

***L'argument du déployeur universel***  
**(UDA : *Universal Dovetailer Argument*)**  
**de Bruno Marchal**

*Jean-Paul Delahaye*

*Professeur à l'Université des Sciences et Technologies de Lille*  
*Laboratoire d'informatique fondamentale de Lille*

*(LIFL, UMR CNRS 8022)*

*USTL, Bat M3-Ext, 59655, Villeneuve d'Ascq CEDEX France*

*E-mail : [jean-paul.delahaye@lifl.fr](mailto:jean-paul.delahaye@lifl.fr)*

*<http://www2.lifl.fr/~delahaye/>*

**3 mai 2010**

***Résumé.** Nous étudions «l'argument du déployeur universel» (UDA : Universal Dovetailer Argument) de Bruno Marchal. Nous utilisons la version en huit étapes publiée en 2004. Nous en examinons les différents points, un par un, en énumérant tout ce qui nous semble sujet à débat ou à doutes. La conclusion est un désaccord dû, entre autres choses, à l'absence d'un traitement détaillé et convaincant des questions de probabilités utilisées et centrales à l'argument.*

## **Introduction**

L'article de Bruno Marchal auquel nous nous référons est :

***The Origin of Physical Laws and Sensations***

*System Administration and Network Engineering, 2004*

*<http://www.sane.nl/events/sane2004/index.html>*

*<http://iridia.ulb.ac.be/~marchal/publications/SANE2004MARCHAL.pdf>*

Certaines des difficultés que je vais évoquer sont certainement connues de Bruno Marchal. Certaines sont même évoquées dans son texte et balayées d'un revers de manche. Je les évoque cependant car elles sont parfois beaucoup plus graves qu'il ne semble prêt à l'admettre.

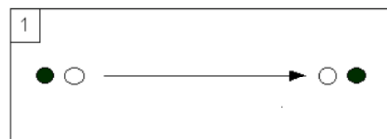
S'il s'agissait d'un raisonnement mathématique, un seul point faible suffirait à invalider tout le raisonnement. Je ne crois pas qu'il s'agisse d'un raisonnement mathématique, et je pense même que bien des concepts évoqués par Bruno Marchal sont vagues et n'ont rien de concepts scientifiques, et ne sont pas, en tout état de cause, utilisés dans un contexte scientifique suffisamment délimité et précis. Il serait raisonnable de prendre l'intéressant raisonnement qu'il propose pour ce qu'il est : une sorte de jeu philosophique intrigant et stimulant.

Je me permets de remarquer que personne (à ma connaissance) ne reprend le raisonnement de Marchal pour en adopter la conclusion. Cela ne prouve pas, bien sûr, qu'il est faux —on peut avoir raison tout seul— mais cela prouve, pour le moins, que Marchal ne réussit pas bien à persuader ceux qui le lisent. Marchal a parfois évoqué l'idée que si personne n'adopte les conclusions de son raisonnement c'est que pour le comprendre il faudrait avoir des connaissances dans une série de domaines différents et difficiles (théorie de la calculabilité, mécanique quantique, logiques modales, théorie de l'esprit, etc.) ce qui n'est le cas que de très rares personnes. Je ne crois pas que cela soit exact. Le raisonnement est élémentaire et ne nécessite que peu de connaissances techniques dans chaque domaine concerné. Il faut chercher ailleurs l'absence de soutien affirmé au raisonnement.

En un mot, voici le jugement où me mènera l'examen de l'argument : pour qu'un raisonnement aboutissant un renversement aussi radical que celui de la conclusion de Marchal —la physique doit se déduire de l'arithmétique—, il faudrait que l'argumentaire soit bien plus serré et minutieux, et particulièrement, comme on va le voir, à propos des probabilités évoquées en plusieurs endroits qui posent des problèmes graves. Bien sûr, si la reconstruction de la physique à partir de l'arithmétique réussissait, on pourrait y voir une raison nouvelle très forte pour considérer que l'argument de Marchal est fondamentalement juste (malgré les critiques auxquelles je le soumets). Malheureusement, cette reconstruction se fait attendre.

Examinons les 8 points du raisonnement un par un.

## Point 1



Comp makes possible (only *in principle* but that is all we need), the use of classical [8] teleportation. [...]

Voir le détail en : <http://iridia.ulb.ac.be/~marchal/publications/SANE2004MARCHAL.pdf>

Marchal propose de raisonner *en principe*. Nous allons accepter cette position, mais nous remarquons cependant que c'est justement en cherchant à faire coïncider ce que nous acceptons *en principe* avec ce que nous pouvons faire *en réalité* que bien des progrès sont faits en science.

Pour illustrer cette idée, je pourrais mentionner la mécanique quantique qui tente de comprendre pourquoi alors qu'*en principe*, l'espace et la matière semblent indéfiniment sécables et continus, ils ne le sont pas *en réalité*. Je prendrai un autre exemple. Avant la théorie des classes de complexités (P, NP, PSPACE, etc.) on considérait comme une bonne approximation de «calculable *en principe*» tout ce qui était «calculable par une machine de Turing». On s'est rendu ensuite compte que les fonctions calculables *en principe* devaient plutôt être assimilées aux

«fonctions calculables en temps polynomial». La prise en compte des algorithmes *probabilistes* et maintenant *quantiques* conduit encore à revoir la notion de *calculable en principe* (classes BPP, QP etc.). Il semble donc bon de considérer avec attention les affirmations qu'on accepte *en principe* et il ne faut pas oublier que parfois on a à les réviser.

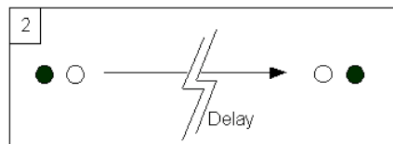
En négligeant d'aller voir le «*en principe*» évoqué dès la première ligne du point 1, nous prenons donc déjà un risque. D'ailleurs, l'acceptation du *en principe* nous précipite immédiatement dans un univers de science-fiction où la téléportation est possible et praticable, alors que personne aujourd'hui n'est en mesure de se prononcer sérieusement sur la question, même pour dans dix siècles.

Avant d'abandonner ce point, il faut encore évoquer qu'en mécanique quantique on considère que certaines informations pourraient ne pas être accessibles (ou même, ne pas exister avant qu'on en prenne connaissance), ce qui rend difficile la téléportation évoquée dans les expériences de pensée de Marchal, puisqu'elle se fonde sur une analyse moléculaire (ou plus) de ce qui est téléporté et dont on fait un plan codé numériquement.

Le «no-cloning theorem» de Wootters, Zurek et Dieks en mécanique quantique pourrait même être vu comme un argument fort bloquant toute la suite : je pourrais être une machine digitale, mais cette machine digitale pour une raison fondamentale de mécanique quantique pourrait ne pas être duplicable ou téléportable. Précisons que la téléportation quantique ne résout pas le problème puisqu'elle oblige à détruire l'original et ne rend donc toujours pas possible la duplication.

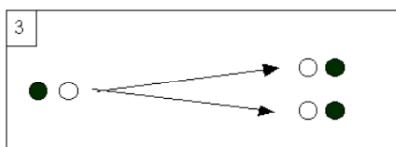
N'insistons pas sur ces considérations et la méfiance qu'on peut ressentir à l'évocation d'expériences de pensée trop éloignées de la réalité technologique et scientifique réelle, et passons au point 2.

## Point 2



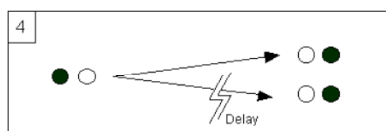
Pas de problème majeur pour suivre Bruno Marchal sur le point 2, sauf toujours que si on peut négliger *en principe* le temps nécessaire à la "lecture" du sujet au départ, et le temps nécessaire pour reconstituer le sujet à l'arrivée, il se pourrait bien qu'*en réalité* même dans dix mille ans, on ne puisse mener de telles opérations *en pratique* (voir même *en principe* pour des raisons non encore reconnues aujourd'hui liées à la quantité d'informations concernée, à la mécanique quantique, ou à d'autres théories physiques).

### Point 3



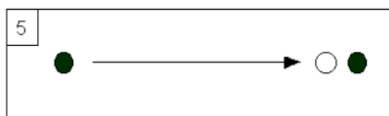
Pas de problème nouveau ici : oui du point de vue de la *première personne*, il y a un indéterminisme fondamental dans les situations évoquées, alors que du point de vue de la troisième personne tout pourrait être parfaitement déterministe. C'est l'idée déjà rencontrée en mécanique quantique avec l'interprétation d'Everett / De Witt, que vu d'un œil extérieur l'univers (le "multivers") pourrait être déterministe, alors que dans chaque "branche" de l'univers tout semble indéterministe.

