LE BAYESIANISME MÊME À VISAGE HUMAIN EST-IL DÉFENDABLE ?

Le but de cette intervention est de montrer que si le bayesianisme (même) à visage humain à la mode de Jeffrey fournit une image acceptable de l'état d'un agent, il est incapable même en principe – contrairement à ce que prétend Jeffrey – de rendre compte de façon minimalement plausible des mécanismes qui entrent en jeu dans le processus de délibération rationnelle. Le bayesianisme de Jeffrey est définitivement statique, il décrit un état d'équilibre et ne peut asseoir une cinématique des probabilités sans subir des modifications profondes.

Ce texte comprend deux parties. La première est une présentation succincte de l'essentiel du bayesianisme à la Jeffrey. La seconde est une évaluation critique des notions de délibération et d'acte telles qu'elles se présentent dans un tel cadre ; je montre que la cinématique jeffreyienne est au mieux triviale.

Le bayesianisme à la Jeffrey

Il n'y a pas unanimité dans la communauté savante sur l'existence d'un ensemble de conditions nécessaires et suffisantes qui caractériserait le bayesianisme. Jeffrey donne trois propriétés qui sont le plus souvent invoquées pour le caractériser[1]. La moins

1. R. Jeffrey, « Bayesiannism with a Human Face », in *Testing Scientific Theories*, J. Hamon (éd.), Minneapolis, University of Minnesota Press, « Minnesota Studies in the Philosophy of Sciences Series », vol. X, 1983.

contestable de ces propriétés est celle qui prescrit, dans un contexte de décision, de maximiser « une utilité attendue relativement à une certaine assignation de probabilité sous-jacente aux états de la nature ». Être rationnel, c'est maximiser l'utilité attendue.

La seconde, que Jeffrey va contester ou plutôt généraliser en présentant son bayesianisme à visage humain, est que la fonction de probabilité de l'agent évolue par conditionalisation : la nouvelle fonction de probabilité d'un agent qui acquiert une nouvelle information est la fonction de probabilité obtenue à partir de l'ancienne en conditionalisant sur la nouvelle information. Si j'apprenais qu'il pleut, ma nouvelle fonction de probabilité attribuerait une valeur nulle aux états du monde ou il ne pleut pas et aux autres leur probabilité actuelle divisée par un facteur constant qui est la probabilité actuelle de « Il pleut ».

Enfin, la troisième, qui ne nous concerne pas ici, est que la fonction de probabilité de départ est la fonction $m^\star$ de Carnap.

Jeffrey propose une façon bien particulière de satisfaire la première condition. Il rompt, en effet, avec l'approche traditionnelle de von Neuman-Morgenstern, reprise par Savage, qui adopte la trichotomie Actes/Conséquences/État du monde. Chez von Neuman-Morgenstern[2], par exemple, décider d'accomplir un Acte A_i, c'est simplement accomplir parmi les actes accessibles A_j, celui qui maximise la fonction d'utilité suivante :

$$U(A_j) = \mathrm{o}(\Sigma ; {}_{Ek})\ Prob(E_k)\, u(C_{jk})$$

où les E_k constituent une partition des états du monde, C_{jk} est la conséquence de l'acte A_j si l'état du monde se révèle être E_k.

On suppose que tout agent possède une fonction de préférence définie sur les conséquences et si cette fonction satisfait à quelques postulats relativement plausibles (transitivité, principe de dominance, etc.), on montre que U est unique aux transformations linéaires positives près. En d'autres termes, l'observation (idéale) du classement préférentiel des actes d'un agent rationnel par deux interprétants conduira à l'élaboration de deux fonctions d'utilité qui seront une transformation linéaire positive l'une de l'autre. La fonction u peut être interprétée comme une sorte de désirabilité

2. En fait, chez von Neuman-Morgenstern il n'est pas à strictement parler question d'acte mais de ticket de loterie. Les deux notions sont équivalentes pour ce qui nous concerne.

intrinsèque des conséquences[3]. C'est le fameux théorème de représentation.

