

## Qu'est-ce qu'un acte jugé faisable ?

François Lepage

Volume 28, Number 2, Fall 2001

URI: <https://id.erudit.org/iderudit/005671ar>

DOI: <https://doi.org/10.7202/005671ar>

[See table of contents](#)

Publisher(s)

Société de philosophie du Québec

ISSN

0316-2923 (print)

1492-1391 (digital)

[Explore this journal](#)

Cite this article

Lepage, F. (2001). Qu'est-ce qu'un acte jugé faisable ? *Philosophiques*, 28(2), 369–380. <https://doi.org/10.7202/005671ar>

Article abstract

Among the set of all propositions on which a rational agent entertains beliefs and desires, some express acts considered as possible options for the agent. The aim of this paper is to characterize this subset of propositions in terms of their probability and their desirability.

# Qu'est-ce qu'un acte jugé faisable ?<sup>1</sup>

FRANÇOIS LEPAGE

Université de Montréal

François.Lepage@umontreal.ca

**RÉSUMÉ.** — Parmi l'ensemble de toutes les propositions à propos desquelles un agent rationnel entretient des croyances et des désirs, certaines correspondent à des actes que l'agent juge faisables. Le but de mon intervention est de caractériser ce sous-ensemble de propositions en termes de leur probabilité et de leur désirabilité.

**ABSTRACT.** — Among the set of all propositions on which a rational agent entertains beliefs and desires, some express acts considered as possible options for the agent. The aim of this paper is to characterize this subset of propositions in terms of their probability and their desirability.

*Tout en finissant son troisième café, Pierre regarde par la fenêtre et se demande s'il va prendre son auto pour aller au bureau. Dehors le vent souffle et la première tempête de neige de l'hiver vient de commencer. La météo annonce 30 cm de neige et des rafales pouvant aller jusqu'à 70 km/heure. Prendre l'autobus n'est, par ailleurs, pas très attrayant. Il sera bondé de gens qui se bousculent et d'élèves armés d'hostiles sacs à dos. Pierre pèse le pour et le contre, hésite et décide de prendre son auto, ce qu'il fait après s'être brossé les dents.*

Ce morne et banal scénario illustre un phénomène non moins morne et banal : un agent accomplit un acte suite à un processus de délibération. Cet acte peut s'exprimer par une proposition : *Pierre prend son auto*.

Le fait qu'une certaine proposition *A* représente pour un agent *a* un acte jugé faisable dépend de la subjectivité de cet agent et que cet acte soit réellement faisable n'est pas pertinent pour caractériser cette propriété. Si je crois que je peux éteindre le soleil en levant le bras, alors éteindre le soleil est un acte que je considère faisable. Si je crois que je vous contrôle, alors lever votre bras est un acte que je juge faisable. Quelles propriétés possèdent donc ces propositions pour pouvoir jouir de ce statut d'actes jugés faisables ? Le but de mon intervention est d'essayer de réduire ou d'éliminer la notion « *a* croit que *A* est un acte jugé faisable » au profit de contraintes sur les deux fonctions qui représentent la subjectivité de l'agent, soit sa fonction de croyance et sa fonction de désirabilité. Plus précisément, il s'agit de fournir un ensemble de conditions nécessaires et suffisantes que doivent posséder les

---

1. Je remercie Christine Tappolet pour ses remarques judicieuses suite à la lecture d'une première version de ce texte. Je remercie également deux évaluateurs anonymes de *Philosophiques* pour leurs précieuses remarques.

fonctions de croyance et de désirabilité d'un agent pour qu'une proposition soit considérée, de son point de vue, comme un acte jugé faisable.

L'intérêt d'une telle caractérisation, si elle s'avérait réalisable, est évident du point de vue de l'ontologie de la théorie de la décision. Classiquement, la théorie de la décision utilise trois notions irréductibles : actes, états du monde et conséquences. Or ces trois types d'entités sont représentées dans le langage par une même catégorie sémantique, celle de proposition.

Richard Jeffrey [Jef83] dans sa théorie telle qu'exposée dans son ouvrage « The Logic of Decision » rejette carrément cette trichotomie et ne fait pas de distinction entre différents types de propositions : l'ensemble des propositions forme une algèbre de Boole et les fonctions de croyance et de désirabilité sont définies partout sur cette algèbre. J'ai montré dans [Lep00] que cette théorie rencontre des problèmes d'adéquation à la modélisation de la notion de délibération. En gros, les deux problèmes principaux sont 1) tout acte jugé faisable par l'agent possède, au moment où il calcule l'utilité de l'accomplir, une probabilité très près de 1 et 2) tout acte jugé faisable par l'agent possède, au moment où il calcule l'utilité de l'accomplir, une désirabilité très près de celle qu'aurait une tautologie quelconque si l'agent apprenait qu'il a accompli cet acte<sup>2</sup>.