### Point 4



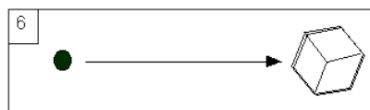
Nous approuvons Marchal sur l'idée que l'attribution de la probabilité 1/2 à chaque cas de l'expérience du point 3, ne va pas de soi, et qu'il vaut mieux s'en passer. Nous allons revenir sur ces drôles de probabilités rencontrées dans ces expériences de pensée.

### Point 5



Rien de choquant (en négligeant toujours le no-cloning theorem).

### Point 6



La formulation du point 6, ne nous pose pas de problème grave, sauf pour ce qui est dit à propos du calcul des probabilités.

Bruno Marchal a l'air de tenir pour acquis qu'il y a une mesure de probabilité bien définie pour le sujet de l'expérience. Il ne la précise pas et l'imagine quelconque mais l'évoque comme si son existence ne posait pas de problème. Je ne pense pas que ce soit le cas.

Il me semble que nous sommes dans un cas *d'indéterminisme sans probabilité*.

(a) D'abord, que peut vouloir dire probabilité à la première personne pour une expérience qui n'est pas répétée ? Il s'agit là d'un problème général de philosophie des probabilités, mais vu que le point est crucial pour la suite, il est essentiel de le reposer dans ce cadre particulier et d'en proposer une solution ou au moins une analyse. Pour l'instant, je n'arrive pas à saisir ce que pourrait être la probabilité d'un événement qu'on ne répète pas, quand on raisonne à la première personne.

(b) Plus grave, nous pouvons démontrer qu'il n'y a pas de probabilité (au sens de la théorie mathématiques des probabilités) dans ce type de situations ou dans le type de situations évoquées aux points 1-5.

Voici la preuve. Imaginons l'expérience de pensée suivante.

Le sujet S est dupliqué à partir d'un état E0 et on le reconstitue au bout d'un an dans l'état E0 (réellement ou dans une machine). On ne laisse vivre qu'une minute (subjective) cette copie de S sans lui communiquer aucune autre information que le nombre 1. On reconstitue S toujours dans l'état E0 au bout de deux ans sans lui communiquer aucune autre information que le nombre 2. etc. On continue comme cela indéfiniment (ce que nous supposons possible).

Pour S, toutes les reconstitutions sont équivalentes les unes aux autres (même si les reconstitutions sont éloignées les unes des autres d'une année, ce qu'il ne sait pas). S'il existe une mesure de probabilités qui ait un sens dans ce contexte pour lui, alors la probabilité de la reconstitution  $i$  ne doit donc pas dépendre de  $i$ , car pour le sujet S, les cas possibles sont équivalents (à la place de lui communiquer  $i$ , on pourrait imaginer lui communiquer n'importe quelle information spécifique à la reconstitution  $i$ , la distinguant des autres sans l'informer du code ; on pourrait par exemple lui communiquer  $f(i)$  où  $f$  est une injection quelconque fixée de  $\mathbf{N}$  dans  $\mathbf{N}$ ). S'il y avait une mesure de probabilité (à la première personne ou à la troisième personne) associée à cette expérience de pensée, nous aurions alors une mesure uniforme de probabilités sur les nombres entiers. On sait que cela n'existe pas.

Il faut donc bien admettre que le type d'expériences qu'envisage Bruno Marchal crée des situations où il existe bien un *indéterminisme sans probabilités*, ou alors que les "probabilités" mentionnées dans ces expériences de pensée ne sont pas celles dont on parle en mathématiques. Dans ce dernier cas, il faut en développer la théorie et ne pas se satisfaire de la considérer déjà disponible.

L'erreur de Marchal, je crois, provient ici qu'il oublie que l'indéterminisme n'est pas nécessairement lié à une mesure de probabilité. C'est un point important (que nous allons retrouver plus loin) et, assez curieusement, c'est aussi un des points

centraux mis en avant récemment par John Conway et Simon Kochen avec ce qu'ils ont appelé *Free will theorem* :

<http://www.ams.org/notices/200902/rtx090200226p.pdf>

(voir aussi Jean-Paul Delahaye, "Libre arbitre et mécanique quantique", *Pour la science, édition française du Scientific American*, N°386, décembre 2009, 96-101, <http://www2.lifl.fr/~delahaye/dnalor/LibreArbitre.pdf>).

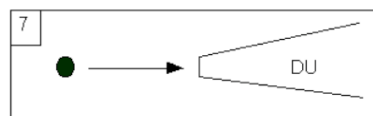
Il semble qu'en mécanique quantique la compréhension approfondie de certaines situations et des conséquences de certaines expériences de pensée comme celles que décrivent Conway et Kochen nécessite d'envisager (et de modéliser) la notion d'indéterminisme sans probabilité (ou indéterminisme fonctionnel).

Marchal termine le point 6 par :

What matter here, is that whatever measure of the comp 1-indeterminacy we choose, that measure will not change in the case where the reconstitution are virtual. Even if the simulation does not last, each first person will take any personal reconstitution as confirming its anticipation, i.e. its bets on its consistent extensions. The probability calculus is again invariant for such a change. This follows directly from our earlier comp assumption that a correct substitution level exists, and that we are Turing emulable.

Ce n'est pas nous qui choisissons la mesure de probabilité. Il faut d'abord prouver qu'elle existe (et nous avons établi que ce n'était pas le cas). L'hypothèse qu'il y a un niveau correct de substitution ne suffit pas à établir l'existence de cette mesure de probabilité. Un manque de rigueur et de précision (dans l'introduction et la manipulation des concepts) entache gravement le point 6.

## Point 7



(a) La notion de déployeur universel ne pose pas de problème *en principe*.

Marchal mentionne qu'il faut juste supposer que l'univers est "robuste", autrement dit qu'il peut vraiment contenir un déployeur universel que l'on ferait calculer TOUJOURS.

Je me demande si Marchal réalise bien à quel point cette hypothèse est problématique.

L'univers visible contient de l'ordre de  $10^{120}$  bits d'information (Seth Lloyd). Un tel morceau d'univers ne peut donc déployer qu'au plus  $10^{120}$  programmes (et même beaucoup moins). Or c'est ridiculement petit et ne donnera aucun programme simulant quoi que ce soit d'intéressant et surtout pas des sujets conscients équivalents à des humains. Pour que le déployeur universel puisse prendre corps et faire quelque chose d'intéressant dans le cas de nos expériences

de pensée, il faut envisager un univers physique très différent de ce que nous savons du nôtre. En clair, il faut envisager un univers qui n'est pas celui que la physique nous montre. C'est assez ennuyeux !

(b) Vient ensuite la difficulté due à l'infini de l'espace nécessaire pour l'existence réelle du déployeur universel. Aujourd'hui, rien n'indique de manière plausible que l'univers est infini et s'il ne l'est pas, aucun déployeur universel ne pourra jamais fonctionner, ni aujourd'hui ni demain.

L'idée, prise au sérieux, d'implanter un déployeur universel qui fonctionnerait sans jamais s'arrêter, n'est pas seulement de la science-fiction risquée (comme la téléportation) c'est quelque chose qui doit être considéré comme totalement improbable et sans doute définitivement inenvisageable. D'après Barrow et Tipler seuls des modèles cosmologiques très particuliers permettent d'envisager "en vrai" une machine de Turing universelle fonctionnant TOUJOURS.

Je passe sur l'idée très étrange que nous ou nos descendants puissions sérieusement souhaiter mettre en place une telle chose : cela me semble un peu près aussi absurde que d'entreprendre l'impression de tous les livres de la Bibliothèque de Babel de Borges, qui eux, pourtant, sont en nombre fini !

Il faut insister sur le fait que nous sommes ici face une exponentielle (et même à quelque chose de pire) et que dans le monde réel toute exponentielle finit par rompre... rapidement.

Je crois que jamais il n'y aura de déployeur universel mis en route et fonctionnant un temps suffisant pour qu'on en tire quoi que ce soit. Je crois que si on en mettait un en marche, même avec la meilleure volonté du monde et tous les moyens technologiques imaginables, il serait interrompu pour une raison ou une autre avant d'avoir produit le moindre état de conscience d'un être équivalent à un humain.

Cela bloque le raisonnement de Marchal.

(c) Mais admettons quand même que nos descendants, un jour, s'appliqueront à mettre en place un déployeur universel et que celui-ci fonctionnera TOUJOURS. Le problème de *l'indéterminisme sans probabilité* se pose à nouveau. Cette fois, c'est plus grave.

Je ne vois pas, et je ne crois pas qu'on puisse construire une mesure de probabilités comme Marchal l'envisage. Surtout, je ne crois pas qu'on puisse affirmer qu'elle existe sans donner aucune construction de cette mesure. **Le problème est mathématique.** Si une telle mesure existe une preuve mathématique est nécessaire. L'absence de cette preuve et d'aucune idée mathématique sérieuse pour la définir est une défaillance grave de l'argumentation.

