

Le libre arbitre

Perspectives contemporaines

Jean-Baptiste Guillon (dir.)

DOI : 10.4000/books.cdf.4937

Éditeur : Collège de France

Lieu d'édition : Paris

Année d'édition : 2017

Date de mise en ligne : 22 décembre 2017

Collection : Philosophie de la connaissance

ISBN électronique : 9782722604797

<http://books.openedition.org>

Référence électronique

GUILLON, Jean-Baptiste (dir.). *Le libre arbitre : Perspectives contemporaines*. Nouvelle édition [en ligne].

Paris : Collège de France, 2017 (généré le 25 décembre 2017). Disponible sur Internet : <<http://books.openedition.org/cdf/4937>>. ISBN : 9782722604797. DOI : 10.4000/books.cdf.4937.

Ce document a été généré automatiquement le 25 décembre 2017.

© Collège de France, 2017

Conditions d'utilisation :

<http://www.openedition.org/6540>

Dualisme épistémique, croyances sellarsiennes et compatibilisme

Jean-Baptiste Rauzy

1. Introduction

- ¹ Dans cet article, je soutiens que l'analyse du compatibilisme donne l'occasion de défendre par surcroît une forme de dualisme épistémique. Il s'agit d'une thèse sellarsienne que je propose de reprendre en la modifiant. Elle *porte à la fois* sur le compatibilisme et sur ce « dualisme épistémique » que j'entends préciser ; elle a donc une double *aboutness* et c'est à mes yeux un trait positif. Mais c'est sans doute aussi ce qui a valu à cette partie de l'héritage sellarsien, qui constituait avant l'heure une métaphysique intégrée¹, de connaître un destin moins enviable que, par exemple, la critique du mythe du donné ou les bases des conceptions inférentielles de la signification. Commençons par en donner les grands traits.
- ² Le compatibilisme est le nom que nous donnons à une position dans une querelle philosophique vénérable : la position selon laquelle le libre arbitre est compatible avec le déterminisme. Dans l'énoncé de cette compatibilité, « le libre arbitre » désigne une capacité de l'agent, alors que « le déterminisme » est le nom d'une autre thèse philosophique : la thèse selon laquelle les faits passés déterminent les faits futurs. Agent, capacité, détermination : comme tous les concepts utilisés dans la querelle peuvent être entendus de multiples manières, l'examen du compatibilisme passe par un effort de précision et de distinctions : de quel genre de capacité est-il question ? Une capacité de contrôle ? Un accès à des possibilités ouvertes ? Quel genre de détermination est à l'œuvre dans le déterminisme ? En répondant à ces questions, en rendant ce qu'on nomme « compatibilisme » plus explicite, on produit *a priori* une famille de thèses par la simple déclinaison du sens des termes. Le compatibilisme est donc un *compendium loquendi* d'un genre spécial, à la fois l'abréviation d'une thèse qu'on aimerait discuter et l'arborescence à laquelle donne lieu l'analyse des concepts qui y sont impliqués, sur le versant du libre arbitre comme sur celui du déterminisme.

- 3 Sur le premier versant, Sellars était intégralement préfrankfurtien. Il pensait que nous sommes en général placés devant des possibilités alternatives lorsque nous agissons – c'est le *fait* des possibilités alternatives – et qu'une action libre compte, parmi ses conditions nécessaires, l'existence de ces possibilités alternatives – c'est le *principe* des possibilités alternatives. Quoique nous pensions de ce fait et de ce principe, l'important était pour lui de constater qu'ils font l'un et l'autre partie de ce qu'il appelait « l'image manifeste de l'homme dans le monde ». Si, à présent, nous demandons : « pour un événement quelconque, existe-t-il un événement ou un complexe d'événements t - δ tel que, si t - δ advient, alors nécessairement t advient ? », nous ne sommes plus dans l'image manifeste ou plus seulement, car nous attendons une réponse qui repose sur des évidences fournies par la science. Pour cette raison, la question de compatibilité qui donne son nom au « compatibilisme » renferme une difficulté de principe. Elle implique qu'on soit parvenu au moins sur ce point particulier à construire des contenus de croyance dont la justification se prolonge dans l'une et l'autre des deux images concurrentes. Impossible donc de prendre position dans la métaphysique du libre arbitre sans avoir préalablement distingué, et en partie fusionné, ces deux images du monde, l'image manifeste et l'image scientifique.
- 4 Sellars proposait donc un pas de côté. Mais c'était un grand pas. « Vous m'interrogez sur le libre arbitre et le déterminisme ? Je vais vous répondre sur tout autre chose. Et puis je vais vous montrer que la réponse que vous attendez sur le libre arbitre porte en réalité sur ce tout autre chose. J'espère – à mots couverts – que le gain que je vais vous faire miroiter du côté du libre arbitre me donnera argument pour vous inviter à me suivre dans le tout autre chose, où j'ai une position dont j'admets volontiers qu'elle est très singulière et qu'elle mérite examen. » Difficile de ne pas reconnaître que c'était un coup risqué, surtout dans un milieu très scolastique où l'on aimait – et où l'on aime encore – ramener la complexité à une version assignable du *pro* et *contra*.
- 5 J'appelle « dualisme épistémique » les formes de dualisme dont la distinction sellarsienne des deux images du monde est une instance. La distinction a été présentée en 1960 devant un public d'historiens des sciences ; deux exposés, publiés deux ans plus tard sous le titre fameux de « Philosophy and the Scientific Image of Man² ». Je n'espère pas convertir qui que ce soit à la lettre de la distinction sellarsienne qui est, comme on va voir, traversée de convictions diverses et contestables. Je soutiens néanmoins qu'on peut procéder par généralisation et en éviter les aspects contestables, tout en conservant les bases de la distinction, à savoir la thèse antiréductionniste selon laquelle (i) le savoir obtenu dans le cadre de l'enquête scientifique n'est pas de même nature que notre savoir ordinaire et n'est pas dérivable à partir de lui. Ainsi conçu, ce dont la distinction sellarsienne est une instance est acceptable et permet de mettre au jour un trait important de certaines de nos croyances : (ii) il existe des croyances qui sont déférentielles par soi ou *nécessairement déférentielles* ; ce sont les croyances qui empruntent leur contenu à l'enquête scientifique mais dont l'*aboutness* n'en relève pas. Tel que je l'interprète, Sellars a fait voir que ce trait est particulièrement saillant dans le débat sur le compatibilisme. En reprenant la distinction des deux images de l'homme dans le monde à l'occasion du débat sur le compatibilisme, on est conduit à insister au moins autant sur la relation des deux images que sur leur distinction et à envisager la fusion au niveau des croyances elles-mêmes. C'est dans ce but, fusionner les images au niveau des croyances, que j'introduis des croyances bi-aspectuelles ou « croyances sellarsiennes ».

