

Peter van Inwagen *Essai sur le libre arbitre*

Chapitre III Trois arguments pour l'incompatibilisme

3.1 La principale question disputée dans les discussions en cours sur le libre arbitre n'est pas, comme on pourrait s'y attendre, de savoir si nous *avons* le libre arbitre. C'est de savoir si le libre arbitre est compatible avec le déterminisme. Il me semble que libre arbitre et le déterminisme sont incompatibles, et dans ce chapitre j'essaierai de démontrer cette incompatibilité.

Les discussions de cette question n'atteignent pas d'habitude un niveau très élevé. Dans la grande majorité des cas, elles sont l'œuvre de compatibilistes et consistent largement à attribuer tel ou tel sophisme infantile aux incompatibilistes (confusion des lois « descriptives » et « prescriptives » ; manque à distinguer entre la nécessité causale et la compulsion/contrainte ; assimilation de la liberté au pur hasard). Donald Davidson se place lui-même dans cette tradition quand il écrit :

Je ne serai pas concerné directement par [les arguments pour l'incompatibilité de la liberté et de la détermination causale], car je n'en connais aucun qui soit plausible autrement qu'en apparence. Hobbes, Locke, Hume, Moore, Schlick, Ayer, Stevenson, et une foule d'autres ont fait ce qui peut être fait, ou ce qui aurait toujours dû être fourni, pour écarter les confusions qui font penser que le déterminisme s'oppose à la liberté¹.

Il n'est pas dans mon intention, dans ce livre, de défendre un auteur antérieur contre l'accusation de sophisme. Mes propres arguments seront explicites, et tout sophisme qu'ils commettraient devrait ainsi être visible. (Il est douteux que qui que ce soit ait jamais été trompé par les sophismes dont on accuse habituellement les incompatibilistes ; si quiconque a jamais atteint un tel niveau d'incompétence philosophique, en ce qui me concerne, j'y échappe.) La ligne qui sépare l'argumentation en faveur d'une position et l'attribution d'un raisonnement fallacieux à ceux qui s'opposent à cette position, bien qu'elle soit claire en théorie, est souvent obscurcie en pratique. Au milieu des accusations de sophisme et de confusion qu'on trouve dans les écrits des philosophes mentionnés par Davidson, il y a des arguments positifs pour la compatibilité du libre arbitre et du déterminisme. Pour autant que ces arguments puissent être séparés des torts qui ont été diagnostiqués et dont je ne suis pas l'héritier – si tant est que ces torts existent bien – je les examinerai au chapitre IV².

3.2 Au chapitre I j'ai mentionné le simple argument suivant en faveur de l'incompatibilisme, l'Argument de la Conséquence :

Si le déterminisme est vrai, alors nos actions sont les conséquences des lois de la nature et d'événements du passé éloigné. Mais ce qui est arrivé avant notre naissance

¹. « Liberté d'agir », art. cit. XXX, p. 139.

². Je ne connais qu'un sophisme dont les incompatibilistes ont été accusés et qui n'est pas infantile. Dans *Will, Freedom and Power* Oxford, Blackwell, 1975, p. 155-6, Anthony Kenny accuse les incompatibilistes de faire usage de la règle d'inférence suivante, dont il prétend qu'elle est invalide :

Je ne peux pas faire ceci

Faire ceci revient, dans cette situation, à faire cela

Donc, je ne peux pas faire cela

(L'incompatibiliste, dit Kenny, procède en substituant 'violer une loi de la nature' à 'ceci' et une expression désignant un acte non réalisé à 'cela'). Que cette forme d'inférence soit ou non valide, et je ne crois pas que Kenny lui donne un contre-exemple clair, je ne peux pas voir qu'aucun des arguments du présent chapitre en dépende, bien que ces arguments dépendent de notre incapacité à rendre fausse toute proposition qui est une loi de la nature.

ne dépend pas de nous (*not up to us*), pas plus que ce que sont les lois de la nature ne dépend de nous. Les conséquences de ces deux choses (y compris nos actes présents) ne dépendent donc pas de nous.

C'est, je crois, un *bon* argument. Mais je dois admettre qu'il est plutôt schématique. Le présent chapitre est une tentative pour compléter dans les détails ce schéma, de plusieurs façons. (J'ai appelé ce chapitre « Trois arguments pour l'incompatibilisme », mais le principe d'individuation pour les arguments est loin d'être clair ; j'aurais pu l'appeler « Trois versions d'un argument pour l'incompatibilisme »). Ces trois arguments, ou versions d'un seul argument, ou quoi qu'ils soient, visent à se soutenir mutuellement. Même s'ils ont essentiellement le même objet, ils sont très différents dans leur structure et leur vocabulaire. Je voudrais exploiter leurs différences de la manière suivante : si quelqu'un ressent un vague sentiment de malaise avec l'un de ces arguments – parce que sa formulation n'a rien à voir avec le langage ordinaire, ou, parce que sa structure logique est plutôt compliquée et peut servir à masquer une confusion – il pourra constater (j'espère qu'il constatera) que ce sentiment de malaise ne se produit pas quand il considère l'un des deux autres arguments, ou les deux. Cela, je crois, tendra à montrer que la cause du sentiment de malaise est un *accident lié à un mode de présentation particulier* des traits communs aux trois arguments et ne réside pas dans ces traits communs eux-mêmes. Bien sûr, il ne s'ensuit pas logiquement qu'il en soit ainsi : la cause du sentiment de malaise que quelqu'un éprouve quand il considère l'un des trois arguments peut être présente mais inactive quand il en considère un autre, en raison du fait que cette cause, qui peut être ou non un défaut réel dans les arguments, est masquée par des traits particuliers à cet autre argument.

Je fais ces remarques en pensant à les appliquer seulement aux cas de « vague sentiment de malaise ». Si un philosophe a une objection spécifique et détaillée à faire à l'un des arguments, alors je ne propose pas de me cacher derrière les autres. En fait, je crois que ce serait impossible : je suis bien certain que toute objection spécifique et détaillée à l'un des arguments pourra être appliquée aux autres ; et je crois que toute objection à l'un des arguments sera une bonne objection à *cet* argument si et seulement si les objections équivalentes aux autres arguments sont de bonnes objections à *ces arguments*. Je peux me tromper sur ce point. Peut-être un des arguments échoue-t-il d'une manière qui ne correspond à aucun défaut dans les arguments correspondants ; en fait, peut-être tous les arguments sont-ils défectueux, mais pour différentes raisons. Je ne crois, cependant, vraiment pas que ce soit le cas. Nous allons bien voir.

Il y a un second avantage dialectique dans ma présentation de ces arguments, pourvu que j'aie raison de penser qu'ils tiennent ou s'effondrent ensemble. Si je ne m'appuyais que sur l'un d'eux, un compatibiliste pourrait raisonner ainsi : « puisque la conclusion de cet argument est fondamentalement erronée, l'argument contient une erreur fondamentale. J'admets que je n'arrive pas à mettre le doigt sur cette erreur, bien que je sois sûr d'y arriver si j'étais prêt à y mettre le temps et l'effort requis. Mais la vie est courte... » J'ai de la sympathie pour ce genre d'attitude, du moins tant que le philosophe qui l'exprime ne prétend pas être un expert du problème du libre arbitre et du déterminisme. La vie *est* courte. Mais je crois que cette attitude est *moins* raisonnable face aux trois arguments qui semblent intuitivement marquer le même point et qui sont pourtant très différents dans le vocabulaire employé et dans leur structure logique. Assurément s'il y avait une erreur fondamentale commune aux trois arguments, ne serait-il pas au moins vraisemblable qu'elle se manifeste dans l'un des trois, aussi cachée soit-elle dans les deux autres ?

Chacun des trois arguments a ses vertus et ses vices particuliers. La vertu du premier argument est son vocabulaire, proche de celui des discussions traditionnelles sur le problème du libre arbitre. Son vice est sa structure extrêmement complexe. La vertu du deuxième

argument est sa structure extrêmement simple. Son vice est son vocabulaire, radicalement différent du vocabulaire traditionnel des arguments sur le libre arbitre. Le troisième argument est un cas intermédiaire quant à la complexité de la structure logique et à l'écart avec la terminologie traditionnelle. Son vice logique particulier est une vertu dialectique : tandis que les deux premiers arguments sont valides en logique extensionnelle du premier ordre, le troisième emploie des principes modaux spéciaux. C'est un vice logique parce que je ne dispose d'aucune sémantique extensionnelle pour l'opérateur modal qui est à l'œuvre dans l'argument, et, par conséquent, ma défense de la validité des principes modaux que j'emploie dans l'argument repose seulement sur l'intuition. Mais l'usage d'un opérateur modal est une vertu dialectique, puisque beaucoup de gens semblent penser qu'un sophisme modal est d'une certaine façon sous-jacent à l'incompatibilisme, et cette accusation est plus facile à évaluer si on l'applique à un argument explicitement modal.

3.3 Dans cette section et la suivante, je vais expliquer ce que j'entends par certains termes utilisés dans le premier argument pour l'incompatibilisme.

Je commencerai par essayer de dire ce que j'entends par *déterminisme*. Pour définir le déterminisme, j'aurai besoin de trois notions subordonnées : la notion de *proposition* (et des notions reliées comme *vérité*, *négation*, *conjonction* ainsi que *implication*), la notion d'*état du monde physique entier à un instant*, et la notion de *loi de la nature*.

Les propositions (c'est-à-dire les porteurs non linguistiques de valeur de vérité) ont été introduites dans la Section 2.4. J'ai peu à ajouter à ce qui a été dit alors et en 2.7 sur les propositions, la vérité et la fausseté. Je supposerai seulement que les propositions ont les quatre propriétés suivantes. (i) Pour toute manière possible dont le monde pourrait être, il lui correspond au moins une proposition, une proposition qui est nécessairement telle qu'elle est vraie si et seulement si le monde *est* de cette façon. Puisqu'il y a sans doute des manières dont les choses pourraient être qui sont trop complexes pour être décrites en un quelconque langage naturel, cela signifie qu'il y a des propositions qui ne peuvent pas être exprimées dans un langage naturel. (ii) Toute proposition est ou bien vraie ou bien fausse. (iii) La conjonction d'une proposition vraie et d'une proposition fausse est une proposition fausse. (iv) Les propositions obéissent à la Loi de la Contradiction quant à l'implication. C'est-à-dire que pour tout x et pour tout y , si x et y sont des propositions, et s'il est impossible que x soit vrai et y faux, alors il est impossible que la négation de y soit vraie et la négation de x fausse.

Voici un « modèle » pour les propositions qui leur donne ces caractéristiques : les propositions sont identifiées avec des ensembles de mondes possibles³. Une proposition en ce sens est *vraie* si elle contient le monde actuel ; la *négation* d'une proposition est son complémentaire dans l'ensemble de tous les mondes possibles ; la *conjonction* de deux propositions est leur intersection ; une proposition *implique* une autre si la première est un sous-ensemble de la seconde. Je ne suis pas lié à ce modèle. Il a des caractéristiques inappropriées aux propositions qu'il modélise pour certaines des tâches traditionnellement allouées aux choses que l'on appelle 'propositions'. Par exemple, dans la Section 2.4, les « propositions » ont été introduites comme des choses que l'on pouvait accepter, asserter, nier etc. Mais les objets de telles activités peuvent difficilement être modélisés par des ensembles de monde possibles, car dans ce modèle, par exemple, la proposition que quelque barbier albanais rase tous et seulement les Albanais qui ne se rasent pas eux-mêmes est identique à la proposition qu'il y a des géomètres grecs que ne savent pas réaliser la trisection de l'angle : chacune est l'ensemble vide. Pourtant si ces propositions sont objet d'assertion et de négation, elles doivent certainement être distinctes. Cela dit, ce modèle donne tous les traits des

³. A la section 3.6 j'expliquerai en détail ce que j'entends par *monde possible* et des notions qui en dépendent comme celle de vérité « dans » un monde.

propositions dont j'ai besoin pour mes buts présents, et montre, en supposant la consistance des notions d'*ensemble* et de *monde possible*, que ces traits sont consistants.