Il est vraiment trop facile d'affirmer des propriétés de cette mesure de probabilité qui n'a pas été définie :

"the invariance of the uncertainty measure, notably for the arbitrary delay--- including the null one "

"It can be argued that finite computations are of measure null, and that the only way to a measure on the states will consist in finding a measure on the set of maximally complete computational history going through those states, with obviously a rather hard to define equivalence relation among computations"

etc.

On est dans un contexte purement mathématique (on manipule des machines abstraites dont la théorie est depuis Turing une théorie liée à la logique et à l'arithmétique, donc incluse dans les mathématiques ; personne ne le conteste), on parle de *mesure* qui est une notion mathématique précise (à la base de la théorie des probabilités telles que Kolmogorov les a axiomatisées en 1933), on parle d'ensembles de mesure nulle qui est une notion de cette théorie, et pourtant on ne donne pas de démonstration, pas même d'esquisse des affirmations qu'on avance. On évoque aussi une "équivalence difficile à définir" sans explication ni aucun détail.

On ne peut pas prétendre faire de la science et, à l'endroit où une démonstration mathématique est nécessaire et attendue, se contenter d'affirmations dans le vide.

Comme pour s'excuser, Marchal écrit :

It is not necessary to be more precise here, giving the non constructivism of the collection of those consistent extensions, and the fact that we will make things utterly precise, by directly interviewing a universal machine on those extensions, and this by taking into account the 1/3 person point of view distinction.

L'argument de la *non-constructivité* pour se dispenser de définir précisément la mesure de probabilité qui joue un rôle si central est pour le moins surprenant. Marchal sait-il que même les objets non constructifs reçoivent des définitions précises et qu'on peut alors les étudier ? Le nombre oméga de Chaitin est un exemple hautement non constructif d'objet mathématique qui est pourtant parfaitement bien défini et dont on peut démontrer beaucoup de propriétés (par exemple qu'il est transcendant, et normal.).

Quant à l'affirmation que, plus loin dans le texte, les choses sont rendues précises, elle est fautive : la suite ne donne pas la définition de la mesure de probabilité qui est pourtant le pivot de l'argument !

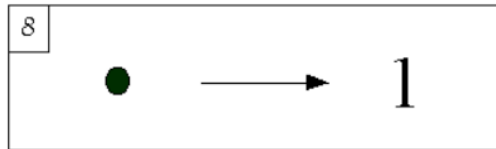
Marchal semble ignorer l'existence d'un *indéterminisme sans probabilité*. Il présuppose donc une mesure de probabilité qui ne peut pas exister (voir la démonstration donnée plus haut) et il lui attribue les propriétés extravagantes qui l'arrangent. Je pense sincèrement que la déficience du raisonnement en ce point est fatale et rien de la conclusion du point 7 n'a finalement été établi :

So, if we grant a sufficiently robust universe, we are completely done: physics, as the "correct" science for the concrete relative predictions must be given by some measure on our consistent relative states. Physics is, in principle reduced to a

measure on the collection of computational histories, as seen from some first person point of views. We can say that in principle, physics has been reduced to computer fundamental psychology.

Mais admettons quand même cette conclusion et passons au point suivant.

## Point 8



Conscient de certaines des objections formulées à propos du point 7, Marchal commence par :

Yes, but what *if we don't grant* a concrete robust physical universe? Up to this stage, we can still escape the conclusion of the seven preceding reasoning steps, by postulating that a “physical universe” really “exists” and is too little in the sense of not being able to generate the entire UD\*, nor any reasonable portions of it, so that our usual physical predictions would be safe from any interference with its UD-generated “little” computational histories.

Nous sommes d'accord.

Such a move can be considered as being *ad hoc* and disgraceful.

Sauf que ce sont les sciences physiques nous le suggèrent aujourd'hui !

It can also be quite weakened by some acceptance of some conceptual version of Ockham's Razor,

Un peu de précision aurait été utile !

and obviously that move is without purpose for those who are willing to accept comp+ (in which case the UDA just show the necessity of the detour in psychology, and the general shape of physics as averages on consistent *I-histories*).

Sauf qu'en l'absence de précisions mathématiques sur la mesure de probabilité sur laquelle on s'appuie (et dont je pense qu'elles n'existent pas) on est coincé.

L'hypothèse comp+, sauf erreur de ma part, n'est pas définie dans l'article !

But logically, there is still a place for both physicalism *and* comp, once we made that move. Actually the 8th present step will explain that such a move is nevertheless without purpose.

Je ne sais pas si cela signifie que les 7 premières étapes sont en fait logiquement inutiles. Si c'est ce que veut dire Marchal et que seul, à ses yeux, le point huit

suffisait pour sa conclusion c'est qu'il s'est moqué de nous... et de moi en particulier en me demandant lequel des points je trouvais fautif.

Vient alors l'évocation de l'argument du graphe filmé et de l'argument analogue de Maudlin.

For any given precise running computation associated to some inner experience, you can modify the device in such a way that the amount of physical activity involved is arbitrarily low, and even null for dreaming experience which has no inputs and no outputs. [...]

Now this shows that any inner experience can be associated with an arbitrary low (even null) physical activity, and *this* in keeping counterfactual correctness. And *that* is absurd with the conjunction of both comp and materialism.

Ce que je trouve étonnant dans cette partie du raisonnement —qui ressemble à un tour de passe-passe de music-hall où le magicien après avoir fait monter son assistante debout sur une table, enlève la table, laissant l'assistante en suspension dans l'air—, c'est qu'on n'y évoque pas le réalisme arithmétique.

Si on admet que le déployeur universel existe physiquement et qu'il existe aussi "mathématiquement" et que les deux types d'existence se valent (si cela à un sens !), alors je suis d'accord qu'on peut enlever la table sous l'assistante sans qu'elle tombe, et donc qu'une expérience à la première personne continue d'être défendable "soutenue" par l'arithmétique (avec les problèmes qui restent concernant les probabilités).

Malheureusement, si je n'accepte pas le réalisme arithmétique ou si je considère que l'être physique est différent de l'être mathématique et que c'est seulement sur le premier sur *survient l'esprit* (je préfère le verbe *survenir* à *supervenir* ou *supervener*), alors le passage du déployeur physique au déployeur arithmétique ne marche pas, et pour le coup l'assistante du magicien tombe par terre.

Je ne suis donc pas d'accord avec l'escamotage de la réalité physique effectuée au point 8, sauf pour celui qui adopte un réalisme arithmétique fort (qui attribue aux objets arithmétiques une existence aussi forte qu'aux objets physiques). Mais alors si on pose cela, on ne démontre pas que la physique doit se déduire de l'arithmétique, on démontre seulement qu'une ontologie arithmétique posée permet de se passer d'une ontologie physique différente !

Le réalisme arithmétique présupposé pour l'escamotage est d'ailleurs si fort qu'il est peu défendable. Je rappelle à Marchal que la grande majorité des physiciens ne sont pas réalistes concernant les objets mathématiques, et que le réalisme mathématique est donc le plus souvent considéré comme faux.

Je passe sur le fait que, quitte à adopter un réalisme mathématique, d'autres options sont possibles et nombreuses (réalisme de l'analyse, des ensembles, des catégories, des grands cardinaux, etc.) et qu'il manque donc des arguments pour "choisir" le réalisme arithmétique plutôt qu'un autre. Même pour un réaliste, on a donc là un choix. Il paraît arbitraire.

Là encore, le raisonnement ne marche pas à mes yeux, ou ne marche que sous des hypothèses très fortes que je ne peux pas accepter sans arguments.

Je voudrais aussi faire la remarque que l'argument du graphe filmé au centre du point 8, montre seulement que le prédicat "avoir un esprit" est un *prédicat vague* (comme être chauve, être intelligent, etc.). Qu'il y ait des situations où nous ne sachions pas s'il faut considérer qu'il y a un esprit ou non, ne me gêne pas plus que le fait de ne pas savoir dire si certaines personnes sont chauves ou non. Oui, il y a des états semi-conscients. Oui je ne sais pas dire si quand je dors mon esprit est vraiment actif, ni si un chien, un oiseau, une mouche, un ver de terre possède un esprit. Croire qu'avoir un esprit est une question *tout ou rien* c'est être victime d'une illusion. Le monde social pour fonder les concepts de responsabilité (par exemple) a besoin de cette illusion, mais ni la philosophie, ni la science ne doivent en être victime. L'argument du graphe filmé ne montre pas du tout que l'esprit peut se passer de la réalité physique. Il montre seulement des cas bizarres et linguistiquement indéterminés (parce qu'ils ne se posent pas dans le monde social) concernant le prédicat "avoir un esprit", comme il y en a pour le prédicat "être chauve". Les questions que pose *le graphe filmé* sont analogues à la question : «est-ce qu'un homme qui n'aurait qu'une touffe dense de poils sur le haut du pavillon de son oreille gauche est chauve ?».

