.

Sans entrer plus dans les détails, on peut noter que rien n'exclut que des membres de l'ensemble des schémas d'arguments de l'ensemble qui unifie le mieux nos croyances ne puissent être instanciés par des simulations – bien au contraire. Un modèle computationnel fournit de par sa forme un schéma déductif qui peut s'appliquer rigoureusement à une gamme large de conditions initiales : en cela, il a un pouvoir d'unification potentiellement important. De plus, les simulations permettent l'application de types d'arguments déjà existants à des situations nouvelles et augmentent par là le pouvoir unifiant des schémas d'arguments explicatifs déjà existants. Par exemple, la simulation précise des trajectoires des comètes est une application de ce que Kitcher décrit par ailleurs comme « le schéma newtonien pour fournir des explications des mouvements sur la base des forces sous-jacentes » [Kitcher 1993, 121, 179]. Enfin, les simulations permettent le développement de nouvelles familles de pratiques, de méthodes, de modèles et donc de schèmes explicatifs³⁸ qui permettent de couvrir des champs de phénomènes qui tomberaient autrement hors du domaine de notre science [Barberousse & Imbert 2013a]. En résumé, grâce aux simulations, le pool des schémas d'argument potentiellement utilisables dans notre entreprise d'explication est donc *de facto* plus large, ce qui ne peut nuire au caractère unificatif de l'ensemble.

6.3 Explication, pouvoir d'unification propre et simulations

Je souhaite maintenant souligner que l'activité explicative produit *par elle-même*, et sur la base de critères *locaux*, des généralités qui sont distinctes des lois utilisées dans les explications et qui ont une valeur unificatrice propre, différente de celle apportée par les lois et le matériau scientifique sur lequel reposent ces explications et qui, comme nous venons de le voir, doit permettre d'unifier des activités scientifiques et des types de phénomènes distincts.

Comme nous l'avons vu dans la section 5.1, le critère de pertinence explicative, qui est spécifique aux explications, incite à exfiltrer du matériau scientifique utilisé (et y compris de sa composante nomologique) tous les aspects non pertinents. Prenons l'exemple de la loi des aires ou deuxième loi de Kepler, qui indique que le mouvement de chaque planète est tel que le segment de droite reliant le soleil et une planète balaie des aires égales pendant des durées égales.

³⁸ Kitcher est l'un des premiers à avoir conceptualisé la notion de pratique scientifique en pointant l'importance de cette notion dans la réponse aux questions centrales de la philosophie des sciences [Kitcher 1983, 1993]. Même si ce point est rarement souligné, il y a une grande proximité entre ce qu'il nomme pratique scientifique et ce qu'il décrit ailleurs comme des schèmes explicatifs [*explanatory patterns* ou *schemas*] [Kitcher 1989].

L'explication de Newton utilise le matériau scientifique suivant (voir [Imbert 2011, 2013] pour une analyse philosophique plus détaillée) :

- I : le système peut être considéré comme un système à 2 corps où une seule force centrale s'exerce ;
- L_{a2} : la dynamique du système x est telle qu'une force produit un incrément de quantité de mouvement dans la direction de cette force ;
- L_{a1} : la dynamique du système considéré est telle qu'en l'absence de force extérieure, un corps persévère dans un mouvement rectiligne uniforme.

Sur la base de ces énoncés, on peut montrer que les planètes instancient la loi des aires et on obtient par ailleurs une *généralité explicative* G_E qui dit que si un système x vérifie les énoncés I , L_{a1} et L_{a2} , alors il obéit à la loi des aires. Une telle généralité explicative identifie explicitement une régularité qui unifie une classe C de systèmes partageant des traits et instanciant un même comportement – ce qu'on peut appeler le *domaine d'invariance stricte* identifié par l'explication et correspondant aux cas dans lesquels l'antécédent de la généralité explicative G_E est vrai. Plus on satisfait le critère de pertinence explicative, plus ce domaine est large et plus on se trouve en mesure de répondre à des questions visant à savoir si le comportement expliqué continuerait à être instancié si on modifiait tel ou tel aspect du système – ce qu'on peut nommer des « *would-still* » questions [*would-the-behavior-still-be-instantiated*] [Imbert 2013]. Il s'agit d'une connaissance contrefactuelle propre apportée spécifiquement par l'activité explicative. La terminologie adoptée vise à distinguer cette notion de celle d'invariance, que je qualifie de « large », proposée par Jim Woodward [Woodward 2003], qui correspond à l'ensemble des cas dans lesquels une généralité permet de répondre à une question sur ce qu'il se serait passé si la valeur de telle ou telle variable apparaissant dans la généralité avait été modifiée (c'est-à-dire à une *what-if* ou *what-if-things-had-been-different question*).