Considérons maintenant l'*état du monde physique entier* (désormais l'*état du monde*) à un instant. Je laisserai cette notion largement inexpliquée, puisque mon argument est assez indépendant de son contenu. Pourvu que les deux conditions suivantes soient satisfaites, on peut concevoir 'l'état du monde' de toutes les manières que l'on voudra.

(i) Notre concept d'*état* doit être tel que, étant donné que le monde est dans un certain état à un certain instant, rien ne *suit* quant à son état à tout autre instant : si x et y sont deux « états » quelconques, et si un monde possible est dans l'état x à t_1 et dans l'état y à t_2 , il y a un monde qui est en x à t_1 et *non* en y à t_2 . Par exemple, nous ne devons pas choisir un concept d'*état* qui autoriserait comme partie d'une description d'un état momentanée du monde, la phrase '... et, à T, le monde est tel que la main gauche de quelqu'un sera levée dix secondes après T'.

La théorie de la relativité a pour conséquence que cette notion – l'état du monde à un instant – n'a pas d'application aux choses telles qu'elles sont, et est même peut-être incorrecte. Mais il y a un concept acceptable du point de vue de la théorie – l'état de choses à la surface d'un cône de lumière – qui pourrait être utilisé à sa place. Ce raffinement ne pourrait être introduit que par une élaboration pénible et philosophiquement sans pertinence (dans le contexte présent) de définitions et d'arguments qui sont déjà plus élaborés que je ne le voudrais. Je vais donc ignorer les considérations liées à la Relativité dans la suite. Le deuxième et le troisième arguments ignoreront également cette sophistication.

(ii) S'il y a un changement observable dans la manière dont les choses sont – si un vêtement blanc devient bleu, un liquide chaud froid, ou si un homme lève la main – ce changement doit impliquer un changement dans l'état du monde. C'est-à-dire que notre concept d'état ne doit pas être si théorique, si divorcé de ce qui est vrai de manière observable que, par exemple, il soit possible que le monde soit dans le même état à t_1 et t_2 , bien qu'un homme ait la main levée à t_1 et non à t_2 . On peut soutenir que ce réquisit est incompatible avec notre interprétation d'un état comme état quantique, car on peut soutenir qu'il y a un moment t et des mondes possibles w_1 et w_2 tels que w_1 et w_2 sont dans le même état quantique à t , bien qu'un certain chat soit vivant à t dans w_1 et mort à t dans w_2 ⁴. Mais je ne dirai rien sur les problèmes profonds et difficiles que soulève cette remarque.

Ayant ainsi défini – ou, en tout cas, discuté – les notions de *proposition* et d'*état du monde à un instant*, nous allons maintenant les combiner. Disons qu'une proposition *exprime l'état du monde à t* si c'est une proposition vraie qui affirme d'un état que, à t , le monde est dans cet état. (Nous pouvons formuler ce point dans les termes de notre modèle des propositions comme « ensembles de monde possibles » : une proposition p vise à exprimer l'état du monde à t s'il y a un état tel que p contient tous et seulement les mondes qui sont dans cet état à t ; p exprime l'état du monde à t si elle vise à exprimer l'état du monde à t et est, en plus, vraie.)

Enfin, j'ai besoin de la notion de *loi de la nature*. Je n'ai pas idée de la manière d'expliquer le terme, encore moins de le définir. Mais je peux dire ce que je n'entends *pas* par ce terme : comme je l'ai dit dans la section 1.3, ce n'est pas un terme d'*épistémologie*. Ontologiquement, une loi de la nature est une proposition : certaines propositions ont les traits d'une loi de la nature, d'autres pas, et ce qui les met dans l'une ou l'autre catégorie est

⁴. Ce point n'a rien à voir avec l'interprétation dite « des mondes multiples » de la mécanique quantique. Ce que j'appelle mondes possibles sont des objets abstraits, des manières dont l'univers pourrait être. Les « mondes » de l'interprétation « des mondes multiples » de la mécanique quantique sont quant à eux des objets concrets : des univers.

absolument indépendant de l'état présent de la connaissance scientifique et de l'histoire de la connaissance scientifique. Les lois de la nature seraient exactement ce qu'elles sont, même s'il n'y avait jamais eu d'êtres humains ou d'autres animaux rationnels. *Loi de la nature*, au moins dans l'usage que j'en fais, n'est pas plus un terme d'épistémologie que ne l'est *étoile*. (Ontologiquement, une étoile est un corps matériel : certains corps matériels ont les caractéristiques des étoiles, d'autres pas, et ce qui les met dans l'une ou l'autre catégorie est absolument indépendant de l'état présent de la connaissance astronomique et de l'histoire de l'astronomie. Les étoiles seraient exactement ce qu'elles sont même s'il n'y avait jamais eu d'êtres humains ou d'autres animaux rationnels.)

Malgré les parallèles que je veux dresser entre les étoiles et les lois de la nature, je veux aussi souligner une différence importante quant à nos *pouvoirs* à l'égard de ces deux classes d'objets. Il est tout à fait concevable que le pouvoir humain grandira jusqu'à ce que nous puissions altérer la course des étoiles. Mais nous ne serons jamais capables de faire quoi que ce soit sur les lois de la nature. Il y a sans doute de nombreuses propositions qui sont vraies en fait mais que nous pouvons néanmoins rendre fausses. Sans doute la proposition que personne n'a jamais lu et ne lira jamais à voix haute le *Oxford English Dictionary* en entier est vraie. J'aimerais penser, cependant, que quelqu'un *pourrait* falsifier cette proposition s'il voulait consacrer une grande part de sa vie à cette tâche inutile. Mais supposons que quelqu'un ait décidé de falsifier le Principe de Conservation du Moment Angulaire. C'est-à-dire, supposons que quelqu'un tente de construire un appareil de laboratoire dont le comportement violerait ce principe. Si le principe est une loi de la nature, il ne peut pas y arriver. S'il *peut* y parvenir (même s'il ne le fait pas), c'est-à-dire, s'il est en son pouvoir de réussir, alors ce n'est pas une loi. Nous pourrions même imaginer que le principe est en fait *vrai*, mais n'est pas une loi, simplement parce que quelqu'un peut le falsifier. Supposons qu'un certain physicien conçoive un appareil et que tous les physiciens compétents soient d'accord pour dire que si cet appareil était construit et mis en marche, il violerait le Principe de Conservation du Moment Angulaire. Supposons que les membres d'une entreprise d'ingénierie respectée, après avoir examiné les explications du physicien, concluent qu'il serait possible, étant donné l'état des connaissances, de construire cet appareil. Supposons que les physiciens et les ingénieurs ont raison. Mais supposons que la construction de l'appareil demanderait une *énorme* dépense de ressources, ce qui explique qu'il n'a jamais été construit. Nous pouvons ajouter de manière consistante à ces hypothèses celle que le moment angulaire est conservé, c'est-à-dire, que le Principe de Conservation du Moment Angulaire est *vrai*. Ce principe ne serait tout simplement pas une loi de la nature si un tel appareil *pouvait* être construit : si les êtres humains *peuvent* (ont en leur pouvoir de) conduire une expérience ou de construire un appareil qui falsifierait une certaine proposition, alors cette proposition n'est pas une loi de la nature. Une loi de la nature doit être protégée d'une telle réfutation possible. Cela est bien sûr consistant avec notre affirmation que, pour toute loi donnée, nous pourrions *concevoir* une expérience qui l'infermerait.

La conclusion de l'argument précédent peut être reformulée ainsi : les lois de la nature imposent des limites à nos capacités : elles déterminent en partie ce qu'il nous est possible de faire. Et, en fait, cette conclusion est à peine plus qu'une tautologie. La bizarrerie de sa négation peut être mise en évidence en imaginant quelqu'un qui donnerait à un subordonné l'ordre de violer une loi de la nature. Supposons qu'un bureaucrate du futur ordonne à un ingénieur de construire un vaisseau spatial capable de voyager plus vite que la lumière. L'ingénieur dit au bureaucrate que c'est une loi de la nature que rien ne peut aller plus vite que la lumière ; le bureaucrate admet la difficulté, mais recommande la persévérance : « je suis sûr », dit-il, « que si vous travaillez assez dur et êtes très intelligent, vous trouverez un moyen d'aller plus vite que la lumière, même si c'est une loi de la nature que rien ne le fait ».

Sa demande est tout simplement incohérente.

Je devrais souligner que la conclusion précédente ne repose pas sur la prémisse que les lois de la nature sont des propositions *vraies*. Je ne suis pas en train de dire qu'il est impossible pour nous d'altérer les lois de la nature simplement en raison du fait qu'il est logiquement impossible de faire qu'une proposition satisfasse simultanément les conditions 'x est une loi de la nature', et 'x est fausse' (les lois de la nature étant vraies par définition)⁵. Si cet argument était valide, on pourrait facilement dériver le fatalisme de la prémisse qu'il est logiquement impossible de faire qu'une proposition satisfasse simultanément les conditions 'x est une proposition vraie' et 'x est fausse'. Mais comme je l'ai dit au chapitre II, de nombreuses propositions vraies sont telles que l'on peut les rendre fausses : le schéma 'Si P est une proposition vraie, personne ne peut rendre P fausse' a de nombreux contre-exemples, selon moi en tout cas. Remplacer 'P' par 'la proposition que personne ne lit jamais tout le *OED* à voix haute' nous en donne un. Aucune loi de la nature n'est cependant telle que qui que ce soit puisse la rendre fausse : le schéma 'Si P est une loi de la nature, alors personne ne peut rendre P fausse' n'a *pas* de contre-exemples. Ou, dit d'une autre façon : il est vrai et trivial, et ce n'est pas ce que je suis en train de soutenir, de dire « Il est impossible qu'il y ait une personne *x* et une proposition *y* telles que *x* puisse faire que *y* soit une loi de la nature et que *y* soit fausse ». Appelez ce principe, le principe *de dicto*. Il est également vrai de dire « il est nécessaire que, pour toute personne *x* et pour toute proposition *y*, si *y* est une loi de la nature, alors *x* ne peut pas rendre *y* fausse ». Voilà qui est, au minimum, *moins* trivial et qui est ce en faveur de quoi j'ai argumenté. Appelez ce principe le principe *de re*. Du principe *de re* et de 'la proposition que le moment est conservé est une loi de la nature' nous pouvons déduire valablement 'Feynman ne peut pas rendre fausse la proposition que le moment est conservé'. Mais de cette prémisse et du principe *de dicto* aucune conclusion semblable ne suit. Remarquez que de telles applications du *modus ponens* aux instances du principe *de re* peuvent donner des vérités contingentes, puisque la proposition qu'une certaine proposition est une loi de la nature peut très bien être une vérité contingente. Feynman, par exemple, peut être incapable de rendre fausse la proposition que le moment est conservé parce que cette proposition est une loi de la nature. Mais, pour autant que je sache, il y a des mondes possibles où il existe et dans lesquels cette proposition (quoique peut-être vraie) n'est pas une loi de la nature, et où il est capable de la rendre fausse.