Jeffrey rejette donc la trichotomie Actes / Conséquences / États. Selon lui, une seule catégorie suffit, celle de proposition. Une bonne raison à l'appui de ce rejet est que la trichotomie Actes/Conséquences/États semble bien d'ordre méthodologique plutôt qu'ontologique. Ce qui dans un certain contexte est un état du monde devient dans un autre une conséquence. Ce qui dans un certain contexte est un acte devient dans un autre un état du monde. Par exemple, je décide de vous frapper avec un bâton, c'est un acte. Mais cet acte est une conséquence d'un autre acte, celui que vous avez accompli en me menaçant avec un revolver, ce qui pour moi est désormais un état du monde.

Une théorie qui ferait l'économie de cette trichotomie tout en préservant l'essentiel, c'est-à-dire que la proposition qui exprime l'acte qui est accompli est celle qui maximise une certaine fonction d'utilité, aurait une simplicité et une élégance bien supérieure à une théorie qui ne le ferait pas, toutes choses étant égales par ailleurs.

Un problème à la fois technique et philosophique se pose rapidement à cette nouvelle approche. Un des principes fondamentaux de l'approche classique est en effet celui de la dominance : si un acte A_j est préférable à un acte A_i dans tous les états du monde de la partition (E_k) alors A_j est préférable à A_i. Ce principe présuppose qu'il y a indépendance totale entre les actes et les états du monde. Plus précisément, ce principe n'est valable que si la probabilité de chacun des états du monde de la partition est exactement la même que l'acte soit accompli ou non. C'est d'ailleurs cette indépendance qui sert de critère pour distinguer les conséquences des actes des états du monde eux-mêmes. Je préfère, par exemple, fumer à ne pas fumer sachant que je n'aurai pas le cancer du poumon et je préfère fumer à ne pas fumer sachant que j'aurai le cancer. Donc je fume ! Ce sophisme repose sur le fait (que je crois) qu'il y a dépendance causale entre le fait de fumer et de développer un cancer.

Jeffrey renonce au principe de dominance ou plus précisément restreint trivialement son usage aux cas d'indépendance. Tous les problèmes ne sont pas résolus pour autant. En rejetant la trichotomie,

3. Voir L. J. Savage, *The Foundations of Statistics*, New York, Dover Publications, 1972, chap. 5 ; R. D. Luce et H. Raiffa, *Games and Decisions*, New York, Dover Publications, 1985.

Jeffrey doit adopter une forme de holisme en ce qui concerne les fonctions de probabilité et d'utilité. En effet, le rejet de la trichotomie ne permet plus de définir la fonction d'utilité comme ci-dessus basée sur des désirabilités intrinsèques des conséquences. À quel sous-ensemble de l'ensemble des propositions une telle fonction devrait-elle s'appliquer ? Répondre à cette question reviendrait à caractériser le sous-ensemble des propositions qui expriment des conséquences potentielles et donc de réintroduire sans la nommer la distinction états / conséquences ! L'approche de Jeffrey suppose donc que désirabilités et probabilités sont intimement liées et que toute mesure de préférence sur des propositions est déjà une mesure d'utilité.

Jeffrey renonce ainsi au principe de dominance comme contrainte générale et présente une axiomatisation différente qui rend compte de ce holisme (c'est en fait Bolker qui, en 1964, démontrera le théorème d'existence et d'unicité). Rappelons brièvement l'essentiel de la construction. On se donne d'abord un ensemble de propositions M (c'est-à-dire un ensemble d'ensembles de mondes possibles) et une relation de préférence $\leq$ sur cet ensemble. La relation de préférence est caractérisée par les axiomes[4] suivants :

Ax. 1 : $\leq$ est transitive et totale sur M qui forme une σ-algèbre complète ;

Ax. 2 : $\mathbf{F} \approx \mathbf{V}$ (les contradictions et les tautologies sont équidésirables) ;

Ax. 3 : pour un certain B (le bon) $\neg B < \mathbf{V} < B$;