## Conclusions.

Pour une multitude de raisons, l'argument présenté en huit points par Marchal n'entraîne pas du tout mon adhésion. Il y a de nombreuses échappatoires, mais il n'y a rien d'étonnant à cela et avec un peu d'attention tout raisonnement un peu complexe en philosophie (et particulièrement en philosophie de l'esprit) peut être traité de la même façon.

Comme je l'ai déjà expliqué à Marchal depuis bien longtemps, et comme j'ai voulu le montrer ici à sa demande, son raisonnement ne peut pas être qualifié d'incontournable comme l'est un raisonnement mathématique, car nous ne sommes pas en mathématiques mais en *philosophie de l'esprit*. Les raisonnements deviennent incontournables quand ils sont formalisés. C'est bien loin d'être le cas ici.

Marchal préférerait qu'on dise qu'il s'agit non pas de *philosophie* (pas même de *philosophie de l'esprit*) mais d'un domaine particulier à «*l'intersection des sciences cognitives et physiques*». Pourtant, son texte ne mentionne rien concernant les sciences cognitives telles que les psychologues, les neurologues, les linguistes, les spécialistes d'I.A. etc. pratiquent ces sciences, et rien concernant la physique — telle que les physiciens font de la physique. Drôle d'intersection !

Je n'ai aucune illusion sur le fait que Marchal changera d'opinion concernant son raisonnement suite à la lecture de ce texte. C'est bien le problème de la philosophie : chacun peut indéfiniment argumenter pour défendre sa position et cela même en restant totalement honnête. Je dois admettre qu'ici Marchal est dans la position la plus difficile, car il doit soutenir huit points qui sont presque tous délicats, et que la position de l'attaquant — que j'ai adoptée à sa demande — est donc assez aisée.

Le point le plus grave pour moi, si je dois en désigner un, est le traitement parfaitement insatisfaisant des probabilités évoquées, qui ne sont pas construites alors pourtant que telles que Marchal les introduit, elles devraient posséder une définition mathématique précise tirée de la notion de déployeur universel. Je pense que cette probabilité supposée par Marchal n'existe pas et que l'indéterminisme de la première personne dans un déployeur universel (à supposer qu'il puisse en exister "pour de vrai" ce dont je doute beaucoup) est un *indéterminisme sans probabilité*. L'impossibilité de prouver l'existence de cette mesure de probabilité est un coup fatal porté à tout l'argument du déployeur universel.

Je précise que, comme en 1998 quand j'ai réuni un jury interdisciplinaire pour que Bruno Marchal soutienne sa thèse à l'Université de Lille, je trouve que le raisonnement qu'il propose et le travail qu'il a produit autour est intéressant et novateur et que c'est une belle trouvaille. Je trouve utile ce type de raisonnements. L'argument de la simulation de Nick Bostrom ou l'argument de John Leslie (*Doomsday argument*) qui ont reçu plus d'attention et que je considère de la même façon —irrecevables mais intéressants— sont d'un niveau équivalent à *l'argument de Marchal*. Il serait souhaitable l'argument de Marchal soit mieux connu et plus discuté car sa richesse et la stimulation qu'on ressent quand on en prend connaissance sont loin d'avoir été pleinement exploitées.

Je remercie Bruno Marchal pour les discussions que nous avons eues récemment et qui m'ont permis d'approfondir certaines questions qui pour moi étaient restées un peu mystérieuses ; ces discussions m'ont amené à étudier ses nouveaux travaux, à préciser mes positions —ici résumées— et à lever quelques hésitations. Le désaccord sur les conclusions qui existait déjà en 1998, —et qui je le crains risque de persister—, ne doit pas être vu comme un désaccord sur l'intérêt du travail et du domaine de réflexion que Bruno Marchal a en quelque sorte créé.

## Réponse de Bruno Marchal

21 mai 2010

From: Bruno Marchal <marchal@ulb.ac.be>  
To: Jean-Paul Delahaye <delahaye@lifl.fr>  
Subject: Réponse UDA 2010  
Date: Fri, 21 May 2010 11:21:49 +0200

[...]

Je ne vois pas de mention d'erreurs dans ta note sur l'argument UDA. Si l'argument du clonage était valide, ou si le réalisme arithmétique était faux, ce serait une réfutation de l'étape zéro, c'est-à-dire que ce serait des arguments pour refuter l'hypothèse du travail (le mécanisme digital).

Mais l'argument du non clonage quantique d'un état inconnu ne réfute pas la possibilité de dupliquer les états connus et/ou classique. Par exemple, le dépoyeur universel émule tous les états quantique digitaux accessibles.

Quant au réalisme arithmétique, il est nécessaire pour donner du sens à la thèse de Church, à la notion de système formel, etc. Je le rend explicite pour éviter seulement l'ultrafinitisme en mathématique.

De toute façon je propose une déduction théorique à partir de principes théorique. La question philosophique de la vérité du mécanisme digital est intensionnellement évitée.

Ton argument sur les termes vagues ne tient pas non plus. La validité d'un raisonnement n'est pas déterminée par le domaine auquel s'applique le raisonnement, ni de la nature des termes utilisés. C'est probablement la difficulté apparente du travail: je fais un raisonnement là où peu ont l'habitude de voir un raisonnement. Exemple: le terme Dieu est supervague, mais dans la théorie classique (= on admet la logique classique) avec les axiomes:

- Tous les dieux sont des caméléons
- John est un dieu

On peut dériver de façon valide que John est un caméléon. Par contre on ne peut pas dériver de façon valide que Claude est un caméléon.

Vu que la confusion est naturelle, et qu'elle a été faites à Bruxelles, j'ai toujours insisté que le travail de thèse n'est pas de la réflexion philosophique, mais de l'argumentation déductive, et tu m'as suivi sur ce point à la défense, notamment en ne prenant pas de philosophe (non-analytique) comme membre du jury.

Pour l'usage de la sorite et la notion de conscience partielle (et de zombies

partiels), je te renvoie à nos discussions précédentes, ou alors tu peux aussi lire l'article très clair de Chalmers:

<<http://consc.net/papers/qualia.html>><http://consc.net/papers/qualia.html>

Pourquoi dis-tu que le travail n'aborde pas mathématiquement le "calcul sur l'incertitude", alors que c'est l'objet principal du travail. UDA montre "seulement" que l'hypothèse du mécanisme digitale fait de la physique un calcul d'incertitude sur des histoires computationnelles relatives. Si de la on montre que la mesure n'existe pas: on réfute le computationnalisme. Mais, Il y a un continuum d'histoires, munit d'une relation de proximité non triviale, ce qui laisse la possibilité de l'existence d'une mesure. Cela est quand même irrelevant dans UDA, puisqu'on reste ouvert à ce que ce programme puisse conduire à une réfutation du mécanisme.

Mais ensuite, et c'est l'essentiel de la thèse, et c'est bien expliqué dans l'article de SANE04 que tu mentionnes, l'approche mathématique de la recherche du calcul d'incertitude est donnée par l'interview de la machine Löbienne. L'interview de la machine Löbienne est la méthode choisie pour isoler mathématiquement le calcul des probabilités. Et je montre que la probabilité 1 (ou plutôt la crédibilité maximale) est justifiée et définie par des variantes des logiques de l'autoréférence, et j'obtiens la proposition qu'elles obéissent nécessairement à des logiques non booléennes de type quantique: les logiques S4Grz1, Z1\* et X1\*.

De plus grâce au splitting de Solovay (que je rédémontre dans "conscience et mécanisme"), on obtient les quantas comme sous-logique de la logique des qualias. Tu ne mentionnes pas cette partie, qui bien sûr est la partie technique difficile dont je dis, comme tu le rappelles, qu'elle nécessite la compréhension simultanée du problème du corps et de l'esprit, de la logique mathématique et de la mécanique quantique.

Je pense que tu avais bien compris tout ceci à la défense de thèse. Je ne comprend pas que tu dises que je n'aborde pas le problème mathématique des probabilité car c'est l'objet principal de la thèse.

La suite sont des problèmes ouvert en logique mathématique. Le premier a été résolu par Eric Vandebush

Vandebussche Eric, 2005,

<<http://iridia.ulb.ac.be/~marchal/Vandebussche/AxiomatisationZ.html>> Axiomatisation des logiques Z, Z\*, Z1, Z1\*, Document manuscrit non publié.

<<http://iridia.ulb.ac.be/~marchal/Vandebussche/AxiomatisationZ.html>><http://iridia.ulb.ac.be/~marchal/Vandebussche/AxiomatisationZ.html>

Le travail n'est qu'une preuve que si on admet le mécanisme digital, alors le problème du corps et de l'esprit est partiellement réduit et transformé en une question difficile se situant à l'intersection des domaines de la logique mathématique et de la physique expérimentale. C'est un travail difficile, mais modeste. Je suis souvent critiqué sur des prétentions qu'on m'attribue, mais que je n'ai pas. Mon but premier est d'illustrer que \*certaines\* hypothèses

philosophiques particulières permettent de réduire des problèmes complexes abordés par des philosophes et théologiens en des questions de sciences "exactes".

