



## La Belle au bois dormant, la fin du monde et les extraterrestres

L'analyse de plusieurs énigmes troublantes suggère la notion d'anamorphose probabiliste et fournit la solution du paradoxe de l'Apocalypse.

**D**ans le Laboratoire de psychologie probabiliste de l'Université de Hasardville, les discussions font rage. Elles portent sur le résultat d'une expérience dont nous allons décrire le protocole. Le dimanche soir, alors que *La Belle au bois dormant* (nous la dénommerons *Belle* dans ce qui suit) est endormie, nous lançons une pièce de monnaie pour un tirage à pile ou face. Une grande confiance règne entre *Belle* et les expérimentateurs ; elle est renforcée par un huissier qui contrôle que tout se déroule selon les règles acceptées par tous.

Si la pièce est tombée sur FACE, le lendemain (le lundi) on réveille *Belle* et on a un entretien avec elle. Si c'est PILE, on la réveille le lundi, on a un entretien avec elle, puis on la rendort après l'avoir soumise à un traitement qui lui fait complètement oublier la journée du lundi. Enfin, toujours dans le cas de PILE, on la réveille à nouveau le mardi (elle ignore qu'on est mardi) et on a un autre entretien avec elle.

Ces expériences doivent mesurer le bon sens probabiliste de *Belle*. À la fin d'un entretien, on lui demande quelles probabilités,  $\Pr(\text{FACE})$  et  $\Pr(\text{PILE})$ , elle attribue au tirage réalisé le dimanche soir, alors qu'elle était endormie ?

*Belle* réfléchit : « La pièce est normale et l'huissier qui contrôle le déroulement des opérations me garantit que le résultat obtenu par ce tirage à pile ou face détermine la suite des opérations conformément au protocole : un réveil le lundi si FACE, un réveil le lundi et un autre le mardi si PILE. À chaque réveil je n'aurai aucune information en plus de celles dont je disposais le dimanche soir avant que la pièce ne soit lancée. Avant de m'endormir dimanche, j'attribuais la probabilité  $1/2$  à FACE et à PILE ; donc, quand je suis réveillée lors de l'expérience, je dois attribuer la même probabilité  $1/2$  à chaque éventualité. Ma réponse est donc :  $1/2$  pour FACE, et  $1/2$  pour PILE. » La démonstration semble irréprochable. Et pourtant...

Après réflexion, *Belle* se ravise et propose un autre raisonnement : « Imaginons qu'on fasse l'expérience 100 fois en opérant 100 semaines de suite. Dans environ la moitié des semaines (50), je serai réveillée le lundi après un tirage FACE. Les autres semaines (50 environ) PILE aura été tiré et je

serai réveillée le lundi et le mardi ; dans ces cas-là, je serai donc réveillée 100 fois. Au total, lors des 100 semaines je serai réveillée environ 150 fois et, sur ces 150 réveils, FACE sera la bonne réponse 50 fois, et PILE la bonne réponse 100 fois. À chaque fois que je suis réveillée, la probabilité que la pièce lancée le dimanche soit tombée sur FACE est donc de  $1/3$  et, pour PILE, c'est  $2/3$ . Ma réponse sera donc :  $1/3$  pour FACE, et  $2/3$  pour PILE. » Faites-vous partie, cher lecteur, des partisans du  $1/2 - 1/2$  ou des partisans du  $1/3 - 2/3$  ?

Les organisateurs désirent que *Belle* prenne les choses au sérieux, aussi introduisent-ils un enjeu financier. Si *Belle* indique correctement le résultat du lancer du dimanche soir, elle gagne 1 000 euros ; si elle se trompe, elle n'a rien. *Belle* raisonne de la façon suivante.

Imaginons comme précédemment que le jeu soit organisé 100 semaines consécutives. Lors des 100 expériences en choisissant toujours FACE je gagnerai environ 50 000 euros, alors qu'en choisissant de répondre toujours PILE je gagnerai 100 000 euros environ. Cela confirme que FACE « est » moins probable que PILE ; puisque PILE fait gagner le double de ce que ferait gagner FACE, sa probabilité « est » double de celle de FACE. On a bien  $\Pr(\text{FACE}) = 1/3$ ,  $\Pr(\text{PILE}) = 2/3$ . C'est sûr, c'est le bon résultat, l'argent ne trompe pas ! Non seulement j'indiquerai les évaluations  $1/3$  pour FACE et  $2/3$  pour PILE, mais je parierai toujours pour PILE. Je suis si certaine de mon évaluation que je suis prête à payer 650 euros pour avoir le droit de jouer, car mon gain moyen sera de  $1\,000 \times 2/3 = 666,6$  euros. »

