Les bienfaits d'Epiménide

Alexandre Boisseau
boisseau@lsv.ens-cachan.fr

L'un d'eux, leur propre prophète[1]*, l'a dit :*
Crétois, toujours menteurs, méchantes bêtes, cœurs blasés.
Saint Paul, Nouveau Testament, Epître à Tite (I,12)

Qu'est-ce qu'un paradoxe ? Ou plutôt, qu'est-ce qu'un paradoxe en mathématiques ? On peut en fait accorder à ce mot deux sens différents. Tout d'abord un paradoxe peut simplement être un résultat qui va à l'encontre de l'intuition. On peut citer par exemple le paradoxe de Banach-Tarski : il est possible de partitionner une boule en un nombre fini de morceaux puis de les réassembler différemment afin d'obtenir deux boules de même volume que la première. Un tel paradoxe met en relief certaines différences entre le monde mathématique, idéal sous certains aspects, et le monde réel. Un paradoxe peut également être un résultat qui s'avère être à la fois vrai et faux. Dans la plupart des cas, ceci intervient au cours d'un raisonnement par l'absurde et permet en fait de montrer que l'une des hypothèses faites est fausse. Plus rarement, les seules hypothèses sont celles communément admises et un tel cas de figure oblige à reconsidérer la théorie mathématique utilisée. Ce sont de tels paradoxes que nous nous proposons d'évoquer dans ce qui suit. Plus précisément, nous allons nous concentrer sur l'un deux, le paradoxe du menteur, et nous laisser guider par ses différentes applications, oscillant entre fondements des mathématiques et informatique. Avant d'entrer dans le vif du sujet, signalons que cet article a été rédigé suite à la lecture de [D] que l'on ne peut que conseiller à chacun, logicien ou non.

Le paradoxe du menteur

L'hypothétique Epiménide dont il est question aurait été un grec, crétois de surcroît, qui aurait énoncé la phrase suivante :

« Tous les Crétois sont des menteurs. »

Le caractère (faussement) paradoxal de cet énoncé apparaît lorsqu'on tient le raisonnement suivant :

– soit cette phrase est vraie, dans ce cas les crétois étant des menteurs et Epiménide un crétois, il se doit de mentir, en particulier lorsqu'il prononce cette phrase, qui par conséquent est fausse,
– soit cette phrase est fausse, ainsi les crétois ne sont pas des menteurs, en particulier Epiménide, et cette phrase doit donc être vraie.

Ce raisonnement n'est correct qu'en apparence, en effet il faut prendre garde au fait que si « être un menteur » désigne le contraire de « toujours dire la vérité », alors un menteur est simplement quelqu'un qui ne dit pas toujours la vérité. Ainsi le raisonnement suivant devient exempt de contradiction :

– soit la phrase prononcée par Epiménide est vraie, dans ce cas les crétois sont des menteurs et Epiménide un crétois, or l'hypothèse nous dit qu'il est en train de dire la vérité, ce qui n'est pas contradictoire (une fois n'est pas coutume...)
– soit cette phrase est fausse, ainsi il y a au moins un crétois qui n'est pas un menteur et tout ce qu'on peut en conclure c'est que ce n'est pas Epiménide.

Ce pseudo-paradoxe possède de nombreuses variantes (réellement paradoxales pour leur part) (on pourra consulter [G], relativement ludique), les plus simples, pour aller à l'essentiel, sont :

« Je mens » (sous-entendu : en ce moment même),

« Cette phrase est fausse ».

R. Smullyan présente dans ses livres [S_1, S_2] une autre manière d'obtenir des paradoxes à partir d'énoncés presque semblables à ceux d'Epiménide qu'il n'applique qu'à des personnes qui possèdent la propriété suivante : soit elles disent toujours la vérité (les « purs »), soit elles mentent toujours (les « pires »). Ainsi, il ne rencontre personne pouvant dire :

« Je suis un pire »

et si quelqu'un vient à dire :

« J'ai entendu quelqu'un dire qu'il était un pire »

alors c'est un pire (et il n'a jamais entendu cette phrase...). On peut varier les situations et les raisonnements à l'infini. Le dénominateur commun à tous ces paradoxes est l'autoréférence (considérer le mot « cette » qui fait référence à la phrase elle-même, ou bien le fait qu'Epiménide, en parlant des crétois, parle en particulier de lui-même, *etc*). C'est cette caractéristique qui rend si intéressante et fructueuse leur transposition à des problèmes mathématiques.

[1]Epiménide de Cnosse, poète et philosophe, VII^e et VI^e siècles avant J.C.

Vers la théorie des ensembles

En 1905, Russel construit un ensemble « paradoxal » rappelant le paradoxe du menteur[2]. En effet, l'ensemble considéré par Russel est le suivant :

$$\Omega = \{x \mid x \notin x\}$$

(l'ensemble des ensembles qui ne s'appartiennent pas) et on peut faire un raisonnement semblable à celui que l'on a fait à propos du paradoxe d'Epiménide :

- soit Ω est élément de lui-même (*i.e.* $\Omega \in \Omega$) ce qui veut dire, par définition de Ω que $\Omega \notin \Omega$ ce qui est contradictoire,
- soit Ω n'est pas élément de lui-même (*i.e.* $\Omega \notin \Omega$) et on a alors par définition $\Omega \in \Omega$ ce qui est également contradictoire.