Si mon entreprise se révélait fructueuse, la trichotomie non représentée dans la grammaire des langues naturelles se trouverait ainsi naturalisée au profit des notions de probabilité et de désirabilité des propositions tout en contournant les problèmes engendrés par l'approche de Jeffrey.

## 1. Quelques *a priori*

Un premier *a priori* de mon approche est donc qu'un acte est exprimé par une proposition. Je ne veux bien sûr pas dire que si je lève la main, je produis une proposition mais tout simplement que l'acte que je lève la main est rapporté par un énoncé auquel correspond une proposition. Souvent l'agent est une composante de la proposition : François lève la main, Pierre prend son auto. Dans l'analyse qui suit, l'agent n'est pas nécessairement une composante de la proposition. « François lève la main » sera analysé comme « François accomplit l'acte ou rend volontairement vraie la proposition « La main de François se lève » et « François allume la télévision » comme « François accomplit l'acte ou rend vraie la proposition “ La télévision s'allume ” ».

Intuitivement, l'ensemble des actes que l'agent juge faisables est un sous-ensemble de l'ensemble des propositions sur lesquelles l'agent a une

---

2. Jeffrey, 1977, p. 136, semble lui-même reconnaître qu'il ne s'agit là que d'un simple processus de renormalisation : au fur et à mesure que l'agent agit sur le monde, la tautologie est de plus en plus désirable car toutes les propositions vraies ont la même désirabilité et la même probabilité (c'est-à-dire 1) ce qui est le cas des actes déjà accomplis.

opinion, le sous-ensemble des propositions que l'agent croit pouvoir rendre vraies s'il le décidait<sup>3</sup>. C'est cette propriété simple et naturelle qu'il s'agit de caractériser en termes de fonctions de croyance et de désirabilité.

Il vaut la peine ici de faire une légère digression sur ce que je n'ai pas l'intention de faire. Je n'ai pas l'intention de proposer une analyse de ce que c'est que de rendre vraie, ou plus simplement que de tenter de rendre vraie, une proposition. Je considère, pour les besoins de cette intervention, que cette question est résolue : un agent rend vraie, ou tente de rendre vraie, la proposition qui maximise une certaine fonction d'utilité. Mais ce n'est pas parce qu'une certaine proposition est hautement désirée (comme celle d'être soudainement milliardaire) qu'elle constitue un candidat potentiel comme acte jugé faisable. L'unique question qui m'intéresse ici est celle des propriétés que doit posséder une proposition pour pouvoir être un tel candidat. Ma thèse est que ces propriétés peuvent être explicitement décrites en utilisant les fonctions de croyances et les fonctions de désirs subjectives des agents. Ceci me conduit à mon deuxième point.

Un second *a priori* est que la fonction de croyance est une fonction de probabilité (définie sur un sous-ensemble de l'ensemble des propositions, sous-ensemble formant une algèbre de Boole). Cet *a priori* est banal et n'échappe pas, bien sûr, aux critiques habituelles de l'assimilation d'une fonction de croyance à une fonction de probabilité. C'est une question aux ramifications complexes qui ne nous intéresse pas directement ici.

Un troisième *a priori* est que la fonction de désirabilité de l'agent définit un pré-ordre sur le même sous-ensemble que l'on peut représenter par une fonction numérique (même commentaire que pour la fonction de croyance).

Un quatrième *a priori*, déjà annoncé plus haut, est que ces deux fonctions déterminent une fonction d'utilité : l'acte que l'agent tente d'accomplir est un de ceux qui maximisent cette fonction d'utilité.

Un acte (potentiel) jugé faisable par un agent sera donc représenté par un quadruplet  $\langle a, P_A, D, A \rangle$  où  $a$  est un agent,  $P_A$  est une fonction de probabilité (celle qu'utilise l'agent  $a$  au moment où il délibère pour calculer l'utilité d'accomplir  $A$ ),  $D$  est la fonction de désirabilité au moment de la délibération et  $A$  est une proposition<sup>4</sup>. Tous les quadruplets  $\langle a, P, D, A \rangle$  ne sont pas des actes potentiels de l'agent  $A$  : il s'agit encore une fois de définir un ensemble de propriétés que satisfait un quadruplet  $\langle a, P, D, A \rangle$  si et seulement si  $A$  est un acte que l'agent  $a$  possédant la fonction de croyance  $P$  et la fonction de désirabilité  $D$  juge faisable.