### Le pari du directeur

La solution  $1/3 - 2/3$  semble ainsi confirmée par l'introduction de l'enjeu des 1 000 euros et tout paraît réglé. Cela est-il certain ? Le directeur du laboratoire trouve que ce protocole de pari lui coûte trop cher et décide qu'il ne veut pas risquer plus de mille euros par expérience. Aussi le système de pari est-il redéfini et nous dénommerons le nouveau protocole, le pari du directeur : lorsque PILE tombe, le laboratoire ne propose le pari que le lundi et, toujours dans le cas de PILE, lorsque



**1. Le paradoxe de *La belle au bois dormant*.** Le dimanche soir *La belle au bois dormant* est endormie et on lance une pièce. Si elle retombe sur FACE alors on réveille *La belle au bois dormant* le lundi elle évalue les probabilités de PILE et de FACE et se rendort. Si c'est PILE on la réveille le lundi, elle doit évaluer les probabilités de PILE ou de FACE, et se rendort en oubliant cet épisode, mais on la réveille aussi le mardi et elle réévalue les probabilités avant de se rendormir. Lors des entretiens, on demande à *La belle au bois dormant* quelle est, à son avis, la

probabilité que FACE soit tombée, et quelle est la probabilité que PILE soit tombé ? Elle peut répondre  $1/2$  pour FACE et  $1/2$  pour PILE en argumentant que la pièce a une chance sur deux de tomber sur une des deux faces ou  $1/3$  pour FACE et  $2/3$  pour PILE, car on lui pose la question deux fois plus souvent (le lundi et le mardi) quand c'est PILE. Êtes-vous partisan de  $1/2 - 1/2$  ou partisan de  $1/3 - 2/3$  ? L'illustration est une peinture de la représentation, au théâtre du Vaudeville à Paris vers 1820, de *La belle au bois dormant* d'après le conte de Perrault (1628-1703).

*Belle* est réveillée le mardi, on a avec elle l'entretien prévu, mais on ne lui propose aucun pari. Lorsque FACE tombe on opère comme précédemment avec un réveil lundi et un pari.

*Belle* reprend son raisonnement et ses calculs. « Durant 100 semaines de jeu, si je choisis toujours FACE je gagnerai 50 000 euros environ, les 50 semaines où FACE sera tombée. Si je choisis toujours PILE, je gagnerai 50 000 euros environ : les semaines où PILE sera tombée, car le lundi on m'aura offert le pari. Donc les deux choix se valent : j'ai le même intérêt à parier sur PILE que sur FACE. »

Ce pari confirme donc l'évaluation  $1/2 - 1/2$ , ce qui est gênant puisque le premier pari avait donné l'assurance que  $1/3 - 2/3$  était la bonne évaluation. Le plus paradoxal semble ici que *Belle*, lorsqu'elle est réveillée, évalue les probabilités qu'elle accorde à FACE ou à PILE sans savoir si

on va lui offrir la possibilité de parier. Donc son évaluation de la probabilité de FACE ou PILE avec les règles du protocole de pari initial ou celles du protocole de pari du directeur devrait donc être la même : tout y est exactement identique au moment où on l'interroge sur son évaluation des probabilités. Comment se peut-il que « monétairement » les deux choses soient différentes et donc que, finalement, les probabilités qu'elle doit attribuer à FACE ou à PILE soient différentes : si c'est le protocole de pari initial,  $1/3 - 2/3$ , si c'est le protocole de pari du directeur,  $1/2 - 1/2$  ?

La confusion semble à son comble. Deux raisonnements en apparence rationnels conduisent à deux conclusions différentes. L'introduction d'enjeux ne résout pas le problème, car la façon dont on fait les paris renforce, soit la proportion  $1/3 - 2/3$ , soit la proportion  $1/2 - 1/2$ . Qu'en pensez-vous ?

## Le protocole sans mardi

Je vais présenter un raisonnement qui remet tout en ordre, mais je dois préciser que cette solution ne fait pas l'unanimité des spécialistes. Considérons une autre expérience, l'expérience sans mardi, qui éclaircira le dilemme.