Ax. 4 : si A et B sont incompatibles, alors

(1) si $A < B$, $A < (A \vee B) < B$

(2) si $A \approx B$, $A \approx (A \vee B) \approx B$;

Ax. 5 : si A et B sont incompatibles et $A \approx B$ et $(A \vee C) \approx (B \vee C)$ pour un certain C incompatible avec A et avec B et qu'il n'est pas le cas que $A \approx C$, alors pour tout D tel que D est incompatible avec A et avec B et qu'il n'est pas le cas que $A \approx D$, $(A \vee D) \approx (B \vee D)$;

Ax. 6 : Soit une suite $E_1,..., E_n,...$, telle que $E_i E_{i+1}$ pour tout

4. Voir R. Jeffrey, *The Logic of Decision*, Chicago, University of Chicago Press, deuxième édition, 1983, chap. 6 et 9 et R. Jeffrey, « Axiomatizing the Logic of Decision », *in* A. Hooker, J. J. Leach et E. F. McClennen (éds), *Foundations and Applications of Decision Theory*, vol.1, C, 1978 ; E. Bolker, « A Simultaneous Axiomatization of Utility and Subjective Probability », *Philosophy of Science*, 34, 1966, p. 292-312.

$i \leq n$ et E un supremum de la suite. Si $F \leq E \leq H$, alors il existe un j tel que pour tout $i > j$, $F \leq E_i \leq H$.

Résultat de Bolker :

Pour toute structure satisfaisante les axiomes Ax.1–Ax.6, il existe une paire $< u, P >$ telle que l'utilité attendue

$$(\star)\, U(A) = \frac{1}{P(A)} \int_A u\, dP \quad \text{si } P(A) \neq 0$$

est définie de manière unique à la classe de transformations suivantes près

$$U'(A) = \mathrm{F}(aU(A) + b\,;\, cU(A) + d)$$

et

$$P'(A) = P(A)\,(cU\,(A) + d)$$

où

$$ad - bc > 0$$
$$cU(A) + d > 0$$
$$cU(\mathbf{V}) + d = 1$$

et u est unique au sens où

$$\int_A (\mathrm{u} - \mathrm{u}')\, dP = 0$$

pour tout u' satisfaisant la contrainte.

Jeffrey a apparemment gagné son pari et la trichotomie est désormais inutile.

Certains problèmes vont cependant se poser en ce qui concerne la dynamique des états épistémiques, c'est-à-dire la dynamique des paires $< u, P >$.

Nous en venons ainsi à l'examen du deuxième postulat du bayesianisme, celui-là même que Jeffrey va humaniser.

Selon le bayesianisme classique, la fonction de croyance de l'agent évolue par conditionnalisation : la paire $< u, P >$ de l'agent qui découvre que B devient la paire $< u_B, P_B >$ où P_B est obtenue en substituant $P(\ |B)$ à P dans $(\star)$.

Jeffrey considère que cette manière de changer de fonction de probabilité n'est qu'un cas particulier qui ne peut rendre compte

de la façon dont de nouvelles informations viennent influer sur la fonction de probabilité de l'agent. En effet, dans beaucoup de cas, il n'est pas possible de spécifier, suite à une expérience de l'agent, quelle proposition est soudainement devenue vraie. Considérons l'exemple suivant.

Un paysan se demande s'il doit semer. Il considère trois états du monde comme étant possibles. Le temps va rester au sec (S), il va pleuvoir (M), il va grêler (G). La décision que le paysan doit prendre est celle de semer ou non. Pour son calcul de l'utilité de semer, notre paysan doit évaluer la probabilité que ses semis survivent s'il sème. Supposons que le paysan considère les trois états S, M et G comme équiprobables, c'est-à-dire que $P(S) = P(M) = P(G)$. Soudain, le vent se lève, des nuages noirs roulent à l'horizon et l'air fraîchit. Les probabilités respectives de S, M et G changent tout à coup de valeur. Qu'en est-il de la probabilité que ses semis survivent (Q) ? Selon Jeffrey

$$P'(Q) = P(Q|S)P'(S) + P(Q|M)P'(M) + P(Q|G)P'(G)$$

mais il est douteux que l'on trouve un E tel que

$$P'(Q) = P(Q|E)$$

Jeffrey propose donc une autre manière de réviser les probabilités : généralisant la formule ci-haut et suivant les lignes directrices de notre exemple, on obtient la règle suivante. Soit (E_k) une partition de l'ensemble des mondes possibles[5] telle que suite à une observation, chacun des éléments E_i possède la probabilité $P(E_i)$[6] .

