

Le réalisme modal de David Lewis nous condamne-t-il à la souffrance éternelle ?

Jiri Benovsky*
(Université de Fribourg)

L'œuvre philosophique de David Lewis se distingue par l'importance qu'elle accorde au sens commun et au respect de nos intuitions préthéoriques et métathéoriques, mais aussi pour les conséquences de théories philosophiques qu'elle tire à propos de nos croyances ordinaires. Soutenue par son rejet général de l'existence d'entités abstraites (non spatio-temporelles), sa théorie réaliste de mondes possibles veut préserver elle aussi sa maxime méthodologique « Ne défendez jamais une théorie philosophique que vous ne pouvez pas vous-même croire dans vos moments les moins philosophiques et les plus proches du sens commun » (Lewis, 1986, p. 135). En effet, sa préférence pour un principe de rasoir d'Occam qualitatif (« Il ne faut pas multiplier le *type* d'entités au-delà de ce qui est nécessaire. ») par opposition à un principe quantitatif (« Il ne faut pas multiplier le *nombre* d'entités au-delà de ce qui est nécessaire. ») fait partie des motivations préthéoriques et métathéoriques qui soutiennent l'édifice de sa pluralité de mondes concrets et poussent Lewis à rejeter les théories ersatzistes concurrentes qui font appel à des entités abstraites ainsi qu'à des formules « magiques » (Lewis, 1986, p. 174-191) pour expliquer la notion de représentation, centrale pour ce type de théories.

Vers la fin de sa vie, David Lewis était réellement préoccupé par les conséquences que pouvait avoir l'une des interprétations philosophiques de la mécanique quantique sur notre vie de tous les jours. En effet, lors d'une conférence à Canberra en 2001, intitulée « How many lives has Schrödinger's cat ? », publiée de manière posthume (Lewis (2004), il s'interroge sur les implications pour notre vie (notre survie !) de l'interprétation Everettienne – connue sous le label de « mondes

* Jiri Benovsky est collaborateur scientifique à l'Université de Fribourg (Suisse). Il est auteur de trois livres (*Persistence through time and across possible worlds*, Ontos Verlag, 2006 ; *Le Puzzle Philosophique*, Editions d'Ithaque, 2010 ; *Qu'est-ce qu'une photographie?*, Vrin, 2010) et de nombreux articles. Ses domaines de spécialisation et de compétences incluent la métaphysique, la philosophie du temps, les modalités et mondes possibles, le vague, l'identité personnelle, la méta-ontologie et méthodologie, philosophie de l'expérience temporelle, ainsi que la représentation pictorielle et la philosophie de la photographie. La page web académique de Jiri Benovsky se trouve à l'adresse www.jiribenovsky.org

multiples » – de la mécanique quantique, avec une conclusion effrayante : nous sommes condamnés à vivre une éternité pleine de souffrance. Dans ce qui suit, je vais m'intéresser à cette préoccupation de Lewis, l'expliquer, et l'analyser aussi bien sous l'hypothèse d'une structure de « mondes multiples branchants », semblable à celle que nous offre l'interprétation d'Everett, que sous l'hypothèse de Lewis lui-même de mondes possibles non-branchants. Je vais suggérer que si l'argument de Lewis était bon, il s'appliquerait non seulement au cas d'une structure branchante de mondes possibles, mais également à toutes les théories de mondes possibles, y compris celle de Lewis lui-même, ce qui montre que l'argument, tel que Lewis le formule, présente (heureusement !) un défaut qui nous permet de s'apercevoir que les craintes pour notre futur qui le préoccupaient ne sont pas bien fondées.

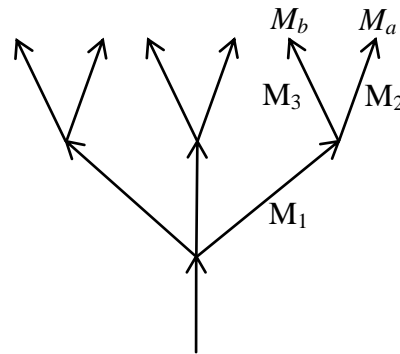
Dans ce qui suit, je vais d'abord discuter et comparer deux ontologies réalistes de mondes possibles : celle de Lewis lui-même, dite de mondes *divergents*, et celle d'une structure *branchante*. Nous verrons également comment comprendre l'actualité dans ces deux cadres, et de quelle manière on peut y concevoir l'identité personnelle – en particulier, je discuterai la théorie lewisienne des contreparties ainsi que sa cousine adaptée à une structure branchante de mondes possibles. Ces acquis nous permettront alors de bien comprendre comment évaluer des énoncés à la première personne à l'intérieur de ces deux versions réalistes de la théorie des mondes possibles, ce qui sera nécessaire pour pouvoir ensuite évaluer l'argument de Lewis concernant l'interprétation Everettienne de la mécanique quantique. Nous verrons que, *contra* Lewis, l'hypothèse branchante n'y joue aucun rôle, et que l'on peut aisément échapper à sa terrifiante conclusion.