– Le dimanche soir, après avoir endormi *Belle* comme précédemment, on lance la pièce de monnaie non truquée.

– *Belle* est toujours réveillée le lundi quand la pièce tombe sur FACE. Si elle tombe sur PILE, on relance la pièce et si elle tombe encore sur PILE, *Belle* est alors réveillée. Sinon elle n'est pas réveillée. Lors de l'entretien, les organisateurs lui demandent alors de parier sur FACE ou PILE en lui disant que si elle devine correctement le résultat du tirage du dimanche, elle gagnera 1 000 euros. Le mardi, rien ne se passe.

– *Belle* est ensuite endormie et elle oublie ce qu'elle a fait le lundi, sauf que, lorsqu'elle a gagné, elle garde dans sa poche les mille euros qu'elle trouve à son réveil.

– Dans le cas des lancers PILE-FACE où les expérimentateurs ne la réveillent pas, *Belle* dort le lundi toute la journée et, à son réveil un des jours suivants, elle trouve bien sûr sa poche vide sans savoir si c'est parce qu'elle a été réveillée et a perdu son pari, ou parce qu'elle a dormi.

Les lundis où *Belle* est réveillée, elle peut raisonner ainsi, car elle connaît le protocole : « La pièce n'est pas truquée, donc la probabilité de PILE est la même que la probabilité de FACE. Mon pari est indifférent, je vote donc toujours pour PILE. » Pourtant, sur 100 expériences, elle gagnera environ 25 000 euros, alors qu'elle en aurait gagné 50 000 si elle avait parié sur FACE : pour *Belle*, il n'est pas vrai que la pièce tombe une fois sur deux sur PILE. On comprend ici l'erreur : le fait qu'on ne la réveille pas systématiquement le lundi lorsque PILE est tombé fausse sa vision du tirage du dimanche soir : *Belle* voit plus souvent FACE, car la moitié des tirages PILE lui sont cachés ! Sa vision est soumise à un « filtre probabiliste » qui change le  $1/2 - 1/2$  réel de la pièce en  $2/3 - 1/3$  en faveur de FACE.

Le raisonnement consistant à dire que « La pièce n'est pas truquée, donc la probabilité de PILE est la même que la probabilité de FACE » est faux. *Belle* doit prendre en compte l'effet de filtre auquel elle est soumise quand elle est engagée dans le protocole expérimental. L'histoire du pêcheur décrit figure 2 est un autre exemple d'effet de filtre. Les effets de sélection observationnelle sont fréquents et toujours dangereux. Les éviter est l'objectif principal des techniques de sondages.

## Dangereux effets de filtre

Une célèbre erreur commise à cause de l'effet de filtre s'est produite à propos de l'élection de Franklin Roosevelt en 1936. Le *Literary Digest* avait interrogé certains électeurs en se fondant principalement sur les listes des propriétaires de téléphones et de véhicules automobiles, et en avait déduit que Alf Landon serait élu président des États-Unis contre Roosevelt. Or ce dernier fut élu. L'erreur, due au fait que les électeurs pauvres n'avaient ni téléphone, ni voiture, et qu'ils soutenaient plutôt Roosevelt, permit à George Gallup de prouver le bien-fondé des méthodes de son Institut de sondage : il avait, quant à lui, émis le pronostic en n'interrogeant qu'un petit nombre de gens, mais en évitant les effets de sélection (de filtre). Une erreur similaire relevée par Hervé le Bras dans les études démographiques concerne le nombre moyen de frères et de sœurs par famille. Vous savez que le nombre moyen d'enfants par famille est de 2.

En revanche, si vous demandez aux enfants combien ils ont de frères et sœurs y compris eux-mêmes, et que vous faites la moyenne des réponses obtenues, le résultat sera bien supérieur à 2... Logique, car vous n'avez pas interrogé les enfants des familles où il n'y en a pas et qui auraient fait baisser la moyenne ! De plus, les enfants de

## 2. Effet de filtre et effet de loupe

Lorsqu'un pêcheur sait qu'un poisson sur deux du lac où il pêche mesure moins de 5 centimètres, la moitié des poissons qu'il pêchera ne dépasseront pas 5 centimètres. Cependant, si les mailles de son filet sont trop larges et qu'elles laissent passer la moitié des poissons de moins de 5 centimètres, alors le pêcheur trouvera dans son filet 2/3 de poissons de plus de 5 centimètres et 1/3 de poissons de moins de 5 centimètres. Son filet filtre les poissons et cette sélection fausse sa vision du contenu du lac. Les effets de sélection observationnelle sont fréquents et un des principaux objectifs des techniques de sondages est de les éviter.