Avant d'entrer dans le vif du sujet, une deuxième digression sur les deux autres termes de la trichotomie, soit la notion d'état du monde et celle

---

3. Je laisse de côté la question des actes auxquels correspond des propositions ineffables.

4. Selon les contextes, je me permettrai l'abus de langage qui consiste à parler de  $A$  comme d'un acte pour ne pas alourdir le texte.

de conséquence. Encore une fois, pour les besoins de la cause, nous pouvons caractériser un état du monde comme étant une proposition dont la valeur de vérité (qu'elle soit connue ou inconnue) est, du point de vue subjectif de l'agent, totalement et absolument hors de son influence. Cette « définition » mérite d'être examinée et le sera plus loin ; il suffit de reconnaître qu'elle constitue un noyau dur de la notion d'état du monde même si elle s'applique aux propositions des mathématiques par exemple, plus généralement aux propositions que l'agent croie valides, analytiques ou nécessaires et non exclusivement à celles qu'il croie causalement indépendantes de ces actes qu'il croit pouvoir accomplir. Quant à la notion de conséquence, je suivrai Jeffrey : une conséquence d'un acte est simplement la conjonction de l'acte avec les états du monde qui résultent de l'accomplissement de l'acte. Pour simplifier, nous ne considérerons comme conséquences que les états qui ont été modifiés par l'accomplissement de l'acte. Cela a pour conséquence que l'acte est une conséquence de lui-même. Cela pourrait à première vue sembler bizarre mais n'entraîne pas de paradoxe. C'est par ailleurs compatible avec l'approche maximisatrice de l'utilité. Imaginons un acte qui consiste à se mettre dans un état mental agréable (penser à un être cher, par exemple) et qui, du point de vue de l'agent, n'a aucune autre conséquence que de procurer un état de bien être, celui justement d'être dans cet état mental.

L'intérêt de cette simplification est que l'ensemble des conséquences d'un acte est simplement le complément ensembliste de l'ensemble des états du monde relativement à l'univers de toutes les propositions. Mon problème est donc de caractériser la différence entre un acte véritable et ses conséquences. Donc, si mon entreprise de caractérisation des actes se révèle être un succès, la trichotomie est dissoute.

## 2. Deux contraintes naturelles

Soit  $P$  la fonction de probabilité de l'agent  $a$  qui n'a pas encore commencé le processus de délibération sur l'utilité d'accomplir l'acte  $A$ .  $P$  est une fonction qui n'est pas nécessairement définie partout, l'agent n'ayant pas une opinion sur toutes les propositions. On admet, c'est là notre deuxième *a priori*, que l'ensemble des propositions pour lesquelles la fonction  $P$  est définie forme une algèbre de Boole.

Il n'y a aucune raison de penser que, d'une manière générale,  $P(A)$  ait une valeur particulière ou même soit définie comme l'illustrent les exemples suivants.

Pierre est exceptionnellement dans un bar avec des collègues après le bureau et il a déjà bu deux bières. Il ne s'est pas posé la question s'il va rendre une troisième bière ou non.  $P_{pA}(A)$  n'est tout simplement pas définie.

Jean est dans un bar avec des collègues après le bureau et il a déjà bu deux bières. Il va régulièrement prendre un verre après le bureau, mais ne

prend jamais plus de deux bières. Il ne s'est pas posé la question s'il va rendre une troisième bière ou non, mais pourtant  $P_{jeA}(A) \equiv 0^5$ .

Jacques est dans un bar avec des collègues après le bureau et il a déjà bu deux bières. Il va régulièrement prendre un verre après le bureau et prend toujours de quatre à cinq bières. Il ne s'est pas posé la question s'il va rendre une troisième bière ou non, mais pourtant  $P_{jaA}(A) \equiv 1$ .

Bref, avant que ne débute le processus de délibération,  $P(A)$  peut prendre à peu près n'importe quelle valeur, y compris ne pas être définie. Appelons  $P_A$  la fonction de probabilité qu'utilise l'agent pour le calcul de l'utilité attendue d'accomplir l'acte  $A$ .  $P_A$  n'est certainement pas identique à  $P$ . Si l'agent croit véritablement que  $A$  est un acte jugé faisable,  $P_A(A)$  doit être très élevée. Il fait en effet partie du concept même de «  $A$  est un acte que l'agent  $a$  juge faisable » que si l'agent décide d'accomplir  $A$ , il pense que  $A$  deviendra presque certainement vraie. Par ailleurs, l'agent qui délibère et calcule l'utilité d'accomplir  $A$  se projette dans un monde où il a rendu  $A$  vraie. Par exemple, si je crois que l'acte d'aller au cinéma est faisable, lorsque je calcule l'utilité d'aller au cinéma, la proposition que je vais au cinéma ainsi que ses conséquences ont une probabilité très grande alors que les propositions incompatibles comme « Je reste à la maison et je termine mon article » ont une probabilité très faible.