David Lewis aborde rapidement la notion d'une structure branchante de mondes possibles dans *On the plurality of worlds* (Lewis, 1986, p. 206-208), sans la défendre, mais seulement pour la comparer à sa propre structure de mondes possibles « divergents ». Cette dernière est une structure statique : les mondes possibles sont ici conçus comme étant causalement et spatio-temporellement disjoints, il ne partagent aucune partie en commun, et de ce fait il y a un très grand nombre de mondes possibles qui contiennent des segments initiaux qui sont des duplicatas qualitatifs. En effet, il y a ici de nombreux mondes qui sont qualitativement identiques pendant des milliards d'années jusqu'au moment où une petite différence apparaît entre eux. Par exemple, il se pourrait que je vienne de taper « pourrait » sur mon ordinateur, évitant ainsi une faute d'orthographe dans mon manuscrit, et il existe donc un monde possible qualitativement identique au monde actuel dont l'histoire diverge uniquement à partir du

moment où je fais cette erreur de frappe – jusque là, les deux mondes sont des duplicatas parfaits. Bien entendu, il ne s’agit pas uniquement de deux mondes : il y a de nombreuses manières dont j’aurais pu commettre (ou pas) une erreur de frappe, et de façon générale il y a de nombreux événements qui auraient pu se produire autrement dans le monde actuel à ce moment, et il y a ainsi un nombre très grand (une infinité) de mondes possibles existant indépendamment les uns des autres mais étant des répliques parfaites qui divergent uniquement à partir de cet instant.

On a souvent reproché à Lewis de postuler un nombre d’entités plus grand que celles que nous croyons communément exister – en effet, selon Lewis, il y a des ânes qui parlent (lorsque le quantificateur existentiel est utilisé de manière non-restreinte). Tout comme Lewis lui-même, je ne suis aucunement impressionné par ce reproche confus : les théories ersatzistes concurrentes postulent autant d’entités abstraites que Lewis postule d’entités concrètes (le rasoir d’Occam quantitatif ne procure donc d’avantage à aucun des partis), et la théorie lewisienne, n’endossant que l’existence d’entités concrètes, est plus respectueuse du rasoir d’Occam qualitatif. Mais s’il est incorrect d’accuser Lewis d’être ontologiquement trop dépensier ici, il est en revanche vrai, comme nous venons de le voir, qu’il postule un très grand nombre de mondes possibles qui sont en grande partie des duplicatas qualitatifs, et s’il y a donc une dépense ontologique inutile dans le système lewisien, c’est bien celle-ci. C’est ici que l’ami d’une structure branchante de mondes possibles trouve le bon moment pour faire son entrée en scène.

En effet, une structure branchante de mondes possibles est telle qu’un monde vient à exister seulement lorsque son existence est nécessaire pour rendre compte d’une possibilité, évitant ainsi de dupliquer des segments initiaux qualitativement identiques. Sous cette hypothèse, tous les mondes possibles partagent le même commencement, mais ensuite ils évoluent différemment : à chaque instant de l’histoire d’un monde, il y a de nombreuses possibilités qui peuvent se réaliser et qui correspondent à toutes les manières possibles dont le monde peut évoluer, et à chacun de ces instants le monde subit alors une fission donnant lieu à des branches différentes qui réalisent ces diverses alternatives. Ainsi, une structure branchante de mondes possibles évite de postuler de nombreux duplicatas de mondes « parallèles » (ceci est une métaphore, car « parallèle » pourrait faussement suggérer l’existence d’une relation spatiale entre les mondes) et en ce sens il s’agit donc d’une structure dynamique car le nombre de mondes possibles augmente avec le temps.



Une telle structure de mondes possibles est ici conçue en termes perdurantistes, proche de la manière dont Lewis lui-même conçoit la persistance de particuliers à travers le temps. En effet, un monde (tel que M_a , M_b , ...) n'est ici rien d'autre qu'un gros particulier qui est constitué de parties temporelles qui sont des segments temporels du monde (tels que M_0 , M_1 , M_2 , pour M_a ...) – un monde est un agrégat ordonné de parties temporelles qui forment une branche complète et qui ont une origine commune. Tout comme sous l'hypothèse lewisienne de mondes possibles divergents, les mondes possibles branchants ont tous le même statut ontologique : toutes les branches sont également réelles et concrètes, y compris le monde actuel qui se trouve être l'une de ces branches. L'actualité n'est pas ici conçue comme étant une propriété absolue du monde actuel, car tout comme sous l'hypothèse divergente, « actuel » doit être ici compris comme étant un terme indexical, pareillement que « ici », « maintenant » ou « je », qui fait référence simplement au monde dans lequel l'énoncé qui contient ce terme a lieu (« le monde actuel » veut simplement dire « notre monde »). Chaque habitant de chaque monde peut ainsi dire avec vérité : « je suis actuel, et j'habite le monde actuel ».