- 6 Admettons que Sellars ait eu raison d'attirer l'attention sur les deux images du monde. Il y a, disons, deux schèmes conceptuels remarquables, plus robustes, plus permanents que les autres, dont on peut donner quelques traits fondamentaux sans pour autant prétendre en avoir analysé tous les recoins. Qu'allons-nous en faire ? La première manière d'accueillir la nouvelle, la voie conservatrice ou moniste, consiste à remarquer que, supposant qu'on soit parvenu à identifier convenablement les deux images, la question de savoir quel rôle épistémique leur revient relève d'une enquête empirique. Et même si, comme le demandent les sellarsiens, nous devons *supposer* que les deux images ont un rôle épistémique central, le conservateur répond que nous sommes largement équipés dans l'épistémologie et la philosophie de l'esprit pour faire face à une situation de ce genre. En particulier, si certains contenus semblent mobiliser les deux images – une croyance portant sur un courant électrique domestique, sur la maladie d'un proche dont on nous fait un tableau savant accessible, sur la maladie vécue en première personne – nous saurons mettre en œuvre les différences entre croire et accepter³, savoir et savoir par déférence, croire et tenir pour vrai⁴, croire au premier et au deuxième ordre. Nous avons tout ce qu'il faut pour éviter de bouleverser le paysage, de sorte que la demande sellarsienne sera, non seulement incertaine, mais surtout ramenée finalement à pas grand-chose : rien de plus qu'une position parmi d'autres, dans l'ensemble des positions que les concepts sous-jacents – la déférence, l'acceptance, etc. – nous ont appris à distinguer. Au contraire, la deuxième manière de voir les choses consiste à prendre au sérieux la dualité des deux images, comme un trait fondamental de notre conception des croyances. C'est la voie dualiste. Elle consiste à faire voir que nos croyances ont, pour ainsi dire, deux jambes, l'une dans chaque image, et qu'en ce sens la distinction des deux images du monde ne décrit pas deux schèmes conceptuels concurrents mais deux dimensions de la croyance. C'est cette voie sur laquelle je me propose de faire quelques pas à l'occasion de l'analyse du compatibilisme.
- 7 Dans la première partie de l'article je rappelle les grandes lignes de PSIM et j'interprète la thèse de Sellars par opposition aux thèses relatives à la réduction interthéorique (section 3). Dans une deuxième partie (sections 4 & 5), je donne les grandes lignes du dualisme sellarsien et en tire des conséquences quant à la nature des croyances. Enfin, en guise de conclusion (section 6) je montre que le dualisme sellarsien est éclairant dans l'explication du phénomène de sensibilité à la description décrit et discuté, à propos du compatibilisme, depuis Nichols et Knobe (2007).

2. Les deux images de l'homme dans le monde

- 8 Assez tôt, déjà dans sa période constructionniste, Carnap avait défendu l'idée selon laquelle la science présente un fort caractère d'unité, en dépit des aspects dispersés et très divers sous lesquels elle se présente à nous. Il revenait à la philosophie, toujours selon Carnap, de faire voir cette unité en dressant une espèce de carte de tout ce que nous savons et en montrant comment on peut se déplacer dans cette géographie du savoir par le simple jeu des dérivations conceptuelles⁵. Sellars prétend au contraire que, lorsqu'il s'essaie à l'exercice carnapien de reconstruction du savoir et pour peu que l'être humain soit également considéré comme un objet de savoir⁶, le philosophe rencontre non pas une image complexe et multidimensionnelle, mais deux images, qui sont l'une comme l'autre complètes ou qui au moins se donnent comme telles⁷. Il introduit donc le terme « d'image » d'une manière assez vague, en se gardant d'en donner la définition explicite,

mais dans l'intention de se démarquer de l'héritage carnapien. Il y a deux images, nous dit en substance Sellars ; et si vous ne savez pas exactement ce que c'est qu'une image, j'espère bien que vous le saurez assez quand vous aurez compris qu'il y en a deux.

- 9 Quels sont les principaux traits des deux images du monde ? 1. Sellars insiste sur le fait que ce sont des idéalizations. Il ne prétend pas que les images existent comme telles dans l'histoire culturelle, pas plus qu'existent un mouvement sans frottement ou un gaz parfait : elles décrivent des règles dans notre manière de produire du savoir. Pour un échantillon du stock épistémique de tel sujet ou de telle communauté, on dira qu'il appartient *plutôt* à une image qu'à l'autre et on évitera un usage historiographique littéral de la distinction, d'autant plus que les deux images ne sont pas des entités statiques mais ont une histoire qui entretient une relation privilégiée à l'histoire de la philosophie⁸. 2. Leur différence n'est pas réductible à une différence dans les points de vue, par exemple à un point de vue « interne » pour l'une et un point de vue « externe » pour l'autre ; ou à une différence dans les bases ou fondements, par exemple une base phénoméniste pour l'une et une base physicaliste pour l'autre. Il est vrai que Sellars insiste sur l'usage des entités inobservables dans l'explication scientifique, mais c'est surtout pour faire voir que la différence des deux images renvoie à une asymétrie dans la manière de catégoriser et d'expliquer. L'image manifeste s'appuie plutôt sur la manière de catégoriser. Comme le monde décrit par Strawson, elle comporte de manière essentielle parmi ses catégories fondamentales le concept de personne et ses dérivés :

Roughly, the manifest image corresponds to the world as conceived by P.F. Strawson – *roughly* it is the world as we know it to be in ordinary experience, supplemented by such inductive procedures as remain within the framework. The manifest image is, in particular, a framework in which the distinctive features of persons are conceptually irreducible to features of nonpersons, e.g. animals and merely material things⁹.

- 10 L'image scientifique s'appuie de son côté sur une manière d'expliquer : l'explication y a recours de manière essentielle à des entités inobservables, des particules, des champs, des relations biochimiques, selon une ontologie ouverte qui n'est limitée par aucun schéma catégoriel. J'admets qu'il n'est pas facile de comprendre *une* différence – celle des deux images – fondée sur deux *discrimen* – la manière de catégoriser et la manière d'expliquer. Pour éviter d'entrer dans une longue exégèse sur ce point, je propose de m'en tenir à une approximation. Dans l'image manifeste, on se pose plutôt la question « qu'est-ce qui fonde quoi ? » et on utilise massivement le concept de personne et ses dérivés dans nos réponses. Dans l'image scientifique on demande : « qu'est-ce qui explique quoi ? » et l'on se donne une grande liberté d'invention catégoriale dans nos réponses¹⁰.

- 11 3. Cette asymétrie des deux images est encore accentuée lorsqu'il s'agit d'y placer l'être humain lui-même. Sellars répète que l'image manifeste est ce par quoi l'être humain se pense lui-même dans le monde :

Man is that being which conceives of itself in terms of the manifest image. To the extent that the manifest does not survive in the synoptic view, to that extent man himself would not survive¹¹.

- 12 Pour cette raison, l'image manifeste possède une forte rémanence. Ses modifications impliquent non seulement une révision des croyances mais aussi une modification dans l'autoportrait que l'être humain dresse de lui-même. Et l'autoportrait agit en retour sur la manière de se comporter et de penser. Le fait, par exemple, de se concevoir comme un être possédant le libre arbitre est une représentation de soi qui engendre une manière d'être. Autre exemple : savoir que nos actions satisfont des intentions collectives nous

conduit à concevoir l'activité humaine comme un ensemble de coopérations et l'activité conceptuelle elle-même comme une forme de jeu représentationnel¹² dans lequel la communication et la relation aux autres occupent une place centrale¹³.

- 13 On peut bien entendu objecter que cette fonction d'auto-représentation ne devrait pas être réservée à l'image manifeste. Quel sens donner à une thèse de ce genre s'il est bien entendu qu'elle *ne doit pas* consister à dénier à la science la capacité à porter sur les interactions sociales ou sur l'esprit humain ? La réponse, telle que je la comprends, passe encore par l'attitude à l'égard des catégories. Les deux images ont l'une comme l'autre la fonction classique d'un cadre conceptuel autonome. On ne peut pas aller chercher un concept forgé dans un cadre pour le faire jouer dans un autre sans en modifier la teneur¹⁴. Sellars ajoute :

The categories of theoretical physics are not essences distilled from the framework of perceptual experience, yet, if the human mind can conceive of *new* categories, it can also refine the old; and it is just as important not to over-estimate the role of creativity in the development of the framework in terms of which you and I experience the world, as it is not to under-estimate its role in the scientific enterprise¹⁵.

- 14 Dans l'image manifeste, on raffine, modifie, dérive des catégories à l'œuvre primitivement dans l'expérience perceptuelle, alors que l'image scientifique laisse libre court à l'invention catégoriale. Si donc l'être humain se conçoit lui-même dans les termes de la première, c'est que « se concevoir soi-même » est plutôt une manière de se *reconnaître*, ou d'admettre les limites que l'expérience, même indéfiniment travaillée, impose à l'autoportrait que la philosophie en dresse continûment. L'autoportrait dont il est question est kantien dans son inspiration profonde : c'est l'expérience, d'abord perceptuelle, puis plus généralement subjective, qui donne le cadre de ce que nous sommes à nos propres yeux¹⁶.