Pour progresser il faut axiomatiser les logiques  $X$ ,  $X^*$ ,  $X1$ ,  $X1^*$ , et  $S4Grz1$ , et vérifier qu'elles vérifient des conditions de cohérences (von Neumann), afin d'étendre les probabilités 1 au probabilités quelconques. Je suis ouvert à toutes idées.

Comme tu rends public la discussion, je te remercie d'avance d'inclure cette courte réponse dans ta page web. Merci.

[...]

## Réponse de Jean-Paul Delahaye à Bruno Marchal

21 juin 2010

### Partie A Réponse sur les détails

[...]

Je ne vois pas de mention d'erreurs dans ta note sur l'argument UDA. Si l'argument du clonage était valide, ou si le réalisme arithmétique était faux, ce serait une réfutation de l'étape zéro, c'est-à-dire que ce serait des arguments pour refuter l'hypothèse du travail (le mécanisme digital).

Etrange. Bruno Marchal me demande de pointer des erreurs dans son UDA ; j'en pointe cinq majeures (que je prends la peine d'expliquer à nouveau dans la partie B de cette réponse). Chacune des erreurs invalide sa conclusion. Lui ne voit rien ! Examiner les hypothèses de travail est tout à fait légitime, et mon texte ne se satisfaisait pas seulement de ça !

Mais l'argument du non clonage quantique d'un état inconnu ne réfute pas la possibilité de dupliquer les états connus et/ou classique. Par exemple, le déployeur universel émule tous les états quantique digitaux accessibles.

Qu'on puisse dupliquer les états classiques ou connus est vrai. La question est posée pour les états quantiques inconnus, et là la réponse est non. Je trouve ennuyeux de se placer sous des hypothèses classiques quand on prétend reconstruire la mécanique quantique à partir de l'arithmétique. En clair, Bruno Marchal néglige la mécanique quantique pour ses expériences de pensée destinées entre autres choses à retrouver la mécanique quantique !

Quant au réalisme arithmétique, il est nécessaire pour donner du sens à la thèse de Church, à la notion de système formel, etc. Je le rend explicite pour éviter seulement l'ultrafinitisme en mathématique.

Il est faux que la thèse de Church ne prenne un sens que sous l'hypothèse du réalisme arithmétique. Les intuitionnistes qui adoptent la thèse de Church n'adoptent pas le réalisme arithmétique. Marchal devrait étudier ses classiques et par exemple lire :

Beeson, M. Foundations of constructive mathematics. Metamathematical Studies, Springer-Verlag, Berlin., 1985.

qui contient beaucoup de choses précises et profondes sur la thèse de Church.

Admettre le réalisme arithmétique n'est pas faire une hypothèse bénigne comme tente de le faire croire Marchal, c'est une hypothèse forte que bien des physiciens refusent d'adopter. Puisque c'est une hypothèse forte, il faut insister dessus et ne pas la glisser sous le tapis.

De toute façon je propose une déduction théorique à partir de principes théorique. La question philosophique de la vérité du mécanisme digital est intensionnellement évitée.

## L'argument du déployeur universel de Bruno Marchal

Ton argument sur les termes vagues ne tient pas non plus. La validité d'un raisonnement n'est pas déterminée par le domaine auquel s'applique le raisonnement, ni de la nature des termes utilisés. C'est probablement la difficulté apparente du travail: je fais un raisonnement là où peu ont l'habitude de voir un raisonnement.

Exemple: le terme Dieu est supervague, mais dans la théorie classique (= on admet la logique classique) avec les axiomes:

- Tous les dieux sont des caméléons
- John est un dieu

On peut dériver de façon valide que John est un caméléon. Par contre on ne peut pas dériver de façon valide que Claude est un caméléon.

Marchal nous explique ce qu'est un formalisme et le principe de la méthode axiomatique. Merci de sa leçon. Il est dommage qu'il ne l'applique pas. Si son raisonnement était formalisé, si les axiomes qu'il utilise étaient précisés, il pourrait dire que son raisonnement est *juste ou faux* et il suffirait de l'examiner pour être en accord ou non avec lui. Rien dans l'UDA n'est formalisé, tout est assez vague. Les concepts sont utilisés en dehors de contextes scientifiques précis : tout l'inverse d'une «*déduction théorique à partir de principes théoriques*». La remarque de Marchal sur la formalisation reste donc une vue de l'esprit. On a rêvé d'une formalisation de UDA. On l'attend !

Vu que la confusion est naturelle, et qu'elle a été faite à Bruxelles, j'ai toujours insisté que le travail de thèse n'est pas de la réflexion philosophique, mais de l'argumentation déductive, et tu m'as suivi sur ce point à la défense, notamment en ne prenant pas de philosophe (non-analytique) comme membre du jury.

Dans le jury de thèse, il y avait le philosophe Paul Gochet qui dans son rapport de présoutenance écrit :

*«La thèse de B.Marchal est une contribution profondément originale à un sujet interdisciplinaire, à l'intersection de la logique et de la philosophie de l'esprit.»*

En effet, pour tout le monde —sauf peut-être pour Marchal ?— le travail présenté dans sa thèse était, pour sa partie la plus importante, un raisonnement dans le domaine de la philosophie de l'esprit. Ce raisonnement manipulait des concepts de domaines variés (calculabilité, logique modale, expérience de pensée, idée de la téléportation, graphe de calcul représentant un cerveau, supervénience, réalisme arithmétique, etc.) mais restait un raisonnement de nature philosophique et aucun des membres du jury n'y a vu un pur raisonnement déductif *vrai ou faux*, et aucun des membres du jury n'a d'ailleurs endossé l'argumentation déductive.

Le rapport de la soutenance du 2 juin 1998 de la thèse de Marchal à Lille, rédigé par le Professeur Max Dauchet quelques jours après (et signé par tous les membres du jury qui en ont validé la formulation) contient cette phrase :

*«la question n'est évidemment pas pour le Jury d'être convaincu ou non des conclusions, mais elle est de juger de la qualité de la démarche et de la fécondité des questionnements.»*

Est-ce clair ?

Merci à Marchal de cesser de dire que son jury a validé son argumentation comme le jury d'une thèse de mathématiques valide la démonstration d'un théorème. Ce n'est pas vrai.

Pour l'usage de la sorite et la notion de conscience partielle (et de zombies partiels), je te renvoie à nos discussions précédentes, ou alors tu peux aussi lire l'article très clair de Chalmers:  
<<http://consc.net/papers/qualia.html>><http://consc.net/papers/qualia.html>

Nos discussions précédentes n'ont pas répondu à mes questions et doutes. Concernant l'article de Chalmers que je connais, je ferai deux remarques élémentaires. D'abord, Chalmers défend dans cet article une conception qui n'est pas partagée par tout le monde. John Searle évidemment ne partage pas l'analyse de Chalmers. C'est le cas aussi de Jacques Mallah qui a écrit un article sur l'argumentation de Chalmers :

Jacques Mallah. *The partial brain thought experiment: partial consciousness and its implications*. Voir :<http://cogprints.org/6321/1/PBA-09.pdf>

Aussi intéressant que soit l'argumentation qu'utilise Chalmers, il est faux de croire que cela règle la question des *fading qualia* une fois pour toutes, et qu'il n'y a plus rien à discuter car tout le monde est tombé d'accord !

Deuxième point : Chalmers qui utilise des arguments assez ressemblants à ceux de Marchal dans son "graphe filmé" (type d'arguments qui, d'après Chalmers, remonte à Pylyshyn 1980) est beaucoup moins définitif sur ce qu'il faut en tirer. Je le cite :

*«If the arguments succeed, we have good reason to believe that absent and inverted qualia are impossible, and that the principle of organizational invariance is true. **These arguments do not constitute conclusive proof of the principle of organizational invariance. Such proof is generally not available in the domain of conscious experience**, where for familiar reasons one cannot even disprove the hypothesis that there is only one conscious being. But even in the **absence of proof**, we can bring to bear arguments for the plausibility and implausibility of different possibilities, and not all possibilities end up equal. I use these thought-experiments as a **plausibility argument** for the principle of organizational invariance, by showing that the alternatives have implausible consequences. If an opponent wishes to hold on to the possibility of absent or inverted qualia she can still do so, but the thought-experiments show that the cost is higher than one might have expected.*

*Perhaps it is useful to see these thought-experiments as playing a role analogous to that played by the "Schrödinger's cat" thought-experiment in the interpretation of quantum mechanics. **Schrödinger's thought-experiment does not deliver a decisive verdict** in favor of one interpretation or another, but it brings out various plausibilities and implausibilities in the interpretations, and it is something that every interpretation must ultimately come to grips with. In a similar, any theory of consciousness must ultimately come to grips with the Fading and Dancing Qualia scenarios, and some will handle them better than others. **In this way, the virtues and drawbacks of various theories are clarified.** »*

Je serais heureux que Marchal prenne modèle sur Chalmers dont il recommande la lecture et qu'il soit moins catégorique sur le caractère impératif des conclusions qu'il tire de ses propres raisonnements et expériences de pensée. Je crois que nos désaccords seraient bien moins grands si Marchal avait cette sagesse.