On pourrait être tenté d'imposer la contrainte que  $P_A(A) = 1$ . Mais alors, il n'y a plus de place pour la croyance d'une possibilité même infime d'un échec contrairement à ce qui est le cas pour (presque ?) tous les actes que nous jugeons faisables. Lorsque Pierre décide de prendre son auto, il sait très bien que le moteur peut refuser de partir, qu'il peut se briser une jambe en tombant dans l'escalier, etc. Et c'est là toute la difficulté : il ne semble pas que, dans la plupart des cas, on puisse faire une distinction très nette entre une conséquence extrêmement probable d'un acte et l'acte lui-même. Notre problème pourrait ainsi être reformulé de la manière suivante : peut-on trouver un ensemble de conditions nécessaires et suffisantes à poser sur  $P_A(A)$  et  $D(A)$  pour que la conséquence  $A$  puisse être raisonnablement considérée comme un acte jugé faisable ? On pourrait discuter de la nature de ce « raisonnablement ». Nous verrons que ce que je propose est une caractérisation assez radicale même si la question de trancher si tel ou tel cas particulier satisfait aux conditions nécessaires et suffisantes reste floue et suggestive : le fait de savoir qu'un ascenseur ne peut supporter plus de 980 kg ne nous est d'aucun secours pour déterminer si la masse totale des gens qui l'attendent dépasse ou non cette limite ! Nous pouvons poser une première contrainte.

C1 : Si  $\langle a, P, D, A \rangle$  est un acte jugé faisable, alors  $P_A(A) \equiv 1$ .

---

5. L'usage que je fais de «  $\equiv$  » est assez souple.  $x \equiv y$  ssi  $x$  et  $y$  ne diffèrent pas de manière significative pour les calculs d'utilité.

Ainsi, si certains « actes » ne possèdent pas cette propriété, on considérera que cette appellation d'acte est un abus de langage. Si Pierre sait que son auto ne démarre qu'une fois sur deux, l'acte sur lequel il délibère n'est pas de prendre son auto ou non mais de tenter de la faire démarrer. Si cette contrainte apparaît assez naturellement comme nécessaire, elle est loin d'être suffisante : toutes les propositions que l'agent croit vraies sous l'hypothèse qu'il décide de les rendre vraies la satisfont ! En particulier, toutes les propositions que l'agent croit vraies et hors de son influence — donc qui expriment des états du monde — seraient des actes jugés faisables. Il nous faut des contraintes supplémentaires. On pourrait penser que  $A$  est un acte jugé faisable non seulement si la probabilité de  $A$  devient très élevée sous l'hypothèse que l'agent décide de la rendre vraie mais que la probabilité de  $A$  reste faible sous l'hypothèse que l'agent décide de ne pas la rendre vraie. Ceci nous conduit à la contrainte suivante :

C'2 : Si  $\langle a, P, D, A \rangle$  est un acte jugé faisable, alors  $P_{\neg A}(\neg A) \cong 1$  (ou, de façon équivalente,  $P_{\neg A}(A) \cong 0$ ).

Cette contrainte ne résiste pas à l'analyse. Considérons l'exemple suivant dans lequel l'agent croit en son pouvoir de rendre une proposition  $A$  vraie, mais croit également que s'il décidait de ne pas le faire,  $A$  deviendrait quand même vraie. Je suis dans une pièce éclairée uniquement par la lumière naturelle. Le soleil se couche. Je crois que de faire de la lumière est un acte jugé faisable (en actionnant le commutateur). Supposons par ailleurs que je crois que si je décidais de ne pas faire de la lumière, alors quelqu'un d'autre le fera. Avons-nous alors  $P_{\neg A}(\neg A) \cong 1$  ? Je ne le pense pas. Au contraire  $P_{\neg A}(\neg A)$  est probablement plus près de 0 que de 1. Il nous faut une contrainte plus faible.

Une version plus faible de cette contrainte est que la probabilité de  $\neg A$  s'il décide de faire  $\neg A$  est plus grande que celle de  $\neg A$  s'il décide de faire  $A$  et vice versa.

$$C2. P_{\neg A}(\neg A) > P_A(\neg A) \text{ et } P_A(A) > P_{\neg A}(A)$$

Cette contrainte est-elle elle-même trop forte ? Je crois que non. Imaginons qu'un certain dispositif soit branché sur mon système nerveux, dispositif tel que :

- 1) Si je décide au temps  $t_0$  de faire de la lumière, le dispositif ne fait rien.
- 2) Si je décide au temps  $t_0$  de ne pas faire de la lumière, le dispositif fait de la lumière.