Une telle généralité explicative s'obtient, contrairement aux types d'unification décrits plus haut, par une analyse *locale* de l'explication. Elle ne correspond à aucun des énoncés (particuliers ou nomologiques) constituant le matériau explicatif puisqu'elle naît en partie de leur combinaison. Son domaine est également différent de celui des lois sur lesquelles elle repose. Dans le cas présent, L_{a2} est une version appauvrie du principe fondamental de la dynamique (qui vaut pour tous les systèmes classiques) ; G_E vaut donc aussi pour des systèmes non classiques qui n'obéissent pas à certains aspects du principe fondamental de la dynamique. Cela signifie aussi que l'unification ici considérée, qu'on peut nommer *unification explicative propre*, n'est pas simplement une unification des phénomènes rencontrés dans *notre*

monde physique mais met en rapport des classes de systèmes physiques possibles, obéissant ou non aux lois de notre monde. Elle se distingue en cela clairement de l'unification apportée par les théories dont le critère de succès est exclusivement relatif à des propriétés de la représentation des phénomènes *de notre monde*.

Cette unification, même si elle s'obtient localement, reste transversale au sens où des systèmes physiques obéissant à des dynamiques significativement différentes, et potentiellement étudiés par des disciplines et des théories distinctes, peuvent être identifiés comme partageant un même comportement, pourvu qu'ils partagent les traits structurels mentionnés dans la généralité explicative [Imbert 2008b] – et c'est une raison suffisante, même si on veut à tout prix considérer que ce pouvoir d'unification explicatif propre n'est qu'un aspect local du pouvoir d'unification global des théories, pour distinguer ce type d'unification de ceux présentés plus haut, qui sont, eux, intra-disciplinaires.

Qu'en est-il maintenant des simulations relativement à cette vertu d'unification locale et propre aux explications ? Comme nous l'avons vu plus haut, les simulations reposent en général sur des descriptions informationnellement replètes des systèmes physiques, à la fois du point de vue de leur dynamique (même si la dynamique du système peut être simple) et des circonstances particulières auxquelles cette dynamique s'applique (même si ces conditions initiales ou aux limites sont simples, elles restent particulières et restrictives). En conséquent, les généralités explicatives correspondant aux simulations explicatives ont un domaine d'invariance stricte restreint, ce qui signifie que les simulations ont par elles-mêmes un pouvoir d'unification propre limité.

On voit donc ici en quoi les simulations sont plus ou moins satisfaisantes selon les types d'unification qu'on considère puisqu'elles contribuent en même temps activement à l'unification globale associée au matériau scientifique utilisé, permettent d'appliquer davantage des schèmes explicatifs unifiants et, puisqu'on peut faire varier leurs conditions initiales, fournissent des schèmes explicatifs applicables à un large ensemble de situations, ce qui leur confère un domaine d'invariance large important (voir *supra* sur cette notion).

Cette distinction entre ces différents types d'unifications permet par ailleurs de dissiper une bizarrerie apparente du modèle de Hempel. Comme nous l'avons vu dans la discussion de ce qu'il nomme « explications génétiques » (qui peuvent être vues comme couvrant les simulations), Hempel accepte les *narrations* comme potentiellement explicatives, pourvu que leurs étapes soient nomologiquement ordonnées par des lois générales (ayant donc déjà par elles-mêmes un fort pouvoir unificateur). De ce point de vue, l'extrême particularité d'une narration à visée explicative n'est nullement gênante car les lois utilisées dans une telle narration peuvent être les mêmes que celles

qu'on utilise dans une explication non narrative (même si elles sont employées différemment). En revanche, une narration explicative sera associée à une généralité explicative avec un domaine d'invariance stricte très restreint. Il convient donc de nettement distinguer entre la *généralité des lois* utilisées de la *généralité (ou invariance stricte) de la généralité explicative* produite. Toutes choses égales par ailleurs, *quand cela est possible*, il est préférable de produire des explications ayant un pouvoir unificateur propre et produisant des généralités explicatives avec une forte invariance stricte. Néanmoins, comme nous l'avons vu, une explication peut avoir un faible pouvoir d'unification explicative propre et être informationnellement optimale (e.g., quand, pour expliquer en détails un événement particulier, on doit irréductiblement faire appel à beaucoup d'information) ; de plus, une explication qui a un trop faible pouvoir d'unification propre (car elle contient trop d'informations non pertinentes) peut être déjà suffisante, car elle peut déjà nous apporter des connaissances explicatives précises sur des cas dans lesquels l'explanandum survient. En revanche, une explication qui ne s'enracine pas dans des lois bien établies (et leur pouvoir d'unification global) n'est-elle pas suffisante car elle ne nous renseigne avec certitude sur aucun cas dans lequel l'explanandum doit arriver – d'où de nouveau le caractère nécessaire de la présence d'une loi dans l'explanans.

