ROBERT NOZICK ET LE LIBRE ARBITRE

The real trouble with this world of ours
is not that it is an unreasonable world,
nor even that it is a reasonable one.
The commonest kind of trouble is that
it is nearly reasonable, but not quite.
Life is not an illogicality;
yet it is a trap for logicians.
It looks just a little more mathematical and regular than it is;
its exactitude is obvious,
but its inexactitude is hidden;
its wildness lies in wait.

G. K. Chesterton, *Orthodoxy* (1908).

My first act of free will
shall be to believe in free will.

William James, *Entrée de journal personnel datée du 30 avril 1870.*

Un article rédigé par le philosophe américain Robert Nozick en 1969 [5] portant sur un paradoxe logique et appartenant à une branche des mathématiques qu'on appelle *théorie de la décision* fut à l'origine de l'une des chroniques *Mathematical Games* ayant suscité le plus de réactions de la part de lecteurs du magazine *Scientific American*.

Au cours du 20e siècle, la logique mathématique, la théorie des jeux, la théorie de la décision et de la philosophie des sciences suscitèrent et profitèrent abondement dans leurs développements respectifs d'une pléthore de paradoxes tous plus décoiffants les uns que les autres. En 1960, alors qu'il cogitait sur un célèbre paradoxe de la théorie des jeux appelé le dilemme du prisonnier, William A. Newcomb (1927-1999), un professeur de physique au laboratoire Lawrence Livermore de l'Université de Californie, formula un exemple à la fois amusant et déroutant de paradoxe de prédiction [1 ; 3]. Par l'intermédiaire d'un ami commun, le mathématicien et physicien Martin David Kruskal, le paradoxe identifié par Newcomb finit par se rendre jusqu'aux oreilles du professeur de philosophie à l'Université de Harvard, Robert Nozick[1], qui en fit l'objet d'un article savant paru en 1969 [5]. Cet article attira l'attention de Martin Gardner, qui y fit écho dans sa chronique *Mathematical Games* de juillet 1973 [1 ; 3].

Considérons l'expérience de pensée suivante, qui prend la forme d'un jeu : deux boîtes opaques et fermées – que l'on désignera dans ce qui suit par B1 et B2 respectivement – sont posées sur une table. On sait avec une absolue certitude que la boîte B1 contient 1 000 $. Quant à la B2, on sait qu'elle contient soit rien, soit un million de dollars (on ignore toutefois dans lequel des deux cas de figure l'on se trouve). Deux avenues s'offrent à nous et il nous faut faire un choix irrévocable entre celles-ci :

- <u>Choix 1</u> : Prendre la boîte B1 (contenant 1 000 $) et la boîte B2 (contenant soit 0 $, soit 1 000 000 $) ;
- <u>Choix 2</u> : Prendre seulement la boîte B2 (contenant soit 0 $, soit 1 000 000 $).

Quelque temps avant le début du jeu, un être supérieur a fait une prédiction au sujet de notre décision. S'il s'attendait à ce que l'on opte pour prendre les deux boîtes, alors il aura laissé la boîte B2 vide. En revanche, s'il s'attendait plutôt à ce que l'on choisisse de ne prendre que la boîte B2, alors il aura mis un million de dollars dans la boîte B2. Enfin, si l'être supérieur s'attendait à ce que, plutôt que de faire un choix, on remette notre sort entre les mains du hasard (en jouant à pile ou face par exemple), alors il aurait laissé la boîte B2 vide.

Précisons que l'être supérieur a fait sa prédiction au sujet de notre décision il y a de cela plusieurs jours. Il a alors rempli les boîtes en fonction de sa prédiction, puis les a scellées. Leur contenu doit être considéré comme immuable puisque l'être supérieur n'a pas le pouvoir

1. Robert Nozick est né à Brooklyn le 16 novembre 1938 au sein d'une famille d'immigrants provenant d'un shtetl russe. Suivant la complétion, en 1959, d'une formation universitaire de premier cycle à l'Université Columbia, Nozick réalisa des études de deuxième (1961) et de troisième (1963) cycles à l'Université de Princeton. Lauréat de l'une des prestigieuses bourses Fulbright, le jeune homme fréquenta ensuite l'Université d'Oxford au cours de l'année pédagogique 1963-1964. Nozick fut nommé professeur de philosophie à l'Université Harvard à l'âge de 30 ans et demeura au sein de cette institution pour l'ensemble de sa carrière. Le philosophe fut emporté, le 23 janvier 2022, par un cancer de l'estomac qu'on lui avait diagnostiqué dès 1994. De nos jours, Robert Nozick est souvent associé au mouvement libertarien en raison de l'ouvrage qu'il fit paraître en 1974 en réponse à *A Theory of Justice* (1971) de John Rawls. Dans cet ouvrage, intitulé *Anarchy, State, and Utopia* [6], Nozick développe et étaye la thèse voulant que les droits individuels sont primordiaux et que rien de plus qu'un État minimal (soit un État en mesure d'assurer une protection à ses citoyens contre la violence et le vol) ne saurait être justifié. Ce livre remporta le National Book Award 1975 dans la catégorie « Philosophie et Religion ».

d’en modifier le contenu *a posteriori* (tout type de rétrocausalité est exclu). De plus, notons que si cet être supérieur se révèle extrêmement doué pour prédire nos comportements, son pouvoir prédictif ne doit pas pour autant être considéré comme absolument infaillible.