Je propose d'autres remarques sur ce problème que je préfère appeler "problème des prédicats vagues" dans la partie B de ce commentaire.

Pourquoi dis-tu que le travail n'aborde pas mathématiquement le "calcul sur l'incertitude", alors que c'est l'objet principal du travail. UDA montre "seulement" que l'hypothèse du mécanisme digitale fait de la physique un calcul d'incertitude sur des histoires computationnelles relatives. Si de la on montre que la mesure n'existe pas: on réfute le computationnalisme. Mais, Il y a un continuum d'histoires, munit d'une relation de proximité non triviale, ce qui laisse la possibilité de l'existence d'une mesure. Cela est quand même irrelevant dans UDA, puisqu'on reste ouvert à ce que ce programme puisse conduire à une réfutation du mécanisme.

UDA ne prend sens que si on définit la probabilité associée au déployeur universel. Si on ne la définit pas, on est dans un monde indéterministe sans probabilité (comme quand en informatique théorique on définit la classe NP). Voir plus bas d'autres remarques sur ce point.

Mais ensuite, et c'est l'essentiel de la thèse, et c'est bien expliqué dans l'article de SANE04 que tu mentionnes, l'approche mathématique de la recherche du calcul d'incertitude est donnée par l'interview de la machine Löbienne. L'interview de la machine Löbienne est la méthode choisie pour isoler mathématiquement le calcul des probabilités. Et je montre que la probabilité 1 (ou plutôt la crédibilité maximale) est justifiée et définie par des variantes des logiques de l'autoréférence, et j'obtiens la proposition qu'elles obéissent nécessairement à des logiques non booléennes de type quantique: les logiques S4Grz1, Z1\* et X1\*.

De plus grâce au splitting de Solovay (que je rédémontre dans "conscience et mécanisme"), on obtient les quantas comme sous-logique de la logique des qualias. Tu ne mentionnes pas cette partie, qui bien sûr est la partie technique difficile dont je dis, comme tu le rappelles, qu'elle nécessite la compréhension simultanée du problème du corps et de l'esprit, de la logique mathématique et de la mécanique quantique.

Je pense que tu avais bien compris tout ceci à la défense de thèse. Je ne comprend pas que tu dises que je n'aborde pas le problème mathématique des probabilité car c'est l'objet principal de la thèse.

La suite sont des problèmes ouvert en logique mathématique. Le premier a été résolu par Eric Vandebush  
Vandenbussche Eric, 2005, <<http://iridia.ulb.ac.be/~marchal/Vandenbussche/AxiomatisationZ.html>>

Axiomatisation des logiques Z, Z\*, Z1, Z1\*, Document manuscrit non publié.

<<http://iridia.ulb.ac.be/~marchal/Vandenbussche/AxiomatisationZ.html>><http://iridia.ulb.ac.be/~marchal/Vandenbussche/AxiomatisationZ.html>

Désolé, mais ces développements ne donnent pas la définition de la mesure de probabilité nécessaire à faire fonctionner UDA. Les logiques propositionnelles obtenues dans cette partie du travail ne définissent pas de probabilités et c'est bien dommage ! Aucun calcul des probabilités n'est isolé contrairement à ce Marchal prétend. Les arguments «*c'est la partie technique et difficile*» «*il faut comprendre le problème du corps et de l'esprit, la logique mathématique et la mécanique quantique*» sont une tentative d'intimidation qui permet peut-être parfois à Marchal de faire illusion, mais de tels "arguments" n'ont pas de place dans une discussion sur le fond. Je l'invite à ne pas utiliser cette méthode : je comprends aussi bien que lui ces théories et j'affirme que ce qu'il en fait ne répond en rien aux points faibles détaillés dans mon texte.

On pourra s'étonner du comportement étrange de Marchal : il lance le défi qu'on trouve des réfutations de UDA et demande qu'on lui indique les points précis parmi les huit de UDA où il y a des erreurs. Je lui en désigne 5 (voir plus bas le

rappel des erreurs, et quelques détails complémentaires). Plutôt que d'y répondre, il se contente d'affirmer : vous savez, c'est bien compliqué.

Tout le monde peut voir que UDA est un argument philosophique assez simple. J'en pointe une série de faiblesses qui sont autant de réfutations. Il faut répondre et non pas se cacher derrière de prétendus développements qui ne résolvent pas les questions posées par UDA lui-même, et en particulier qui ne répondent pas à l'objection principale : tant que la définition mathématique de la probabilité associée au déployeur n'est pas démontrée exister et spécifiée (et elle ne l'est pas) l'argument UDA ne tient pas.

Le travail n'est qu'une preuve que si on admet le mécanisme digital, alors le problème du corps et de l'esprit est partiellement réduit et transformé en une question difficile se situant à l'intersection des domaines de la logique mathématique et de la physique expérimentale.

On attend des précisions pour mener les expériences de physique que Marchal évoque ici.

C'est un travail difficile, mais modeste. Je suis souvent critiqué sur des prétentions qu'on m'attribue, mais que je n'ai pas. Mon but premier est d'illustrer que \*certaines\* hypothèses philosophiques particulières permettent de réduire des problèmes complexes abordés par des philosophes et théologiens en des questions de sciences "exactes".

Pour progresser il faut axiomatiser les logiques  $X$ ,  $X^*$ ,  $X1$ ,  $X1^*$ , et  $S4Grz1$ , et vérifier qu'elles vérifient des conditions de cohérences (von Neumann), afin d'étendre les probabilités 1 aux probabilités quelconques. Je suis ouvert à toutes idées.

Ceci est une façon de dire que le problème de la définition d'une probabilité que j'évoque n'est pas du tout résolu. Merci de cet aveu. Malheureusement cette absence détruit UDA. Dernière remarque sur ce point : je ne vois pas de réponse à l'argument que je donne et qui prouve que cette probabilité n'existe pas.

Marchal adore donner des leçons et expliquer à quel point ses conceptions sont difficiles, mais quand on l'interroge et qu'on lui soumet des problèmes précis, il se défile, se comportant avec désinvolture vis-à-vis de son interlocuteur auquel il envoie une réponse, pas même relue et comportant plus de 20 fautes d'orthographe en quelques lignes.

Comme tu rends public la discussion, je te remercie d'avance d'inclure cette courte réponse dans ta page web.

[...]

## Partie B Réponse sur le fond

Après avoir abordé les détails de la «réponse» du 21 mai, je souhaite approfondir certains aspects de notre désaccord et faire le bilan de chaque difficulté soulevée.

Mon petit article évoquait cinq points principaux. Chacun attendait une réplique détaillée de sa part. Je les reformule rapidement et analyse à chaque fois ce que Marchal répond.

*- 1 Les raisonnements "en principe" sont dangereux car ce que l'on considère acceptable en principe à une époque (par exemple du temps de la physique Newtonienne) ne l'est plus nécessairement plus tard (par exemple quand la mécanique quantique est devenu la théorie qu'il faut adopter par défaut). Les "en principe" de Marchal ne sont pas compatibles en particulier avec le no-cloning theorem ce qui est franchement ennuyeux pour une théorie qui se donne comme but de "reconstruire la physique".*

Marchal ne répond pas ou plutôt semble accepter le risque des "en principe" Newtoniens qui sont au point de départ de son raisonnement, sans mesurer que cela place son raisonnement philosophique dans un cadre conceptuel désuet.

Si chaque étape de son raisonnement UDA est l'étape d'un raisonnement démonstratif assimilable à une preuve mathématique, Marchal doit défendre chaque étape. Je ne comprends pas pourquoi il ne répond pas à la première réfutation que je propose de UDA.

*- 2 Il n'est pas sérieusement envisageable de faire fonctionner un déployeur universel dans le monde physique que nous connaissons (problème cosmologique, problème de l'exponentielle, problème de l'infini). De plus, même si cela était possible l'idée de le faire est psychologiquement invraisemblable, car aussi absurde et délirante que d'imprimer tous les livres de la bibliothèque de Babel de Borges.*

Marchal ignore la question. Il est certain que c'est là une méthode efficace pour répondre aux objections, et que cela permet facilement de se persuader d'avoir raison de tous les contradicteurs.

Comment Marchal peut-il oser écrire : « Je ne vois pas de mention d'erreurs dans ta note sur l'argument UDA. ? ».

Raisonner en s'appuyant sur des hypothèses considérées comme invraisemblables par tout le monde est une erreur. Négliger le problème que pose l'infini et même déjà l'exponentielle est une autre erreur. Etc.