6.4 Les possibles effets du pouvoir : risques de fracture théorique et de développement d'une science en mosaïque

Je souhaite pour finir montrer que l'affirmation d'une contribution en partie positive des simulations aux différents types d'unification décrits ci-dessus, affirmation qui repose sur une analyse du type logique qu'instancient les simulations, reste par ailleurs compatible avec la thèse selon laquelle les simulations, *du fait de leurs usages*, donnent (souvent, la plupart du temps, dans certains types de cas, etc.) lieu à des pratiques scientifiques qui s'insèrent mal dans une science unifiée, voire qui contribuent à la dés-unification de cette dernière parce que, d'un point de vue explicatif, les usages « problématiques » sont fréquents. Deux types de cas peuvent notamment être envisagés.

6.4.1 Domaines couverts par des descriptions théoriques : le risque de la désunification

L'application des théories physiques, même au sein des domaines qu'elles couvrent en principe, rencontre de multiples problèmes, au premier rang desquels se trouvent la complexité des modèles à résoudre, le manque de données expérimentales sur les sujets étudiés, l'incertitude sur certains aspects du fonctionnement du système

particulier étudié, etc. De tels problèmes surgissent notamment dans l'étude des systèmes complexes, dont l'étude se fait en grande partie par des simulations. Ces dernières offrent en effet un environnement et des ressources pour résoudre certains de ces problèmes, en facilitant leur traitement mathématique exact ou approché mais aussi en permettant de modifier et d'enrichir le contenu des représentations scientifiques de départ (e.g., en s'éloignant de ce que disent les théories, en reformulant des hypothèses locales particulières, en re-théorisant un sous-domaine, en ajoutant des modules phénoménologiques aux modèles, en ajoutant des variables *ad hoc*, en calibrant empiriquement le modèle, etc.).

Dans de tels cas, la priorité scientifique est souvent avant tout de résoudre les problèmes particuliers étudiés, quitte à avoir recours à des justifications et des solutions locales. L'ajout de contenus aux représentations scientifiques explorées par les simulations est donc ici d'un type différent de l'ajout dans les théories d'entités non-observables qui permettent de rendre celles-ci plus performantes *globalement*. Ces solutions et leurs justifications *locales* peuvent donc créer une solution de continuité (ou rupture) avec les théories d'arrière-plan utilisées : dans de tels cas, la condition de possibilité du succès local est une désintertion partielle du cadre théorique existant. Ainsi, les simulations peuvent tendre à développer, pour détourner une expression de Hacking à propos des expériences [Hacking 1983-1992, 307], une vie propre [*a life of their own*] à la fois du point de vue de leur contenu et de leur mode de justification.

Les conséquences de cette possible fracture théorique sont multiples du point de vue explicatif. D'abord, à partir du moment où on s'écarte de ce qu'indique littéralement la théorie de base qui couvre l'investigation considérée, cette investigation ne peut plus bénéficier automatiquement du support justificatif fourni par la théorie et elle nécessite des procédures de justification supplémentaires locales. Typiquement, quand on décide de produire des modèles phénoménologiques simplifiés de formation des nuages dans une simulation climatique, en ignorant en partie la physique sous-jacente, et donc qu'on « re-théorise » à une nouvelle échelle ou localement les phénomènes étudiés, les nouvelles représentations obtenues nécessitent d'être justifiées de nouveau et en partie indépendamment. Cette rupture théorique a également pour conséquence que les gains obtenus, qui sont justifiés de façon locale et particulière, sont moins généralisables et exportables à d'autres applications de la théorie. Enfin, l'appartenance à un même schème conceptuel ou cadre théorique d'arrière-plan de ces investigations scientifiques semi-autonomes persiste mais devient plus lâche. *De facto*, l'unité théorique et conceptuelle d'un logiciel de simulation d'écoulements autour de véhicules qui repose sur l'utilisation d'automates cellulaires, d'un modèle computationnel de

simulation générale de l'atmosphère et d'un modèle numérique de simulation du mouvement des fluides dans les poumons n'est nullement obvie. L'unité existante n'est pas logique au sens où les mêmes contenus théoriques et les mêmes concepts seraient exactement appliqués de la même façon. Ces modèles de simulation ont plutôt un *air de famille* car ils sont construits à partir de briques structurantes plus ou moins fondamentales et plus ou moins récurrentes, comme les principes de conservation ou tout autre résultat connu dont on fait en sorte de préserver la validité dans le modèle³⁹.