3. Le dualisme épistémique et la réduction

- 15 Les deux images du monde sont deux cadres conceptuels distincts à propos desquels il est permis de se demander s'ils ont entre eux des relations ou s'ils doivent être parfaitement imperméables l'un à l'autre. Le monde de l'image manifeste est le corrélat de nos actions. Le monde de l'autre image est un cadre où nous projetons les bases ontologiques de nos explications scientifiques. Il va de soi qu'une position imperméabiliste est ici parfaitement intenable, ne serait-ce que parce que la plupart de nos actions sophistiquées sont élaborées sur la base de considérations conséquentielles dans lesquelles nous intercalons des fragments du savoir scientifique. À cet égard, Sellars n'écarte pas l'hypothèse d'une traductibilité des concepts d'une image dans les concepts de l'autre : exactement, écrit-il, comme on devrait pouvoir en principe reconstruire les concepts de la biochimie dans ceux de la physique subatomique, on devrait pouvoir aussi bien reconduire sans perte les concepts de l'image manifeste dans ceux de l'image scientifique. Et même à propos de la catégorie la personne, qui est au centre de l'image manifeste, Sellars concède qu'il n'est pas impossible qu'elle ait une contrepartie dans les concepts fondamentaux de l'image scientifique¹⁷. La fin de PSIM propose, sur cette relation des deux images, une argumentation double : (i) un argument tiré de la complexité et (ii) l'argument de la normativité. C'est l'argument (i) qui est établi à partir de l'exemple du compatibilisme. Les agents responsables sont placés devant de véritables options et font

ce que, en bien des circonstances, ils auraient pu ne pas faire. Or, on peut éventuellement soutenir qu'une assertion comme « elle aurait pu agir autrement » pourrait être reformulée dans le *framework* de l'image scientifique, mais le niveau de complexité de la proposition exprimée serait alors *beaucoup plus grand* que celui de la proposition à laquelle nous sommes accoutumés, beaucoup plus grand également que le niveau de complexité des propositions qui semblent du même ordre lorsque nous pensons dans l'image scientifique, par exemple lorsque nous pensons que le nombre atomique de l'Hélium est 2. Le passage d'une image dans l'autre, en supposant qu'il soit possible sans perte, implique au moins ce changement dans la complexité dont il convient de ne pas sous-évaluer la portée. L'argument (ii) est plus connu et est plus décisif quant à l'irréductibilité des deux images. Lorsque nous disons d'une personne qu'elle désirait faire A, bien qu'elle eût le devoir de faire B, mais qu'elle a été finalement contrainte de faire C, nous ne décrivons ni la personne en question ni un de ses états, pas davantage une classe de ses états : nous les situons dans l'espace logique des raisons¹⁸.

16 C'est ici que lecteur d'aujourd'hui est le plus susceptible d'éprouver une forme de perplexité. Les exemples proposés par Sellars font partie de ceux qui sont également discutés en philosophie des sciences, autour du problème de la réduction interthéorique¹⁹, où ils font l'objet d'un traitement précis qui est incontestablement plus satisfaisant. On est en droit de se demander si la référence aux deux images ne nous entraîne pas dans une version édulcorée de la réduction. Car une théorie, qu'elle soit conçue comme un ensemble d'énoncés ou comme un ensemble de modèles, est un objet bien déterminé. Nous pouvons répondre avec un haut degré de précision à la question de savoir si la mécanique statistique réduit la théorie thermodynamique des gaz – s'il existe une variété précise de réduction qui se trouve exemplifiée par cette relation interthéorique particulière – parce que nous savons formuler des énoncés fonctionnels dont les valeurs d'entrée sont des concepts de la théorie réductrice (par exemple l'énergie cinétique moyenne distribuée sur les molécules d'un gaz enfermé dans un contenant) et les valeurs de sortie des concepts de la théorie réduite – les conditions de température et de pression de ce gaz. En analysant les conditions exactes de ces fonctions-ponts, on peut savoir si la réduction est possible, si elle est *a priori*, si elle implique une forme de correction ou d'approximation, etc.²⁰. Par comparaison, les deux images sellarsiennes conservent quelque chose d'incertain. Et l'argument de la normativité est tellement tranchant qu'il donne l'impression de clore trop vite un débat à peine entamé. Pourquoi ne pas poser le problème du compatibilisme dans les termes de la réduction ?

17 Je fais l'hypothèse que la distinction des deux images est pertinente *dans l'espace doxastique* et qu'elle ne porte pas directement sur des théories. S'il est vrai qu'une théorie, lorsqu'elle est acceptée ou tenue pour la meilleure, donne lieu à un ensemble de croyances, à l'inverse une croyance n'a pas nécessairement le format qu'impose l'acceptation d'une théorie, *a fortiori* d'une théorie scientifique. On peut supposer que les croyances prises dans l'image scientifique ont nécessairement un cadre théorique. Mais l'image scientifique elle-même n'est ni une théorie ni une classe de théories. Plutôt : l'ensemble de ce qu'on croit – ou de ce qu'on est disposé à croire – quand ces théories sont prises comme base. C'est la raison pour laquelle les images sellarsiennes sont des idéaux-types et non des catégories psychologiques. Nos croyances ne sont pas distribuées dans ces deux images comme dans deux boîtes dont la frontière serait partout parfaitement assignable.

18 Considérons l'exemple du déterminisme. La question « le monde est-il déterministe ? » s'adresse à la science. (Ceux qui pensent différemment considèrent que nos connaissances relatives à l'ameublement du monde ne viennent pas de la science ou ne lui doivent rien). Néanmoins, on ne peut pas entreprendre de répondre à cette question d'un point de vue scientifique sans avoir apporté quelques clarifications générales, par exemple en distinguant entre l'aspect ontologique – le déterminisme proprement dit – et l'aspect épistémologique – la prédictibilité ; entre le déterminisme intégral – un monde est indéterministe en ce sens si une seule de ses séquences l'est – et un déterminisme global – un monde déterministe en ce sens peut abriter des séquences indéterministes ; ou encore entre un déterminisme des processus et un déterminisme des résultats. Supposons que tous ces points aient été distingués et clarifiés et admettons, c'est évidemment une simplification illégitime, que la réponse à la question ainsi clarifiée puisse être apportée par les seules sciences physiques : même dans ce cas, nous ne pouvons pas espérer *une* réponse mais devons-nous attendre à une multiplicité d'appréciations correspondant à plusieurs examens, étagés selon les chapitres de la science physique elle-même²¹. Et rien n'assure que nous pouvons disposer d'une réponse univoque, même dans les limites d'un seul de ces chapitres²². Devant une situation de ce genre, nous avons des raisons de penser que c'est la *question elle-même* qui présente un profil particulier. La question porte sur la valeur de vérité d'une proposition :

O Est-il vrai que Φ ?

19 En l'occurrence : « est-il vrai que le monde est déterministe ? » Mais nous constatons que toute réponse potentielle acceptable à cette question passe par une substitution du type :

RO_{dial} . Il est vrai que Φ au sens de Φ^* et non au sens de Φ^{**}

où Φ^* et Φ^{**} sont des déterminations incompatibles de Φ , autrement dit des déterminés du même déterminable dialectique. Dans notre exemple, le déterminable est le contenu [que le monde est déterministe] et les déterminés les contenus dérivés obtenus en précisant : (i) l'acception des prédicats ; par exemple : [que le monde est déterministe ontologique intégral portant sur les processus] ; et (ii) le cadre théorique ; par exemple : [que le monde est déterministe selon la relativité restreinte].