Un habitant d'un monde, selon la théorie des contreparties de Lewis, ne peut habiter qu'un seul monde – il n'y a jamais deux objets dans deux mondes qui seraient numériquement identiques, ou plus exactement : aucun objet n'existe dans deux mondes. Cette idée peut également être adaptée au cas de la structure branchante réaliste de mondes possibles. Mon objectif est ici non seulement de développer une théorie des contreparties viable en principe, mais également de rapprocher la structure branchante des mondes possibles autant que possible de la vision de Lewis lui-même afin que mon argument qui suivra puisse être pertinent *per se* mais également *ad hominem*.

La théorie des contreparties lewisienne est fondée sur la relation de similarité : x est une de mes contreparties dans un autre monde s'il s'agit d'un individu qui, dans ce monde, me ressemble en ce qui concerne mes

traits les plus importants et qui m’y ressemble davantage que tout autre objet de ce monde. Nous sommes deux individus numériquement différents et nous sommes causalement et spatio-temporellement isolés, mais nous sommes les contreparties l’un de l’autre en vertu de notre ressemblance et c’est en vertu de cette ressemblance que l’individu habitant un autre monde que moi *est* ce que je *serais* si le monde actuel (mon monde) était différent.

Cette théorie des contreparties modales avait subi des attaques qui ont été prises très au sérieux ; je parle ici des difficultés soulevées par Kripke (1972, p. 45) et Plantinga (1973, p. 164-165) qui soutiennent que puisque la théorie des contreparties est une théorie d’individus *unimondains* (qui n’existent que dans un seul monde), elle manque son but, car elle ne nous fournit pas une bonne explication de ce qui pourrait être le cas ou ne pas être le cas pour *nous* (pour les particuliers actuels) ; en effet, les individus qui sont nos contreparties sont des gens qui ne sont reliés à nous d’aucune autre manière que par le simple fait qu’ils/elles nous ressemblent. Ainsi, selon les objecteurs, l’explication de ce qui pourrait m’arriver à *moi* n’est pas du tout à propos de *moi*, mais seulement à propos de quelqu’un d’autre qui habite un autre monde et qui me ressemble. En conséquence, notre croyance que *nous* avons certaines propriétés de manière seulement contingente et qu’il y a de nombreuses manières dont les choses auraient pu être différentes pour *nous* n’a pas été expliquée. Si je viens tout juste d’échapper à un accident mortel, il est central pour mes croyances et ma sensation de soulagement de croire que c’était *moi* qui aurait pu mourir – or, la théorie des contreparties ne semble pas être à même de rendre compte de cette croyance centrale car, d’après cette théorie, l’énoncé selon lequel j’aurais pu mourir n’est pas du tout à propos de moi ! Il s’agit d’un énoncé qui parle de quelqu’un d’autre qui est mort et qui me ressemblait, mais *qui n’est pas moi*. Ainsi, il semble que notre conception des modalités *de re* est violée par la théorie des contreparties qui est inadéquate pour rendre compte de nos croyances.

Cette objection a depuis perdu un peu de sa force, non seulement suite aux réponses que Lewis lui-même y a faites (Lewis (1986, p. 196), mais également suite au travail de ceux qui défendent une théorie de contreparties *temporelles* (en particulier Sider (1996, 2001, à paraître) et au travail d’ersatzistes qui défendent la théorie des contreparties modales mais sans endosser le réalisme modal lewisien (en particulier Heller (1998a, 1998b)). Pour ma part, dans ce qui suit, je ne souhaite pas prendre parti en faveur du défenseur ou de l’objecteur. Ce qui m’intéresse ici, et qui va être pertinent plus loin, est le fait que les *défenseurs* de la théorie des contreparties (modales et/ou temporelles) pensent que, contrairement à ce

qu'affirme l'objecteur, cette théorie *est* vraiment à propos de *nous* et de *nos* propriétés.

Ted Sider, par exemple, lorsqu'il défend sa théorie de la persistance à travers le temps, à savoir la « stage view »¹ qui fait appel de manière centrale à la théorie des contreparties temporelles, présente sa réponse à l'objection ainsi :

« Il est faux de dire que la *stage view* nie que « Tu vas faire ceci ou cela » veut dire que *tu* vas faire ceci ou cela. « Ted était autrefois un petit garçon » attribue une propriété temporelle, la propriété d'avoir été un petit garçon, à *moi* et à personne d'autre. Bien sûr, la *stage view* fournit une analyse de mon exemplification de cette propriété en termes d'un autre objet qui est un petit garçon [c'est-à-dire en termes de ma relation à l'une de mes contreparties temporelles, numériquement distincte de moi], mais *je* suis celui qui a cette propriété temporelle, et c'est cela qui compte. La *stage view* est consistante avec le fait que les différentes contreparties temporelles ont des propriétés temporelles; c'est juste que la possession de propriétés temporelles est analysée en termes de contreparties. » (Sider, 2001, p. 195, mes italiques, ma traduction)

Lewis lui-même résume sa réponse à l'objection ainsi :