En fonction de la perspective choisie, laquelle peut varier selon les objectifs contextuels de l'investigation scientifique menée et des positions philosophiques qu'on adopte, on sera enclin à considérer que les inconvénients ci-dessus sont majeurs ou non eu égard à la valeur de l'explication proposée. Si, comme Kitcher, on est sensible à l'importance de *l'unité des schèmes explicatifs utilisés (comme condition du succès explicatif)*, alors on considèrera que de telles activités explicatives sont moyennement satisfaisantes. Si on considère que les garanties théoriques générales ne sont qu'un *moyen de garantir la valeur du matériau explicatif* – mais qu'il peut exister d'autres moyens locaux pour cela, alors on sera plus enclin à considérer que les explications ainsi produites sont potentiellement valables. On peut aussi considérer que l'unification est une valeur souhaitable pour pouvoir éventuellement *généraliser et exporter les bénéfices* d'une enquête explicative locale et donc préférer, toutes choses égales par ailleurs, des explications fortement unifiantes, mais considérer en même temps que le caractère unificatif n'est pas une condition du succès explicatif mais que c'est un bénéfice qui vient par surcroît. Dans un tel cas, on regrettera le manque de caractère unificatif de ces enquêtes tout en accordant que les explications proposées peuvent être satisfaisantes eu égard au phénomène local étudié.

6.4.2 Domaines non couverts par des descriptions théoriques : le risque du localisme et du particularisme

Des remarques similaires peuvent être faites dans le cas des domaines qui ne sont pas déjà couverts par des descriptions théoriques générales consensuelles, comme c'est par exemple souvent le cas en sciences sociales. Les simulations, et notamment les simulations multi-agents, utilisées dans le cadre des *sciences sociales computationnelles*, constituent un environnement puissant pour explorer des domaines

³⁹ En d'autres termes, stricto sensu, on ne peut pas considérer que l'ensemble de ces modèles d'écoulement ou de fluides forme une théorie unifiée au sens où, pour la conception modèle-théorique des modèles, une théorie correspond à un ensemble de modèles.

d'étude sans être guidé par des théories générales, en analysant les effets d'interactions et de mécanismes particuliers dont on a de bonnes raisons (non théoriques et typiquement empiriques) de croire qu'ils sont à l'œuvre dans les systèmes étudiés. Dans de tels cas, même en cas de succès explicatif réel, les gains obtenus dans de tels contextes riches en hypothèses particulières, en difficultés spécifiques et en solutions locales, risquent d'être peu aisément exportables ; de plus, de telles enquêtes complexes et spécifiques peuvent rendre difficile l'identification de régularités ou de mécanismes généraux, voire la production de théories générales, dont l'identification pourrait, dans un second temps, contribuer à garantir la qualité des matériaux utilisés dans les enquêtes explicatives au sein de ces champs. En bref, dans de tels cas, la contrepartie des opportunités d'investigation fournies par les simulations est que l'unification et l'intégration des résultats des enquêtes menées est difficile et n'offre par une perspective attractive. Le risque de parcellisation des recherches dans de tels champs est donc réel.

7 Conclusion

J'ai essayé dans cet article de proposer une analyse critique de différentes dimensions selon lesquelles on peut évaluer la valeur explicative des simulations. Il convient pour finir de faire les remarques et mises en garde suivantes.

1. Il est important de distinguer les conclusions à propos des possibilités explicatives offertes par les simulations en vertu de leurs caractéristiques liées à leur appartenance à une classe d'activités d'un certain type logique des conclusions à propos des usages effectifs des simulations. Comme nous l'avons vu, certains usages « problématiques » d'un point de vue explicatif peuvent être prédominants, notamment si, de par les possibilités qu'elles offrent, les simulations sont spécifiquement utilisées pour les enquêtes difficiles à propos de systèmes complexes ou dans des domaines n'ayant pas de théorie pour guider les enquêtes.
2. Certaines conclusions, positives ou négatives, valent parfois seulement pour les simulations et parfois pour d'autres représentations scientifiques comme les modèles ou les théories. La spécificité des simulations est qu'elles correspondent à des activités scientifiques effectives, alors que les modèles ou les théories ne sont que des représentations statiques ne correspondant par elles-mêmes à aucune activité scientifique. *Stricto sensu*, un modèle ou une théorie n'ont qu'un pouvoir explicatif potentiel. Par eux-mêmes, ils ne montrent pas comment ils expliquent des phénomènes. En effet, leur contenu n'est pas directement accessible, même si nous