20 Bien entendu, le fait que les réponses acceptables à une question aient une forme dialectique de ce genre n'est pas exceptionnel. Et ça l'est encore moins si la question est de nature philosophique ! Ce qui est remarquable ici, c'est plutôt l'enchâssement de contenus théoriques dans un tel jeu de questions et de réponses. Pour qu'une instance de RO_{dial} soit une réponse acceptable à la question O il faut au moins que :

- son contenu propositionnel Φ^* est un déterminable de Φ et
- il existe une théorie acceptée qui vaut comme vérifacteur de Φ^*
- [...].

où [...] désigne les conditions sémantiques générales²³ selon lesquelles une assertion est une réponse acceptable à une phrase interrogative de la forme « Est-il vrai que Φ ? ». Mais le déterminable Φ n'est pas un déterminable de la théorie qui le rend vrai à travers Φ^* . Il faut se placer d'un point de vue extérieur pour savoir que Φ^* apporte une réponse à la question « est-ce que Φ » ? C'est en ce sens qu'on peut dire que la question *porte* sur des théories mais ne leur *appartient* pas, et c'est en ce sens que j'ai dit en commençant que la question du déterminisme porte sur l'image scientifique mais est énoncée depuis l'image manifeste. Elle met en œuvre, pour ainsi dire, l'articulation des deux images.

21 Un autre trait remarquable des images sellarsiennes peut être établi à l'aide du même exemple. Imaginons que, sur la question qui nous agite, nous disposions des réponses

fournies par un oracle véridique analogue à celui qu'a proposé Burgess pour la métaphysique des abstraits mathématiques²⁴. De même que l'oracle de Burgess nous apprend que les nombres n'existent pas – à nous de concevoir ce qui s'ensuit, de même ici l'oracle nous révèle que, disons, le compatibilisme est faux et le déterminisme est vrai. Apprenant la nouvelle, nous serions en droit de demander des précisions – Quel compatibilisme ? Pour quel déterminisme ? – de dérouler devant l'oracle l'arborescence des concepts impliqués, afin qu'il précise dialectiquement le cadre conceptuel de sa réponse. Mais ce n'est pas tout. Même si celui-ci se prêtait au jeu et acceptait de décliner sa réponse avec autant de nuances, rien de nous garantit que notre stock doxastique s'en trouverait entièrement purgé des croyances contraires auxquelles nous sommes accoutumés ; que nous cesserions, par exemple, de tenir les criminels pour responsables de leurs crimes, d'admirer les courageux et d'honorer la mémoire de ceux qui se sont sacrifiés, y compris dans l'hypothèse où nous serions convaincus que l'oracle dit vrai. Nous verserions sans aucun doute dans l'inconsistance. Dans une inconsistance préférable, au sens où un agent S informée préférera dire :

ID Je sais que Φ et je crois que Φ'

où Φ' est incompatible avec Φ ou avec ce qui suit de Φ , plutôt que d'opérer les coûteux ajustements imposés par la situation informationnelle. Un cas analogue d'inconsistance doxastique est établi à propos du compatibilisme par l'enquête de Knobe & Nichols sur laquelle je reviens plus bas. Mais l'inconsistance doxastique est également sous-jacente dans de nombreux contextes d'attribution de croyance. Supposons par exemple que notre agent S soit une soignante dans une institution judico-psychiatrique, que l'on puisse recueillir par l'observation les détails de sa conduite et que ceux-ci fassent apparaître les traits suivants : les gestes et les paroles de S, lorsqu'elle est au contact d'un criminel cruel, ne diffèrent aucunement des gestes et mouvements observés lorsqu'elle a affaire à un malade ordinaire ou même à une victime. Devrons-nous attribuer à S une forme sévère d'insensibilité ou une forme forte de l'incompatibilisme ? Peut-être que selon S, les actions du criminel ne lui étant pas plus imputables que celles d'un patient ordinaire, il n'y a pas de raison de se conduire différemment avec l'un et avec les autres. Admettons même que S soit une incompatibiliste radicale. Nous pouvons néanmoins penser que S est insensible et que c'est son insensibilité, autant ou plus que la théorie dont elle se targue, qui explique son comportement. Autrement dit : alors que les convictions incompatibilistes attribuées à S sont suffisantes pour expliquer le détail de sa conduite observée, il nous est difficile de ne pas chercher ailleurs la source de sa motivation²⁵. De même que la dualité des images sellarsiennes n'est pas en cause dans tous les cas où nos contenus de croyance sont spécifiés par un déterminable dialectique, de même ici il est clair que toutes les situations dans lesquelles on peut relever une inconsistance doxastique ne mettent pas en jeu la dualité des images sellarsiennes. On peut néanmoins affirmer que le genre d'inconsistance que fait voir l'oracle de Burgess appliqué au compatibilisme est rendu possible par le dualisme épistémique et donc le signale en partie.

4. Effets du dualisme

²² Afin de soutenir cette hypothèse, arrêtons-nous sur les croyances du sujet qui a été informé par l'oracle. Les croyances sont en général représentées dans un espace doxastique de scénarios. Pour un contenu p , l'état doxastique d'un sujet S qui croit que p est une partition de l'espace doxastique. Si S croit que Hespérus est Phosphorus, cette

croissance exclut tous les scénarios dans lesquels Hespérus n'est pas Phosphorus et divise ainsi l'espace doxastique de S en deux parties. Selon le dualisme, une croyance est une partition d'une espace de scénarios *dans une image*. Ces images sont elles-mêmes des classes de scénarios. Les croyances qui nous intéressent portent sur le compatibilisme et le déterminisme. S a rencontré l'oracle dont la parole ne peut être mise en doute, l'oracle lui a annoncé, comme on a dit, que le monde est déterministe et que le déterminisme est incompatible avec le libre arbitre. S sait désormais que le monde est déterministe et que le déterminisme est incompatible avec le libre arbitre. Mais dans quel sens ? Selon quelle version ?

- 23 **Histoires numéro 1** : Comme S est une philosophe avertie, elle a pris soin de demander à l'oracle des précisions, c'est-à-dire de décliner les déterminations dont il était question dans la section précédente. L'oracle les lui a données et S est à présent en relation avec un contenu de croyances déterminé. S pensait que la mécanique quantique était indéterministe et l'oracle lui apprend qu'elle ne l'est pas ; S pensait que la mécanique newtonienne était déterministe l'oracle lui apprend qu'elle ne l'est pas ; S pensait que les tentatives de conciliation du déterminisme et du libre arbitre étaient vouées à l'échec et l'oracle lui apprend qu'elles le sont effectivement. Enfin l'oracle ajoute : le caractère indéterministe de la mécanique newtonienne n'est pas pertinent, car la mécanique newtonienne peut être réduite par une théorie dont elle est une approximation acceptable dans certaines conditions C et la théorie réductrice est déterministe. S tient désormais une partie de ses croyances de l'oracle, dont elle sait d'autre part qu'il est parfaitement fiable. Néanmoins, S peut entendre l'oracle, acquiescer, et maintenir que dans l'image manifeste, la mécanique newtonienne a une place particulière, car les conditions C qui rendent son approximation acceptable sont également les conditions des actions humaines et de la moralité. Par conséquent S peut encore penser de manière fondée que ces informations qu'elle tient d'une source fiable ne ferment pas entièrement à la porte, sinon au compatibilisme, au moins au libre arbitre. J'ajoute que dans ce cas, cette porte entrouverte ressemble à l'inconsistance doxastique, qui est cette fois comme autorisée par la dualité des deux images :

ID_S S sait que Φ [les théories réductrices qui décrivent le monde dans l'image scientifique sont déterministes] et croit que Φ' [les théories qui décrivent le monde dans l'image manifeste ne le sont pas]

Les croyances que S tient de l'oracle sont des connaissances, justifiées par la croyance que l'oracle est fiable et véridique. Mais sur la base de ce savoir et par le jeu des images sellarsiennes, S dérive des croyances propres. Que ces croyances soient hors de portée du jugement univoque de l'oracle est rendu possible par le dualisme épistémique. Sur la base des mêmes données informationnelles, le cadre dualiste et le cadre moniste ne favorisent pas les mêmes thèses philosophiques.