Si nous interprétons 'loi de la nature' de manière très large, il semble qu'il y a des exceptions au principe *de re*, et nous devons en parler. Considérons les lois *psychologiques*, y compris les lois, s'il y en a, sur le comportement volontaire des agents rationnels. S'il y a de telles lois, on peut au moins discuter de leur inclusion parmi les « lois de la nature » ; les agents rationnels sont, après tout, en un sens une partie de la « nature ». Mais il est difficile de voir comment éviter la conclusion que, si nous avons le libre arbitre, il est en notre pouvoir d'agir différemment de la façon dont de telles lois disent que nous agirons. Considérons un cas particulier.

Supposons que les psychologues découvrent que qui a reçu une éducation morale de type A dans sa prime enfance ne propagera jamais des rumeurs mensongères sur ses collègues de profession. Supposons que vous et moi avons reçu en fait une telle éducation. S'ensuit-il que nous *ne pouvons pas* nous engager dans cette odieuse activité ? Je ne vois pas pourquoi on devrait supposer que cette conclusion s'ensuit. (Mark Twain : « je suis moralement supérieur à George Washington. Il n'aurait pas pu mentir. Je le peux et ne le fais pas. ») Supposons en

⁵. Ce n'est pas que je *croie* que les lois de la nature sont vraies par définition. Voyez la section 1.5. Mais j'admettrai dorénavant que la nomicité implique la vérité. On pourrait se dispenser de cette hypothèse moyennant quelques complications mineures.

outre que vous et moi sommes en fait *capables* de répandre des rumeurs mensongères sur nos collègues. S'ensuit-il qu'un énoncé de la régularité dont nous avons imaginé que les psychologues l'avaient découverte, même si elle vraie, n'est pas une loi ? Supposons que l'existence de cette régularité est une conséquence logique d'une théorie bien confirmée du développement moral qui a un grand pouvoir explicatif et prédictif. En ce cas, il serait certainement très tentant d'appeler cet énoncé une 'loi' ; j'aurais du mal à conseiller de résister à cette tentation. « Mais pourquoi, pourrait-on demander, cette régularité du comportement se produit-elle si les gens ne doivent pas s'y conformer ? » Remarquez que seuls sont capables de s'en écarter les gens qui ont reçu l'éducation morale de type A. Peut-être est-ce que ces gens ne voient tout simplement pas l'intérêt de répandre des rumeurs mensongères. Voir l'intérêt à ne pas exercer une capacité que l'on a ne consiste pas à *perdre* cette capacité.

Il semble donc au moins plausible de supposer que le principe *de re* pourrait être faux si 'loi de la nature' était interprétée assez largement. Je vais simplement restreindre l'interprétation de 'loi de la nature' par un simple *fiat* : « lois de la nature » désignera donc par la suite les propositions qui s'appliquent de manière non vide aux choses qui ne sont *pas* des agents rationnels. (Des choses comme les tasses à thé, les électrons, et les galaxies). Pour de telles lois, je soutiens que le principe *de re* est valable. La stipulation a une conséquence importante. Il peut bien se faire, malgré tout ce qui est dit dans ce livre, que le comportement humain soit totalement prévisible sur la base de lois portant sur le comportement volontaire des agents rationnels. Bien plus, je ne vois pas de raison de penser qu'une telle prévisibilité serait incompatible avec le libre arbitre.

Faisons une dernière stipulation sur les lois de la nature : les conséquences logiques de tout ensemble de lois sont aussi des lois. La stipulation produit, en quelque manière, une notion artificielle de loi de la nature. Elle a pour conséquence, par exemple, que la conjonction de la Loi de Snell avec le Principe de Conservation du Moment Angulaire (en supposant que ces propositions sont des lois) est une loi. Si quelqu'un est gêné par cette stipulation, que je n'adopte que parce qu'elle simplifiera l'énoncé de mon argument, il peut remplacer 'les lois de la nature' dans ce qui suit par 'les conséquences logiques des lois de la nature'. Ainsi, selon notre modèle des propositions comme « ensembles de mondes possibles », une loi de la nature est tout ensemble de mondes qui a pour sous-ensemble l'ensemble de tous les mondes dans lesquels les lois de la nature sont les mêmes que celles du monde actuel, ou, pourrions-nous dire, de tous les mondes qui sont *nomologiquement congruents avec* le monde actuel.

Nous pouvons maintenant définir le 'déterminisme'. Nous appliquerons ce terme à la conjonction des deux thèses suivantes :

Pour tout instant, il y a une proposition qui exprime l'état du monde à cet instant⁶ ;

Si *p* et *q* sont n'importe quelles propositions exprimant l'état du monde à deux instants quelconques, alors la conjonction de *p* avec les lois de la nature implique *q*.

Cette définition me semble saisir au moins une thèse que l'on pourrait proprement appeler déterminisme. Le déterminisme est, intuitivement, la thèse que, étant donnés le passé et les lois de la nature, il n'y a qu'un futur possible. Et cette définition a certainement cette

⁶. Cette thèse semble impliquer qu'il y ait au moins autant de propositions que d'instant ; et il y a sans doute autant d'instant que de nombres réels. Postuler une infinité non dénombrable de propositions, dont la plupart ne sont pas des choses auxquelles on puisse penser, étant donnée leur inimaginable complexité, peut sembler extravagant à certains. Mais je ne vois pas comment formuler la thèse du déterminisme sans hypothèse « extravagante ». Voyez la critique que fait Richard Montague, dans les premiers paragraphes de « Deterministic Theories » (dans *Formal Philosophy*, New Haven, Yale University Press, 1974, p. 303 et suiv.), de la caractérisation du déterminisme qu'il attribue à Laplace et à Ernst Nagel.

conséquence. Elle a *aussi* la conséquence que le futur détermine un passé unique. Mais cette conséquence ne me gêne pas. Les seules théories *physiques* connues qui sont déterministes « du passé vers le futur » (la mécanique classique pour deux particules, et certaines théories mathématiquement similaires) sont aussi connues pour être déterministes « du futur vers le passé ». Il y a des théories, dans un sens large du mot qui ne sont déterministes que « dans un sens » - des « théories » qui décrivent le comportement de certaines machines de Turing, par exemple – mais je serais au moins un peu surpris de voir une quelconque théorie *physique* ayant cette caractéristique. Si quelqu'un est gêné par cela, il peut ajouter un « plus tard que » opportun à la définition.

Le lecteur notera que l'horrible petit mot 'cause' n'apparaît pas dans la définition. La causalité est un marais dans lequel je refuse de mettre le pied. En tout cas, tant que je n'y suis pas poussé. Certains arguments pour la compatibilité du libre arbitre et du déterminisme me forceront à dire quelque chose de la relation entre déterminisme et causalité universelle (voir section 4.4)

3.4 Le déterminisme est une thèse sur les propositions, mais la thèse du libre arbitre est une thèse sur les agents. Si nous faire porter l'enquête sur les relations conceptuelles entre ces deux thèses, nous ferions bien d'énoncer la thèse du libre arbitre comme une thèse à propos des agents et des propositions. Je propose de le faire en forgeant une manière de décrire nos pouvoirs d'agir – et, en agissant, de modifier le monde – comme des pouvoirs sur les valeurs de vérité des propositions. Cela peut être fait comme suit. Considérez les propositions que j'exprimerais si je devais prononcer n'importe laquelle des phrases suivantes à l'instant présent :

- (a) $27 \times 15 = 405$;
- (b) Les aimants attirent le fer ;
- (c) La reine Marie d'Ecosse fut mise à mort en 1587 ;
- (d) Je n'ai jamais lu *Les Enseignements de Don Juan* ;
- (e) Personne n'a jamais lu l'*Enquête* de Hume à voix haute ;
- (f) La tasse sur mon bureau n'a jamais été brisée.

(Toutes ces propositions sont, je crois, vraies.) Il y a au moins une différence importante et intéressante entre les relations que j'ai avec (a)-(c) et celle que j'ai avec (d)-(f). La différence que j'ai à l'esprit pourrait être décrite de plusieurs façons : il n'y a rien que je peux faire, ou que j'ai jamais pu faire, quant au fait que (a)-(c) sont vraies, et tel n'est pas le cas de (d)-(f) ; la vérité de (a)-(c) est quelque chose qu'il n'a jamais été en mon pouvoir de changer, bien que la vérité de (d)-(f) est quelque chose qu'il est en mon pouvoir, ou qu'il a été en mon pouvoir, de changer ; (a)-(c) sont vraies et je n'ai pas et n'ai jamais eu le choix à ce propos, mais, bien que (d)-(f) soient vraies, c'est quelque chose à propos de quoi j'ai le choix, ou quelque chose à propos de quoi j'ai eu le choix. (En faisant ces assertions sur la différence entre (a)-(c) d'un côté et (d)-(f) de l'autre, j'admets que j'ai le libre arbitre. Si je ne l'ai pas, alors cette différence apparente entre les deux ensembles d'énoncés est illusoire.)

Je ferai une telle distinction en utilisant une façon de parler introduite dans la section 2.5 : je *peux rendre* ou *aurais pu rendre* chacune des propositions (d)-(f) *fausse* ; je *ne peux pas rendre* et *n'aurais jamais pu rendre* aucune des propositions (a)-(c) *fausse*. Je préférerais le nom *être capable de rendre faux* pour désigner cette relation que j'ai avec (d)-(f) mais pas avec (a)-(c). Mais j'admets que ce nom pourrait être égarant. En fait, il a été égarant. Mon usage de cette expression a parfois créé l'impression que je crois que les êtres humains

peuvent d'une certaine façon entrer dans des relations *causales* avec les propositions⁷. Mais être capable de rendre fausse une proposition n'a rien de si métaphysiquement exotique ; être capable de rendre fausse une proposition c'est être capable d'arranger ou de modifier des objets concrets qui constituent son propre environnement – chaussures, bateaux, morceau de cire – d'une manière suffisante pour la fausseté de cette proposition.

Mais comment comprendre cette suffisance ? Nous pourrions comprendre que 'suffisant' signifie 'logiquement suffisant'. C'est-à-dire que nous pourrions comprendre que 's peut rendre *p* fausse' signifie 'Il est au pouvoir de *s* d'arranger ou de modifier les objets concrets qui constituent son environnement d'une manière telle qu'il n'est pas possible « au sens logique large » qu'il arrange ou modifie ces objets de cette façon et que *p* soit vraie'⁸. Par exemple, je peux, selon cette suggestion, rendre fausse la proposition que cette tasse n'est jamais cassée, car je peux casser cette tasse – au moins si j'ai le libre arbitre – et il n'est pas possible que je casse cette tasse et que cette tasse ne soit jamais cassée. Si je *pouvais* bouger ma main plus vite que la lumière, alors je *pourrais*, dans le sens suggéré, rendre fausse la proposition que rien ne va jamais plus vite que la lumière, car il n'est pas possible que je bouge ma main plus vite que la lumière et que rien n'aille jamais plus vite que la lumière.

Je ne crois pas que cette suggestion saisisse exactement la notion intuitive d'être capable de rendre fausse une proposition, soit la notion d'« avoir le contrôle sur » la valeur de vérité d'une proposition. Supposons qu'en 1550 Nostradamus ait prédit que le Sphinx durerait jusqu'à la fin du monde. Et supposons que cette prédiction était correcte, et qu'en fait *toutes* les prédictions de Nostradamus étaient correctes. Supposons aussi qu'il était au pouvoir de Gamal Abdel Nasser de faire que le Sphinx soit détruit. Alors, selon moi, il était au pouvoir de Nasser de rendre fausse la proposition que toutes les prédictions de Nostradamus étaient correctes. Mais ce ne serait pas le cas, selon la définition proposée dans le paragraphe précédent, puisqu'il est possible au sens logique large que Nasser ait fait que le Sphinx soit détruit et pourtant que toutes les prédictions de Nostradamus aient été correctes. C'est-à-dire qu'il y a des mondes possibles où la proposition que toutes les prédictions de Nostradamus étaient correctes est vraie et dans lesquels Nasser a fait que le Sphinx soit détruit : des mondes où Nostradamus *n'a pas* prédit que le Sphinx durerait jusqu'à la fin du monde, et n'a fait aucune autre prédiction qui aurait été falsifiée par la destruction du Sphinx par Nasser.