Je me répète : pour défendre UDA qui comporte huit points, Marchal doit défendre chacun des huit points et répondre à toutes les réfutations qu'on lui oppose.

J'insiste. Si le raisonnement n'a pas besoin vraiment que les huit étapes soient valides, c'est que son raisonnement peut être rendu plus direct et qu'il est mal formulé. Qu'il le dise ! Si son raisonnement a besoin des huit étapes (ce que j'ai admis puisqu'il demandait qu'on lui indique un point qu'on considérerait comme réfutant UDA parmi les huit) alors Marchal ne peut se dispenser de répondre à chaque point mis en cause dans son UDA.

*- 3 Le problème de la mesure de probabilité à construire mathématiquement à partir du fonctionnement du déployeur universel, qui, si on ne le résout, pas rend impossible de prétendre qu'on peut retrouver la physique.*

Marchal comme je viens déjà de le dire ne veut pas entendre l'objection. Il ne propose pas vraiment d'idée mathématique qui permettrait de définir de manière précise cette probabilité indispensable à son argument —il demande qu'on l'aide—, dont par ailleurs j'ai expliqué qu'elle ne peut pas exister si on exige d'elle le minimum de propriétés de symétrie et d'invariance. Les résultats portant sur les logiques booléennes de type quantique ne permettent pas d'avoir une probabilité et donc comme Marchal a fini par le reconnaître le problème central des probabilités dans UDA reste non traité.

Je précise que j'ai pointé une objection sur la possibilité de définir cette probabilité à Marchal il y a longtemps et qu'elle se trouvait clairement dans l'article de la revue *Pour la science* que j'ai publié en 1998 pour présenter le travail de Marchal auquel j'ai ainsi fait largement de la publicité. J'ai repris cette objection en la détaillant un peu dans mon texte du 3 mai 2010. Avec une exceptionnelle obstination, Marchal ne daigne pas regarder ce trou béant qu'on lui montre. On lui donne une preuve que la probabilité qu'il recherche ne peut pas exister et il répond que tout va bien. Léger mieux en cette année 2010 : il reconnaît que le travail reste à faire à la fin de son message (mais refuse de répondre sur la preuve d'inexistence).

*- 4 Le problème des ontologies (pourquoi s'appuyer sur l'arithmétique et pas sur une ontologie mathématique plus large ?).*

Il apparaît peu raisonnable de nier qu'il y a un problème et de tenter de faire croire que le problème de l'ontologie acceptée ne concerne pas l'argument du déployeur universel alors que l'étape 8 de l'argument est au contraire construite sur une hypothèse ontologique arithmétique forte. Je ne suis pas d'avis qu'on résout les problèmes ontologiques par l'évocation de quelques notions techniques de logique formelle (et donc syntaxique). Ici encore l'aveuglement persistant de Marchal semble incompréhensible : je lui pose cette question depuis plus d'une dizaine d'années sans obtenir aucune réponse sérieuse, ni même simplement l'aveu que tout s'appuie sur une hypothèse ontologique *ad hoc* (car d'autres hypothèses réalistes en mathématiques sont possibles) sans qu'on puisse la justifier et qu'aucune nécessité ne contraint d'accepter ou même de la considérer avec sérieux.

*- 5 L'argument que le graphe filmé de Marchal ne prouve pas ce qu'il prétend, mais seulement que le prédicat "avoir un esprit" est un prédicat vague.*

Marchal par sa réponse montre qu'il croit que le prédicat "avoir un esprit" est un prédicat *tout ou rien*, et s'exprime en présupposant qu'on est nécessairement de cet avis sur ce point. D'où sa petite leçon sur l'axiomatisation, méthode qui, justement, ne traite pas convenablement les prédicats vagues. Comme il ne formalise pas son raisonnement, sa réponse apparaît totalement paradoxale. Il propose une solution (la formalisation) qui est aveugle aux problèmes qu'on lui soumet (le problème des prédicats vagues ne se laisse enfermer dans aucun formalisme axiomatique qui fasse l'unanimité) et, sourd cette fois au conseil qu'il vient de donner, il ne formalise même pas son raisonnement !

### **Le corps et l'esprit, émergence, physique.**

Sa stratégie de réponse à propos du problème du corps et de l'esprit consiste à décréter que ceux qui ne veulent pas penser le problème comme lui commettent une erreur fondamentale (le plus souvent pour lui : identifier l'un avec l'autre). En clair, toutes les options naturalistes, physicalistes, éliminationnistes ou qui refusent d'accorder à l'esprit un statut de première catégorie sont par avance décrétées erronées. Dit autrement encore, Marchal croit que la question du sexe des anges est une bonne question et est dans l'incapacité d'examiner les objections de ceux qui lui disent que ce n'est peut-être pas une bonne question. Sur les prédicats vagues et leur importance, nous nous permettons de lui conseiller :

Roy Sorensen, *Vagueness and Contradiction*, Oxford University Press, 2001

Vann MaxGee. *Truth, Vagueness and Paradox*, Hackett Publishing, 1991.

Il faut ajouter ici quelques remarques sur l'usage du verbe *émerger*. Une conception assez générale et largement partagée dans le domaine de l'intelligence artificielle (et cette position me semble une des meilleures) est que dès qu'un système matériel est assez complexe, possède des capacités de raisonner et de manipuler de l'information, et est en interaction étroite avec son milieu et en particulier dispose de données sur lui-même (self-awareness) alors *émerge* l'intelligence, l'esprit et la conscience. Certes l'utilisation du verbe *émerger* rend la conception peu précise et hypothétique, surtout tant que nous n'aurons pas réussi à construire réellement de tels systèmes physiques complexes. Nous comprenons qu'on puisse la trouver peu satisfaisante. L'usage du verbe *émerger* est une facilité et peut-être même une forme de tricherie. Il se trouve que malheureusement Marchal, lui aussi, utilise cette facilité qui est peut-être une forme de tricherie.

The interview of the sound Löbian machine will suggest a notion of closeness, a priori independent of any space-time, and, contrarily, will explain how a notion of space time can **emerge** from the closeness relation, in concordance with the conclusion of the UD reasoning. (SANE page 7).

If comp is correct, the appearance of physics must be recovered from some point of views **emerging** from those propositions. (SANE page 11)

Given that the UDA reasoning has shown that physics should **emerge** from a probabilistic structure bearing on its maximal consistent extensions; it is natural to interrogate the machine on its consistent extensions. (SANE page 12)

Du coup les récusations par Marchal des théories de l'esprit (qui font *émerger*

l'esprit des systèmes matériels complexes) perdent beaucoup de poids. Chez Marchal ce n'est pas l'esprit qui émerge, c'est la physique, et cela sans expliquer ce que signifie précisément ce verbe magique et trop commode. On ne gagnera donc rien à échanger une théorie naturaliste de l'esprit contre sa théorie de *l'émergence de la physique*. Loin d'éclaircir et de résoudre le problème du corps et de l'esprit, la conception de Marchal ne fait que le déplacer, tout en proposant un schéma général parfaitement inefficace s'appuyant, comme pour en augmenter le flou, sur la magie d'une "probabilistic structure" tout aussi mystérieuse que l'émergence elle-même.

Pour le dire simplement, on a l'impression qu'on serait largement perdant dans l'échange. Après avoir vu fonctionner des systèmes d'intelligence artificielle (aussi imparfaits et limités que soient ceux d'aujourd'hui) nombreux sont ceux qui acceptent de croire que l'esprit pourrait survenir/émerger dans ce type de dispositifs. En revanche, ce que nous indique Marchal reste tellement vague qu'on ne réussit pas à se persuader que la physique pourrait survenir/émerger de «l'interview des machines Lobiennes». Comment, pourquoi l'interview fait survenir/émerger la physique ? Interview par qui ? Où ça ? L'émergence dans le renversement que propose Marchal est autrement plus mystérieuse que celle qu'elle propose d'éviter ! La prétendue solution du problème du corps et de l'esprit de Marchal engendre plus de problèmes et des bien plus difficiles que ceux qu'on a sans elle !

Revenons pour terminer à l'opposition classique entre *dualisme* et *monisme*. Le dualisme est confronté à une difficulté grave qui est de décrire l'interaction entre les deux sortes d'entités postulées (en général les entités physiques et les entités mentales, mais ce pourrait être aussi les entités physiques et des entités mathématiques). Le monisme doit rendre compte de son côté des deux sortes d'entités "apparentes" à partir d'une seule. Le physicalisme s'en tire —ou tente de s'en tirer— en faisant survenir (émerger) l'esprit sur la matière (*matière* pris dans un sens large). Ce n'est pas facile, il existe un grand nombre de variantes et aujourd'hui peut-être qu'aucune n'est pleinement satisfaisante. Le travail se poursuit. La double insatisfaction ressentie face au dualisme et au monisme (de type physicaliste) fait dire à certains —dont Bruno Marchal— que le problème du corps et de l'esprit n'est pas résolu et attend encore sa résolution.