- 24 **Histoire numéro 2** : Supposons à présent que S soit encore une philosophe avertie, qu'elle ait pris soin de demander à l'oracle de lui donner les précisions qui s'imposent, mais que l'oracle ait été dans un mauvais jour. Lorsque S lui a demandé de décliner l'arborescence des concepts impliqués, l'oracle a commencé de bonne grâce puis, assez vite, s'est lassé, déclarant que c'était déjà bien qu'elle, puisse jouir de cette réponse plus générale dont la vérité lui était assurée. La situation de S est à présent un mélange de savoir et d'ignorance, caractéristique du savoir déférentiel : S sait que quelqu'un sait que le monde est déterministe et que le déterminisme est incompatible avec le libre arbitre [Φ], mais le grain de son savoir que Φ est déterminé, non par le détail de son enquête,

mais par les conditions informationnelles et communicationnelles de la relation de S au porteur du savoir. Et l'oracle a choisi de lui donner un savoir seulement *déterminable*. Si S est moniste, il lui revient éventuellement d'envisager les déterminés correspondants. Mais à quoi bon ? Pour chercher la fondation d'une réponse qu'elle possède déjà et qu'elle tient pour assurée ? Si S est moniste, l'enquête n'a pas à être poursuivie plus avant. Elle peut l'être. La réponse de l'oracle peut être précisée, mais le sens de la réponse ne s'en trouvera pas modifié. En revanche, si S est dualiste, il lui revient de distribuer le savoir communiqué par l'oracle sur les deux images et, dans la mesure où la distribution avancée dans la première histoire est encore à sa disposition, S peut prétendre que la réponse de l'oracle est insuffisante en l'état : le scénario dualiste compatibiliste est compatible avec l'information dont elle dispose. Donc le dualisme modifie éventuellement le sens de la réponse et surtout la relation du savoir communiqué à l'ignorance.

5. Croyances sellarsiennes

- 25 J'appelle « croyances sellarsiennes » le genre auxquelles appartiennent les croyances engendrées dans notre exemple par la confiance dans la parole oraculaire. Les croyances sellarsiennes sont bidimensionnelles. Elles sont composées d'un contenu pris dans chaque image, ou d'un contenu pris dans une image et de sa contrepartie dans l'autre. Sellars lui-même, dans « Fatalism and Determinism », n'envisage pas les croyances de cette manière, mais indique que les pensées et les actions humaines ont une « contrepartie scientifique », c'est-à-dire une contrepartie dans l'image scientifique, et que c'est la version scientifique du prédicat « prédictible » qu'on doit chercher à leur appliquer dans l'enquête sur le compatibilisme :

The question I want to raise concerns the force of the term « predictable » in these two radically different conceptual frameworks. Thus we have supposed it to be a framework principle of the scientific image that every episode, including the scientific counterparts of human thoughts and actions, is predictable. It is, correspondingly, a framework principle of the manifest image that all human actions be predictable ? Could it perhaps be a framework principle of the manifest image that not all actions are predictable, although the very same actions that are « manifestly » unpredictable are, as projected in the scientific image, in principle predictable²⁶ ?

- 26 Nos croyances relatives au compatibilisme ont un contenu qui est comme distribué sur un écran divisé. Une version du compatibilisme est ainsi :

Compatibilisme sellarsien < [les actions_m sont imprédictibles_m] & [les actions_s sont prédictibles_s] >

où les indices m et s servent à désigner, pour un item dans une image, sa contrepartie dans l'autre. Plus généralement on notera :

Croyance sellarsienne que $[\Phi] < [\Phi]_m \& [\Phi]_s >$

où les constituants de $[\Phi]_m$ et de $[\Phi]_s$ sont deux à deux des contreparties.

- 27 Il va sans dire que cette caractérisation des « croyances sellarsiennes » soulève de nombreuses difficultés et requiert des précisions. J'en mentionne quelques-unes immédiatement, dans le seul but de donner une esquisse de l'interprétation du dualisme en termes de croyances.

- 28 (1) Si le dualisme est assez fort, on doit ménager une place à ces contenus de croyance décrits dans une image et qui n'ont pas de contrepartie dans l'autre. Par exemple, selon

Sellars, nos croyances normatives exprimées par des phrases en « ought to » n'appartiennent pas à l'image scientifique. Si c'est bien le cas il y aura, outre la division en deux écrans, une classe spéciale, elle-même enjambant éventuellement les deux images ou incluse dans une seule, rassemblant ces contenus particuliers. On notera que ces contenus d'une image qui n'ont pas de contrepartie dans l'autre ne sont pas pour autant insensibles à l'existence des deux images. On dira plutôt que ce sont des contenus sellarsiens *liés* à une image dont la forme générale est :

Croyance sellarsienne que $[\Phi]$ *liée* à l'image manifeste $\langle [\Phi]_m \& [\emptyset]_s \rangle$

Croyance sellarsienne que $[\Phi]$ *liée* à l'image scientifique $\langle [\emptyset]_m \& [\Phi]_s \rangle$

où $[\emptyset]$ est un contenu vide. Il est important de noter que les croyances de ce genre ne sont pas le propre de l'image manifeste. Au contraire, dans la mesure où l'image scientifique est construite sur la base d'une liberté catégoriale étendue, il y a de fortes chances pour que le produit de cette liberté soit plus souvent inexportable dans l'autre image et constitue des croyances sellarsiennes *liées*. Les théories scientifiques produisent plus de croyance que nous ne sommes capables d'en exprimer dans l'image manifeste.

29 (2) Se pose alors la question du privilège épistémique de l'image manifeste. Une chose est de dire que certaines au moins de nos croyances sont distribuées dans un cadre double, autre chose de défendre la thèse selon laquelle disposer d'une contrepartie manifeste est une condition pour accéder au statut de connaissance. Selon cette thèse – disons, le kantisme sellarsien – les contenus qui prétendent accéder au statut de connaissance doivent s'enraciner, d'une manière ou d'une autre, dans l'image manifeste, soit parce qu'ils y possèdent une contrepartie, soit, dans le cas de termes théoriques complexes, parce que l'explication qu'on en donne est ancrée plus médiatement dans l'image manifeste. Bien que Sellars ait probablement défendu une thèse de ce genre, il est important de noter que celle-ci est différente du dualisme proprement dit et n'est pas impliqué par lui. On peut aussi défendre un dualisme dans lequel la force épistémique est assurée par des contenus de l'image scientifique. Dans la discussion du compatibilisme telle que je la conçois, ce n'est pas cet aspect du dualisme qui importe. Nous n'avons pas à chercher une réponse à la question d'un côté ou de l'autre de l'écran divisé mais d'abord à reconnaître que la question elle-même est posée dans les termes de l'image manifeste et porte sur des croyances prises dans l'autre côté.