pouvons essayer de les utiliser dans des enquêtes ou des activités explicatives à l'occasion desquelles nous explorons leur contenu. En revanche, les simulations, en tant qu'activités explicatives effectives, sont *de facto* marquées par toutes les difficultés qu'on rencontre dans les enquêtes scientifiques réelles, comme celle de devoir réellement accéder à des solutions (exactes ou approchées) de problèmes logico-mathématiques. Analyser le pouvoir explicatif des modèles, c'est adopter une perspective abstraite et tronquée sur l'activité explicative, et donc écarter *ab initio* une grande partie des problèmes rencontrés dans cette activité, et notamment les difficultés épistémologiques qui sont liées à l'exploration et l'usage de ces modèles dans le cadre d'une activité explicative adaptée à nos moyens scientifiques. La différence entre ces deux perspectives est un cas particulier de la différence, soulignée par des auteurs comme William Wimsatt [Wimsatt 2007] ou Paul Humphreys [Humphreys 2004] entre une approche en principe (dans laquelle on ne tient pas compte des contraintes liées à la réalisation effective des activités) et une approche en pratique (où on réintègre ces contraintes en s'attachant à garder à l'esprit ce qu'il est possible de faire). Il faut donc prendre garde à ne pas attribuer aux simulations des spécificités, et notamment des défauts (eu égard à l'activité explicative) qui sont forcément absents des discussions relatives à la valeur explicative d'autres items dont l'étude ne permet d'analyser que de façon très partielle (et donc potentiellement trompeuse) l'activité explicative.

3. Dans tous les cas, il convient d'être sensible au contexte scientifique dans lequel une simulation est menée, lequel oriente en général la discussion sur sa valeur explicative. Ainsi, dans les domaines (e.g., les sciences cognitives) dans lesquels les hypothèses sur le fonctionnement précis de l'objet cible (e.g., le cerveau) ne sont pas consensuelles, la discussion sur la valeur explicative des simulations risque de se réduire pour l'essentiel à la discussion sur la vérité des hypothèses théoriques utilisées. À cela se conjugue le fait que les théories de l'explication sont multiples et que, selon les domaines, elles peuvent sembler plus ou moins adaptées (e.g., la théorie de Salmon des processus causaux, décrits en termes de quantités invariantes, cadre bien avec la physique et ce que peuvent aider à faire des simulations). Au final, un risque important est que les discussions sur le pouvoir explicatif des simulations s'enlisent dans des discussions régionales qui privilégient certaines normes disciplinaires, utilisent une théorie de l'explication *ad hoc* et méconnaissent les aspects généraux du

problème et les riches acquis des discussions en philosophie des sciences générales sur ce sujet – c'est *de facto* ce qui se passe déjà en partie dans les discussions sur l'explication qui se spécialisent dans des domaines comme la physique, l'astronomie, les sciences sociales, la biologie ou les sciences cognitives. J'ai pour ma part essayé dans la mesure du possible de produire des arguments généraux – ce qui reste tout à fait compatible avec la thèse selon laquelle certains aspects de la discussion sont propres aux domaines étudiés.

4. Je souhaite pour finir souligner que, comme nous l'avons vu, les raisons pour lesquelles les simulations semblent problématiques d'un point de vue explicatif sont peut-être, dans de multiples cas, moins logiques qu'épistémologiques. Bien expliquer nécessite en effet de fournir des explications qui satisfont des critères logiques : reposer sur des représentations dont la partie du contenu qui joue un rôle effectif dans l'explication est vraie (condition de vérité) ; ne présenter que des informations explicativement pertinentes (condition de pertinence ou de sobriété informationnelle) ; présenter des explications dont la mise en œuvre est peu coûteuse (condition de sobriété inférentielle). Ces conditions sont compatibles avec le fait que des représentations en partie fausses, informationnellement riches, ou inférentiellement coûteuses, comme le sont souvent les simulations, puissent être de bonnes explications. Par défaut, il n'est pas illégitime d'avoir un soupçon méthodologique envers des explications qui sont en partie fausses, sont informationnellement riches, ou sont inférentiellement coûteuses. Ainsi, pour affirmer catégoriquement qu'une activité explicative est de bonne qualité, il conviendrait d'être en mesure de déterminer au cas par cas dans quelle mesure un écart donné à la vérité, une abondance d'information ou un coût inférentiel élevé sont problématiques et nous privent de bénéfices explicatifs authentiques. Le problème est que nous sommes souvent dans l'incapacité de déterminer ce qu'il en est, c'est-à-dire de *reconnaître* les explications (non) problématiques quand on nous les présente et de savoir si une explication donnée est *trop* fausse, *trop* riche ou *trop* coûteuse – ce qui est pour le coup un problème épistémologique. En définitive, d'un point de vue explicatif, le malheur des simulations est peut-être moins d'être souvent de mauvaises explications que de ne pouvoir être authentifiées comme de bonnes explications (qui satisfont les critères logiques que doit satisfaire une bonne explication) dans les cas où elles le sont. Les autres explications posent assurément les mêmes problèmes quand, à notre insu, elles

présentent des informations non-pertinentes (mais vraies) ou sont informationnellement et inférentiellement sous-optimales (mais sont à taille humaine) ; néanmoins, sans doute à tort, elles ne déclenchent pas une telle suspicion. L'étude du pouvoir explicatif des simulations est donc une invitation à la vigilance épistémique envers les explications qui, parce qu'elles ont une forme plus usuelle, ont des propriétés épistémiques par ailleurs appréciables (e.g., la vérité) ou sont adaptées à nos capacités cognitives, nous satisfont et ne suscitent pas le soupçon – potentiellement à tort.