La nouvelle fonction P' est liée à l'ancienne par la formule suivante :

$$P'(A) = \sum_{Ek} P'(E_k)\, P(A|E_k)$$

C'est le bayesianisme à visage humain.

Cette généralisation de la conditionnalisation est très intéressante et très convaincante à la fois d'un point de vue intuitif et du

5. Donc $\sum_{E_k} P(E_k) = 1$.

6. On a bien sûr $\sum_{E_k} P'(E_k) = 1$.

point de vue formel. En effet, si la conditionnalisation semble bien décrire la bonne manière de modifier la fonction de probabilité en apprenant qu'un certain énoncé est vrai, la formule de Jeffrey semble également décrire la bonne manière de modifier la fonction de probabilité suite à la modification de la fonction de probabilité de l'agent dans un contexte où l'agent lui-même est incapable d'exprimer par une proposition les modifications de sa connaissance du monde. De telles modifications de la fonction de probabilité suite à l'apport de nouvelles informations ineffables sont courantes et le fait que l'on puisse en rendre compte simplement en généralisant la conditionnalisation augmente la crédibilité de l'approche de Jeffrey. Mais outre le caractère intuitif de cette approche, il y a de bonnes raisons formelles de l'accepter. En effet, si $P(E)$ et $P'(E)$ sont toutes deux positives, la conjonction de

(1) $P(\ |E) = P'(\ |E)$
et de
(2) $P'(E) = 1$
est équivalente à
(3) $P' = P(\ |E)$

Si après révision la nouvelle fonction de probabilité se comporte exactement comme l'ancienne lorsqu'on conditionnalise sur une certaine proposition E et que la nouvelle fonction de probabilité attribue à E la probabilité 1, alors la nouvelle fonction est celle obtenue en conditionnalisant l'ancienne sur E (et vice versa).

On montre aussi facilement le résultat suivant.

Si (E_k) est une partition de la tautologie et P et P' deux fonctions de probabilité telle que pour tout E_k de la partition $P(E_k)$ et $P'(E_k)$ sont non nulles, la conjonction de

(1') $P(\ |E_k) = P'(\ |E_k)$ pour tout E_k et de
(2') $o(\Sigma ; E_k)\ P'(E_k) = 1$ (trivial, je le mets pour respecter la symétrie)
est équivalente à
(3') $P' = o(\Sigma ; E_k)\ P'(E_k)\ P(\ |E_k)$

Conclusion : il semble bien que nous ayons là la seule façon cohérente possible (la seconde étant une généralisation de la première) de réagir à de nouvelles informations.

Conditionnaliser sur un acte

Laissons de côté pour l'instant le visage humain et redevenons conditionnaliseurs.

Une des conséquences de l'abandon de la trichotomie par Jeffrey est que l'agent qui délibère pour maximiser sa fonction d'utilité conditionnalise sur ses actes. Le schéma est simple : soit A_1, ..., A_n, n actes accessibles. Lequel accomplir ? Un de ceux qui maximisent l'utilité. Quelle fonction de probabilité faut-il utiliser ? Celle qui conditionnalise sur l'acte envisagé.

Écoutons Jeffrey.

> Supposez que vous croyez en votre pouvoir de faire que A soit vrai si vous le désirez [...] et que vous êtes convaincu que A est préférable à toutes vos autres options. Alors $P(A) = 1$, car vous savez que vous allez rendre A vrai[7].