Dans le cas de l'expérience de *La belle au bois dormant*, il se produit un effet opposé à l'effet de filtre, l'effet de loupe. Lorsque, dans la première expérience, la pièce, lancée le dimanche soir, tombe sur PILE, cela double l'observation du résultat PILE.

Ainsi, de même que l'effet de filtre, l'effet de loupe modifie les probabilités observées. En astrophysique, le phénomène comparable de loupe gravitationnelle conduit à voir deux fois



certaines objets célestes qui pourtant sont uniques. Ces effets de filtre et de loupe (regroupés sous le nom d'effets d'anamorphose probabiliste) sont la clef de plusieurs paradoxes et ont donné lieu à de nombreux échanges ces dernières années.

### 3. La formule de Bayes

La formule de Bayes indique comment, en fonction d'une information nouvelle, il faut réévaluer des probabilités précédemment attribuées.

On dispose de la probabilité  $Pr(A)$  pour l'événement  $A$ . La formule de Bayes est une formule compliquée qui exprime un fait simple : quand un événement apparaît, les théories où cet événement est le plus probable sont renforcées. Comme nous allons le voir en chiffres, si les tricheurs gagnent toujours et que vous perdez contre un adversaire, il devient plus probable que c'est un tricheur, que vous ne le pensiez *a priori*.

En termes abstraits, un événement  $E$  se produit et on veut connaître la probabilité, notée  $Pr(A/E)$ , d'avoir  $A$  sachant que  $E$  s'est produit. Il existe une formule qui exprime cette probabilité  $Pr(A/E)$  que l'on cherche à connaître en fonction de  $Pr(A)$ ,  $Pr(E/A)$ ,  $Pr(nonA)$ ,  $Pr(E/nonA)$  qui sont connues. Cette formule (qui ne nous apprend évidemment rien quand les événements  $A$  et  $E$  sont indépendants), indique que :

$$Pr(A/E) = \frac{Pr(A) \cdot Pr(E/A)}{Pr(A) \cdot Pr(E/A) + Pr(nonA) \cdot Pr(E/nonA)}$$

#### Exemple 1. Le tricheur

Je joue avec un partenaire qui pourrait tricher. J'évalue *a priori*

que la probabilité que mon partenaire triche est  $1/10$  :  $Pr(A) = Pr(\text{tricheur}) = 1/10$ . Lorsque je joue avec un tricheur, il gagne dans  $2/3$  des cas :  $Pr(\text{je perds}|\text{tricheur}) = 2/3$ . Comme je suis expert, je sais que, face à un joueur honnête, je gagne 99 fois sur 100 :  $Pr(\text{je perds}|\text{non tricheur}) = 1/100$ . Je viens de jouer et j'ai perdu.

Quelle est la probabilité *a posteriori* pour que, en face de moi, il y ait un tricheur ? On applique la formule de Bayes avec  $A = \text{tricheur}$ ,  $E = \text{je perds}$  :  $Pr(\text{tricheur}|\text{je perds}) = Pr(A) \cdot Pr(E/A) / [Pr(A) \cdot Pr(E/A) + Pr(nonA) \cdot Pr(E/nonA)] = 1/10 \cdot 2/3 / [1/10 \cdot 2/3 + 9/10 \cdot 1/100] = 88,1\%$ .

La probabilité, après avoir perdu, que j'ai joué avec un tricheur est bien plus forte maintenant : elle est passée de  $10\%$  à  $88,1\%$ . Avant que je ne perde il était peu probable que le joueur en face de moi soit un tricheur, maintenant c'est presque certain. C'est le glissement bayésien.

#### Exemple 2.

##### Le pari du directeur

Cas du pari du directeur :

$A = \text{FACE}$  ;  $E = \text{offre de pari}$ .

$Pr(A) = 1/3$  (avant l'offre de pari en prenant en compte l'effet de loupe probabiliste).

$Pr(A/E) = 1/3 \cdot 1 / [1/3 \cdot 1 + 2/3 \cdot 1/2] = 1/2$



familles à deux enfants sont deux fois plus souvent interrogés que ceux de familles à un enfant, etc..