ACHINSTEIN, Peter [1983], *The Nature of Explanation*, New York: Oxford University Press.

ARNOLD, Eckhart [2013], Experiments and simulations: Do they fuse?, dans *Computer Simulations and the Changing Face of Scientific Experimentation*, édité par J. M. Durán & E. Arnold, Newcastle upon Tyne: Cambridge Scholars Publishing, 46–75.

ARNOLD, Eckhart [2014], What's wrong with social simulations?, *The Monist*, 97(3), 359–377, doi: 10.5840/monist201497323.

ASPERTI, Andrea, GEUVERS, Herman, & NATARAJAN, Raja [2009], Social processes, program verification and all that, *Mathematical Structures in Computer Science*, 19(5), 877–896, doi: 10.1017/S0960129509990041.

BARBEROUSSE, Anouk, FRANCESCHELLI, Sara, & IMBERT, Cyrille [2009], Computer simulations as experiments, *Synthese*, 169(3), 557–574, doi : 10.1007/s11229-008-9430-7.

BARBEROUSSE, Anouk & IMBERT, Cyrille [2013a], Le tournant computationnel et l'innovation théorique, dans *Précis de philosophie de la physique*, édité par S. Le Bihan, Paris : Vuibert, 244–273.

BARBEROUSSE, Anouk & IMBERT, Cyrille [2013b], New mathematics for old physics: The case of lattice fluids, *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3), 231–241, doi: 10.1016/j.shpsb.2013.03.003.

BARENBLATT, Grigory Isaakovich [1996], *Scaling, Self-Similarity, and Intermediate Asymptotics*, Cambridge Texts in Applied Mathematics, Cambridge: Cambridge University Press.

BATTERMAN, Robert W. [2002], *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*, Oxford: Oxford University Press.

BEISBART, Claus [2012], How can computer simulations produce new knowledge?, *European Journal for Philosophy of Science*, 2(3), 395–434, doi: 10.1007/s13194-012-0049-7.

BROMBERGER, Sylvain [1966], Why questions, dans *Mind and Cosmos: Essays in Contemporary Science and Philosophy*, édité par R. Colodny, Pittsburgh: University of Pittsburgh Press, 86–111.

DE MILLO, Richard A., LIPTON, Richard J., & PERLIS, Alan J. [1979], Social processes and proofs of theorems and programs, *Commun. ACM*, 22(5), 271–280, doi: 10.1145/359104.359106.

DE REGT, Henk W. & DIEKS, Dennis [2005], A contextual approach to scientific understanding, *Synthese*, 144(1), 137–170, doi: 10.1007/s11229-005-5000-4.

DESCARTES, René [1684], *Regulae Ad Directionem Ingenii*, t. X, Paris : Vrin, édité par Ch. Adam & P. Tannery, 1996.

DOWE, Phil [1992], Westley Salmon’s process theory of causality and the conserved quantity theory, *Philosophy of Science*, 59(2), 195–216, doi: 10.1086/289662.

DOWE, Phil [1995], Causality and conserved quantities: A reply to Salmon, *Philosophy of Science*, 62(2), 321–333, doi: 10.1086/289859.

DOWE, Phil [2000], *Physical Causation*, Cambridge: Cambridge University Press.

DRAY, William [1957], *Laws and Explanation in History*, Londres: Oxford University Press.

DUHEM, Pierre [1981], *La Théorie physique. Son objet. Sa structure*, Paris : Vrin, première édition, 1906.

DURÁN, Juan M. [2013], The use of the “materiality argument” in the literature for computer simulations, dans *Computer Simulations and the Changing Face of Scientific Experimentation*, édité par J. M. Durán & E. Arnold, Newcastle upon Tyne: Cambridge Scholars Publishing, 76–98.

EDWARDS, Paul N. [2010], *A Vast Machine: Computer models, climate data, and the politics of global warming*, Cambridge, Ma: The MIT Press.

EPSTEIN, Brian & FORBER, Patrick [2013], The perils of tweaking: How to use macrodata to set parameters in complex simulation models, *Synthese*, 190(2), 203–218, doi: 10.1007/s11229-012-0142-7.