Supposons, de plus, que j'accorde à la capacité du dispositif de « voir » ma décision une crédibilité presque absolue. Dans ce cas, on a  $P_{\neg A}(\neg A) = P_A(\neg A) = 0$  et  $P_A(A) = P_{\neg A}(A) = 1$ . Cela contredit C2. Je n'ai pas d'autre argument que de dire que dans de telles situations ni  $A$  ni  $\neg A$  ne sont des actes jugés faisables. Lorsqu'un agent est certain qu'une certaine proposition sera vraie qu'il décide ou non de la rendre vraie, cette proposition ne peut plus

être considérée par un agent rationnel comme un acte jugé faisable, elle exprime tout simplement un *état du monde*. Nous touchons ici un point essentiel. On pourrait objecter que le fait que la lumière se fasse par l'intermédiaire de l'agent plutôt que par l'intermédiaire du mécanisme correspond à deux scénarios qui conduisent au même état, la lumière se fait, mais dans le premier cas seulement l'agent accomplit l'acte de faire la lumière. Je répondrais que l'agent qui choisit de se lever pour faire de la lumière *n'a pas délibéré* en calculant l'utilité de faire ou non de la lumière ou non. Il peut se lever pour se dégourdir les jambes, pour prouver qu'il n'est pas paresseux mais pas pour faire de la lumière qui se fera de toute façon.

Avons-nous avec C1 et C2 une condition nécessaire et suffisante pour qu'une conséquence puisse être considérée comme un acte jugé faisable ? Il semble bien que non comme le suggère le scénario suivant.

### 3. La prise d'otage

Supposons qu'un agent  $a_1$  soit engagé dans un jeu (au sens de la théorie des jeux) avec un agent  $a_2$  et que  $a_1$  connaisse parfaitement la fonction d'utilité de  $a_2$ . L'agent  $a_1$  sait que s'il accomplit  $A_1$ , l'agent  $a_2$  accomplira  $A_2$  et que s'il accomplit  $B_1$ , l'agent  $a_2$  accomplira  $B_2$ . Selon C1 et C2, on peut aussi bien dire que si  $a_1$  accomplit  $A_1$ , il accomplit  $A_2$  et que si  $a_1$  accomplit  $B_1$ , il accomplit  $B_2$ . Est-ce intuitivement acceptable ?

Supposons que  $a_1$  soit le responsable de la sécurité d'un état où un terroriste  $a_2$  a pris en otage les passagers d'un avion. Si  $a_1$  ne libère pas les compagnons de  $a_2$  emprisonnés pour avoir commis des actes terroristes d'une barbarie inouïe, tous les passagers de l'appareil seront exécutés. Mais la politique de cet État est de ne jamais céder au chantage, les conséquences à long terme du « laxisme » étant d'encourager ce type d'action et de mettre en danger un nombre important de vies d'honnêtes citoyens. Soit  $A$  la proposition « Les terroristes sont libérés » et  $B$  « Les passagers ont la vie sauve » et  $C$  « Un nombre important d'honnêtes citoyens seront tués ». Peut-on dire que  $a_1$  délibère pour décider s'il va accomplir  $B$  ou  $\neg B$  au même titre qu'il délibère pour décider qu'il va accomplir  $A$  ou  $\neg A$  ?

On aurait tort de penser que c'est le caractère direct (pour  $A$  ou  $\neg A$ ) qui s'oppose au caractère indirect (pour  $B$  ou  $\neg B$ ) qui fait ici la différence. Certains actes indirects sont reconnus comme des actes authentiques, par exemple, l'officier qui ordonne à un subalterne de faire sauter un pont : autant l'officier que l'exécutant rendent vraie par leur action la proposition que le pont saute<sup>6</sup>.

---

6. Ce fait est reconnu entre autres par les cours de justice ! Il s'agit ici de deux actes différents, les agents étant différents, mais ils ont le même contenu propositionnel. Notre représentation des actes comme des quadruplets a pour conséquence que plusieurs agents peuvent décider de rendre vraie la même proposition.

Le problème est ailleurs. En fait, ce qui distingue  $A$  ou  $\neg A$  de  $\neg B$  ou  $B$  c'est que  $A$  ou  $\neg A$  ont des conséquences extrêmement importantes que n'ont pas  $\neg B$  ou  $B$ .

Pour  $a_1$  la situation est la suivante :  $a_1$  calcule l'utilité de  $A$  et celle de  $\neg A$ .

$$U(A) = P_A(B)D(B) + P_A(\neg B)D(\neg B) + P_A(C)D(C) + P_A(\neg C)D(\neg C)$$

et

$$U(\neg A) = P_{\neg A}(B)D(B) + P_{\neg A}(\neg B)D(\neg B) + P_{\neg A}(C)D(C) + P_{\neg A}(\neg C)D(\neg C).$$

comme  $P_A(B) \cong P_A(C) \cong P_{\neg A}(\neg B) \cong P_{\neg A}(\neg C) \cong 1$  et  $P_A(\neg B) \cong P_A(\neg C) \cong P_{\neg A}(B) \cong P_{\neg A}(C) \cong 0$