Le problème, (posé à Jeffrey par Bolker dans une lettre) est qu'une des conséquences triviales de (⋆) est que

$$U(A \vee B)P(A \vee B) = U(A)P(A) + U(B)P(B)$$

(pour A et B incompatibles).

Or, en remplaçant B par $\neg A$ on obtient

$$U(\mathbf{V}) = U(A)$$

En clair, sachant que la proposition qui exprime un acte potentiel va devenir vraie si l'agent décide de l'accomplir, cet acte possède la même désirabilité que la tautologie. Ou encore : sachant par conditionnalisation sur les actes (crus) accessibles lequel maximise l'utilité attendue, l'agent n'a nul besoin de délibérer, il se voit accomplir (ou apprend qu'il va accomplir) celui qui maximise l'utilité. La rationalité bayesienne tue en quelque sorte le désir, parce qu'on ne peut désirer ce que l'on possède déjà.

La défense de Jeffrey consiste à nier que la fonction de probabilité utilisée par l'agent pour calculer l'utilité de l'acte A soit $P(\ |A)$. La fonction de probabilité d'un agent ne devient $P(\ |A)$ qu'au moment où l'agent accomplit effectivement A. Appelons P_A la

7. R. Jeffrey, « A Note on the Kinematics of Preference », *Erkenntnis*, vol. 11, 1977, p. 136.

fonction de probabilité que l'agent utilise lorsqu'il délibère c'est-à-dire pour le calcul de l'utilité attendue de l'accomplissement de l'acte A.

Le «paradoxe» de Bolker nous force à admettre que $P_A \neq P(\ |A)$, sinon tous les actes potentiels ont la même utilité, celle de la tautologie.

Par ailleurs, il n'est pas vrai en général que $P_A = P$, P étant la fonction de probabilité de l'agent avant qu'il n'amorce le processus de délibération. L'exemple suivant illustre le caractère peu plausible de cette identité. Supposons que je n'ai pas encore réfléchi sérieusement à ce que je vais faire ce soir. Les deux options sont «Je vais au cinéma» (C) «Je reste à la maison pour terminer un article promis pour demain» (T).

Supposons qu'avant d'amorcer le processus de délibération on ait

$$P(C \vee T) \approx 1$$

car je sais que je vais très probablement choisir une des deux options et que j'exclus pratiquement toute autre option. Mais je ne sais pas laquelle je vais choisir n'ayant pas encore pesé le pour et le contre. Ceci peut s'exprimer par le fait que

$$P(C) \approx P(T) \approx 0{,}5$$

J'amorce maintenant le processus de délibération et je calcule $U(C)$ et $U(T)$.

Il est assez évident que non seulement ce n'est pas la fonction P que j'utilise pour les calculs des utilités parce ce n'est pas la même fonction que j'utilise pour le calcul de $U(C)$ et pour le calcul de $U(T)$. Il serait, en effet, pour le moins bizarre que dans l'hypothèse où je décide de rester à la maison je continue d'accorder à la proposition «Je vais au cinéma» la même probabilité qu'à la proposition «Je reste à la maison pour terminer un article promis pour demain». Donc, $P \neq P_A$, $P \neq P_B$, et $P_B \neq P_A$. L'exemple que nous venons d'examiner suggère fortement que P_A doit se situer quelque part entre P et $P(\ |A)$, probablement plus près de $P(\ |A)$ que de P.

Indépendamment de l'argument de Bolker, une raison qui nous fait douter que P_A soit $P(\ |A)$ est qu'il y a une certaine possibilité que l'agent échoue dans l'accomplissement de l'acte qu'il a décidé d'accomplir.

Ceci nous amène à discuter des conditions que doit satisfaire une proposition A pour être envisagée par un agent rationnel comme un acte accessible. Ces conditions sont très sensibles au contexte et sont, bien sûr, subjectives. Effectuer un transit vers Jupiter était probablement considéré comme un acte accessible par les adeptes de l'Ordre du Temple Solaire mais pas pour nous. Pour qu'une proposition A puisse, dans un contexte de délibération, être considérée comme un acte accessible, l'agent doit croire que s'il décide de l'accomplir, A deviendra très probablement vraie. Une première condition qui semble s'imposer de façon naturelle serait la donc suivante.