« [...] il y a d'autres mondes (véritables ou ersatz) selon lesquels Humphrey – *lui-même* ! (tapez du pied, tapez du poing sur la table) – gagne l'élection. [...] Humphrey – *lui-même* – n'est pas une partie de ces autres mondes. D'une manière ou d'une autre, peut-être parce qu'ils contiennent des composants adéquats ou peut-être par magie, mais en tout cas pas parce qu'ils contiendraient Humphrey lui-même, les autres mondes *le représentent lui-même* comme gagnant l'élection. » (Lewis, 1986, p. 196, mes italiques, ma traduction)

Le point sur lequel je veux attirer l'attention ici, je le répète, est que les *défenseurs* de la théorie des contreparties croient que l'objection a été traitée de manière satisfaisante et que les énoncés modaux *de re* à propos de

¹ L'étiquette anglaise « stage view », devenue standard en anglais, désigne ici la théorie de la persistance d'objets matériels à travers le temps selon laquelle un objet existe uniquement à un instant, et jamais à plus d'un instant. Un objet persiste à travers le temps non pas en existant à plusieurs instants (endurantisme), ni en ayant des parties temporelles à différents instants (la « worm view » perdurantiste), mais en entretenant une relation de contrepartie temporelle avec d'autres objets (numériquement distincts) à d'autres instants. Un objet comme une table ou un âne est ainsi appelé « stage », c'est-à-dire une « étape » ou « tranche » temporelle, mais cette appellation est trompeuse car elle suggère qu'il s'agit d'une « étape » ou « tranche » *de quelque chose* – ce qui n'est, justement, pas le cas ici, car la totalité de l'objet 'se résume' à ce qui existe à un instant donné. Je n'envisage pas donc ici de tenter une traduction de cette étiquette anglaise, trompeuse et inexacte.

moi sont réellement à propos de moi, même si leurs conditions de vérité impliquent des objets numériquement distincts de moi.

La théorie des contreparties peut être adaptée également à la structure de mondes possibles branchants. Rappelons-nous le schéma de cette structure ci-dessus. On pourra alors dire qu'un individu qui existe dans une partie temporelle d'un monde, par exemple M_2 sur le schéma, est une contrepartie d'un autre individu qui existe dans une autre partie temporelle d'un monde, par exemple M_3 , si et seulement si les deux individus ont été un seul et même individu dans la partie temporelle M_1 qui avait, par fission, donné lieu à M_2 et M_3 . L'idée ici est que ce qui est vrai des mondes est également vrai des individus, car après tout les mondes ne sont rien d'autre que de gros individus.

Pour reprendre l'exemple que nous avons à l'esprit, disons que je suis un individu dans M_1 , disons que l'individu dans M_2 est celui qui aura écrit « pourrait » au lieu de « aurait » et disons que l'individu dans M_3 saura éviter cette faute. M_2 et M_3 résultent de la fission de M_1 , et il en est de même pour les occupants de ces parties temporelles de mondes : l'individu dans M_1 aura fissionné en deux, donnant lieu aux individus de M_2 et M_3 . De cette manière, les deux individus de M_2 et M_3 , l'un commettant une faute l'autre pas, sont des contreparties l'un de l'autre car ils ont tous les deux été le même individu avant la fission (dans M_1).

Afin qu'une telle idée soit plausible, il faut ici aussi l'entendre d'une oreille perdurantiste, proche de la vision de Lewis lui-même : il y a ici en réalité *deux* individus temporellement étendus et composés de parties temporelles ; le premier, appelons-le « Individu_a » (similairement au monde appelé « M_a »), est l'agrégat de parties temporelles d'individus existants dans M_1 , M_2 , ... ; le second, appelons-le « Individu_b », est l'agrégat de parties temporelles d'individus existants dans M_1 , M_3 , Individu_a et Individu_b *ont une partie en commun* (à savoir, leur partie temporelle de M_1). C'est de cette manière que Lewis comprend les cas de fission d'individus et elle est ici aisément adaptée pour faire fonctionner la théorie des contreparties dans une structure branchante de mondes possibles.

En ce qui concerne l'objection à la théorie des contreparties que nous avons vu ci-dessus, cette théorie des contreparties modifiée et adaptée à la structure branchante de mondes possibles semble avoir un avantage par rapport à la théorie des contreparties de Lewis et sa structure divergente statique de mondes. Il semble en effet que si je viens d'éviter de justesse un accident mortel, l'énoncé qu'il a été possible pour *moi* de mourir il y a un instant est réellement à propos de *moi*. Rappelons-nous ce que dit Kripke et considérons un cas formulé à la première personne et du point de vue