30 (3) Les contenus sellarsiens liés à une image doivent être distingués de ceux qui sont éventuellement identiques dans les deux images. Je pense par exemple à des contenus superassertables, pour lesquels nous savons que l'augmentation du savoir ne modifie pas la valeur de vérité de la proposition associée. Ainsi le contenu de croyance que $[2+2=4]$ peut être pensé dans les deux images, par le marchand de Spinoza qui calcule très vite sans rien savoir des nombres – image manifeste –, par un mathématicien d'aujourd'hui qui connaît l'algèbre et la théorie des ensembles – image scientifique – ou encore par le grand calculateur Jean Suisset que Leibniz admirait et dont Gauss s'est amusé à relever les erreurs : je ne pense pas qu'il y ait un sens à situer sa pensée ici ou là. Ils ont les uns et les autres des justifications différentes pour le même contenu. On peut éventuellement le situer dans l'un ou l'autre des cadres en considérant la nature des justifications. Mais dans la mesure où un contenu de croyance peut être isolé du réseau de ses justifications, on aura ici des contenus sellarsiens *insensibles* à l'image :

Croyance sellarsienne que $[\Phi]$ *insensible* à l'image $\langle [\Phi]_m = [\Phi]_s \rangle$

31 On dira à l'inverse qu'une croyance est *sensible* à l'image si sa version dans une image est assez différente de sa contrepartie dans l'autre. C'est ce genre de sensibilité à l'image que

Sellars a en vue à propos de la croyance exprimée par « elle aurait pu agir autrement ». Le contenu exprimé par « elle aurait pu agir autrement » est sensible à l'image car sa contrepartie dans l'image scientifique, si elle existe, occupe une place reculée dans l'architecture du savoir à laquelle nous avons difficilement accès. Le niveau de complexité est un bon indicateur de la sensibilité à l'image.

- 32 (4) Une dernière remarque. Nous sommes capables de désigner nos contenus de croyance au moyen d'un désignateur fonctionnel comme « la contrepartie de XYZ dans l'image_s ». Ou inversement: « la contrepartie de XYZ dans l'image_m ». Cela signifie que certaines de nos croyances sont descriptives: c'est la description par laquelle on s'y réfère qui est liée à une image, plus que le contenu lui-même. De plus, toujours avec ces désignateurs, nous pouvons former des abstractions d'un genre spécial. Sellars est friand de ce genre d'abstraction, qu'il signale en général par la fiction des échecs texans. Un joueur peut visualiser un coup sur son échiquier et le communiquer à son adversaire par la poste ou par téléphone. Si, par extraordinaire, l'adversaire joue aux « échecs texans », c'est-à-dire avec un échiquier démesuré où les pièces sont des automobiles de différentes tailles, cela n'aura aucune incidence, ni sur le mode de communication de notre joueur – il peut parfaitement l'ignorer – ni sur l'issue de la partie. On fait quelque chose d'analogue aux échecs texans lorsqu'on fait référence à un contenu de croyance en faisant abstraction des images sellarsiennes. Par exemple en disant: « le a dont la contrepartie dans l'image_x est p » je peux chercher à désigner l'entité d'ordre élevé dont les contenus sellarsiens sont des instances. Les croyances insensibles à l'image ont la particularité d'être semblables à leur version abstraite; pour les autres, on les obtient par congruence sur les propriétés des composants des croyances sellarsiennes de premier ordre. Il s'agit d'une distinction importante dans la mesure où certaines pensées ne portent pas sur des composantes des croyances sellarsiennes, mais plutôt sur ce qu'on peut appeler les projections, dans une image, des abstractions de ce type. Nous savons que certaines entités de la physique statistique, par exemple les fermions, ont des conditions d'identités différentes de celles des objets ordinaires. Les fermions-dans-l'image-manifeste sont difficilement concevables. Mais savoir – dans l'image manifeste – qu'il y a des fermions signifie quelque chose. Les fermions dans l'image manifeste sont des projections. Le rôle des projections de ce genre, dans la genèse et dans l'histoire de nos croyances manifestes ne doit pas être sous-estimé.

6. Knobe & Nichols (2007) en contexte sellarsien

- 33 Je ne suis pas sûr que l'apparatus sellarsien donne un avantage décisif à l'une ou l'autre des positions défendues dans la métaphysique du libre arbitre, qu'il soit en particulier de nature à nous convaincre du bien-fondé du compatibilisme, comme Sellars l'a prétendu. Mais je pense qu'il est très utile dans la perspective des attributions de croyance, lorsque celles-ci portent sur la perception du libre arbitre et de la responsabilité morale. Dans un article de 2007 souvent commenté depuis, Nichols et Knobe ont voulu savoir si nos intuitions morales étaient plutôt compatibilistes ou plutôt incompatibilistes; si des sujets réels, qui se reconnaissent dans des théories métaphysiques et morales variées, sont plutôt enclins à penser que le libre arbitre est compatible avec le déterminisme ou si c'est l'inverse. Ils ont d'abord interrogé un échantillon d'étudiants en leur proposant de se représenter deux univers: l'un – l'univers A – dans lequel « tout ce qui advient est complètement causé par ce qui est advenu antérieurement » et l'autre – l'univers B – où

tout ce qui advient est également causé par ce qui est advenu antérieurement à l'exception de nos décisions d'action²⁷. La première question qui était posée aux sujets de l'enquête était : lequel de ces deux univers est selon vous plus semblable au nôtre ? » et 90 % des participants ont répondu que c'était plutôt l'univers B. Nos intuitions sont donc, selon cette enquête, massivement indéterministes. Les auteurs ont ensuite demandé aux participants de se situer dans un univers A, un univers fortement déterministe et de dire si, dans ce type d'univers, les personnes étaient ou non responsables de leurs actions. C'est surtout cette seconde partie de l'enquête qui a retenu l'attention. Ils ont en effet proposé successivement deux versions de cette question. La première contenait une description exemplifiant le concept d'une action moralement répréhensible et la deuxième ne contenait pas de description.

Question 1. In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family. Is Bill fully morally responsible for killing his wife and children ?

Question 2. In Universe A, is it possible for a person to be fully morally responsible for their actions?

³⁴ 72% des personnes interrogées donnent une réponse *compatibiliste* à la question 1 et pensent que Bill est bien responsable moralement de ses actions, bien qu'on soit dans un univers entièrement déterministe. Mais 86 % des personnes interrogées donnent une réponse *incompatibiliste* à la question 2 autrement dit, pensent que, si l'univers est entièrement déterministe, alors nous ne sommes pas moralement responsables de nos actions. Bien que la question 1 et la question 2 soient en un sens la même question, les réponses données sont sensiblement différentes. Les intuitions morales qui sont mobilisées par la première question ne sont pas les mêmes que les intuitions morales mobilisées par la seconde, elles sont même exactement opposées pour un nombre important de sujets, alors que la question 1 et la question 2 sont les mêmes, à une description près.

³⁵ Les auteurs de l'enquête ont cherché à expliquer ces résultats au moyen d'hypothèses psychologiques. Ils proposent trois modèles explicatifs 1. Selon le premier modèle ils supposent que les participants à l'enquête éprouvent des émotions assez intenses et que l'interférence des émotions et des concepts empêche les sujets d'appliquer la théorie de la responsabilité qu'ils appliquent d'ordinaire ou dans laquelle ils/elles se reconnaissent. Selon cette lecture, la description concrète est simplement plus émotiviste que la description abstraite. 2. Une autre possibilité interprétative consiste à attribuer un rôle *massif* à l'affect de telle sorte que les émotions ne sont pas une source de distorsion mais contribuent de manière essentielle aux jugements ordinaires de responsabilité : les sujets possèdent des théories mais, dans leur pratique effective, l'attribution de responsabilité s'appuie sur des processus d'ordre émotionnel. La différence de la réponse 1 et de la réponse 2 implique que la description plus concrète nous rapproche de la pratique ordinaire de l'exercice des concepts moraux. Enfin 3. Il existe un dernier cadre interprétatif, celui de la compétence concrète, qui postule que la réponse dans la condition concrète – dans le cas où la description mentionne un sujet moral particulier, lui donne un nom propre, lui attribue un acte assez précis – serait engendrée par un module inné. L'information serait élaborée selon des processus automatiques et inconscients dans un module isolé lorsqu'elle est délivrée par une description concrète et

le sujet n'aurait pas accès à la théorie de la responsabilité qu'il ou elle adopte en général et qu'il ou elle adopte en particulier pour répondre à la question dans la condition abstraite.