L'expérience sans mardi est claire : le protocole est tel qu'un effet de filtre opère sur ce que *Belle* « voit » du lancer de pièce du dimanche soir. Pour elle, malgré le fait incontestable que la pièce n'est pas truquée, on la réveille toujours pour *FACE*, alors que pour *PILE* on ne la réveille qu'une fois sur deux seulement. Donc un réveil correspond à *FACE* deux fois sur trois et *PILE* une fois sur trois : il n'y a pas de paradoxe, juste un effet de filtre. Dans le paradoxe initial, la distorsion du protocole est un phénomène analogue qui, au lieu de provoquer un effet de filtre sur les données (en masquant certains tirages), opère un effet de loupe : lorsque *PILE* tombe, *Belle* voit deux fois *PILE*, une fois le lundi et une autre fois le mardi. Le protocole fausse la vision que *Belle* a du tirage de la pièce. Du fait de cette déformation, elle voit un monde où la pièce tombe deux fois plus souvent sur *PILE* que sur *FACE*. Au total, il n'y a rien de mystérieux : un effet de loupe grossit la probabilité de *PILE*, comme un effet de filtre diminue celle de *PILE* dans l'expérience sans mardi. Ces phénomènes qu'on peut regrouper sous le nom d'anamorphoses probabilistes sont la clef du paradoxe de *La Belle au bois dormant*.

## Oui, c'est bien $1/3 - 2/3$

Tout semble être rentré dans l'ordre à l'exception du protocole de pari du directeur, qui reste à élucider. Cette dernière difficulté va être résolue en évoquant ce que l'on dénomme le glissement Bayésien, connu depuis longtemps.

La clef de la solution est que dans le protocole du directeur, il faut distinguer l'évaluation que doit faire *Belle*

quand on l'interroge, et le pari qu'elle doit faire quand on lui propose de parier, car le fait qu'on lui propose un pari est, avec ces règles, une information nouvelle. Lorsqu'elle donne son évaluation pour *FACE* et pour *PILE*, elle doit proposer  $1/3 - 2/3$ , car au moment même où elle fournit cette évaluation tout est exactement identique au protocole initial pour lequel elle prend en compte l'effet de loupe probabiliste identifié plus haut. En revanche, lorsqu'on lui propose de parier, les choses changent, car elle sait alors que l'on n'est pas mardi (dans le protocole du directeur, aucun pari n'est proposé le mardi). Elle doit prendre en compte cette information nouvelle, ce qui a pour effet d'annuler l'effet de loupe et la conduit à l'évaluation  $1/2 - 1/2$ .

Les calculs des gains moyens avec le protocole de pari du directeur confirment ce  $1/2 - 1/2$ . Un calcul numérique basé sur ce qu'on appelle la formule de Bayes (voir la figure ci-dessus) montre que l'évaluation de  $1/3 - 2/3$  passe à  $1/2 - 1/2$  entre le moment où on a l'entretien avec *Belle* sans lui offrir de pari et le moment où on lui offre le pari. La confusion est levée : le pari du directeur, en introduisant une information, faussait le  $1/3 - 2/3$  qui est la bonne évaluation que doit faire *Belle* à cause de l'effet de loupe probabiliste. Notons les leçons de cette histoire :

– Le fait qu'une pièce de monnaie ne soit pas truquée ne signifie pas que nous devons observer autant de *FACE* que de *PILE*. Lorsque nous sommes pris dans un protocole expérimental qui biaise par anamorphose probabiliste nos observations – soit par effet de filtre, soit par effet de loupe – une autre vision des tirages peut nous être offerte.

– Le seul fait de proposer un pari constitue parfois une information qui change les évaluations de probabilité.

## 4. Le paradoxe de l'Apocalypse

Dans l'expérience de pensée qui va suivre, deux théories, que nous désignerons par  $A$  et  $B$ , sont en compétition.

**Théorie A. La fin de l'humanité se produira avant 2050 et au total il y aura eu environ cent milliards ( $10^{11}$ ) d'êtres humains.**

**Théorie B. La fin de l'humanité se produira bien plus tard et au total il y aura eu cent mille milliards ( $10^{14}$ ) d'êtres humains.**

Supposons que l'on sache par des analyses provenant des diverses sciences fondamentales que :  $\Pr(A) = 1/2$  et  $\Pr(B) = 1/2$ . L'information « Je suis un humain dont le rang de naissance est inférieur à cent milliards » (événement  $E$ ) permet grâce à la formule de Bayes de réévaluer les probabilités de  $A$  et de  $B$ .