d'avant l'accident (ceci sera également utile plus tard) : supposons que je décide d'égaliser l'exploit de Ueli Steck de gravir la face nord de l'Eiger en solo en 2 heures et 47 minutes. Etant donné mon manque d'entraînement et de capacités physiques et techniques, il est non seulement très probable que je n'y arrive pas mais il est également possible que je meurs suite à une chute fatale en faisant cette folle tentative. Toutefois, il *est métaphysiquement possible* que j'arrive au sommet sain et sauf, après exactement 2 heures et 47 minutes d'ascension. Selon la théorie des contreparties adaptée au système branchant de mondes, lorsque je dis tout cela (qu'il est possible que je meure, ou qu'il est possible que j'arrive au sommet), c'est en effet bien de *moi* que je parle (et non pas, comme cela serait le cas selon la théorie des contreparties lewisienne, de quelqu'un d'autre qui me ressemble mais qui n'est pas moi). La situation est la suivante : dans le segment temporel M_1 il y a une seule personne, moi, qui contemple sa future ascension et se dit qu'il va soit mourir soit être victorieux ; au moment critique de la chute, M_1 fissionne alors en M_2 et M_3 , où (pour simplifier) dans l'une de ces branches, je meurs sur le coup et dans l'autre, j'arrive au sommet. Lorsque *je* me trouverai au sommet, c'est très justement que j'aurai alors le sentiment de soulagement car *c'était bien moi* qui, avant la fission, aurais pu faire la chute fatale et mourir. Bien sûr, ma contrepartie qui a eu moins de chance que moi est, *après* l'accident, une personne entièrement (numériquement) différente de moi mais elle *était (numériquement) identique à moi* avant la fission (dans M_1).

Ainsi, il semble que la théorie des contreparties branchante possède un avantage en ce qui concerne l'objection kripkéenne et qu'elle justifie nos sentiments de préoccupation et de soulagement quand il s'agit de modalité *de re* à la première personne. Mais – et c'est ici que nous arrivons au cœur du sujet de cet article – Lewis soulève une conséquence terrible et terrifiante découlant de l'hypothèse d'une structure branchante de mondes possibles : nous allons tous être immortels et vivre une éternité pleine de souffrance.

Ici aussi, il est important de formuler l'argument de Lewis à la première personne – je vais le faire ici pour moi, faites de même pour vous. Je m'imagine au point où je commence l'ascension de la face nord de l'Eiger, et je m'imagine (j'anticipe) quelles expériences je vais avoir dans les heures qui suivent, sachant qu'étant donné mon manque de capacités il y a de fortes chances que je fasse une chute grave. Je peux donc m'attendre à mourir, et je peux m'attendre à survivre. Etant donné la structure branchante de mondes possibles, il y aura de nombreuses fissions successives qui vont donner lieu à de nombreuses branches (mondes) où dans certaines je serai mort et dans d'autres je serai vivant. Etant donné

l'hypothèse réaliste en ce qui concerne l'existence de mondes, je sais avec certitude, avant d'entamer l'ascension, que toutes ces branches (celles où je meurs de même que celles où je survis) vont exister. Etant donné tout ceci, que dois-je attendre de mon futur ? Quelles expériences futures dois-je m'attendre à avoir ? De prime abord, il pourrait sembler que je dois m'attendre à avoir également l'expérience de vivre et celle de mourir. Mais voici ce que constate Lewis :

« Mais ceci n'a aucun sens ! Il n'y a aucune expérience de « être mort » [There's nothing it's like to be dead]. (La vraie mort, je veux dire. La « vie après la mort » est la vie, et non pas la mort.) *On ne doit jamais s'attendre à avoir l'expérience d'être mort car il n'y a pas de telle expérience.* » (Lewis, 2004, mes italiques, ma traduction)

L'idée de Lewis est ici claire : lorsque je me demande à quoi je dois m'attendre dans mon futur, je me demande quelles expériences je vais avoir, mais puisque je ne vais jamais avoir l'expérience d'être mort (je ne vais jamais constater « je suis mort »), je ne dois pas m'attendre à avoir une telle expérience, et donc je devrais tout simplement ne pas prendre en considération les branches dans lesquelles je meurs, et inclure dans mes délibérations expectatives uniquement les branches futures dans lesquelles je survis – en conséquence, je dois m'attendre avec certitude à survivre.

Ce même raisonnement doit bien entendu être appliqué de manière générale, encore et encore, dans toutes les situations où je pourrais mourir – par exemple, lorsqu'après avoir gravi (avec succès !) le sommet de l'Eiger, je me hasarde à des exploits encore plus dangereux comme celui de traverser une rue à Paris en heure de pointe – et puisque je dois à chaque fois m'attendre à survivre, et jamais à être mort, je dois ainsi m'attendre à être *immortel*. En effet, aussi improbable que cela puisse être, il est toujours *métaphysiquement possible* que j'échappe d'une manière ou une autre à la mort, et en conséquence il va toujours y avoir au moins une branche dans laquelle je vais survivre de sorte que suite à chaque situation de danger de mort je dois toujours m'attendre à me trouver vivant après l'incident en question.