FEYNMAN, Richard Phillips, LEIGHTON, Robert B., & SANDS, Matthew Linzee [1963], *The Feynman Lectures on Physics*, Reading: Addison-Wesley.

FRIGG, Roman & HARTMANN, Stephan [2012], Models in science, *Stanford Encyclopedia of Philosophy*.

FRIGG, Roman & REISS, Julian [2009], The philosophy of simulation: Hot new issues or same old stew?, *Synthese*, 169(3), 593–613, doi: 10.1007/s11229-008-9438-z.

GRICE, Paul [1975], Logic and conversation, dans *Syntax and Semantics. 3: Speech Acts*, édité par P. Cole & J. Morgan, New York: Academic Press, 41–58.

HACKING, Ian [1983-1992], Do thought experiments have a life of their own?, dans *PSA*, édité par A. Fine, M. Forbes & K. Okruhlik, East Lansing: The Philosophy of Science Association, t. 2, 302–310.

HARDWIG, John [1991], The role of trust in knowledge, *Journal of Philosophy*, 88(12), 693–708, doi: 10.2307/2027007.

HEMPEL, Carl G. [1942], The function of general laws in history, *The Journal of Philosophy*, 39(2), 35–48, doi: 10.2307/2017635.

HEMPEL, Carl G. [1965], *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: Free Press.

HEMPEL, Carl G. [2001], *The Philosophy of Carl G. Hempel: Studies in Science, Explanation, and Rationality: Studies in Science, Explanation, and Rationality*, Oxford: Oxford University Press, édité par J. H. Fetzer.

HEMPEL, Carl G. & OPPENHEIM, Paul [1948], Studies in the logic of explanation, *Philosophy of Science*, 15(2), 135–175, doi: 10.1086/286983, réimprimé dans Hempel1965.

HITCHCOCK, Christopher Read [1995], Salmon on explanatory relevance, *Philosophy of Science*, 62(2), 304–320, doi: 10.1086/289858.

HUMPHREYS, Paul [2004], *Extending Ourselves. Computational Science, Empiricism, and Scientific Method*, Oxford: Oxford University Press.

HUMPHREYS, Paul [2008], Computational and conceptual emergence, *Philosophy of Science*, 75(5), 584–594, doi: 10.1086/596776.

HUMPHREYS, Paul [2009], The philosophical novelty of computer simulation methods, *Synthese*, 169(3), 615–626, doi: 10.1007/s11229-008-9435-2.

IMBERT, Cyrille [2008], *L'Opacité intrinsèque de la nature : théories connues, phénomènes difficiles à expliquer et limites de la science*, Thèse de doctorat, Université Paris 1, <http://www.theses.fr/2008PA010703>.

IMBERT, Cyrille [2008b], Sciences différentes, explications similaires, régularités transversales, dans *L'Unité des sciences*, édité par Th. Martin, Paris: Vuibert, 27–44, sélection d'articles provenant du 2e congrès de la SPS.

IMBERT, Cyrille [2011], Explication et pertinence : du sel ensorcelé à la loi des aires, *Dialogue*, 50(4), 689–723, doi: 10.1017/S0012217312000170.

IMBERT, Cyrille [2013], Relevance, not invariance, explanatoriness, not manipulability : Discussion of Woodward's views on explanatory relevance, *Philosophy of Science*, 80(5), 625–636, doi: 10.1086/674002.

IMBERT, Cyrille [2017], Computer simulations and computational models in science, dans *Springer Handbook of Model-Based Science*, édité par L. Magnani & T. Bertolotti, Springer, chap. part G « Models and the Computational Turn ».

KITCHER, Philip [1983], *The Nature of Mathematical Knowledge*, Oxford ; New York : Oxford University Press.

KITCHER, Philip [1989], Explanatory unification and the causal structure of the world, dans *Scientific Explanation*, édité par W. C. Salmon & P. Kitcher, Minneapolis : University of Minnesota Press, *Minnesota Studies in the Philosophy of Science*, t. 13, 410–505.

KITCHER, Philip [1993], *The Advancement of Science : Science Without Legend, Objectivity Without Illusions*, New York : Oxford University Press.

KYBURG, Henry E. Jr. [1965], Salmon's paper, *Philosophy of Science*, 32(2), 147–151, doi : 10.1086/288034.

LEHTINEN, Aki & KUORIKOSKI, Jaakko [2007], Computing the perfect model : Why do economists shun simulation ?, *Philosophy of Science*, 74(3), 304–329, doi : 10.1086/522359.

LENHARD, Johannes [2016], Holism, or the erosion of modularity—a methodological challenge for validation, <http://philsci-archive.pitt.edu/12381/>.