(I) Sans prise en compte de l'effet de loupe.

Les calculs, analogues à ceux du cas du tricheur, donnent :  
 $\Pr(A/E) = \Pr(E/A) \Pr(A) / [\Pr(E/A) \Pr(A) + \Pr(E/B) \Pr(B)] = 1 \times 1/2 / [1 \times 1/2 + 1/1\,000 \times 1/2] = 0,999 = 99,9\%$

$\Pr(B/E) = 0,1\%$ .

La probabilité d'une fin proche de l'humanité est donc 99,9%. C'est le troublant paradoxe de l'Apocalypse défendu par B. Carter et J. Leslie.

(II) Avec prise en compte préalable de l'effet de loupe.

En fait, les évaluations  $\Pr(A) = \Pr(B) = 1/2$  sont faites dans un contexte atemporel alors que nous, humains, sommes situés dans l'Univers. C'est la même différence qu'entre la situation de *Belle* avant l'expérience et sa situation quand on l'interroge durant l'expérience. Prendre en compte que nous sommes situés dans l'Univers, consiste à évaluer l'effet de loupe probabiliste. Puisqu'il y a mille fois plus d'humains lorsque la théorie B est satisfaite, la réévaluation donne :  $\Pr'(A/E) = 1/1001$  ;  $\Pr'(B/E) = 1\,000/1001$

L'utilisation de la formule de Bayes à partir de  $\Pr'(A)$  et  $\Pr'(B)$  fournit une seconde réévaluation qui donne :  $\Pr(A/E) = \Pr(B/E) = 1/2$   
Le paradoxe de l'Apocalypse a disparu.



### Le paradoxe de l'Apocalypse

La conséquence remarquable de l'étude du paradoxe de *La belle au bois dormant* (qui n'en est plus un si vous acceptez le point de vue des filtres et des loupes probabilistes) est qu'il résout le paradoxe de l'Apocalypse. Ce paradoxe introduit en 1983 par l'astrophysicien Brandon Carter, et sur lequel le philosophe canadien John Leslie a attiré l'attention au début des années 1990, a donné lieu à un grand nombre d'articles.

En 1993, je l'ai mentionné dans cette chronique et des dizaines de lettres ont été envoyées à *Pour la Science*, proposant des idées pour sa résolution. Parmi les solutions proposées, beaucoup sont normatives : elles consistent à décréter qu'on n'a pas le droit de raisonner avec les probabilités quand le temps est impliqué, ou parce qu'on se réfère à une notion trop vague d'être humain. Les solutions normatives sont insatisfaisantes, car elles limitent le champ d'application des concepts probabilistes et ne correspondent pas à ce qui est en général admis. Se rassurer à bon compte en s'interdisant de raisonner est trop facile.

Rappelons le paradoxe. Imaginons que deux théories,  $A$  et  $B$ , sur le destin ultime de l'humanité, sont en compétition :

**Théorie A. La fin de l'humanité se produira avant 2050 et au total il y aura eu environ cent milliards ( $10^{11}$ ) d'êtres humains.**

**Théorie B. La fin de l'humanité se produira bien plus tard et, au total, il y aura eu cent mille milliards ( $10^{14}$ ) d'êtres humains.**

Imaginons que l'on détermine *a priori*, par des analyses provenant de diverses sciences fondamentales (mathématiques, physique, chimie, biologie, etc.), que :  $\Pr(A) = 1/2$  et  $\Pr(B) = 1/2$ . Le paradoxe de Brandon Carter défendu avec acharnement par John Leslie consiste à proposer l'argumentation suivante. Les évaluations  $\Pr(A) = 1/2$  et  $\Pr(B) = 1/2$  sont faites dans le cadre général des sciences, mais je dispose d'une information particulière qui me permet de rééva-

luer ces probabilités : « Je suis un humain dont le rang de naissance est inférieur à cent milliards » (événement  $E$ ). Dans la théorie  $A$ , l'événement  $E$  est banal (inévitabile même), en revanche, dans la théorie  $B$ , cet événement est étonnant car cela me place dans le premier millième des êtres humains, chose statistiquement peu probable (la probabilité de  $E$  sachant  $B$  est  $1/1\,000$ ). Si une information est plus probable dans le cadre d'une théorie  $A$  que dans le cadre d'une théorie  $B$ , et que j'apprends alors que cette information est vraie, cela renforce ma confiance en la théorie  $A$  et diminue celle que j'ai concernant la théorie  $B$ . La formule de Bayes permet de rendre cela précis. Le calcul détaillé sur la figure ci-dessus conduit à :

– Probabilité de  $A$  sachant que  $E$  est vrai =  $0,9990 = 99,90\%$ .