$$U(A) = D(B) + D(C)$$

et

$$U(\neg A) = D(\neg B) + D(\neg C)$$

Si l'agent  $a_1$  accorde à  $\neg C$  une utilité si grande que l'utilité de  $\neg A$  dépasse celle de  $A$ , alors  $a_1$  choisira de faire  $\neg A$  même si cette décision rend  $\neg B$  quasi certain. Mais même si  $\neg B$  est quasi certain, nous ne pourrions rapporter la situation en disant que  $a_1$  a accompli l'acte  $\neg B$ . La situation est différente pour le terroriste qui, lui, se trouve dans une des deux situations suivantes : ou bien  $a_1$  a accompli l'acte  $A$  et c'est un état du monde et  $a_2$  tient compte dans le calcul de  $U(B)$  et de  $U(\neg B)$  et ou bien  $a_1$  a accompli l'acte  $\neg A$  et  $a_2$  tient compte dans le calcul de  $U(B)$  et de  $U(\neg B)$ .

$A$  est une proposition que l'agent  $a_1$  croit pouvoir rendre vraie ou fausse,  $C$  et  $B$  sont des conséquences pratiquement certaines de  $A$  et  $\neg C$  et  $\neg B$  sont des conséquences pratiquement certaines de  $\neg A$ . Qu'est-ce qui fait que  $A$  ou  $\neg A$  peuvent être considérés comme des actes jugés faisables et  $\neg B$  seulement comme une conséquence de  $\neg A$  ? Une réponse qui se dégage fortement de l'exemple développé plus haut est que les conséquences de  $\neg A$  qui ne sont pas des conséquences de  $\neg B$  sont loin d'être négligeables, ici  $\neg C$ . On peut tenter de formaliser cette propriété de la manière suivante.

*Définition* :  $B$  est une conséquence immédiate de  $A$  ssi pour toute conséquence  $C$  de  $B$ ,  $P_A(C) \cong P_B(C)$ .

*Définition* : Si  $A$  est un acte jugé faisable et  $B$  est une conséquence immédiate de  $A$ , on dira que  $B$  est un acte jugé faisable au même titre que  $A$  si pour toute conséquence  $C$  de  $A$  qui n'est pas une conséquence de  $B$ ,  $U_A(C) = P_A(C)D(C) \cong 0$ .

En d'autres termes, une conséquence immédiate  $B$  d'un acte  $A$  peut être considérée comme un acte si l'utilité des conséquences de  $A$  qui ne sont pas des conséquences de  $B$  est considérée par l'agent comme négligeable.

En particulier, on aura

*Proposition* : Si  $B$  est un acte jugé faisable au même titre que  $A$ , alors l'utilité attendue d'un échec à rendre  $B$  vrai en accomplissant  $A$  est négligeable, c'est-à-dire que  $U_A(\neg B) \cong 0$ .

*Preuve* : Si  $\neg B$  n'est pas une conséquence potentielle de  $A$  et  $B$  est une conséquence potentielle de  $A$ , alors  $P_A(\neg B) = 0$ . Sinon  $\neg B$  est une conséquence potentielle de  $A$  qui n'est pas une conséquence potentielle de  $B$  et est négligeable.

Nous n'avons cependant pas entièrement résolu notre problème. Si nos deux définitions sont utiles pour caractériser le fait qu'un acte est jugé faisable au même titre qu'un autre, elles sont impuissantes à caractériser ce qu'est un acte jugé faisable tout court.

On peut, en effet, imaginer une suite d'actes jugés faisables

$$\dots A_i \dots A_j \dots$$

Rien n'interdit que nous ayons une régression infinie sur la gauche. On pourrait, en effet, imaginer qu'on puisse trouver, pour n'importe quel candidat  $B$ , un autre candidat  $A$  dont  $B$  serait la conséquence immédiate, et ce, *ad infinitum*. Dans ce cas, les définitions sont au mieux circulaires et ne sont susceptibles d'expliquer quoi que ce soit. La possibilité d'une telle régression est cependant complètement contre-intuitive. Prenons un acte simple. Je suis dans mon auto et je tourne à gauche. Cet acte est la conséquence immédiate de celui de mouvoir le volant dans le sens contraire des aiguilles d'une montre. Cet acte est à son tour la conséquence immédiate du fait que j'empoigne le volant et lui impose une force caractéristique. On pourrait continuer, mais les propositions sont de plus en plus difficiles à exprimer : elles deviennent de plus en plus de l'ordre du privé, de l'ineffable. Nous sommes incapables, par exemple, de formuler les actes qui consistent à « donner des ordres » à nos organes moteurs. Je sais comment me mettre dans l'état où mon bras va se lever, mais je suis incapable de formuler la proposition qui exprime cet acte. Il est cependant naturel de postuler l'existence d'actes jugés faisables primitifs qui sont tels que décider, c'est faire<sup>7</sup>.