La meilleure manière d'exclure de tels contre-exemples semblerait être d'« introduire le passé dans » notre définition. Plus précisément, nous pouvons définir 's peut rendre *p* fausse' comme suit :

Il est au pouvoir de *s* d'arranger ou de modifier les objets concrets qui constituent son environnement d'une manière telle qu'il n'est pas possible au sens logique large qu'il arrange ou modifie ces objets d'une certaine façon et que le passé ait été exactement le même et que *p* soit vraie.

Je crois que cette définition saisit la notion suggérée par les mots 'être capable de rendre fausse une proposition', sauf, peut-être, le cas de propositions fausses portant sur le passé. Par exemple, elle a pour conséquence que je peux rendre fausse la proposition que Socrate est mort à un âge avancé, puisqu'il n'est pas possible que le passé ait été exactement le même et que Socrate soit mort à un âge avancé. Il y a trois raisons pour lesquelles ce trait de notre définition ne doit pas nous gêner cependant. Premièrement, il serait facile de l'écarter par une stipulation *ad hoc*. En outre, le premier argument pour l'incompatibilité du libre arbitre et du déterminisme n'envisagera que des propositions vraies. Enfin, le langage du 'peut rendre

⁷. Voir Jan Narveson, « Compatibilism Defended », *Philosophical Studies*, 1977.

⁸. L'expression 'possible au sens logique large' vient d'Alvin Plantinga. Voir *The Nature of Necessity*, Oxford, Clarendon Press, 1974, p. 1 et suiv.

faux' n'apparaîtra pas dans la *conclusion* de l'argument. Ainsi les conséquences curieuses de la définition ne peuvent affecter la conclusion que si elles résultent d'un défaut (comme la fausseté) dans au moins une des prémisses.

3.5 Maintenant le premier argument. Je vais imaginer une situation où un certain individu, après la délibération opportune, s'est retenu de réaliser une certaine action qu'il avait envisagée. Je soutiendrai ensuite que, si le déterminisme est vrai, cet homme *n'aurait pas pu* réaliser cette action. Parce que cet argument ne dépendra d'aucune des caractéristiques particulières à la situation imaginée, l'incompatibilité du libre arbitre et du déterminisme en général sera établi, puisque, comme ce sera évident, un argument parallèle pourrait être aisément construit pour le cas de n'importe quel agent et de n'importe quelle action non réalisée.

Supposons qu'il y a eu une fois un juge qui n'avait qu'à lever sa main droite à un certain moment, T, pour prévenir l'exécution d'une sentence de mort pesant sur un certain criminel, un tel mouvement de la main étant le signe, selon les conventions du pays du juge, que l'on accorde une forme de clémence. Supposons en outre que le juge – appelons-le 'J' – se soit retenu de lever la main à T, et que cette inaction ait conduit à la mise à mort du criminel. Nous pouvons également supposer que J était libre de tout lien, en pleine santé, et exempt de toute paralysie de ses membres ; qu'il a décidé de ne pas lever la main à T seulement après un temps de délibération calme, rationnelle et pertinente ; qu'il n'a été le sujet d'aucune « pression » pour décider dans un sens ou dans un autre de la mort du criminel, qu'il n'était sous l'influence d'aucune drogue, hypnose, ou rien de tel ; et, finalement, qu'il n'y avait aucun élément dans sa délibération qui aurait eu un intérêt spécial pour qui étudie les anomalies psychiques. Je soutiendrai que, malgré tous ces avantages, J n'aurait pas pu lever sa main à T si le déterminisme est vrai.

Mon argument en faveur de cette conclusion prendra la forme de commentaires sur les prémisses d'un « argument » dans le sens des manuels de logique : une séquence de propositions numérotées, dont toutes sauf la dernière sont les *prémisses* de l'argument et dont la dernière est la *conclusion*. Un critique a suggéré que mon usage d'un argument dans le sens des manuels de logique constituait ma tentative de fournir une « preuve formelle » pour une thèse philosophique, et a tourné en dérision la seule possibilité d'une telle entreprise⁹. J'applaudis à ses valeurs, mais je déplore son exégèse. On ne saurait donner de preuves formelles de thèses philosophiques, et je serais un insensé si je me risquais à cette entreprise. Mon critique s'est mépris sur ce qui n'est rien de plus qu'un instrument de comptabilité pour un argument, et, en fait, pour une preuve. La série numérotée de propositions ci-dessous n'est pas mon premier argument pour l'incompatibilisme. Mon premier argument pour l'incompatibilisme a plutôt la forme d'un commentaire sur les prémisses de l'« argument dans le sens des manuels de logique ». Mon argument, en fait, est coextensif avec la section 3.5 de ce livre. Je vais le distinguer de la séquence numérotée de propositions qu'il discute en appelant le premier le Premier Argument, et le second le Premier Argument Formel. Ce sont des étiquettes de pure convenance qui devraient être comprises à la lumière du présent paragraphe.

Dans le Premier Argument Formel et le commentaire que j'en donne à la suite, j'utiliserai 'T₀' pour dénoter un instant choisi arbitrairement avant la naissance de J, 'P₀' pour dénoter une proposition qui exprime un état du monde à T₀, 'P' pour dénoter une proposition qui exprime l'état du monde à T, et 'L' pour dénoter la conjonction en une seule proposition de toutes les lois de la nature. Tous ces symboles doivent être considérés comme des « désignateurs rigides ». Ainsi, si je discute certaines situations contrefactuelles ou

⁹. Naveson, *op. cit.*

possibilités non réalisées, et si j'utilise, par exemple, 'P' dans le cours de cette discussion, je veux dire 'P' pour désigner une proposition qui *en fait* exprime l'état du monde à T, qu'il exprime ou non également l'état du monde à T dans le cas où ces situations se produiraient ou dans le cas où ces possibilités seraient réalisées.

Le Premier Argument Formel consiste en sept propositions, dont la septième suit des six premières :

- (1) Si le déterminisme est vrai, alors la conjonction de P_0 et L implique logiquement P
- (2) Il n'est pas possible que J ait levé la main à T et que P soit vraie
- (3) Si (2) est vraie, alors, si J avait pu lever sa main à T, J aurait rendu P fausse
- (4) Si J avait pu rendre P fausse, et si la conjonction de P_0 et L implique P, alors J aurait pu rendre fausse la conjonction de P_0 et L
- (5) Si J avait pu rendre fausse la conjonction de P_0 et L alors J aurait rendu L fausse
- (6) J n'aurait pas pu rendre L fausse
- (7) Si le déterminisme est vrai, J n'aurait pas pu lever la main à T.

Que (7) suive de (1)-(6) peut facilement être établi par la logique des fonctions de vérité. Remarquez que les conditionnels qui apparaissent dans (1)-(7) sont des conditionnels matériels : le 'aurait pu' qui y figure a le sens d'un simple indicatif passé de 'pouvoir' (*pouvait, a pu, avait pu*). (« John n'a pas lu cette lettre, n'est-ce pas ? » « Je ne sais pas, mais j'espère que vous l'avez gardée sous clé. Vous savez quel fieffé curieux il est. S'il a pu la lire, il l'a lue »).

Un critique a suggéré que les prémisses de cet argument sont supposées être des vérités nécessaires¹⁰. Mais tel n'est pas le cas. Si les prémisses de l'argument étaient des vérités nécessaires, alors sa conclusion serait une vérité nécessaire. Mais sa conclusion est fausse dans tous les mondes possibles où le déterminisme est vrai et où J a levé sa main à T. Bien sûr, dans de tels mondes, l'une au moins des prémisses sera fausse. C'est une simple conséquence de la validité formelle de l'argument. Prenez, par exemple, un monde où L est la conjonction de toutes les lois de la nature, le déterminisme est vrai, et J a levé sa main à T. Dans un tel monde, (5) est faux, puisque son conséquent est faux et son antécédent vrai. L'antécédent de (5) est vrai dans un tel monde parce que J aurait pu faire, dans ce monde, quelque chose d'incompatible avec la vérité de la (fausse) conjonction de P_0 et L : lever sa main. Il aurait pu le faire dans ce monde, du simple fait qu'il l'a fait.

Donc, je ne dis pas que toutes les prémisses du Premier Argument Formel sont des vérités nécessaires. Mais je dis qu'elles sont vraies dans tous les mondes possibles où l'histoire que nous avons racontée à propos de J est vraie. C'est-à-dire que les prémisses du Premier argument Formel, et donc sa conclusion, *suivent de* notre histoire sur J. L'histoire de J est vraie, ou du moins nous prétendons que cette histoire est vraie. Mais elle n'est évidemment vraie que de manière contingente : il y a des mondes possibles où J a levé sa main à T ; il y a des mondes possibles où J n'a jamais existé. Et dans plusieurs de ces mondes, (7) sera fausse et l'une au moins des prémisses (1)-(6) sera donc fausse. Et c'est bien ainsi que doivent être les choses, car la conclusion du Premier Argument Formel *ne doit pas* être une conséquence logique de l'incompatibilisme seul.

Examinons maintenant les prémisses du Premier Argument Formel. Le paragraphe précédent devrait montrer que nous pouvons, dans cet examen, nous appuyer sur tous les faits

¹⁰. André Gallois, « Van Inwagen on Free Will and Determinism », *Philosophical Studies* (1977).

concernant J et sa situation qui ont été présentés dans l'histoire racontée à propos de J.

Prémisse (1). Cette prémisse suit de notre définition du déterminisme et de nos spécifications des désignations de 'P₀', 'L', et 'P'.

Prémisse (2). Le symbole 'P' est notre nom pour la proposition qui exprime ce qui était en fait l'état du monde à T, un temps où la main de J n'était pas levée. Il est donc impossible que P soit vraie si la main de J était levée à T, ou, pour tout dire, si les choses étaient différentes en quoi que ce soit à T de ce qu'elles ont effectivement été.

Prémisse (3). La phrase 'J aurait pu lever sa main à T' est ambiguë. En utilisant le système des parenthèses pour lever l'ambiguïté, comme nous l'avons introduit à la section 2.6, nous pouvons représenter cette ambiguïté comme suit : cette phrase pourrait signifier ou bien 'J aurait pu (lever sa main à T)' ou 'J aurait pu (lever sa main) à T'. J'entends que cette phrase soit comprise au premier sens¹¹.

Quant à la *vérité* de (3) : il est évident que si (2) est vraie et si J a pu lever sa main à T, alors il y a un certain « arrangement ou modification des objets concrets qui constituent l'environnement de J » - la main de J qui se lève à T – tel que (i) il n'est pas possible que J arrange les choses de cette manière et que P soit vraie, et (ii) J aurait pu arranger les choses de cette manière.

Prémisse (4). Cette prémisse est une instance, moyennant une modification du temps verbal, du principe général suivant :

Si *s* peut rendre *r* fausse, et si *q* implique *r*, alors *s* peut rendre *q* fausse.