– Probabilité de  $B$  sachant que  $E$  est vrai =  $0,001 = 0,10\%$ .

Il y a un « glissement bayésien » en faveur de l'hypothèse d'une fin proche de l'humanité. Un tel glissement a lieu même si plus de deux théories sont en compétition et se produit quelles que soient les hypothèses numériques formulées. La fin de l'humanité serait donc plus proche que ce que l'on peut croire en se fiant uniquement aux risques évalués scientifiquement. On comprend que ce raisonnement ait troublé des lecteurs !

Il se trouve qu'en regardant bien et aux valeurs près des paramètres numériques, la situation décrite dans le paradoxe de l'Apocalypse est exactement la même que celle du paradoxe de *La belle au bois dormant* lorsque, dans le protocole de pari du directeur, elle apprend qu'on est lundi parce qu'on lui propose un pari.

Lorsque *Belle* raisonne, elle oppose deux théories : celle qui affirme que *FACE* est tombée et dans laquelle son existence au sein de l'expérience sera unique (puisqu'elle ne sera pas réveillée le mardi), et celle qui affirme que *PILE* est tombée et dans laquelle son existence au sein de l'expérience sera double. Raisonner comme Leslie et Carter le proposent, consiste à dire que les probabilités de *FACE* et *PILE* sont  $1/2 - 1/2$  et lorsque *Belle* apprend que c'est

## La vie dans l'Univers

lundi (l'équivalent d'apprendre que je suis dans le premier millième des humains) je dois réévaluer les probabilités de FACE et PILE, ce qui renforce FACE (un calcul avec la formule de Bayes donne  $2/3$  pour FACE et  $1/3$  pour PILE). En résumé : on part de  $1/2 - 1/2$  pour FACE et PILE, et en utilisant l'information qu'on est lundi, on arrive à  $2/3 - 1/3$  en faveur de FACE. Mais raisonner ainsi c'est commettre un grave oubli : l'effet de loupe. Le raisonnement correct, lui, se déroule en trois phases.

(a) Le dimanche soir *Belle* attribue les probabilités  $1/2$  pour FACE et  $1/2$  pour PILE.

(b) Le lundi matin, avant qu'on lui offre de parier (c'est-à-dire avant qu'elle apprenne qu'on est lundi), elle prend en compte l'effet de loupe engendré par le protocole dans lequel elle est située, et elle attribue donc les probabilités  $1/3$  pour FACE et  $2/3$  pour PILE.

(c) Le lundi matin un peu plus tard, quand on lui offre le pari et qu'elle sait donc qu'on est lundi (information nouvelle), elle réexamine ses évaluations et, utilisant la formule de Bayes, retombe sur l'évaluation  $1/2$  pour FACE et  $1/2$  pour PILE qu'elle avait obtenue le dimanche soir.

Dans le cas du paradoxe de l'Apocalypse, le raisonnement correct calqué sur le précédent se formule ainsi.

(a) Utilisant diverses analyses scientifiques, on attribue les probabilités  $1/2$  pour la théorie A et  $1/2$  pour la théorie B. Cette phase se place avant que je ne prenne en compte mon existence dans l'Univers. C'est l'équivalent de l'examen de la pièce de monnaie que *Belle* effectue le dimanche soir qui lui indique que la pièce n'est pas truquée et que les probabilités sont donc  $1/2$ , pour FACE comme pour PILE.

(b) Je prends en compte ma situation dans l'Univers, ce qui est équivalent à prendre en compte que je participe à une expérience du type de celle organisée pour *Belle*. Je sais que, dans de telles expériences, je subis un effet de loupe probabiliste. Ici, l'effet est très sensible puisque les êtres humains sont mille fois plus nombreux dans la théorie B que dans la théorie A. Je réévalue mes attributions de probabilités, ce qui donne ici :  $1/1001 = 0,1\%$  pour A et  $1000/1001 = 99,9\%$  pour B.