- 36 Les auteurs suggèrent que la relation qui unit les deux questions est celle du concret à l'abstrait. Elles seraient deux manières différentes de dire la même chose sur un mode spécial : non pas *synonyme* mais *paronyme*. Dans la question 2 dite « abstraite » : (i) l'agent est non déterminé, (ii) son action est non déterminée et (iii) elle est le résultat de l'état antérieur de l'univers (Univers de type A) ; (iv) elle est susceptible de louange ou de blâme (non déterminé) et enfin (v) susceptible de lui être imputée sous la forme d'une responsabilité (c'est l'objet de la question). Dans la question 1 : (i) l'agent est désigné par un nom propre (déterminé) et sa situation est précisée – il a une famille, un travail, une secrétaire, une maison – (ii) son action est donnée sous une description : elle appartient à un type d'action – un meurtre – et certains détails en font un meurtre déterminé ; (iii) cette action est encore le résultat de l'état antérieur de l'univers (type A) ; (iv) le type auquel l'action appartient est en général sujet de blâme – surtout lorsqu'elle est accompagnée du genre de circonstance qui font l'objet de la courte description ; (v) enfin, comme dans la question 1, l'action est susceptible d'être imputée sous la forme d'une responsabilité (objet de la question). C'est donc la relation du déterminable au déterminé qui me semble ici pertinente, plus que la relation de l'abstrait au concret qui est beaucoup trop équivoque.
- 37 S'il faut néanmoins préciser le genre d'abstraction qui se trouve ici impliqué, il faudrait pouvoir en premier lieu distinguer une abstraction catégoriale ou, pourrait-on dire, « goodmanienne ». Dans une situation de connaissance déterminée, nous disposons pour décrire un événement, une action ou un objet, de moyens conceptuels appartenant à des catégories différentes et hiérarchisées. L'axe du concret et de l'abstrait décrit en ce sens une variation dans les catégories que nous renseignons dans la description, ou, mieux : le rapport des catégories que nous renseignons relativement aux catégories dont nous disposons. Dans le cas d'une description, cette dimension systémique est particulièrement importante. La description nous donne certains traits de la chose décrite – par exemple qu'il s'agit d'un meurtre épouvantable – mais elle nous renseigne aussi sur les catégories que nous sommes autorisés à employer pour concevoir et éventuellement évaluer la chose décrite : il s'agit, par exemple ici, d'une action effectuée dans un univers déterministe. Cette abstraction de nature systémique ne doit pas être confondue avec d'autres types d'abstraction qui portent, par exemple, sur la nature de l'identité, les propriétés congruentes ou la ressemblance à un exemplaire. Qu'en est-il à présent des émotions ? Celles-ci font éventuellement, mais non nécessairement, partie du réglage catégorial. En général, les individus ont des émotions et les émotions interviennent dans l'évaluation morale des actions. L'évaluateur est-il autorisé à considérer que les émotions sont des ingrédients du monde décrit ? Qu'elles peuvent intervenir dans l'évaluation ? Le destinataire d'une description fictionnelle cherche les réponses à ces questions métafictionnelles dans l'échantillon de fiction auquel il est confronté et dans son évaluation contextuelle. Si la description est celle d'un meurtre atroce, elle comporte pour ainsi dire par soi une invitation à considérer les émotions comme des ingrédients à part entière. Si c'est une description strictement médicale ou clinique, la situation épistémique qu'elle motive a de fortes chances d'être différente.
- 38 Mais n'a-t-on pas également affaire à une abstraction sellarsienne au sens que j'ai proposé dans la section précédente ? Supposons que S fait partie des sujets qui ont fait preuve

d'inconsistance et ont donné une réponse compatibiliste à la question 1 et incompatibiliste à la question 2. S'est comme le personnage de notre section 4, placé dans la situation oraculaire. On lui demande de croire que le monde est déterministe ou de faire comme si on le lui avait appris. On lui donne la moitié de la réponse de l'oracle et on lui demande de fournir l'autre moitié en activant des croyances conditionnelles : si le monde est comme ceci et comme cela, croyez-vous que... ? Si ces croyances conditionnelles sont des croyances sellarsiennes, il est aisé de voir qu'elles n'ont pas le même profil selon qu'elles sont mobilisées par la question 1 ou par la question 2. Dans la question numéro 2, on fait appel à une croyance sellarsienne abstraite dans le sens que nous avons dégagé. On demande au sujet d'envisager le contenu que [le monde est déterministe] dans l'une et l'autre des images et de faire de cette croyance la base des inférences par lesquelles il faut passer pour fournir une réponse. Par la question numéro 1 au contraire, on ancre la situation dans l'image manifeste et on demande au sujet d'y faire intervenir un contenu abstrait, toujours le même, donné par l'antécédent du conditionnel. On peut supposer que c'est cette fois la *projection* de la croyance sellarsienne abstraite qui est mise en jeu et non la croyance sellarsienne proprement dite, avec ses deux composantes. Que cette projection ait un pouvoir limité dans l'ordre logique des raisons ne doit pas nous étonner. C'est pour le moins quelque chose qu'un inférentialiste est certainement prêt à accepter.

BIBLIOGRAPHIE

- Carnap R. (1924) « Dreidimensionalität des Raumes und Kausalität : Eine Untersuchung über den logischen Zusammenhang zweier Fiktionen », *Annalen der Philosophie und philosophischen Kritik*, 4/3, pp. 105-130.
- Carnap R. (1927) « Eigentliche und uneigentliche Begriffe », *Philosophische Zeitschrift für Forschung und Aussprache*, 1/4, p. 355-374.
- Chalmers, D. J. (2012), *Constructing the World*, Oxford : Oxford University Press.
- Cohen, L. J. (1992) *An Essay on Belief and Acceptance*, Oxford : Oxford University Press.
- Earman, J. (2004) « Determinism : What We Have Learned and What We Still Don't Know », in : *Freedom and Determinism*, edited by Joseph Keim Campbell, Michael O'Rourke, and David Shier, The MIT Press Cambridge, Massachusetts, pp. 2-46.
- Engel, P. (ed.) (2000) *Believing and accepting*, Dordrecht : Kluwer Academic Publisher.
- Higginbotham J. (1996). « The Semantics of Questions », in S. Lappin (ed.), *the Handbook of Contemporary Semantic Theory*, Malden : Blackwell, ch. 14.
- Kistler, M. (2016) *L'esprit matériel. Réduction et émergence*, Paris : Ithaque.
- Knobe, J. & S. Nichols (2007) « Moral Responsibility and Determinism : The Cognitive Science of Folk Intuitions », *Noûs*, 41/4, pp. 663-685.
- Lewis, D. (1974) « *Radical Interpretation* », *Synthese*, Vol. 27, No. 3/4, *Intentionality, Language, and Translation*, pp. 331-344.

- Peacocke, C. (1999) *Being Known*, Oxford : OUP.
- Richards, R. J. (1973) « Sellars' Kantian Perspective on the Compatibility of Freedom and Determinism », *Southern Journal of Philosophy*, pp. 228-236.
- Schaffer J. (2009) « On What Grounds What » in : David Chalmers, David Manley, and Ryan Wasserman (eds.), *Metametaphysics, New Essays on the Foundations of Ontology*, Oxford : Clarendon Press, p. 348-383.
- Schaffner, Kenneth F. (1967) « Approaches to Reduction », *Philosophy of Science*, 34/2, pp. 137-147.
- Sellars, W. (1966,) « Thought and Action » in : Keith Lehrer (ed.) *Freedom and Determinism*, NY : Random House, pp. 140-174
- Sellars, W. (1966₂) « Fatalism and Determinism » *ibid.* (Cité d'après la version révisée : *Wilfrid Sellars papers, 1899-1990*, ASP191.01, Archives of Scientific Philosophy, Special collection Departement, University of Pittsburgh, Box 30, Folder 02-04.
- Sellars, W. (1963¹, 1991²) *Science, Perception and Reality*, Atascadero : Ridgeview Publishing Company.
- Sellars, W. (1968) *Science and Metaphysics. Variations on Kantian Themes*, London : Routledge and Kegan Paul.
- Szabo Gendler, T. (2008) « Alief and Belief », *The Journal of philosophy*, 104/10, pp 634-663.
- Ullman-Margalit, E. and A. Margalit (1992), « Holding-True and Holding as True », *Synthese*, 92, pp. 167-187.
- Yablo, S. (2000) « Apriority and Existence », in : Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the A Priori*, Oxford : Oxford University Press, pp. 197-228.