(Remplacez '*s*' par 'J', '*r*' par 'P', et '*q*' par 'la conjonction de P₀ et L') Ce principe est une vérité triviale. Car si *q* implique *r*, la négation de *r* implique celle de *q*. Tout ce qui est suffisant au sens logique large pour la fausseté de *r* est donc aussi suffisant pour la fausseté de *q*. Par conséquent, s'il y a un arrangement d'objets que *s* peut produire, et qui est tel que sa production par *s* serait suffisante pour la fausseté de *r*, il y a un arrangement d'objets – le même arrangement – que *s* peut produire, et qui est tel que sa production par *s* serait suffisante pour la fausseté de *q*.

Prémisse (5). Cette prémisse est une instance du principe général suivant :

Si *q* est une proposition vraie qui ne concerne que des états de choses qui se sont produits avant la naissance de *s*, et si *s* peut rendre fausse la conjonction de *q* et *r*, alors *s* peut rendre *r* fausse

Considérons, par exemple, les propositions

L'Armada espagnole fut vaincue en 1588

et

¹¹. De nombreux philosophes pensent qu'une « double référence temporelle » est requise pour rendre parfaitement explicites les attributions de capacité. Ils croient donc que la « forme réelle » des attributions de capacité est du type : *s* aurait pu à *t*₁ faire A à *t*₂. Je ne suis pas convaincu par leurs arguments. Je trouve que « Tom aurait pu lever sa main à midi » est ambiguë mais claire – autrement dit, je crois que les deux choses que cette phrase peut servir à dire sont chacune claire – et 'Tom aurait pu à onze heure lever sa main à midi' me paraît plutôt étrange. Je crois que la meilleure manière de comprendre des phrases de ce type serait quelque chose comme : '*s* aurait pu à *t*₁ faire A à *t*₂' =_{df} '*s* aurait pu (faire A à *t*₂) et à *t*₁ il n'était pas trop tard pour que *s* fasse A à *t*₂'. Autrement dit, je préfère prendre des phrases comme '*s* aurait pu fait A à *t*' pour « basiques » (quoi que cela signifie exactement) et définir des phrases contenant une « double référence temporelle » au moyen de ces phrases basiques, de parenthèses qui lèvent l'ambiguïté et de la notion « pas encore trop tard ». Je ne crois pas que mon argument ait besoin de cette notion plutôt spéciale.

Peter van Inwagen n'a jamais visité l'Alaska

Pour autant que je sache, la conjonction de ces deux propositions est vraie. En tout cas, admettons qu'elle est vraie. Etant donné qu'elle est vraie, il semble tout à fait clair que je peux la rendre fausse si et seulement si je peux visiter l'Alaska. Si, pour une raison ou une autre, il n'est jamais en mon pouvoir de visiter l'Alaska, alors je *ne peux pas* la rendre fausse. C'est une assertion tout à fait triviale, et le principe général dont elle est une instance est à peine moins trivial. Et il semble incontestable que la prémisse (5) est également une instance de ce principe.

Le principe général dont (5) est une instance ne doit pas être défendu seulement par un appel à l'intuition. Supposons que l'antécédent du principe est vrai. C'est-à-dire, supposons la vérité de

- (a) q est une proposition vraie qui ne concerne que des états de choses qui se sont produits avant la naissance de s ,
- et
- (b) s peut rendre fausse la conjonction de q et r

Nous allons dériver le conséquent du principe.

Soit W (pour 'was') la conjonction en une proposition simple de toutes les propositions sur le passé. Nous avons alors, à partir de (b) et de la définition de 's peut rendre p fausse' donnée à la section 3.4 :

- (c) Il y a un arrangement possible a d'objets tel que s peut faire que a soit réalisé et tel que la conjonction de W avec la proposition que a est réalisé (appelons cette conjonction ' $W\&a$ ') implique la négation de la conjonction de q et r (" $\text{non}(q\&r)$ ").

Il suit de (a) que W implique q et donc que $W\&a$ implique q . De cela et de (c) il suit que $W\&a$ implique la conjonction de q et $\text{non}(q\&r)$. Et de là il suit que $W\&a$ implique la négation de r . Et, par conséquent, puisque s peut faire que a soit réalisé, s peut rendre r fausse.

Prémisse (6). Cette prémisse semble être une conséquence évidente de ce qui a été dit de nos pouvoirs à l'égard des lois de la nature en 3.3. Mais un compatibiliste pourrait rejeter cette affirmation. Je peux imaginer un compatibiliste argumentant ainsi :

Supposons que le déterminisme est vrai, et supposons que je ne vais pas, en fait, lever ma main dans une minute. Il s'ensuit qu'il y a un certain arrangement possible d'objets – tout arrangement qui inclut ma main levée ferait l'affaire – tel qu'il n'est pas possible que dans le sens logique large je puisse arranger des objets de cette manière alors que le passé a été exactement ce qu'il a été et alors que L est vraie. Mais puisque je *peux* lever ma main dans une minute, je *peux* donc rendre L fausse, même si, bien sûr, je ne vais pas le faire. Si ce résultat *sonne* bizarrement, ce n'est pas ma faute ; la bizarrerie vient entièrement de votre définition de 's peut rendre p fausse'.

Il y a évidemment *un* sens où je ne peux pas rendre fausses les lois de la nature : je n'ai pas le choix quant à ce que sont les lois de la nature ; il n'y a rien que je puisse *faire* à leur sujet. Il y a de nombreuses propositions dont les valeurs de vérité sont en mon pouvoir, mais les lois de la nature n'en font sûrement pas partie. Mettons donc de côté la définition de 's peut rendre p fausse' donnée à la section 3.4 et demandons-nous : les prémisses du Premier argument Formel ne sont-elles pas évidemment vraies, quelle que soit la manière dont on définit 's peut rendre p fausse' ? Notre compréhension pré-analytique de la notion du pouvoir d'un agent sur une proposition n'est-elle pas suffisamment claire pour que nous puissions tout simplement

voir que les prémisses du Premier argument Formel sont vraies ? Le compatibiliste endurci répondra vraisemblablement à ces questions de la façon suivante :

Je ne suis pas disposé à accorder ce point. Les prémisses du Premier Argument Formel peuvent être plausibles à première vue, mais cela peut être dit de nombreuses propositions fausses. Je ne peux pas dire, sans doute, laquelle de vos prémisses est fausse puisque, maintenant que vous avez écarté la définition donnée en 3.4, ce que vous entendez par *peut rendre fausse* n'est plus clair pour moi. Mais j'ai l'assurance que pour toute spécification assez précise de cette notion – à savoir une spécification aussi claire que celle donnée dans la section 3.4 – je serai capable de montrer qu'au moins une de vos prémisses est fausse quand elle est interprétée en accord avec cette définition. Si vous deviez forger une définition qui rende justice à notre conviction pré-analytique que nul ne peut rendre fausse une loi de la nature, alors je dis que je serai capable de montrer, selon les termes stricts de cette définition, que l'une des autres prémisses de l'argument – je m'attends à ce que ce soit (4) ou (5) – est fausse si le compatibilisme est vrai¹².

Mais maintenant je crois que le compatibiliste ne fait rien de plus qu'attirer notre attention sur le fait que, si le compatibilisme est vrai, alors une prémisses de notre argument est fausse. C'est-à-dire qu'il attire notre attention sur le fait que l'argument est valide.

Le lecteur qui accepte la caractérisation de *peut rendre fausse* présentée dans la section 3.4 relèvera l'ironie de la conséquence de l'argument du compatibiliste (qui est parfaitement correcte) pour la conclusion que le compatibilisme implique la fausseté de la prémisses (6) : c'est le compatibiliste, et non l'incompatibiliste, qui croit dans la « liberté contra-causale ».

Voilà qui complète ma présentation du Premier Argument. Tandis que cet argument contient un grand nombre de détails, l'idée générale en est assez simple. Considérons n'importe quel acte que quelqu'un pourrait (logiquement) avoir réalisé. S'il s'avérait que cet acte était incompatible avec l'état du monde avant la naissance de cette personne joint aux lois de la nature, alors il s'ensuit que cette personne n'aurait pas pu accomplir cet acte. Bien plus, si le déterminisme est vrai, alors *toute* déviation du cours actuel des événements serait incompatible avec n'importe quel état passé du monde joint aux lois de la nature. Par conséquent, si le déterminisme est vrai, il n'a jamais été en mon pouvoir de dévier du cours actuel des événements qui a constitué mon histoire.

Certains philosophes semblent penser que l'énoncé 'si un acte était incompatible avec l'état du monde avant la naissance d'une personne joint aux lois de la nature, alors il s'ensuit que cette personne n'aurait pas pu accomplir cet acte' *doit* être fondé sur une confusion. Mais si l'on examine un cas actuel dans lequel un certain acte est éliminé par l'état du monde avant la naissance de quelqu'un et les lois de la nature, alors cette conclusion *semble* bien s'ensuivre, et il est très difficile de voir sur quelle confusion pourrait être basée la conviction qu'il en est ainsi. Demandons-nous, par exemple, si j'aurais pu visiter l'étoile Arcturus il y a une demi-heure. Ma visite d'Arcturus à ce moment semble éliminée par l'état du monde avant que je sois né et les lois de la nature. Considérons l'état du monde une minute avant ma naissance. A ce moment, j'étais approximativement à 3.6×10^{17} mètres d'Arcturus. L'intervalle temporel séparant le moment une minute avant ma naissance du moment qui s'est produit il y a une demi-heure est d'environ 1.16×10^9 secondes. C'est une loi de la nature – en tout cas c'est ce que nous croyons à présent ; supposons que nous ayons raison – que deux objets n'ont jamais une vitesse relative supérieure à 3×10^8 mètres par seconde¹³. Il s'ensuit,

¹². Cf. Naveson, op. cit. La possibilité d'une telle ligne d'argumentation m'a d'abord été suggérée en conversation par Raymond Martin et Michael Gardner.

¹³. La vitesse de la lumière est juste inférieure à 3×10^8 mètres par seconde. La différence est minuscule,

par simple arithmétique, que je n'ai pas visité Arcturus il y a une demi-heure.

Nous avons ici un cas où la proposition que je *n'ai pas fait* une certaine chose peut être déduite de l'état du monde avant ma naissance joint à une loi de la nature. Et il *semble* très certainement suivre du fait que cette déduction est possible que je *n'aurais pas pu* faire cette chose. En tout cas, il est *vrai* que je n'aurais pas pu faire cette chose (que je n'aurais pas pu visiter Arcturus il y a une demi-heure). Y a-t-il un autre fait, hors celui que ma non-visite d'Arcturus peut être déduite de l'état du monde avant ma naissance et des lois de la nature, auquel nous devrions faire référence pour justifier notre croyance que je n'aurais pas pu visiter Arcturus il y a une demi-heure ? Je ne vois pas ce que pourrait être cet autre fait. Y a-t-il une caractéristique de cette « déduction de non-performance » qui ne soit pas une caractéristique universelle des déductions de non-performance, et sur lequel nous aurions pris appui ? En ce cas, laquelle ? Pour bien me faire comprendre, je vais construire un argument parallèle au Premier Argument Formel pour la conclusion que je *n'aurais pas pu* visiter Arcturus il y a une demi-heure. Utilisons 'T' pour désigner le moment du temps qui s'est produit il y a une demi-heure. Mettons que 'P' désigne la proposition que je n'ai pas visité Arcturus à T. Mettons que 'P₀' désigne la proposition qui exprime l'état du monde une minute avant ma naissance. (Remarquez que P₀ implique la proposition qu'à ce moment Arcturus et moi étions éloignés de 3.6×10^{17} mètres). Mettons que 'L' désigne la proposition que rien ne va plus vite que 3×10^8 mètres par seconde. Nous pouvons maintenant développer l'argument :

- (1) La conjonction de P₀ et L implique P
- (2) Il n'est pas possible que j'ai visité Arcturus à T et que P soit vraie
- (3) Si (2) est vraie, alors si j'avais pu visiter Arcturus à T, j'aurais pu rendre P fausse
- (4) Si j'avais pu rendre P fausse, et si la conjonction de P₀ et L implique P, alors j'aurais pu rendre fausse la conjonction de P₀ et L.
- (5) Si j'avais pu rendre fausse la conjonction de P₀ et L, alors j'aurais pu rendre L fausse
- (6) Je n'aurais pas pu rendre L fausse
- (7) Je n'aurais pas pu visiter Arcturus à T

Voilà qui me semble un argument parfaitement contraignant et sans exception pour la conclusion que je n'aurais pas pu visiter Arcturus à T. Quiconque pense qu'il peut démontrer qu'une des prémisses du Premier Argument Formel est fausse doit ou bien montrer que cet argument ne démontre pas également la fausseté de la prémisse correspondant de l'argument « Arcturus », ou bien il doit accepter cette conclusion et expliquer pourquoi la vérité apparente des prémisses de l'argument « Arcturus » est seulement apparente. Quelqu'un pourra peut-être faire l'une de ces deux choses, mais la perspective ne me semble pas très prometteuse.