Lewis (2004) lui-même a formulé cet argument non pas dans le cadre d'une théorie branchante de mondes possibles, mais plutôt sous l'hypothèse de l'interprétation « de mondes multiples » d'Everett de la mécanique quantique (voir Everett, 1957 ; Wheeler, 1957 ; De Witt, 1970 et 1973). Le but du présent article n'est pas d'entrer dans le délicat débat concernant les diverses interprétations de la mécanique quantique ; je souhaite ici simplement mettre en lumière certaines similarités importantes

entre l'interprétation everettienne et la structure branchante de mondes possibles telle que nous l'avons explorée ci-dessus. La discussion commence ici avec le fameux cas du chat de Schrödinger : un chat est enfermé dans une boîte qui contient une bouteille d'un poison très puissant et volatile, reliée à un mécanisme qui permet de briser la bouteille et qui est lui-même relié à un mécanisme de déclenchement contenant un compteur Geiger et un atome radioactif – le chat est alors laissé là pendant une minute. Si l'atome se désintègre, le compteur Geiger va détecter la radiation et tuer le chat, et si l'atome ne se désintègre pas le chat va rester en vie. Après une minute, étant donné que l'atome suit les lois de la mécanique quantique, il sera dans une superposition d'états quantiques (désintégré et non-désintégré), en conséquence de quoi le mécanisme de déclenchement sera également dans une superposition d'états, de même que le chat lui-même qui sera dans une superposition d'états – mort *et* vivant : une conclusion apparemment absurde.

S'inscrivant en faux contre l'interprétation « classique » de Copenhague de ce qui se passe lorsqu'une superposition d'états quantiques est « résolue » – à savoir que c'est l'observateur qui joue ici un rôle central dans le sens que la superposition va « se fixer » sur l'un des deux états lorsqu'un observateur ouvrira la boîte et regardera le chat (et le trouvera donc soit mort, soit vivant) – l'interprétation d'Everett prédit qu'une *fission* va avoir lieu au moment où la superposition survient, avec une branche où l'atome se désintègre et une autre branche où l'atome ne se désintègre pas, et donc avec une branche où le chat meurt et une autre branche où le chat survit. C'est ici que l'analogie avec la structure branchante de mondes possibles a un sens : à chaque fois qu'il y a des issues possibles différentes à une quelconque situation, des fissions ont lieu et créent des branches alternatives mais toutes également réelles où toutes ces possibilités sont réalisées. Il y a bien entendu des différences importantes entre le système everettien et le système possibiliste branchant. Pour n'en citer que deux, d'une part la possibilité en question est la possibilité physique dans un cas et la possibilité métaphysique dans l'autre cas, et d'autre part, seule la théorie d'Everett inclut la notion de probabilités (certaines issues d'expériences sont plus probables que d'autres). Je ne discuterai ici aucune de ces différences et difficultés car rien dans l'argument que nous avons vu ci-dessus n'en dépend. L'argument de Lewis concerne ce à quoi nous devrions nous attendre étant donné une structure branchante postulée par l'interprétation de mondes multiples de la mécanique quantique, et en ce qui concerne ce point les différences entre cette interprétation et la structure branchante réaliste de mondes possibles n'ont guère de pertinence. A quoi le

chat doit-il s'attendre lorsqu'il entre dans la boîte ? Lewis dit : à l'immortalité, pour exactement les raisons que nous venons de voir ci-dessus. Ni la forme ni l'esprit de l'argument de Lewis n'ont été altérés en formulant son argument sous l'hypothèse de la structure branchante de mondes possibles plutôt que sous l'hypothèse Everettienne.

La conclusion en faveur de laquelle Lewis argumente devrait donc, s'il a raison, être acceptée aussi bien sous l'hypothèse everettienne que sous l'hypothèse d'une structure branchante de mondes possibles : je dois m'attendre à vivre immortel, de même que vous. Est-ce une bonne nouvelle? Malheureusement pas, car l'argument de Lewis ne s'arrête pas ici. Voici le pas final de l'argument : avec le temps, plus je vis de situations où je risque de mourir, plus ma survie va être extraordinaire, et plus je vais subir de conséquences d'incidents quasi-mortels. Par ailleurs, je vais continuer à vieillir, et ma condition physique et mentale va se détériorer naturellement avec le temps dans toutes les branches dans lesquelles je survivrai à toutes les causes accidentelles et naturelles qui me feront mourir dans les branches où je meurs. Après, disons, mon 150^{ème} anniversaire, chaque seconde où je continuerai à vivre sera extraordinaire, et je serai toujours dans un état proche de la mort, sans jamais vraiment mourir. Sur le long terme, seules des branches de ce type vont être celles où je vais survivre, et suivant l'argument de Lewis, ce sont uniquement ces branches dont je dois tenir compte lorsque je me demande à quoi je dois m'attendre en ce qui concerne mon futur. C'est pourquoi je dois m'attendre à une immortalité horrible et pleine de souffrance.

Une telle conséquence serait certainement très troublante, si elle était vraie. Heureusement, je pense qu'elle ne l'est pas. Je vais à présent d'abord montrer que Lewis a accordé trop de force à son argument et que celui-ci s'applique à n'importe quelle théorie de mondes possibles, et non pas uniquement à la théorie branchante. En particulier, je vais montrer que si l'argument de Lewis était correct, la même conséquence serait vraie également sous sa propre hypothèse divergente de mondes possibles. On pourrait alors penser que ceci rend l'argument lewisien encore plus déprimant, mais en réalité ce que cela nous montre c'est qu'il doit y avoir quelque chose d'incorrect dans cet argument. Je vais ensuite examiner ce que c'est qui rend l'argument incorrect.