Donald Davidson [Dav80] a soutenu que les actes primitifs pourraient bien tous être des mouvements volontaires du corps. On pourrait certes expliquer causalement ces mouvements de base (décharges nerveuses causant des contractions musculaires, contractions musculaires causant la levée de mon bras, etc.) mais ces causes ne peuvent être considérées comme les causes des actes en tant qu'actes : « Doing something that causes my finger to move does not cause me to move my finger ; it *is* moving my finger ». Une analyse de ce genre est tout à fait compatible avec mon approche et fournit

---

7. Les situations réelles sont beaucoup plus complexes. Des actes du genre de ceux qui ont été donnés en exemple, comme tourner à gauche, sont la conséquence d'une multitude d'actes primitifs. Il suffit pour régler cette question de considérer que la classe des actes primitifs indépendants les uns des autres est fermée sous les conjonctions finies.

même un critère relativement clair et précis permettant d'arrêter la régression vers la gauche. Le problème est que si ce que dit Davidson est vrai, cela n'en demeure pas moins, en un certain sens, une vérité empirique. Nous pouvons très bien imaginer un monde où certains actes similaires à des actes primitifs comme bouger le doigt portent sur des choses beaucoup moins intimes. Je ne sais pas ce qui se passe exactement dans la tête d'un agent qui croit être doué du don de télékinésie au moment où il décide de faire voler un objet mais je suppose que cela ressemble beaucoup à ce qui se passe dans la mienne quand je bouge le doigt. Par ailleurs, si se mettre dans un certain état mental *est* un mouvement du corps au sens large (comme semble le suggérer Davidson) alors nous avons là des candidats sérieux pour occuper la fonction d'actes primitifs. Une chose est certaine : tout accomplissement d'un acte passe par le fait que l'agent se met dans un certain état mental. On ne peut pas imaginer un monde où accomplir un acte ne passe pas par une phase où l'agent se met dans un certain état mental qui se confond précisément avec l'accomplissement de l'acte primitif. Remarquons que ces actes sont également susceptibles d'échouer. Non seulement je peux échouer à lever le bras, mais également échouer à me mettre dans l'état mental qui fait que je lève le bras comme le funambule qui n'arrive plus à se mettre en état d'équilibre à cause d'un problème de l'oreille interne. Revenons à nos actes primitifs.

*Définition* : Un acte *A* primitif pour un agent *a* est une proposition satisfaisant *C1* et *C2* et qui n'est la conséquence immédiate d'aucune autre proposition satisfaisant *C1* et *C2*.

Nous pouvons maintenant tenter de caractériser la classe des actes jugés faisables.

*Définition* : Un acte *A* est jugé faisable ssi

- a) *A* est un acte primitif ou
- b) *A* est jugé faisable au même titre qu'un acte *B* jugé faisable.

Remarquons que la clause a) a pour unique fonction de fonder la définition d'un acte jugé faisable en général. C'est la clause b) qui permet de distinguer un acte de ses conséquences.

Tel qu'annoncé au début de cette intervention, cette définition élimine la notion d'acte jugé faisable au profit de contraintes sur les fonctions de probabilité et de désirabilité de l'agent.

Il faut maintenant se poser la question du statut de cette réduction. Est-ce une réduction purement méthodologique en ce sens que toute proposition satisfaisant la propriété donnée ci-dessus (appelons-la *F*) est jugée faisable et réciproquement parce que nous sommes incapables de trouver des contre-exemples ? Cette réduction ne posséderait ainsi aucun pouvoir explicatif au niveau sémantico-pragmatique et encore moins au niveau psychologique.

Je pense en fait que c'est tout à fait le contraire. La propriété *F* est la propriété qu'un agent rationnel doit attribuer à une proposition pour qu'il accepte de dépenser de l'énergie délibératoire sur ses conséquences. Une digression s'impose ici sur la notion de rationalité.

Deux notions totalement distinctes entrent en jeu dans la description des processus de délibération. La première est de nature logique. Elle renvoie à la notion de cohérence et à la capacité de l'agent à tirer les conséquences logiques de ses croyances. Si l'expression « logique » est entendue au sens de la logique classique, alors les agents réels sont loin d'être logiques. Mais cette question n'est pas de toute première importance ici. Il suffit de faire l'hypothèse que, même si un agent entretient un ensemble incohérent de croyances au sens de la logique classique, il ne réalise pas cette incohérence (dans les deux sens du terme, soit qu'il ne la voit pas parce qu'il ne la matérialise pas). De ce point de vue, tout agent est logique, il est logiquement omniscient, mais sa logique est plus ou moins puissante. Le mot logique renvoie donc ici à l'ensemble des règles formelles que l'agent utilise effectivement et lorsque l'utilisation de ces règles le conduit à une contradiction, il réaménage ses croyances par un mécanisme qu'il n'est pas pertinent de préciser ici. Bref, il suffit de prêter aux agents l'incapacité d'entretenir explicitement des croyances contradictoires et une certaine capacité à faire des déductions.