Page \$\$\$, j'ai dit que le « compatibiliste endurci » maintiendrait vraisemblablement que, si la prémisse (6) du Premier Argument Formel est vraie, alors ou bien la prémisse (4) ou bien la prémisse (5) est fausse, ou, au moins, inconsistante avec le compatibilisme. (*Laquelle* de (4) ou (5) est fausse, selon le compatibiliste, dépendra de la façon dont on définit exactement 'peut rendre faux'). Cette thèse sur (4) et (5) est fort peu plausible quand on l'applique, *mutatis mutandis*, aux prémisses correspondantes de l'argument « Arcturus ». Le lecteur trouvera peut-être instructif de réfléchir sur ce manque de plausibilité et de se demander si cela n'affaiblit pas la thèse du compatibiliste concernant le Premier Argument Formel.

Considérons la prémisse (4). Supposons que j'aurais pu rendre P fausse, c'est-à-dire, que j'aurais pu arranger les choses de telles façons que j'ai visité Arcturus à T. Si P_0 et L impliquent que je n'ai pas visité Arcturus à T, alors j'aurais pu arranger les choses de telle façon que la conjonction de ces deux propositions soit fausse : mon arrangement des choses de manière à visiter Arcturus à T est logiquement suffisant pour la fausseté de la conjonction.

Considérons la prémisse (5). P_0 est une proposition sur l'arrangement du mobilier du monde avant ma naissance, et donc avant que je sois capable d'arranger les choses. L est une proposition dont la vérité impose une limite supérieure aux vitesses relatives des objets physiques en tout temps. Si j'avais pu faire quelque chose de suffisant pour la fausseté de la conjonction de ces deux propositions, ce serait certainement ceci : obtenir que deux objets manifestent une vitesse relative incompatible avec la vérité de L.

Les arguments pour la vérité de (4) et (5) peuvent être combinés de la manière intuitive suivante. Je crois qu'une minute avant ma naissance, l'étoile Arcturus et moi étions séparés d'une distance d'environ 3.6×10^{17} mètres. Je crois qu'entre le moment correspondant à une minute avant ma naissance et celui qui s'est produit il y a une demi-heure à peu près 1.16×10^9 secondes ont passé. Je crois que jamais deux objets n'ont été en mouvement l'un vers l'autre avec une vitesse supérieure à 3×10^8 mètres par seconde. Supposons que quelqu'un de fiable – Dieu, par exemple – me dise que, bien que ces croyances que j'ai soient vraies, j'aurais tort d'en inférer qu'il n'était pas en mon pouvoir d'avoir visité Arcturus il y a une demi-heure. En effet, Dieu dit que cela *était* en mon pouvoir. Que pourrais-je conclure de cette révélation ? Une seule conclusion me semble possible : il était en mon pouvoir de faire quelque chose qui n'a jamais été fait, à savoir de voyager (relativement à un certain objet) à une vitesse supérieure à 3×10^8 mètres par seconde. Si Dieu me disait aussi que c'est une *loi de la nature* que rien ne dépasse cette vitesse, alors je pourrais seulement conclure qu'il était en mon pouvoir de violer une loi de la nature, c'est-à-dire, de réaliser un miracle. Cela pourrait être une conclusion désagréable, mais pas autant que la conclusion qu'il était en mon pouvoir de naître une année avant ma naissance.

3.10 Le Troisième Argument est un argument modal ; autrement dit, il utilise un opérateur modal, un opérateur qui porte sur des phrases qui ont (ou qui expriment des propositions qui ont) une valeur de vérité, pour former des phrases qui ont une valeur de vérité ; et la valeur de vérité de la phrase formée de cette façon n'est pas toujours fonction de valeur de vérité de la phrase sur laquelle porte l'opérateur modal. L'opérateur est 'N'. Pour toute phrase p , mettre 'N' en préfixe à p peut être considéré comme une abréviation de l'association de 'et nul n'a ou n'a jamais eu le choix quant au fait que...' à des occurrences de p . Ainsi

N Tous les hommes sont mortels

est une abréviation de

Tous les hommes sont mortels et nul n'a, ni n'a jamais eu, le choix quant au fait que tout les hommes sont mortels

'N' est un opérateur très intéressant, et la tâche consistant à construire une logique plausible et complète pour 'N' serait intéressante. Ce serait même plus intéressant si le vocabulaire de cette logique incluait les opérateurs modaux « aléthiques » habituels. Mais il ne sera pas nécessaire à nos objectifs de construire une telle logique¹⁴. Je crois que les deux règles d'inférence suivantes doivent être des règles valides de cette logique :

(α) $\Box p \mid - Np$

(β) $N(p \supset q), Np \mid - Nq$

Je n'aurai besoin d'aucune formule contenant 'N' qui soit logiquement vraie hormis celles que ces deux règles nous permettent de déduire de l'ensemble vide. L'opérateur ' \Box ' dans la Règle (α) représente la « nécessité logique large ».

Le Troisième Argument aura la structure logique suivante : d'une conséquence logique du déterminisme (le déterminisme étant conçu comme dans le Premier Argument), nous déduirons que nul n'a jamais eu le choix à propos de quoi que ce soit.

Dans le Troisième Argument nous utiliserons une fois de plus les symboles ' P_o ' et ' L ', qui sont apparus dans le Premier Argument. Mais ces symboles auront un sens légèrement différent. Dans le Premier Argument, ' P_o ' et ' L ' étaient utilisés comme des noms pour certaines propositions. Dans le Troisième Argument, nous les utiliserons comme des abréviations de phrases exprimant ces propositions. Ainsi, dans leur usage présent, ' P_o ' et ' L ' ont les caractéristiques grammaticales de phrases et non de termes : dans l'usage du Premier Argument, ' P_o est vrai' était grammaticale et ' $P_o \supset L$ ' était incorrect grammaticalement (pour la même raison que 'Quine \supset Davidson' est incorrect) ; dans l'usage présent, ' $P_o \supset L$ ' est grammaticalement correct, et ' P_o est vrai' ne l'est pas (pour la même raison que 'Tous les hommes sont mortels est vrai' est incorrect grammaticalement).

Nous utiliserons une fois de plus la lettre ' P '. Cette lettre servira de figure à laquelle on pourra substituer toute phrase que l'on veut, pourvu qu'elle exprime une proposition vraie – 'Quine trouve que la logique modale quantifiée est problématique', par exemple, ou 'Paris est la capitale de la France'.

Si le déterminisme est vrai, il s'ensuit que

(1) $\Box (P_o \ \& \ L \ \supset \ P)$

est vraie. De (1) Nous pouvons déduire

¹⁴. Je vais suggérer une sémantique pour 'N', mais seulement comme une manière de souligner le fait que toute sémantique vraiment intéressante de cet opérateur sera sujet de controverses.

(2) $\Box (P_0 \supset (L \supset P))$

par la logique modale propositionnelle élémentaire. En appliquant la règle (α) à (2), nous obtenons :

(3) $N(P_0 \supset (L \supset P))$

Nous introduisons maintenant une prémisse

(4) NP_0

De (3) et (4) nous avons par la Règle (β) :

(5) $N(L \supset P)$

Nous introduisons une seconde prémisse

(6) NL

Alors, par (5), (6) et (β) :

(8) NP

Cette déduction montre que si le déterminisme est vrai, alors nul n'a jamais eu le choix sur rien, puisque toute phrase qui exprime une vérité peut remplacer 'P'. Considérons, par exemple, la question de savoir si quiconque a eu le choix quant au fait que Richard Nixon recevrait le pardon pour les offenses qu'il pourrait avoir commis pendant son mandat. Nixon a reçu un tel pardon. Donc, si le déterminisme est vrai

$\Box (P_0 \ \& \ L \ \supset \text{Richard Nixon a reçu le pardon pour toute offense qu'il a pu commettre durant son mandat}).$

Par conséquent, si (α) et (β) sont des règles valides, et si NL et NP_0 , alors nous avons :

Nul n'a jamais eu le choix quant au fait que Richard Nixon recevrait le pardon pour toute offense qu'il a pu commettre durant son mandat

La plupart d'entre nous, je suppose, pensent que cette conclusion est fautive. La plupart d'entre nous pensent que Gerald Ford a eu le choix quant au fait que Nixon serait pardonné. (Même si quelqu'un pense que les actions de Ford ont été complètement contrôlées par un complot, il pense quand même que *quelqu'un* a eu le choix quant au fait que Nixon soit pardonné). Mais la déduction précédente (le Troisième Argument Formel) montre que si nous souhaitons accepter cette conclusion, nous devons rejeter l'une des cinq propositions suivantes

Le Déterminisme est vrai ;

NP_0 ;

NL ;

La Règle (α) est valide ;

La Règle (β) est valide.

Mon choix, bien sûr, est de rejeter le déterminisme. Mais examinons les solutions alternatives.

Je ne vois pas comment quiconque peut rejeter 'NP₀' ou 'NL'. Mes raisons sont essentiellement celles que j'ai données en faveur de MAA et MAB dans le Second Argument et en faveur des prémisses (5) et (6) du Premier Argument Formel. La proposition P_0 porte sur le passé éloigné. Nous pourrions, si nous le voulions, stipuler que c'est une proposition sur la distribution et la quantité de mouvement des atomes et des autres particules dans la nébuleuse

primordiale. Par conséquent, personne n'a eu le choix quant au fait que P_0 . La proposition L est une proposition qui « énonce » les lois de la nature. Si c'est une loi de la nature que le moment angulaire est conservé, alors nul n'a jamais eu le choix quant au fait que le moment angulaire est conservé, et, plus généralement, puisque c'est une loi de la nature que L, nul n'a eu le choix quant au fait que L.

Je ne vois pas comment quiconque pourrait rejeter la Règle (α). Si (α) est invalide, alors quelqu'un pourrait avoir le choix quant à ce qui est nécessairement vrai. Mais quasiment personne, hormis Descartes, n'a voulu concéder une telle capacité, même à Dieu. Personne, pour autant que je sache, n'a jamais suggéré que les êtres humains pourraient avoir le choix sur ce qui est nécessairement vrai. (Je considère que les théories « conventionnalistes » de la modalité, n'impliquent pas *réellement* la thèse que les êtres humains ont le choix quant à ce qui est nécessairement vrai ; ou, du moins, pas dans un sens pertinent pour ce qui nous occupe).