Supposons, pour commencer, que je me trouve dans une structure branchante de mondes possibles. Lorsque je garde à l'esprit les considérations ci-dessus, dois-je considérer l'énoncé suivant comme vrai ?

[1] *J'ai de nombreux futurs*, certains où je suis mort et d'autres où je suis vivant.

Comme nous l'avons vu, selon Lewis nous devrions ne pas tenir compte de tous les futurs où je vais mourir. Ceci est, comme nous le verrons plus loin, l'erreur à cause de laquelle l'argument de Lewis n'est pas acceptable. En effet, lorsque je me demande à quoi je dois m'attendre en ce qui concerne mon futur, je *peux* et je *dois* considérer la possibilité de ma propre mort. Je vais retourner à ce point en détail ci-dessous. Pour le moment, je m'intéresse à un défaut différent de l'argument de Lewis : *si* cela était un bon argument, alors ce serait un bon argument *indépendamment de l'hypothèse branchante*. Considérons pour l'instant *toutes* les branches futures, y compris celles où je meurs. Il y a deux points importants à remarquer ici, que j'ai mis en italique dans l'énoncé [1]. D'une part, « je » joue un rôle important car il est crucial pour l'argument de Lewis que c'est vraiment à *moi* que tous ces futurs vont arriver, autrement je n'aurais aucune raison de m'en préoccuper ; et d'autre part, la mention de « nombreux futurs » est ici hautement importante – et incorrecte. Car indépendamment du fait que l'on se trouve dans une structure branchante ou non branchante, je ne dois jamais m'attendre à vivre quelque chose comme des « vies futures parallèles » où j'aurai « simultanément » des expériences de plusieurs personnes futures ; au contraire, j'aurai uniquement des expériences d'une seule personne. Voici donc ce que je dois dire pour être plus précis et correct :

[2] *J'ai de nombreux futurs possibles*, certains où je suis mort et d'autres où je suis vivant.

Ainsi formulée, avec l'accent mis sur « je » et sur « futurs possibles », mon attente à propos de mon futur est correcte. Qu'en est-il si au lieu d'une hypothèse branchante de mondes possibles nous acceptons l'hypothèse divergente de Lewis lui-même ? Est-ce que ceci est vrai ?

[3] J'ai uniquement *un* futur, et tous les autres futurs dans les autres mondes divergents ne sont pas vraiment *mes* futurs. (Et c'est pour cette raison que je ne dois pas les craindre.)

Ceci semble correct, mais c'est une formulation imprécise, car elle est ambiguë concernant le point mentionné plus haut à propos de l'(im)possibilité de vivre des vies parallèles multiples. Afin de rendre [3]

non-ambigu et d'éviter d'affirmer une fausseté de ce type, nous devons plutôt formuler [3] ainsi :

[4] J'ai uniquement *un* futur, et tous les autres futurs *possibles* dans les autres mondes divergents ne sont pas vraiment *mes* futurs *possibles*. (Et c'est pour cette raison que je ne dois pas les craindre.)

De même que dans [2], il fallait ici insister sur les « futurs possibles ». Mais est-ce que ceci est quelque chose que je devrais dire sous l'hypothèse de la théorie divergente lewisienne de mondes possibles ? Lewis lui-même ne l'accepterait certainement pas, de concert avec tous les défenseurs des diverses variantes de la théorie des contreparties. En effet, comme nous l'avons vu plus haut, les défenseurs de la théorie des contreparties sont convaincus que l'objection à leur égard a été correctement évitée et que des énoncés modaux *de re* sont réellement à propos de *moi*, même si leurs conditions de vérité impliquent des objets numériquement distincts de moi (mes contreparties dans d'autres mondes). Accepter [4] reviendrait alors pour Lewis à concéder que sa théorie des contreparties fournit une analyse incorrecte des énoncés modaux *de re*, et ce n'est certainement pas quelque chose qu'il serait prêt à concéder. En conséquence, ce qu'il doit faire est de modifier [4] dans l'esprit de la théorie des contreparties :

[5] J'ai uniquement *un* futur, mais tous les autres futurs dans les autres mondes divergents sont *mes* futurs *possibles*.

C'est bien ceci qu'il faut affirmer sous l'hypothèse lewisienne divergente de mondes possibles, enrichie de la théorie des contreparties. Mais c'est bien exactement ceci que disait l'énoncé [2] sous l'hypothèse d'une structure *branchante* de mondes possibles ! Ainsi, que ce soit dans le système divergent ou dans le système branchant, il est vrai dans les deux cas que je vais vivre *un* seul futur, mais que j'ai de nombreux futurs *possibles*, qui sont tous possibles pour *moi*. En conséquence, dans les cas des *deux* théories de mondes possibles, je devrais avoir les mêmes attentes en ce qui concerne mon futur, et de ce fait, si Lewis avait raison, je devrais m'attendre à une immortalité horrible dans ces deux cas ! L'on voit donc ici que *l'hypothèse branchante ne joue en réalité aucun rôle dans l'argument de Lewis*.