MACHAMER, Peter, DARDEN, Lindley, & CRAVER, Carl F. [2000], Thinking about mechanisms, *Philosophy of Science*, 67(1), 1–25, doi : 10.1086/392759.

MCMULLIN, Ernan [1968], What do physical models tell us ?, *Studies in Logic and the Foundations of Mathematics*, 52, 385–396, doi : 10.1016/S0049-237X(08)71206-0.

MERTENS, Stephan [2002], Computational complexity for physicists, *Computing in Science & Engineering*, 4, 31–47, doi : 10.1109/5992.998639.

NAGEL, Kai & SCHRECKENBERG, Michael [1992], A cellular automaton model for freeway traffic, *J. Phys. I France*, 2(12), 2221–2229, doi : 10.1051/jp1 :1992277.

OBERKAMPF, William L. & ROY, Christopher J. [2010], *Verification and Validation in Scientific Computing*, Cambridge : Cambridge University Press.

POPE, Stephen B. [2000], *Turbulent Flows*, Cambridge : Cambridge University Press.

PYLYSHYN, Zenon W. [1984], *Computation and Cognition*, Cambridge, Ma : MIT Press, 2^e éd.

RAILTON, Peter [1980], *Explaining Explanation : A realist account of scientific explanation and understanding*, Thèse de doctorat, Princeton University.

RAILTON, Peter [1981], Probability, explanation, and information, *Synthese*, 48(2), 233–256, doi : 10.1007/BF01063889.

[REISS, Julian [2011], A plea for (good) simulations : Nudging economics toward an experimental science, *Simulation & Gaming*, 42(2), 243–264, doi : 10.1177/1046878110393941.

ROHRLICH, Fritz [1990], Computer simulation in the physical sciences, *PSA : Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Two : Symposia and Invited Papers, 507–518.

RUPHY, Stéphanie [2011], Limits to modeling : Balancing ambition and outcome in astrophysics and cosmology, *Simulation & Gaming*, 42(2), 177–194, doi : 10.1177/1046878108319640.

SALMON, Wesley C. [1984], *Scientific Explanation and the Causal Structure of the World*, Princeton : Princeton University Press.

SALMON, Wesley C. [1989], Four decades of scientific explanation, dans *Scientific Explanation*, édité par W. C. Salmon & P. Kitcher, Minneapolis : University of Minnesota Press, *Minnesota Studies in the Philosophy of Science*, t. 13, 3–219, doi : 10.2307/j.ctt5vkdm7.5.

SALMON, Wesley C. [1994], Causality without counterfactuals, *Philosophy of Science*, 61(2), 297–312, doi : 10.1086/289801.

SALMON, Wesley C. [1997], Causality and explanation : A reply to two critiques, *Philosophy of Science*, 64(3), 461–477, doi : 10.1086/392561.

SCHURZ, Gerhard [1991], Relevant deduction, *Erkenntnis*, 35(1), 391–437, doi : 10.1007/BF00388295.

SCRIVEN, Michael [1962], Explanations, predictions, and laws, dans *Scientific Explanation, Space, and Time*, édité par H. Feigl & G. Maxwell, Minneapolis : University of Minnesota Press, *Minnesota Studies in the Philosophy of Science*, t. 3, 170–203.

STAUFFER, Dietrich [1985], *Introduction to Percolation Theory*, Londres : Taylor & Francis.

STREVEN, Michael [2008], *Depth : An Account of Scientific Explanation*, Cambridge, Ma : Harvard University Press.

TABERY, James G. [2004], Synthesizing activities and interactions in the concept of a mechanism, *Philosophy of Science*, 71(1), 1–15, doi : 10.1086/381409.

VAN FRAASSEN, Bastiaan [1980], *The Scientific Image*, Oxford : Oxford University Press.

VAN FRAASSEN, Bastiaan [1994], *Lois et Symétrie*, Paris : Vrin.

WIMSATT, William C. [2007], *Re-Engineering Philosophy for Limited Beings : Piecewise approximations to reality*, Cambridge, Ma : Harvard University Press.

WOODWARD, James F. [2003], *Making Things Happen : A theory of causal explanation*, New York : Oxford University Press.

ZWIRN, Hervé [2006], *Les Systèmes complexes*, Paris : Odile Jacob.

ZWIRN, Hervé & DELAHAYE, Jean-Paul [2013], Unpredictability and computational irreducibility, dans *Irreducibility and Computational Equivalence : 10 Years After Wolfram's A New Kind of Science*, édité par H. Zenil, Berlin ; Heidelberg : Springer, *Emergence, Complexity and Computation*, t. 2, 273–295, 10.1007/978-3-642-35482-3₁₉.