Pour qu'une proposition A soit considérée comme exprimant un acte accessible pour un agent, il faut que la fonction de probabilité que l'agent utilise dans le processus de délibération, soit P_A, soit telle que $P_A(A) \approx 1$ autrement dit, l'agent doit croire fortement que s'il décide de rendre A vrai, A sera très probablement vrai.

Cette contrainte entraîne un certain relativisme – et même une certaine ambiguïté – dans la description de l'acte qui est accompli. On peut représenter cette ambiguïté par une suite de propositions exprimant des conséquences

$$A_1,\ldots,A_n,\ldots$$

Notre condition énoncée ci-haut pourrait s'exprimer de la façon suivante : une condition nécessaire pour que Ai dans la suite des conséquences puisse être considérée comme un acte accessible est que

$$\text{(i)} P_{Aj}(A_i) \approx 1 \text{ pour tout } j \leq i$$

Considérons l'exemple suivant. Je décide d'aller dîner dans mon restaurant préféré. Dans un contexte donné, la probabilité que je réussisse à effectivement dîner dans ce restaurant étant donnée ma décision est suffisamment grande que je peux considérer que l'acte accompli est bien d'aller dîner dans ce restaurant. Dans un autre contexte (nous sommes lundi et je ne sais plus trop si ce restaurant ferme le lundi ou le mardi…), on ne pourra plus considérer que d'aller dîner dans ce restaurant est un acte accessible. C'est plutôt une conséquence plus ou moins probable d'une série d'autres actes comme monter dans l'auto, démarrer, aller jusqu'au restaurant. Une fois sur place, si l'état du monde est que le restaurant ferme le mardi, la proposition redevient un acte accessible. Mais il en est de même pour tous les autres actes. Si par exemple mon auto refuse une fois sur deux de démarrer, ce n'est

qu'après avoir réussi à la démarrer que d'aller au restaurant devient un acte accessible. Bref, pour qu'une proposition exprime un acte accessible, elle doit se voir attribuer une probabilité très grande par la fonction de probabilité qui est utilisée dans le processus de calcul de l'utilité.

Cette condition n'est cependant pas assez forte. On peut imaginer des situations où $P_{Aj}(A_i) \approx 1$ mais que les conséquences d'un échec sont catastrophiques. Considérons les situations suivantes.

Je me promène à la campagne. J'arrive tout à coup à un ruisseau qui entrave ma route, un petit ruisseau d'un mètre de large. Le paysage de l'autre côté est très joli, le terrain est dégagé et je n'ai aucune envie de rebrousser chemin. Je délibère : je saute le ruisseau (S) ou je ne saute pas le ruisseau.

Comme je suis un sportif en bonne santé, que le terrain ne contient visiblement pas d'embûche, etc.,

$$P_S(S) \approx 1$$

Il y a cependant une probabilité très faible que je rate mon coup même si je décide de faire S. Mais dans cet exemple, l'utilité attendue de l'échec est faible, négligeable par rapport à celle de la réussite.

Imaginons la même scène mais au lieu d'un ruisseau c'est un canyon d'un mètre de large et d'un kilomètre de profondeur. On a également

$$P_S(S) \approx 1$$

Mais ici j'hésite à sauter. Pourquoi ? Parce que dans ce dernier cas, et bien que les probabilités soient les mêmes, les conséquences d'un échec sont très désagréables : la probabilité que je rate mon saut est très faible si je décide de sauter mais les conséquences sont désastreuses ici, la mort quasi certaine. Bref, bien que la probabilité soit faible, l'utilité – négative – est loin d'être négligeable. Il faut donc ajouter une condition supplémentaire portant sur l'utilité d'un échec. Je propose la condition suivante. Pour que la proposition A_i de la suite représente un acte accessible à un agent rationnel, il faut que

(ii) $|U_{Aj}(\neg A_i)| << |U_{Aj}(A_i)|$ pour tout $j \leq i$

Cette condition signifie que non seulement l'agent croit que A va être le cas s'il décide que A mais que s'il échouait l'utilité positive ou négative est négligeable comparée à l'utilité attendue de la réussite.