L'autre notion de rationalité qui entre en jeu dans les processus de délibération renvoie à des questions d'ordre épistémiques. Étant donné ce que l'agent croit ou croit savoir factuellement du monde, dans quelle mesure est-il rationnel pour lui de croire en des lois générales reliant entre eux les faits, de croire en l'existence de liens causaux. Ce sont, en effet, des croyances de ce type qui jouent un rôle essentiel dans les processus de délibération. Les conséquences logiques du fait que je prends mon auto plutôt que les transport en commun m'importent peu ; ce sont les conséquences causales appréhendées (confort, retard, embouteillage, etc.) qui servent dans le calcul de l'utilité.

Tout comme pour la rationalité logique tout le monde est parfaitement rationnel au point de vue épistémique. *Anything goes !* Les astrologues, les naturopathes, les acuponcteurs et les homéopathes sont aussi rationnels que les actuaires et les médecins. Être épistémiquement rationnel, c'est être capable de tirer les conséquences causales de ses croyances factuelles, quelles que soient par ailleurs les lois générales exprimant ces croyances en des liens causaux (sous réserve, bien sûr, du respect de la cohérence logique au sens développé plus haut).

Revenons au processus de délibération. Délibérer pour savoir s'il est pertinent de rendre vraie une certaine proposition a un coût, exige un effort. Quelle que soit la rationalité logique ou épistémique de l'agent, l'effort de délibération lui-même possède une utilité négative. Peser le pour ou le contre des choses n'est une activité digne d'intérêt que si le profit attendu compense de façon presque certaine celui qui est dépensé. Ma thèse est que cette condition n'est satisfaite que si la proposition sur laquelle l'agent délibère satisfait la condition *F* relativement aux fonctions de croyance et de désirabilité de l'agent. Aussi fol soit-il, un agent n'accepte sérieusement de délibérer sur la possibilité de rendre vraie une proposition que si cela a du sens. Et cela n'a

du sens que si la condition  $F$  est satisfaite. Une délibération sur une proposition qui ne satisfait pas  $F$  est une pseudo-délibération. Autrement dit, si la réduction de la notion d' « acte jugé faisable » à celle de « acte possédant la propriété  $F$  » est adéquate, elle est analytique : ce que nous voulons dire lorsque nous disons que «  $a$  juge  $A$  faisable » est que «  $a$  juge que  $F(A)$  ».

Pour terminer je voudrais suggérer quelques pistes pour tenter d'invalider ma proposition.

Tout d'abord la notion même d'acte primitif m'apparaît quelque peu *ad hoc*. Je ne vois cependant pas comment on pourrait l'éliminer. Si une régression vers la gauche dans la suite des conséquences se produisait, la décision ne pourrait jamais être prise. Mais cet argument est faible. On peut imaginer que l'agent délibérant s'arrête de régresser par lâcheté ou par paresse et juge faisable de rendre vrai une proposition qui ne satisfait pas C1 et C2. Je pense ici à quelque chose comme l'homicide involontaire qui existe en droit pénal canadien. Un acte pour lequel on reproche à un individu de ne pas avoir suffisamment délibéré, d'avoir eu une attitude frivole. Est-ce vraiment un acte ? L'acte dans ce cas n'est-il pas plutôt de ne pas avoir suffisamment délibéré ? Il faudrait trouver un exemple crédible où C1 ou C2 sont carrément violés et où pourtant l'intuition commune voit un acte véritable.

Un autre problème avec la notion d'acte primitif concerne la possibilité de l'échec. Considérons un agent qui lève le doigt et dont le doigt ne se lève pas. Si cet acte était vraiment primitif, la position de repli qui consiste, du point de vue de l'agent, à justifier l'échec par un argument du style « J'ai tenté de lever le doigt, mais mon doigt ne s'est pas levé » ne serait pas une option acceptable et même concevable, étant en contradiction flagrante avec le caractère primitif de l'acte. La seule conclusion que, me semble-t-il, on puisse tirer est que la notion d'acte primitif est définitivement relative.

## Bibliographie

- Davidson, Donald, *Essays on Actions and Events*, Oxford, Oxford University Press, 1980, 49-50.
- Jeffrey, Richard, « A Note on the Kinematics of Preference », *Erkenntnis*, Vol. 11, 1977, 135-141.
- Jeffrey, Richard, *The Logic of Decision*, Chicago, University of Chicago Press, 1983.
- Lepage, François, « A Problem with Jeffrey's Approach to Bayesianism », in *New Studies in Exact Philosophy : Logic, Mathematics, and Science*, Hermes Science Publication, New York, 2000, 263-272.