(c) Je prends maintenant en compte mon rang de naissance et par utilisation de la formule de Bayes je trouve (voir les calculs figure 4) : Probabilité de A sachant que E est vrai =  $1/2$ , probabilité de B sachant que E est vrai =  $1/2$ .

Je suis retombé sur mes évaluations de départ. Il n'y a plus de paradoxe. Cette compensation exacte entre l'effet de loupe qui grossit la probabilité à attribuer à l'hypothèse que l'humanité va survivre longtemps et l'effet de la formule de Bayes qui grossit la probabilité que l'humanité va bientôt disparaître se produit avec tous les jeux de données imaginables.

Cela règle définitivement le problème du prétendu paradoxe de Leslie-Carter. Il est amusant de noter que lorsqu'il s'agit du paradoxe de *La belle au bois dormant*, les gens ayant participé aux débats se sont répartis à peu près équitablement entre partisans du  $1/2 - 1/2$  et partisans du  $1/3 - 2/3$ , alors que dans le cas du paradoxe de l'Apocalypse, pourtant formellement équivalent, la presque-totalité des avis refusait le mode de raisonnement de Leslie-Carter. Selon les enjeux, la rationalité varie... Encore un effet de sélection.

L'idée que finalement il faut prendre en compte l'effet de loupe probabiliste dans des cas comme celui du paradoxe de l'Apocalypse a malgré tout une conséquence ennuyeuse (qui est l'un des arguments utilisés par ceux qui s'opposent à la solution  $1/3 - 2/3$  du paradoxe de *La belle au bois dormant*). En effet, quand on s'interroge sur la possibilité de la vie extraterrestre, on aboutit à une conclusion qu'on peut juger contre-intuitive. Supposons pour simplifier que deux théories s'opposent.

**Théorie A.** Les lois de la physique rendent peu vraisemblable que la vie apparaisse sur une planète et il n'y a en moyenne qu'une planète donnant naissance à la vie par galaxie.

**Théorie B.** Les lois de la physique rendent assez vraisemblable que la vie apparaisse sur une planète et il y a en moyenne un milliard de planètes portant la vie par galaxie.

Imaginons qu'en raisonnant sur la base de données scientifiques générales (mathématiques, physique, chimie, biologie théorique, etc.), on évalue  $\text{Pr}(A) = 1/2$ ,  $\text{Pr}(B) = 1/2$ . Alors si nous voulons être cohérents avec la solution adoptée pour le paradoxe de l'Apocalypse, nous devons prendre en compte l'effet de loupe probabiliste et conclure que  $\text{Pr}(A)$  et  $\text{Pr}(B)$  doivent être réévaluées par nous humains situés dans l'Univers à un instant donné et non pas à l'extérieur. Les calculs aboutissent alors à la conclusion qu'en recherchant la vie autour de nous dans notre galaxie, la Voie lactée, nous avons un milliard de fois plus de chances de trouver d'autres planètes habitées que de ne pas en trouver. De  $1/2$  pour A,  $1/2$  pour B, nous passons par prise en compte de l'effet de loupe à  $1/1\,000\,000\,001$  pour A et  $1\,000\,000\,000/1\,000\,000\,001$  pour B.

L'hypothèse A est devenue totalement improbable (moins d'une chance sur un milliard) et il y a donc sans doute des extraterrestres proches de la Terre. Cette réévaluation n'est-elle pas troublante ? Êtes-vous toujours aussi sûr que l'effet de loupe doit être pris en compte et qu'il est la solution définitive du paradoxe de l'Apocalypse et du problème de *La belle au bois dormant* ?

Jean-Paul DELAHAYE est professeur d'informatique à l'Univ. de Lille.

J.-P. DELAHAYE, *Logique, informatique et paradoxes*. Le chapitre III traite du paradoxe de l'Apocalypse. Éd. Belin/Pour la science, 1995.

N. BOSTROM, *Anthropic Bias. Observation Selection Effects in Science and Philosophy*. Routledge, New-York, 2002

J. LESLIE, *The End of the World. The Science and Ethics of Human Extinction*. Routledge, New-York, 1996.

P. FRANCESCHI, Thèse : <http://www.univ-corse.fr/~franceschi/index-fr.htm>

*Anthropic Principle Preprint Archive* [site internet donnant accès à de nombreux articles] : <http://www.anthropic-principle.com/preprints.html>

Some « *Sleeping Beauty* » postings. Discussion autour du paradoxe de *Belle au bois dormant* : <http://www.maproom.demon.co.uk/sb.html>