Son horrible conclusion provient d'une autre prémisse, à savoir son affirmation que *nous devons ne pas tenir compte des branches futures (ou, simplement, de mes futurs possibles) où je vais mourir*. Sans ce pas argumentatif, je devrais évidemment m'attendre de manière égale à subir des futurs où je vais mourir ainsi que des futurs où je vais vivre. Sans ce pas argumentatif, je peux soulager mes craintes d'une horrible immortalité en nourrissant l'espoir d'une mort qui va me délivrer de mes souffrances. Il est vrai que, même sans ce pas argumentatif, je dois admettre qu'il est *possible* que je vive une horrible immortalité, mais ceci est plus supportable que la certitude d'un tel avenir. Il y a en effet beaucoup de choses horribles qui sont possibles pour moi même maintenant, à l'instant, mais je ne les crains pas, et ce n'est pas que je sois une personne particulièrement courageuse – car c'est bien normal de ne pas craindre toutes les choses terribles qui peuvent nous arriver, comme par exemple d'être écrasé par une météorite avant d'avoir le temps de finir d'écrire cette phrase (ouf !).

Lewis a-t-il raison de dire que nous devrions ne pas tenir compte des futurs où nous allons mourir lorsque nous formulons nos attentes à propos de notre futur ? Il a certainement raison d'affirmer qu'il n'y a aucune *expérience* d'« être mort », mais comme l'a remarqué Papineau (2004), ceci ne m'empêche pas de contempler un futur où je suis mort – un futur où je n'existe plus. Nous ne pouvons pas avoir, ni formuler dans l'imagination, une expérience de ce que cela serait que d'être mort, mais nous pouvons avoir une idée claire de ce en quoi un tel futur consisterait s'il se produisait. Ainsi, il est correct de dire que je ne dois pas m'attendre à avoir l'expérience d'être mort (et ceci *indépendamment* d'une hypothèse divergente ou branchante de mondes possibles), mais je peux m'attendre à être mort dans un sens tout à fait clair et intelligible.

Nous ne sommes donc pas obligés de concentrer nos attentes concernant notre futur uniquement sur les futurs possibles où nous serons vivants, mais également sur ceux où nous serons morts – et de ce fait, notre peur du futur ne doit pas être celle de la certitude d'une horrible immortalité pleine de souffrance, mais seulement de la possibilité que ceci se produise. Par ailleurs, selon la thèse que j'ai défendue dans cet article, ce résultat est entièrement indépendant du choix d'une structure branchante ou divergente de mondes possibles, voir même du choix d'une théorie ersatziste².

² Je tiens à remercier Julien Christen, Lynda Gaudemard, Richard Glauser, Baptiste Le Bihan, Frédéric Lelong, Martine Nida-Ruemelin, Daniel Schulthess, et Gianfranco Soldati pour leurs précieux commentaires et critiques qui m'ont permis d'améliorer cet article.

Bibliographie

- B. De Witt, « Quantum Mechanics and Reality », *Physics Today* 23 : 30-35, 1970.
- B. De Witt et N. Graham, *The Many-worlds Interpretation of Quantum Mechanics*, Princeton, Princeton University Press, 1973.
- H. Everett, « Relative State Formulation of Quantum Mechanics », *Reviews of modern physics*, 29 : 454-461, 1957.
- M. Heller, « Property Counterparts in Ersatz Worlds », *The Journal of Philosophy*, p. 293-316, 1998a.
- M. Heller, « Five Layers of Interpretation for Possible Worlds », *Philosophical Studies*, p. 205-214, 1998b.
- S. Kripke, *Naming and Necessity*, Harvard, Harvard University Press, 1972
- D. Lewis, « Counterpart Theory and Quantified Modal Logic », *Journal of Philosophy*, 1968, 65 : 113-26.
- D. Lewis, « Survival and identity » in *Philosophical Papers*, volume I, Oxford University Press, 1983.
- D. Lewis, *On the Plurality of Worlds*, Oxford, Blackwell Publishers, 1986.
- D. Lewis, « How Many Lives has Schrödinger's Cat? » *Australasian Journal of Philosophy*, 2004, 82 (1) : 3-22.
- D. Papineau, « David Lewis and Schrodinger's Cat », *Australasian Journal of Philosophy*, 2004.
- A. Plantinga « Transworld identity or worldbound individuals » in *Logic and Ontology*, éd. Milton Munitz, New York University Press, 1973. Repris in *The Possible and the Actual*, éd. Michael Loux, Cornell University Press, 1979.
- T. Sider, « All the World's a Stage ». *Australasian Journal of Philosophy* , 1996, 74 : 433-453.
- T. Sider, *Four-dimensionalism*, Oxford, Clarendon Press, 2001.
- T. Sider, « Beyond the Humphrey Objection », à paraître (http://www.memorialartgallery.art.museum/College/REL/faculty/wierenga/591/counterpart_theory.pdf).
- J. Wheeler, « Assessment of Everett's 'Relative State' Formulation of Quantum Theory », *Reviews of modern physics* 29 : 463-465, 1957.