Revenons au paradoxe de Bolker. Les conditions (i) et (ii), qui sont difficilement contestables comme conditions nécessaires, ont, par un argument de continuité, comme conséquence que si A est un acte accessible

$$U_A(\mathrm{V}) \approx U_A(A)$$

Mince résultat qu'avait déjà concédé Jeffrey ? Pas tout à fait. U_A n'est pas la fonction d'utilité de l'agent après qu'il ait accompli A mais bien celle qu'il utilise dans le processus de délibération. Morale : même si on abandonne l'hypothèse que $P_A = P(\ |A)$, si P_A satisfait (i) et (ii) on retrouve intact le paradoxe de Bolker.

On pourrait croire que ce n'est pas seulement l'approche de Jeffrey qui fait problème mais le principe de maximisation lui-même. Rappelons-nous le résultat élémentaire présenté plus haut, soit que

si $P(E)$ et $P'(E)$ sont toutes deux positives, la conjonction de

(1) $P(\ |E) = P'(\ |E)$ et de

(2) $P'(E) = 1$

est équivalente à

(3) $P' = P(\ |E)$

Si E est A et P'est P_A, et si on remplace = par ≈ on obtient (par le même argument de continuité) :

si $P(A)$ et $P_A(A)$ sont toutes deux positives, la conjonction de

(1) $P(\ |A) \approx P_A(\ |A)$ et de

(2) $P_A(A) \approx 1$

est équivalente à

(3) $P_A \approx P(\ |A)$

Si on rejette (3), il faut rejeter soit (1), soit (2). Nous venons d'argumenter que (2) est incontournable. Où est l'erreur ? C'est celle de croire que $P(\ |A) \approx P_A(\ |A)$. Bien sûr, $P(A|A) \approx P(A|A) \approx 1$, mais en général il n'est pas vrai que $P(X|A) \approx P_A(X|A)$, en particulier lorsque X est une conséquence de A dont l'agent n'est pas conscient avant le processus de délibération mais dont justement il prend conscience au cours du processus de délibération lorsqu'il

examine les conséquences d'accomplir A. Pour dire les choses autrement, l'utilisation de la conditionnalisation dans la description d'un processus de délibération revient à supposer que l'agent qui délibère a déjà dans l'établissement de sa fonction de probabilité et avant que le processus de délibération ne débute, calculé toutes les conséquences de ses actes potentiels. Découvrant au fur et à mesure que le monde se dévoile quels sont les états qui sont réalisés, l'agent découvre en même temps quels sont les actes qu'il va accomplir, leurs probabilités conditionnelles étant devenues leurs probabilités tout court.

Cette difficulté est incontournable pour la théorie Jeffreyienne car elle est une conséquence des hypothèses centrales. Revenons au fameux théorème de représentation de Bolker. Nous avons déjà signalé le caractère holistique de la théorie de Jeffrey. Selon cette théorie, probabilité et désirabilité sont indissociables. La formule (*)

$$U(A) = \mathrm{F}(1\,;P(A))\,\mathrm{I}(A\,;;\,u\,dP) \text{ si } P(A) \neq 0$$

exprime bien ce caractère holiste. À première vue cette formule semble exprimer que l'utilité d'une proposition est inversement proportionnelle à sa probabilité. Il n'en est rien. La présence de la fraction $\frac{1}{P(A)}$ est nécessaire parce que u est déjà une sorte d'utilité qui tient compte de la probabilité.