Il reste à considérer la Règle (β). La validité de (β), selon moi, est la plus difficile à défendre des prémisses du Troisième Argument. Comment pourrait-on faire ? Comment, en général, montrer qu'une règle d'inférence est valide ? Il semble qu'il y a deux méthodes.

D'abord, on pourrait employer les méthodes de la sémantique formelle. Dans le cas présent, puisque 'N' est un opérateur modal, les méthodes de la sémantique des *mondes possibles* peuvent sembler prometteuses. Voici un exemple d'application de ces méthodes à (β)¹⁵. Nous délimitons d'abord un ensemble W de mondes et disons que Np est vrai dans le cas où p est vrai dans tous ces mondes. Cela équivaut à une définition sémantique de 'N'. C'est une définition du même type que celle que nous donnons en disant qu'une proposition est physiquement nécessaire si elle est vraie dans tous les mondes où valent les lois de la nature, ou quand nous disons qu'une proposition est moralement requise si elle est vraie dans tous les mondes moralement permis. Par exemple, nous pourrions dire que Np est vraie si p est vraie à la fois dans le monde actuel et dans tous les mondes non-actuels tels que les êtres humains ont ou on eu le choix quant à leur actualité. De manière intéressante, la question de l'appartenance à W n'a pas d'importance *formelle*. Si nous acceptons n'importe quelle définition de Np de la forme : ' Np est vrai si et seulement si p est vrai dans tous les mondes tels que...', où la condition qui remplit la place vide ne fait pas mention de p , alors (β) se « révélera » valide. (Évidemment si p est vrai dans tout membre de W, et si $p \supset q$ est vrai dans tout membre de W, alors q est vrai dans tout membre de W). Malgré sa trivialité, ce résultat formel n'est pas sans force persuasive. Pourtant, il est loin d'être décisif. Il dépend de l'hypothèse qu'il y a un ensemble de mondes tel que Np *peut* de façon plausible être conçu comme faisant l'assertion que p est vrai dans tout membre de cet ensemble. Bien que cette hypothèse me semble correcte, je n'ai pas d'argument en sa faveur, et qui était déterminé à rejeter (β) pourrait très bien rejeter cette hypothèse.

Deuxièmement, on pourrait tenter de montrer qu'une règle d'inférence est valide en la « réduisant » à des règles d'inférence généralement acceptées, c'est-à-dire, en déduisant sa validité de l'hypothèse que ces règles généralement acceptées sont valides. Mais il semble peu vraisemblable que cette méthode puisse être utilisée pour montrer que (β) est valide : je ne peux penser à aucune règle d'inférence acceptée qui semble pertinente pour (β), et encore moins si (β) devait en être déduite.

Les perspectives d'une *démonstration* que (β) est valide semblent donc faibles, mais peut-être pas plus que celles de démontrer quoi que ce soit qui ait un intérêt philosophique. Je dois

¹⁵. Dans la suite de ce paragraphe, je vais être moins scrupuleux qu'à l'habitude sur la distinction de l'usage et de la mention.

confesser que ma croyance en la validité de (β) a seulement deux sources, l'une n'est pas communicable, l'autre n'est pas conclusive. La première source est ce que les philosophes aiment appeler « intuition » : quand je considère attentivement (β), elle semble être valide. Mais je ne peux pas m'attendre à ce que quiconque soit très impressionné par ce fait. Les intuitions des gens, après tout, les ont conduits à accepter toute sorte de propositions insensées, et plusieurs propositions sensées, mais fausses. (Le Principe de Compréhension Non-retreinte dans la théorie des ensembles et la Loi de Galilée sur l'Addition des Vitesses en physiques, sont de bons exemples de la seconde catégorie). La seconde source est le fait que je ne peux imaginer aucune application de (β) qui ait ou puisse avoir des prémisses vraies et une conclusion fausse. Considérez par exemple ces deux applications de (β) :

Alice a de l'asthme et personne n'a, ni n'a jamais eu, le choix quant au fait qu'elle ait de l'asthme ;

Si Alice a de l'asthme, elle a parfois de la difficulté à respirer, et personne n'a, ni n'a jamais eu, le choix quant au fait que, si elle a de l'asthme, elle a parfois de la difficulté à respirer ;

donc, Alice a parfois de la difficulté à respirer et personne n'a, ni n'a jamais eu, le choix quant au fait qu'Alice ait de la difficulté à respirer

Le soleil explosera en 3000 ap. J.-C et personne n'a, ni n'a jamais eu le choix quant au fait que le soleil explose en 3000 ap. J.-C.

Si le soleil explose en 3000 ap. J.-C., toute la vie sur terre finira en 3000 ap. J.-C., et personne n'a ni n'a jamais eu le choix quant au fait que, si le soleil explose en 3000 ap. J.-C., toute la vie sur terre finira en 3000 ap. J.-C.

donc, Toute la vie sur terre finira en 3000 ap. J.-C., et personne n'a ni n'a jamais eu le choix quant au fait que toute la vie sur terre finira en 3000 ap. J.-C.

Ces arguments sont clairement valides. Les choses ne pourraient tout simplement pas être arrangées de manière à rendre les prémisses vraies et les conclusions fausses. Prenez le second argument. Il est concevable que nous fassions quelque chose pour empêcher l'explosion du soleil. Alors la conclusion de cet argument serait peut-être fausse ; mais sa première prémisse serait alors aussi fausse. Nous pourrions peut-être construire un énorme bouclier qui protégerait la terre de l'explosion du soleil ; et si nous pouvions le faire, la conclusion serait fausse. Mais la seconde prémisse serait également fausse. Peut-être pourrions-nous répandre un poison qui détruirait toute la vie sur terre avant 3000 ; mais en ce cas aussi la seconde prémisse serait fausse. Je ne peux pas m'empêcher de pressentir que le lecteur qui tentera sérieusement de construire un contre-exemple de (β) finira par apprécier, même s'il ne finit pas par partager, l'intuition que j'ai exprimé en disant, « quand je considère attentivement (β), elle semble être valide »¹⁶.

¹⁶. Thomas McKay a proposé un intéressant contre-exemple possible à (β). Nous dirions normalement que, si quelqu'un lance un dé non pipé dans un jeu honnête, nul n'a le choix quant au résultat du jet. Et la plupart d'entre nous pensent que la plupart des joueurs ont le choix quant à la pratique de leur vice. Mais alors, l'application suivant de (β) pourrait aisément avoir de vraies prémisses et une conclusion fausse :

N Alfred tire un 6

N (Alfred tire un 6 \supset Alfred joue aux dés)

Donc, N Alfred joue aux dés

(La seconde prémisse est vraie, parce que le conditionnel qui en fait partie est une vérité nécessaire). *Respondeo* : à strictement parler, Alfred a le choix quant au fait qu'il tire un 6, du moins s'il a le choix quant au fait qu'il joue aux dés. Il peut éviter de tirer un 6 en évitant de jouer aux dés. Ce sur quoi Alfred n'a pas le choix c'est le fait que, *étant donné qu'il joue aux dés*, il tire un 6. Autrement dit – en supposant que nul autre n'a le

Il est intéressant de noter que la règle (β) semble figurer dans des discussions récentes sur les implications philosophiques et sociales de la socio-biologie et sur la question de savoir si certains traits répandus du comportement humain social sont déterminés génétiquement. Voyez par exemple la proposition (P) qu'il y a certains métiers (des métiers que les deux sexes sont physiquement capables de réaliser) tels que, dans toute société, ces métiers reposent presque entièrement sur les femmes. Supposons qu'un socio-biologiste soutienne qu'il y a un certain fait ou ensemble de faits (F) concernant l'histoire évolutionnaire de notre espèce, qui explique pourquoi P est vraie. Quiconque dit cela sera vraisemblablement l'objet de critique comme celle-ci :

Ce que vous dites, c'est que le « rôle des femmes » est déterminé génétiquement, et donc que toutes les tentatives pour changer le rôle des femmes dans cette société ou tout autre sont condamnées à échouer par la biologie. Cette doctrine est pernicieuse. Vous n'êtes pas un scientifique mais un idéologue, et l'idéologie que vous colportez fait de vous un bon propagandiste du système en place.

Quiconque ferait l'objet d'une telle critique se défendrait vraisemblablement d'une manière proche de celle-ci :

Il n'en est rien. Alors que je crois que F explique pourquoi P est vrai, je ne dis pas que F rend P *inévitablement* vrai. Etant donné F, P tend à être vrai, mais les tendances peuvent être empêchées. Je ne dis pas que « la biologie est le destin ». Il peut bien se faire que nous ayons le choix quant au fait que nous allons nous comporter en accord avec la tendance que notre héritage évolutionnaire nous a laissée.

Je ne suis pas tant intéressé à savoir qui a raison dans une dispute de cette forme – après tout, cela peut dépendre de ce qui sera mis pour 'F' et pour 'P' – que je le suis par la logique qui la sous-tend. Je crois que le squelette logique de cette dispute ressemble à ceci :

Critique : Il suit de votre position que les prémisses de l'argument valide suivant sont vraies :

N F est réalisé

N (F est réalisé \supset P est vrai)

Donc, N P est vrai

Socio-biologiste : La première prémisses est certainement vraie, mais la seconde ne suit pas de ma position et pourrait bien être fausse.

Il ne viendrait probablement jamais à l'esprit du socio-biologiste de nier que la conclusion de l'argument qu'on lui attribue suive effectivement de ses prémisses. S'il le faisait, le critique l'accuserait à bon droit de sophisme, car si l'on accorde que nul n'a le choix quant au fait que, si l'histoire est telle, alors nous agissons ainsi, et s'il est accordé que notre histoire *est* bien telle et que nous n'avons pas le choix quant à ce fait, alors il s'ensuit évidemment que nul n'a le choix quant au fait que nous agissions ainsi. Ceci n'implique pas, bien sûr, que (β) soit valide, car il se pourrait qu'alors que cette application de (β) est valide, d'autres applications de (β) soient invalides. Mais je crois que quiconque dirait que, alors que l'argument du critique est valide, (β) n'est *pas* valide, dirait quelque chose de très peu plausible à première vue. La validité de (β), dans toute sa généralité, semble certainement faire partie de la « base commune » dans la dispute socio-biologique que j'ai imaginée. (Malgré le fait que je l'ai imaginée, sa structure logique est typique des disputes sur le déterminisme biologique.) C'est-

choix quant à ce fait, et que Alfred tire bien un 6 – N (Alfred joue aux dés \supset Alfred tire un 6).

à-dire que (β) semble être acceptée, à juste titre, par les deux partis de la dispute ; celle-ci semble tourner simplement sur *ce à propos de quoi* – selon la socio-biologie – nous avons le choix, et non sur la validité d’inférences qui contiennent l’idée d’avoir le choix. Les gens qui acceptent, ou sont accusés d’accepter, le « déterminisme spécial » - c’est-à-dire des théories qui disent, ou sont souvent interprétées comme disant, qu’un aspect important du comportement humain est déterminé par tel ou tel facteur hors de notre contrôle – tendent à se trouver eux-mêmes embarqués dans des disputes sur la liberté de la volonté. J’ai pris un socio-biologiste comme exemple d’une telle personne parce que j’écris maintenant, mais le débat sur « la biologie est notre destin » n’est pas le seul de ce genre. Si j’avais écrit il y a quelques années, j’aurais construit un exemple avec un Freudien ou un Marxiste et le point mis en valeur par mon exemple aurait été le même. Quiconque refuse la validité de (β), me semble-t-il, doit réagir à ces débats de l’une des deux manières suivantes. Il doit ou bien considérer que ce sont de pseudo-débats qui surgissent parce que les deux partis de la dispute ont le tort d’accepter (β), ou il doit montrer que la dispute ne repose pas vraiment sur (β) après tout. Dans le premier cas, il devrait recommander au socio-biologiste de répondre au critique comme ceci :

J’admets que F est un fait concernant l’histoire de notre espèce, et que c’est quelque chose sur quoi nul n’a le choix. J’admets que *si* F est un fait concernant l’histoire d’une espèce donnée, alors, dans les sociétés qui comprennent des membres de cette espèce, certains métiers reviendront presque uniquement aux femmes, et que nul n’a le choix quant à *cela*. Pourtant, nous avons le choix quant au fait que, dans notre société, ces métiers reviennent presque uniquement aux femmes.