N'oublions pas toutefois que tout le monde ne pense pas cela et que certains philosophes, dont Wittgenstein et aujourd'hui Putnam, considèrent que le problème du corps et de l'esprit est simplement un faux problème dû à une mauvaise compréhension de ce qu'est *parler de corps et d'esprit*. Voir :

Ludwig Wittgenstein, *Philosophical Investigations*, Macmillan, 1958 ;

Hilary Putnam, *The Threefold Cord : Mind, Body, and World*. Columbia University Press, 2000.

Le point de vue que je défends en évoquant le prédicat "être chauve" est une variante de ces points de vue "linguistiques". Des livres entiers ont été écrits, encore récemment ayant pour thèse principale que : mal posé comme il l'est par le langage le problème du corps et de l'esprit ne peut pas recevoir de solution. Un

exemple :

Irving Krakow, *Why the Mind–Body Problem Cannot Be Solved*, Lanham, MD: University Press of America, 2002.

Admettons toutefois que le problème du corps et de l'esprit soit un authentique problème, précis, bien formulé et attendant une solution, et revenons à l'opposition moniste/dualiste. La solution que propose Marchal est moniste. Il ne s'agit pas du *monisme du physicaliste*, mais du *monisme de l'esprit* ou, ce qui revient à peu près au même pour lui, du *monisme de l'arithmétique*. Les difficultés qu'il rencontre sont les mêmes que celles d'un physicaliste : construire l'autre "réalité" à partir de celle admise. Pour lui : construire la physique en partant de l'arithmétique. Il se trouve qu'aujourd'hui ce que propose Marchal est de mon point de vue (qui je crois est partagé par tous ceux qui connaissent les idées de Marchal) bien plus imparfait, partiel et insatisfaisant que les options monistes de type physicaliste. La nécessité d'utiliser un concept (plus ou moins bien défini) d'émergence dans les deux monismes n'est pas une surprise, mais choisir entre les deux monismes doit se faire en considérant leurs succès respectifs.

Chaque programme d'I.A. qui réussit quelque chose d'intéressant est un point marqué par les monistes de type physicaliste. De l'autre côté, tant que la façon dont la physique émerge de l'arithmétique n'aura pas été rendue plus précise (en particulier en proposant une solution au problème des probabilités), tant que plus de physique n'aura pas été prouvée émerger de l'arithmétique, le monisme de Marchal ne suscitera que scepticisme. Marchal, contrairement à ce qu'il dit, ne propose pas la solution tant attendue du problème du corps et de l'esprit, mais propose une solution si peu avancée et si peu satisfaisante qu'elle ne peut prétendre entrer sérieusement en compétition avec les autres solutions aujourd'hui discutées, y compris celles de Wittgenstein et Putnam qui consistent à considérer que le problème provient d'une illusion de type linguistique.

Mon texte omettait d'aborder un autre point gênant dans l'argumentation générale de Marchal qui est l'incertitude qui persiste sur ce qu'il nomme *la physique* (quand il défend l'idée que l'arithmétique pourrait retrouver la physique en la faisant "émerger"). La définition qu'il en donne (par exemple dans l'article de SANE) ne permet pas de savoir si les lois de la gravitation font partie de *la physique*. Lui-même semble ne pas pouvoir répondre à cette question. Cela ne le gêne pas, moi si ! Accordons lui un point, si effectivement derrière ce qu'il appelle *la physique*, il ne met rien ou très peu, alors oui, on peut adopter l'idée que ce rien ou ce très peu se déduisent de l'arithmétique. Dans un tel cas cependant, il y aurait *tromperie sur la marchandise*, un peu comme dans l'attrape-nigaud qu'on propose aux enfants : «d'une seule main, je peux soulever une maison... oui une maison dessinée sur une feuille de papier». On peut se poser une question : n'est-il pas malhonnête de parler de *reconstruire la physique* (ou de la faire émerger), lorsque l'on n'est pas même en mesure de dire soi-même ce qu'est cette *physique* et qu'on utilise donc un mot (qui possède un sens assez précis pour tout le monde) dans un sens différent de l'usage courant, sens qui en définitive pour Marchal est *parfaitement indéterminé* ?

## Conclusion

On aura compris que le message de Marchal me laisse profondément insatisfait, car il ne répond en rien aux objections et réfutations que je lui présente.

J'aimerais que, contrairement à son habitude, Marchal ne prétende pas avoir résolu tous les problèmes que je lui soumets, et m'avoir convaincu par ses réponses. Marchal ne répond à rien de ce qui me gêne dans ses arguments et mon scepticisme d'il y a dix ans est devenu une certitude aujourd'hui : l'argument UDA de Marchal et la solution qu'il en tire du problème du corps et de l'esprit, à moins de progrès nombreux et importants dans sa formulation et les conséquences prouvées, ne peuvent être acceptés par personne.

En 1998, quand j'ai réuni un jury pour la soutenance de sa thèse, le jury trouvait intéressant et original le travail fait, qui semblait prometteur et proposait de nombreuses perspectives. J'ai été heureux d'aider Marchal à faire reconnaître ce travail initial comme thèse de doctorat et je ne le regrette pas, car une des fonctions de l'Université est de donner leurs chances aux idées même les plus singulières, et d'autoriser ainsi leur approfondissement pour qu'elles conduisent à de travaux aussi sérieux que possibles et reconnus internationalement.

Je trouve navrant que Marchal n'ait jamais publié d'article à partir de sa thèse (l'article SANE à la base de cette discussion n'a jamais été vraiment publié et n'est d'ailleurs disponible que sur les pages personnelles internet de Marchal).

Il me semble par ailleurs que le travail entamé dans sa thèse n'a presque pas avancé depuis 1998, et que la façon dont maintenant Marchal le présente et le défend (par exemple dans l'article SANE ou dans les réponses qu'il consent à ceux qui l'interrogent) doit être considérée comme un recul. Ses conceptions semblent aujourd'hui fossilisées et transformées en une sorte de dogme dont il fait la promotion avec une farouche obstination sur divers forums plutôt que de publier un bon texte dans une revue reconnue. Publier sur ce sujet est parfaitement possible, c'est ce qu'ont fait Tegmark, Leslie, Bostrom, Tipler, etc. qui ont proposé des arguments de même nature que le sien, utilisant des concepts associant les mêmes disciplines que celles concernées par UDA, arguments qui (contrairement à UDA) sont largement discutés dans les meilleurs journaux internationaux. Voici une petite liste de tels journaux où il est possible à Marchal de publier :

- Minds and Machines, International Journal of Foundation of Computer Science, Philosophy of Science, Foundations of Science, Foundation of Physics, Physical Cosmology and Philosophy, International Journal of Theoretical Physics, Annals of Physics, Philosophical Quarterly, American Philosophical Quarterly, Journal of Applied Philosophy, Mind, Nature, Scientific American, etc.

Marchal croit et tente de faire croire que ses jugements personnels sont des jugements admis par tous (particulièrement à propos du problème du corps et de l'esprit) mais surtout, il prétend mener un travail purement scientifique alors qu'il mène une réflexion philosophique utilisant par endroits certains appareillages conceptuels scientifiques hors de leurs champs disciplinaires (ce qui n'est pas

déshonorant !). Marchal risque de perdre tout crédit s'il persiste dans son attitude et la confusion qu'il fait entre (a) *raisonnements spéculatifs de nature philosophique* et (b) raisonnements rigoureux contrôlables et conduisant nécessairement à l'unanimité, dont chacun sait qu'ils n'existent que dans les disciplines mathématiques, ou à propos de domaines étroits et bien formalisés des sciences.

Il doit cesser d'insister pesamment en prétendant qu'un argument risqué et incertain (UDA) est une «*déduction théorique à partir de principes théoriques*» dont on doit nécessairement dire s'il est *exact* ou *faux* (ce que d'ailleurs, utilisant une méthode d'intimidation, il traduit toujours par *approuvé* ou *incompris* !).

J'ai établi dans mon premier texte complété par celui-ci, que UDA est un raisonnement s'appuyant sur des hypothèses non parfaitement explicitées, une vision classique de l'information (alors qu'une vision quantique s'impose aujourd'hui), des hypothèses physiquement irréalistes, l'hypothèse qu'une certaine probabilité existe alors que ce n'est pas le cas, un réalisme mathématique arbitraire et excessivement fort, une erreur d'interprétation de l'expérience de pensée du graphe filmé (il n'est pas vrai du tout que l'expérience de pensée décrite contraigne à jeter le monde physique par dessus bord !), un usage abusif du mot "physique", etc.

L'argument du déployeur universel (UDA) est intéressant, mais n'est pas ce qu'en dit Marchal qui, en définitive, fait du tort à son idée en la défendant mal.