L'utilité de l'univers jeffreyien est de 1, et cette utilité se répartit sur A et sur $\neg A$, pour tout A. Donc, plus une proposition est probable, plus son utilité s'approche de 1. Cette propriété est incompatible avec la notion même d'acte. Est-il soutenable que la proposition exprimant que je vais mourir un jour possède la même utilité que celle d'aller encaisser le million de dollars que je viens de gagner avec le ticket de loterie que j'ai en poche ? Ces deux propositions ont pour Jeffrey la même utilité parce que leur probabilité, qui est très près de 1, échappe en quelque sorte à l'influence de l'agent. La première parce qu'elle exprime un état du monde sur lequel l'agent ne croit pas avoir d'influence, la seconde parce qu'elle est la plus utile et donc que c'est elle que l'agent se verra accomplir. Le fait que les propositions exprimant les actes potentiels d'un agent aient le même statut que n'importe quelle autre proposition, c'est-à-dire qu'elles sont vraies ou fausses mais que cette vérité ou fausseté est inconnue de l'agent qui en est réduit à leur attribuer une probabilité, doublé du fait que cette probabilité dépend directement, pour un acte accessible, de l'utilité (parce que l'agent va justement accomplir

la plus utile des propositions) font de l'univers jeffreyien un univers sans surprise en ce qui concerne les actions des agents rationnels. Les seules surprises concernent les états du monde, les actions, elles, sont déjà connues (conditionnellement aux états du monde, bien sûr).

La théorie de Jeffrey ne permet pas de rendre compte de l'intuition très simple qu'un acte est une intervention dans le monde qui bouleverse le cours des choses, que ce bouleversement est la somme des conséquences de l'acte. Ce phénomène ne peut tout simplement pas être représenté par la conditionnalisation sur l'acte. L'agent jeffreyien est un robot programmé, totalement rationnel, qui découvre – au fur et à mesure que les états du monde se déploient – quelles actions il va accomplir parmi toutes les actions dont il a *déjà* calculé l'utilité conditionnelle. Cette construction n'est pas un modèle crédible de la prise de décision rationnelle.

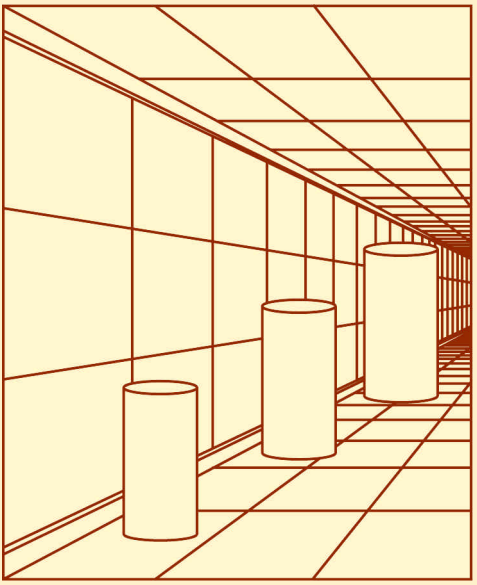
Reste une possibilité, formelle, celle du bayesianisme à visage humain. Cette possibilité consisterait à dire que dans un processus de délibération ce n'est plus un acte A qui est accompli ni une fonction P_A qui est utilisée par l'agent mais qu'il existe une partition E_k dont les éléments subissent un changement de probabilité. Je doute qu'une telle approche soit immunisée contre le paradoxe à la Bolker bien que la reconstruction d'un argument similaire à celui ci-dessus ne semble pas possible. Elle doit quand même être rejetée. Nous avons vu que le bayesiannisme à visage humain ne s'applique que lorsqu'il n'y a pas de proposition sur laquelle l'agent pourrait conditionnaliser pour rendre compte de la modification intervenue dans sa fonction de probabilité. L'adoption de cette position de repli aurait ainsi pour conséquence triviale que *tout* acte accompli par un agent rationnel est, pour cet agent, ineffable.

Conclusion : un agent jeffreyien ou bien se voit accomplir ce qu'il savait déjà maximiser l'utilité et donc ne délibère pas, ou bien est incapable de s'exprimer à lui-même ce qu'il vient de décider. Pour ces raisons, cette théorie doit être rejetée.

François LEPAGE
Université de Montréal

Cahiers de Philosophie
de l'Université de Caen

Philosophie analytique



1997-1998 N° 31-32

Presses Universitaires de Caen