Je ne peux imaginer personne qui dise cela sans ciller. Dans le second cas, il nous doit une interprétation de ces disputes qui montrent que l’apparente acceptation de (β) par les deux partis n’est qu’une apparence.

Le point de cette discussion peut être résumé par une simple question : Pourquoi aucun des participants aux débats sur le déterminisme biologique n’est-il compatibiliste ? La réponse est peut-être que les participants à ces débats prennent l’idée de déterminisme biologique beaucoup plus au sérieux que les philosophes ne sont enclins à prendre au sérieux les idées de déterminisme ‘universel’ ou ‘Laplacien’, et que le compatibilisme à l’égard d’un type particulier de déterminisme n’est possible que pour les gens qui ne prennent pas ce type de déterminisme très au sérieux¹⁷.

¹⁷. Il y a un argument concernant le fatalisme qui est l’exact parallèle de cet argument à propos du déterminisme biologique. Nous pouvons formuler ce qu’est exactement l’argument pour le fatalisme examiné à la section 2.5 de manière à faire apparaître explicitement l’usage de la règle (β). Soit S la proposition (vraie) que je me raserai demain, et soit t un moment du passé éloigné. Utilisant lui-même la règle (β), le fataliste argumente ainsi :

N S était vrai à t
 N (S était vrai à $t \supset$ je me raserai demain)
 Par conséquent, N je me raserai demain

A la section 2.5, j’ai soutenu que la première prémisse de cet argument ou bien n’avait pas de sens, ou bien, moyennant une certaine stipulation sur ce qu’elle pourrait signifier, était douteuse. (Evidemment, un tel coup ne nous est pas permis relativement aux arguments de ce chapitre. La proposition que P_0 est réellement à propos du passé et par conséquent à propos de ce qui échappe à notre contrôle. La proposition que S était vraie à t est faite pour qu’un tour linguistique la présente comme étant à propos du passé. C’est pourquoi le problème concernant le fatalisme est un problème qui doit être dissous par une attention soigneuse au langage du temps et de la vérité, tandis que le problème du libre arbitre est profond, intraitable et qu’il rend fou). C’est une thèse à propos de laquelle le fataliste et moi pourrions avoir une discussion philosophique sérieuse. Mais supposons qu’au lieu de dire cela, j’avais simplement récusé la validité de l’instance de (β) ci-dessus. Alors, je crois, le fataliste aurait pu m’accuser à juste titre de refuser d’apporter une contribution sérieuse à la discussion, comme l’aurait fait n’importe quel anti-fataliste qui aurait un réel intérêt pour atteindre le fond du problème du fataliste.

J'ai dit plus haut que je ne pourrais penser à aucune instance de (β) qui aurait, ou pourrait avoir, des prémisses vraies et une conclusion fausse. Je voulais dire, bien sûr, que je ne pourrais penser à aucune instance de (β) qui pourrait être tenue pour avoir des prémisses vraies et une conclusion fausse indépendamment de la question de savoir si le libre arbitre est compatible avec le déterminisme. Si le libre arbitre est compatible avec le déterminisme, et si le déterminisme est vrai, alors, sans doute, au moins l'une des deux instances suivantes de (β) a de vraies prémisses et une conclusion fausse :

$N(P_0 \supset (L \text{ Nixon} \supset \text{a été pardonné}))$

$N P_0$

Par conséquent, $N (L \supset \text{Nixon é été pardonné})$

$N (L \supset \text{Nixon a été pardonné})$

$N L$

Par conséquent, $N \text{ Nixon a été pardonné}$

Mais il serait mieux de trouver un contre-exemple à (β) qui ne présuppose pas la compatibilité du libre arbitre et du déterminisme. Après tout, les exemples que j'ai donnés en faveur de (β) ne présupposaient pas l'*incompatibilité* du libre arbitre et du déterminisme. Je penserais plutôt que s'il y avait des contre-exemples à (β) , on pourrait en montrer certains, au moins, indépendamment de la question de savoir si le libre arbitre et le déterminisme sont compatibles.

Cela peut être difficile à croire, mais il y a presque certainement des philosophes qui diraient que le fait que le déterminisme et le compatibilisme pris ensemble impliquent qu'il y a des contre-exemples à (β) montre que mon usage de (β) pour un argument en faveur de l'incompatibilisme commet une pétition de principe. Mais si cette accusation était correcte, il est difficile de voir comment un argument pourrait éviter de commettre une pétition de principe. Si l'on présente un argument en faveur d'une certaine proposition, alors, si cette proposition est fausse, *certaines* étapes du raisonnement sont incorrectes (je compte l'usage d'une fausse prémisse comme une étape incorrecte) ; et on peut croire d'une étape particulière de l'argument que *si* une étape est incorrecte, *celle-là* l'est. Mais il ne s'ensuit pas pour autant que l'on « commet une pétition de principe » du fait de cette étape. On peut commettre une pétition de principe – quoi que ce soit exactement – mais que l'on commet une pétition de principe n'est pas une conséquence de la simple existence d'un chaînon faible dans la chaîne de raisonnement de quelqu'un.

Ces questions sur la pétition de principe et sur la charge de la preuve sont très difficiles. Considérons les sous un autre angle. Supposons qu'un compatibiliste dise quelque chose comme : « Le troisième Argument dépend de la validité de (β) . Mais cette règle est invalide. Je le prouve ainsi :

Le Compatibilisme est vrai

Par conséquent, (β) est invalide.

Vous admettez vous-même que la conclusion de cet argument suit de ses prémisses » (comme moi). « Vous pouvez ne pas accepter ses prémisses, mais c'est alors *votre* problème, car cette prémisse est *vraie*. Bien plus, vous pouvez difficilement objecter à mon petit argument en tirant parti du fait qu'il commettrait une pétition de principe. Il n'est pas pire, à cet égard, que *votre* argument, qui revient essentiellement à ceci :

(β) est valide

Par conséquent, le Compatibilisme est faux. »

Que répondre à cela ? Je suppose que je ne peux pas faire plus qu'en appeler aux intuitions des lecteurs. Voici ce qu'il m'en semble : Si deux principes sont en conflit, que l'un d'eux semble évidemment correct mais pas l'autre, alors, si l'on doit choisir, on devrait accepter celui qui semble évidemment correct.

Supposons-donc que le compatibiliste dise qu'il trouve le compatibilisme évidemment correct. Sans doute cette assertion est-elle ou bien fondée sur une relation immédiate et intuitive avec le compatibilisme – il prétend *voir* que celui-ci est vrai, tout comme je prétends voir que (β) est valide – ou bien fondée sur un argument en faveur du compatibilisme. Considérons le cas du philosophe qui prétend voir la vérité du compatibilisme intuitivement. Les arguments, comme les explications, doivent finir quelque part. S'il y a un tel philosophe, lui et moi formons peut-être un véritable cas de conflit ultime d'intuitions. Mais je dois dire que je trouve impossible de croire une thèse comme celle que je viens d'imaginer. Le compatibilisme me semble être le genre de choses que l'on ne peut croire que si l'on a un argument en sa faveur. Je ne peux simplement pas voir ce qui se passerait dans la tête de quelqu'un qui prétendrait le connaître intuitivement. Je ne sais pas ce que cela ferait de le penser.

Le philosophe qui accepte le compatibilisme sur la base d'un *argument* ne recèle pas pour moi le même mystère. Mais je voudrais savoir quelles sont les prémisses de son argument. Et je poserai la question suivante à leur propos : sont-elles intuitivement plus plausibles que (β) ? Je trouve difficile de croire qu'il y ait des propositions qui impliquent le compatibilisme et qui soient plus plausibles que (β). Mais qui sait. Il y a en a peut-être. Comment faire autrement que de procéder cas par cas ? Dans la section 4.3, nous allons comparer (β) avec les prémisses d'un argument populaire pour le compatibilisme, l'argument que j'ai appelé 'Argument de l'Analyse Conditionnelle' au chapitre 1¹⁸.

Voilà qui complète ma présentation du Troisième Argument, et ma défense de ses prémisses. J'ai donné trois arguments pour l'incompatibilité du libre arbitre et du déterminisme. Si le compatibiliste souhaite réfuter ces arguments – et, bien sûr, rien ne l'oblige à le faire – voici ce qu'il devra accomplir : il devra produire un ensemble de propositions intuitivement plus plausibles que les prémisses de ces arguments ou bien il devra former des arguments en faveur de la fausseté de certaines des prémisses employées dans ce chapitre, des arguments qui puissent être évalués et dont la correction puisse être constatée indépendamment de la question de savoir si le libre arbitre et le déterminisme sont compatibles.

3.11 Certains compatibilistes, quand ils sont confrontés aux arguments en faveur de l'incompatibilité du libre arbitre et du déterminisme, disent quelque chose comme : « votre argument démontre simplement que quand vous utilisez une expression comme 'pouvoir agir autrement', ou 'avoir le choix', vous leur donnez une signification différente de celle qu'elles ont dans nos débats sur la responsabilité morale ». Cette critique est également applicable, *mutatis mutandis*, aux trois arguments en faveur de l'incompatibilisme. Et ma réponse est essentiellement la même dans chaque cas. Mais cette réponse peut être présentée de manière très compacte et efficace dans les termes du vocabulaire employé dans le Troisième Argument. Par conséquent, je répondrai à cette seule accusation : « Quand vous utilisez

¹⁸. Au chapitre IV, le lecteur s'en souviendra, nous examinerons deux autres arguments pour le compatibilisme, l'« Argument du Cas Paradigmatique » et l'« Argument *Mind* ». Je ne comparerai pas les prémisses de ces arguments avec (β). On verra que la version ultime de l'Argument *Mind* – il y en a trois – compte la validité de (β) parmi ses prémisses.

l'expression 'avoir le choix' vous lui donnez un sens différent de celui qu'elle a dans les débats sur la responsabilité morale », et je laisserai au lecteur la tâche mécanique d'adapter cette réponse aux besoins des Premier et Deuxième Arguments. Ma réponse réside simplement dans une réinterprétation de 'N' :

$Np =df p$ et, dans le sens de *avoir le choix* qui est pertinent dans les débats sur la responsabilité morale, nul n'a ni n'a jamais eu le choix quant au fait que p .

S'il y a quelque chose dans l'objection que nous considérons, alors des quatre propositions suivantes

NP_0 ,

NL,

La Règle (α) est valide,

La Règle (β) est valide,

l'une au moins est fausse, étant admis que 'N' – qui a une occurrence dans (α) et (β) – est interprété comme ci-dessus. Mais cela ne semble pas être le cas. Si l'on reprend soigneusement les étapes du Troisième Argument, on trouvera, je crois, qu'aucune étape ne devient douteuse dans notre nouvelle interprétation de 'N'. La conclusion du présent chapitre est donc que si le déterminisme est vrai, alors nul n'a le choix sur rien, exactement dans le sens de *avoir le choix* qui est pertinent pour les débats sur la responsabilité.