Le paradoxe réside dans le fait que l’on est en mesure d’avancer de solides arguments pour l’un ou l’autre des deux choix qui s’offrent à nous :

- <u>Argument du principe de maximisation de l’utilité espérée</u> : Si on prend les deux boîtes, l’être supérieur l’aura presque assurément anticipé ; la boîte B2 sera donc selon toute vraisemblance vide et nos gains se limiteront à 1 000 $. Si on se restreint plutôt à prendre que la boîte B2, l’être suprême l’aura, une fois de plus, presque assurément anticipé et y aura placé un million de dollars. Cet argument peut être formalisé mathématiquement comme suit. Soit p la probabilité que l’être suprême prédise correctement nos comportements. Alors, l’utilité espérée du choix 1 (soit le fait de prendre les deux boîtes) est donnée par :

  $$p \times (1\,000) + (1 - p) \times (1\,001\,000).$$

  Quant à l’utilité espérée du choix 2 (ne prendre que la boîte B2), elle est donnée par :

  $$p \times (1\,000\,000) + (1 - p) \times 0.$$

  Or, un calcul fort simple nous permet de vérifier que pour toute valeur supérieure à $\frac{1001}{2000}$ (c’est-à-dire 50,05 %), l’utilité espérée du choix 2 est supérieure à celle du choix 1. Comme l’être supérieur est réputé être très bon prédicteur, il est raisonnable de supposer qu’il prédit correctement nos comportements avec un taux de succès considérablement supérieur à 50,05 %. Par conséquent, il apparaît avantageux selon cette logique d’opter pour le choix 2 et de ne prendre que la boîte B2.

- Argument du principe dominance : Puisque l’être supérieur a déjà fait sa prédiction, qu’il a rempli les boîtes en fonction de sa prédiction et qu’il n’a pas le pouvoir d’en modifier le contenu *a posteriori,* la situation est pour ainsi dire fixe et déterminée. Ou bien un million de dollars se trouvent dans la deuxième boîte, ou bien ils ne s’y trouvent pas. Mais dans un cas ou dans l’autre, rien de ce que l’on choisira de faire n’y changera quoi que ce soit.

> Dans ce cas, pourquoi ne pas prendre les deux boîtes et ainsi s'emparer de tout l'argent qui se trouve à notre portée ? Si la boîte B2 contient un million de dollars, alors tant mieux ! Nous engrangerons donc des gains cumulés de 1 001 000 $. Si, au contraire, la boîte B2 est vide alors nous n'avons rien perdu (elle était vide de toute façon) et nous repartons avec 1 000 $. Si l'on raisonne ainsi, le fait d'opter pour ne prendre que la boîte B2 apparaît absurde puisqu'on engrangera 1 000 000 $ dans le meilleur des cas et rien du tout dans le pire, soit deux issues moins favorables que celles où l'on prend les deux boîtes. Il convient donc de faire preuve d'un peu d'audace et de prendre les deux boîtes.

Il appert donc que deux des principes les plus fondamentaux de la théorie de la décision – à savoir le principe de maximisation de l'utilité espérée et le principe de dominance – donnent lieu à des recommandations qui semblent incontestables. Elles sont toutefois contradictoires !

Suivant la parution, en juillet 1973, de sa chronique portant sur le paradoxe de Newcomb, Gardner reçut une montagne de lettres de lecteurs analysant ce paradoxe et y allant de leurs propres explications quant à sa résolution. Le vulgarisateur se fit un malin plaisir à les rediriger vers Nozick [2 ; 3]. Au terme d'un examen minutieux des quelque 650 pages d'écrits rédigés par des lecteurs du magazine *Scientific American*, Nozick produisit une analyse que Gardner reproduisit dans sa chronique *Mathematical Games* de mars 1974 [2 ; 4]. Il en ressort toutefois que le paradoxe demeurait, aux yeux de Nozick à tout le moins, non résolu.