NOTES

1. Au sens de « l'integration challenge » de Christopher Peacocke : « I call the general task of providing, for a given area, a simultaneously acceptable metaphysics and epistemology, and showing them to be so, the *Integration Challenge* for that area. » Peacocke, C. (1999) p. 3.
2. Désormais PSIM. Les références sont prises dans Sellars W. (1963). J'emploie « image » ou « image du monde », désormais sans marqueur citationnel, à la manière de Sellars, sauf précision contraire.
3. Cohen J. (1992) et Engel P. (ed.) 2000.
4. Ullman-Margalit, E. and A. Margalit (1992).
5. Tous les concepts légitimes des sciences empiriques doivent pouvoir être indiqués dans un plan unique « comme est indiqué un lieu à la surface de la Terre par la latitude et la longitude. » Carnap, R. (1927) p. 358. Voir aussi Carnap, R (1934) : « The opinion is generally accepted that the various sciences named [philosophy, formal sciences, natural sciences, social sciences] are fundamentally distinct in respect of subject matter, sources of knowledge, and technique. Opposed to this opinion is the thesis defended in this paper that science is a unity, that all empirical statements can be expressed in a single language, all states of affairs are of one kind and are known by the same method. » avec le commentaire dans Chalmers, D.J. (2012), p. 301-305.
6. Les deux images « de l'homme dans le monde » sont en ce sens un fragment de philosophie néo-kantienne et une tentative de synthèse de l'usage pratique et de l'usage théorique de la raison. Cf. Richard R. J. (1973) qui insiste sur cet aspect.

7. « For the philosopher is confronted not by one complex many-dimensional picture, the unity of which, such as it is, he must come to appreciate ; but by *two* pictures of essentially the same order of complexity, each of which purports to be a complete picture of man-in-the-world, and which, after separate scrutiny, he must fuse into one vision. » PSIM, p. 4. Une forme de dualisme épistémique est présente dans les premiers textes de Carnap. Cf. Carnap R. (1924) où il est question, à propos des dimensions de l'espace, de deux fictions distinctes « Auch der dreidimensionale Raum ist schon eine fiktive Erweiterung des zweidimensionalen Raumes der (primären) Erfahrung » p. 106.

8. Toute une tradition, celle de la « perennial philosophy », n'est pas autre chose que « l'image manifeste considérée comme réelle » : « Let me elaborate on this theme by introducing another construct which I shall call [...] the perennial philosophy of man-in-the-world. This construct, which is the 'ideal type' around which philosophies in what might be called, in a suitably broad sense, the Platonic tradition cluster, is simply the manifest image endorsed as real, and its outline taken to be the large-scale map of reality to which science brings a needle-point of detail and an elaborate technique of map-reading. » PSIM, p. 8.

9. Sellars, W. (1966₂).

10. PSIM, p. 9 : « A fundamental question with respect to any conceptual framework is 'of what sort are the basic objects of the framework ?' This question involves, on the other hand, the contrast between an object and what can be true of it in the way of properties, relations and activities ; and on the other, a contrast between the basic objects of the framework and the various kinds of groups they can compose. [...] Now to ask, 'what are the basic objects of a (given) framework ?' is to ask not for a list, but a classification. And the classification will be more or less 'abstract' depending on what the purpose of the inquiry is. » Comparer : Schaffer J. (2009).

11. PSIM, p. 18.

12. « It is current practice to compare the inter-subjective standards without which there would be no thinking, to the inter-subjective standards without which there would be no such thing as a game. It is worth noting, however, that conceptual thinking is a unique game in two respects: (a) one cannot learn to play it by being told the rules ; (b) whatever else conceptual thinking makes possible – and without it there is nothing characteristically human – it does so by virtue of containing a way of representing the world. » PSIM, p. 17.

13. « A group exists in the way in which members of the group represent themselves. Conceptual thinking is not by accident that which is communicated to others, any more than the decision to move a chess piece is by accident that which finds an expression in a move on a board between two people. » *Ibid.*

14. Sellars ajoute à cet égard une considération assez radicale : il n'y a pas à proprement parler de croyance sur le cadre, seulement des croyances dans le cadre. Le contenu completif ne peut pas être un trait du cadre. C'est au point que certains contenus que nous tenons pour des croyances sont invités à sortir du stock doxastique. Impossible, par exemple, de *croire* que le triangle est une figure plane. « One cannot be said to believe that a triangle is a plane figure. When primitive man ceased to think of what we called trees as persons, the change was more radical than a change in belief; it was a change in category. » *Ibid.* p. 10

15. PSIM, p. 10.

16. La pensée de Sellars est donnée sur ces questions avec toutes ses nuances dans Sellars W. (1968) : cf. en particulier les §§ 58-64.

17. PSIM, p. 38.

18. PSIM, pp. 39-40.

19. Cf. Schaffner K. F. (1967) donne, cinq ans après l'édition de PSIM, une tentative de classification des variétés de la réduction.

20. L'exemple, classique, est clarifié par Max Kistler dans Kistler (2016).

21. Cf le panorama proposé dans : Earman, J. (2004).

22. Ce point est particulièrement saillant dans le cadre de la mécanique newtonienne. John Earman écrit par exemple : « Consider a finite number of point mass particles interacting via Newton's $1/r^2$ force law. Let's simply ignore problems about collision singularities by focusing on solutions that are collision-free. Nevertheless, after many decades of work, it has been established that noncollision singularities can occur; that is, drawing on the infinitely deep $1/r$ potential well, the particles can accelerate themselves off to spatial infinity in a finite amount of time [...]. The time reverse of such a process is an example of 'space invaders', particles appearing from spatial infinity without any prior warning. To put it crudely, you can't hope to have Laplacian determinism for open systems, and for the type of interaction under discussion, the entire universe is an open system. » Earman, J. (2004), p. 25

23. Higginbotham J. (1996).

24. Cf. Yablo (2000) p. 200.

25. Pour attribuer la thèse compatibiliste, nous devons parcourir à rebours le chemin que David Lewis a décrit à propos de la traduction radicale (Lewis D. 1974) : (i) établir la thèse compatibiliste comme un ensemble de phrases exprimant des propositions dont nous connaissons les conditions de vérités ; (ii) savoir reconnaître ces propositions dans des attitudes propositionnelles exprimées dans notre langage et dans le langage de celui ou de celle à qui nous l'attribuons ; enfin (iii) relier ces attitudes propositionnelles à leurs manifestations dans un système physique observable sur la base duquel l'attribution est susceptible d'être justifiée. C'est dans cette dernière phase que l'attributeur aura tendance à ajouter la motivation seconde et redondante : S croit que Φ et je agit selon Φ' , où agir selon Φ selon Φ' sont indiscernables.

26. « Fatalism And Determinism », p. 4.

27. « In Universe A every decision is completely caused by what happened before the decision—given the past, each decision *has to happen* the way that it does. By contrast, in Universe B, decisions are not completely caused by the past, and each human decision *does not have to happen* the way that it does » (Nichols, S. & J. Knobe [2007], p. 669).

INDEX

Mots-clés : bidimensionnalisme, compatibilisme, image manifeste et image scientifique, prédictibilité, Wilfrid Sellars

AUTEUR

JEAN-BAPTISTE RAUZY

Université Paris-Sorbonne –SND