Parmi les nombreuses lettres reçues par Gardner et analysées par Nozick, on retrouve un message particulièrement savoureux de la part d'Isaac Asimov. Ce dernier, qui soutient pourtant adhérer à la conception déterministe, affirme qu'il *choisirait* (quoi que cela puisse signifier) les deux boîtes, et ce, sans la moindre hésitation. Le prolifique auteur de romans de science-fiction et d'essais de vulgarisation scientifique russo-américain explique son raisonnement ainsi : tout être humain digne de ce nom devrait se porter avec hardiesse à la défense du libre arbitre, si tant est qu'une telle chose existe. Supposons que l'on prenne les deux boîtes et qu'il s'avère que l'être supérieur, ayant prévu cela, n'a rien placé dans la deuxième boîte. Alors on n'aura peut-être pas remporté un million de dollars, mais au moins on aura exprimé avec conviction notre volonté de parier sur sa non-omniscience et sur notre propre libre arbitre. Et si d'aventure l'être supérieur avait erré et placé un million de dollars dans la boîte, alors non seulement aura-t-on gagné un million, mais on pourra aussi célébrer la réfutation de l'omniscience de l'être supérieur ainsi que la démonstration définitive de l'existence d'une telle faculté que le libre arbitre. Celui qui ne prend que la deuxième boîte aura beau empocher un million de dollars, il n'en ressortirait pas moins perdant puisque pour obtenir ce million il lui aura fallu démontrer qu'il consent à aliéner sa liberté. Toute personne qui se vend ainsi se rend indigne, affirme l'homme de lettres, d'être considérée comme un humain.

Dans un addendum qui suit la reproduction de sa chronique de mars 1974 dans le recueil de textes publié en 1986 sous le titre *Knotted Doughnut and Other Mathematical Entertainments* [4], Gardner livra le fond de sa pensée au sujet du paradoxe de Newcomb. À son avis, l'apparente contradiction entre les recommandations découlant de l'application des principes de dominance et de maximisation de l'utilité espérée est la démonstration patente que – de la même manière qu'il ne pourrait exister un barbier rasant tous les habitants masculins du village qui ne se rasent pas eux-mêmes et seulement ceux-ci – il ne saurait exister un être suprême en mesure de faire des prédictions interagissant causalement avec les événements visés par les prédictions.

Eût-il vécu quelques années de plus que Gardner aurait pu connaître le fin mot de l'histoire, car le paradoxe de Newcomb est, de l'avis de plusieurs, aujourd'hui considéré comme étant résolu. Sa résolution fut l'œuvre conjointe du mathématicien, physicien et informaticien David Wolpert et du physicien et auteur de science-fiction Gregory Benford. Dans une analyse détaillée publiée en 2013 [7], les deux hommes montrent que le paradoxe de Newcomb naît de ce que l'on peut tirer deux interprétations incompatibles d'une situation vaguement exposée, chacune d'elle suggérant une formalisation mathématique différente et menant, par voie de conséquence, à une réponse différente ; le tout en laissant croire qu'on est tout au long en présence de la même interprétation. Il n'y a donc pas, dans le paradoxe de Newcomb, de réel conflit entre le principe de dominance et le principe de maximisation de l'utilité espérée ; il y a simplement une imprécision dans la spécification de la structure probabiliste du jeu auquel on se prête. Dès lors que l'on prend soin de spécifier la structure probabiliste qui sous-tend le jeu, l'illusion d'un conflit entre les deux importants principes de la théorie de la décision se dissipe et la question admet alors une unique réponse optimale bien définie.

Il s'agit là d'une énième démonstration de ce que les paradoxes tendent à devenir lorsque les détails pertinents à la compréhension d'un problème ne sont pas tous spécifiés et qu'il existe plus d'une façon *intuitivement évidente* de combler les manquements informationnels. « La leçon du paradoxe de Newcomb, nous disent Wolpert et Benford, n'est qu'un rappel de l'ancienne vérité selon laquelle il faut définir soigneusement tous ses termes ».

Références

[1] Gardner, M. (1973, juillet). « Mathematical Games: Free will revisited, with a mind-bending prediction paradox by William Newcomb ». *Scientific American, 229* (1), 104-109. [www.jstor.org/stable/24923152]

[2] Gardner, M. (1974, mars). « Mathematical Games: Reflections on Newcomb's problem : a prediction and free-will dilemma ». *Scientific American, 230* (3), 102-109. [www.jstor.org/stable/24950033]

[3] Gardner, M. (1986). « Newcomb's Paradox ». Ch. 13 dans *Knotted Doughnut and Other Mathematical Entertainments*. W. H. Freeman & Co.

[4] Gardner, M. (1986). « Reflections on Newcomb's Paradox ». Ch. 14 dans *Knotted Doughnut and Other Mathematical Entertainments*. W. H. Freeman & Co.

[5] Nozick, R. (1969). « Newcomb's problem and two principles of choice ». Dans *Essays in honor of Carl G. Hempel*, 114-146. Springer, Dordrecht.

[6] Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.

[7] Wolpert, D. H., et Benford, G. (2013). « The lesson of Newcomb's paradox ». *Synthese, 190* (9), 1637-1646.