Ainsi, la théorie mathématique considérée à cette époque était contradictoire puisqu'elle autorisait la définition d'un tel ensemble. Par la suite, différentes restrictions furent énoncées afin déviter de tels paradoxes, citons en seulement deux :

- dans la théorie de Zermelo-Fraenkel (qui est celle communément utilisée), on ne peut construire l'ensemble :

  $$\{x \mid P(x)\}$$

  (P désignant une propriété de x) que lorsqu'il a été établi au préalable qu'il existe un ensemble A tel que tout x satisfaisant P est dans A (ce qui n'était pas le cas pour l'ensemble Ω),
- en analysant le paradoxe, Russel et Whitehead ont observé qu'il découlait du cercle vicieux qui apparaît lorsqu'on considère l'expression $x \notin x$. Pour éviter la formation de telles expressions, ils ont donc mis au point une théorie (dite « théorie des types ») dans laquelle chaque objet possède un type (qui est un entier), les objets de type 0 étant appelés « objets de base » et les objets de type $n+1$ étant par définition les ensembles d'objets de type n. L'appartenance d'un objet à un autre est alors une (famille de) relation(s) entre objets de type n et objets de type $n+1$ et l'expression $x \notin x$ n'est pas une expression correcte.

Russel avait déduit sa construction de l'ensemble paradoxal Ω à partir de la démonstration du résultat de Cantor qui énonçait en 1890 que tout ensemble est strictement plus petit que l'ensemble de ses parties[3] (démonstration qui utilisait elle aussi une variante du paradoxe du menteur). De ce résultat, Cantor déduisait entre autres que l'ensemble des réels n'est pas dénombrable et qu'il existe des suites infinies strictement croissantes d'ensembles (en considérant pour un ensemble A quelconque :

$$A < \mathcal{P}(A) < \mathcal{P}(\mathcal{P}(A)) < \ldots)$$

Nous n'allons pas donner ici la démonstration du fait que $A < \mathcal{P}(A)$, mais nous allons prouver le résultat suivant, qui est assez semblable :

Théorème (du point fixe). — ***Pour un ensemble A quelconque, s'il existe une bijection de A dans A^A (l'ensemble des applications de A dans A) alors toute application de A dans A possède un point fixe.***

En fait, lorsque l'ensemble A est infini, $\mathcal{P}(A)$ est en bijection avec A^A et ce résultat est donc équivalent à $A < \mathcal{P}(A)$.

Considérons donc une bijection de A dans A^A (on désignera par $[a]$ l'image de $a \in A$ par cette bijection) ainsi qu'une application :

$$F: \quad A \quad \to \quad A$$

et montrons que F possède un point fixe. Pour $a \in A$, $[a]$ est une application de A dans A et on peut donc considérer $[a](a)$ qui est un élément de A. On pose ensuite $G(a) = F([a](a))$ ce qui définit une application G de A dans A que l'on écrit $G = [g]$ avec $g \in A$. On pose enfin $a_0 = G(g)$ et on obtient :

$$a_0 = G(g) = F([g](g)) = F(G(g)) = F(a_0)$$

ce qui montre que a_0 est un point fixe pour F.

Bien entendu, comme il existe clairement des applications de A dans A qui ne possèdent pas de point fixe (dès que A n'est pas un singleton) on en déduit que A n'est pas en bijection avec A^A. Remarquons que pour construire son ensemble paradoxal, Russel considère les x tels que :

$$\neg(x \in x)$$

tandis que nous considérons la fonction qui à x associe :

$$F([x](x)).$$

En fait, les constructions $x \in x$ et $[x](x)$ sont très semblables, à chaque fois on tente d'« appliquer x à lui-même ». D'autre part, ce qui joue le rôle de F

[2]Les développements historiques évoqués dans cette section et les suivantes sont extraits de [B_5]

[3]On dit q'un ensemble A est strictement plus petit qu'un ensemble B (et on note $A < B$) si et seulement si il existe une injection de A dans B mais pas de bijection, à ne pas confondre avec A strictement inclus dans B

dans la construction de Russel est l'opérateur de négation. De manière très informelle, on peut donc dire que l'obtention du paradoxe découle de la construction d'un point fixe pour la négation.

Ce résultat de point fixe peut paraître peu intéressant puisque les prémisses ne semblent pas pouvoir être satisfaites, excepté dans des cas triviaux. En fait, il prend toute sa valeur si on ne considère pas l'ensemble A^A tout entier mais seulement un sous-ensemble $\mathcal{F} \subseteq A^A$ qui se trouve être en bijection avec A (et posséder en plus certaines propriétés de stabilité). On obtient alors naturellement un théorème d'existence de point fixe pour les éléments de $\mathcal{F}$. De telles constructions sont essentiellement considérées en informatique, dans le cadre de la sémantique du λ-calcul[4]. Contentons-nous ici d'esquisser deux exemples :

- en 1969, D. Scott considère un espace topologique A et montre que sous certaines conditions l'espace des fonctions continues de A dans A est isomorphe à A,
- en 1972, G. Plotkin utilise P_ω (l'ensemble des parties de $\mathbf{N}$), l'ensemble des fonctions associé étant l'ensemble des fonctions $f : P_\omega \to P_\omega$ telles que pour $A \subseteq \mathbf{N}$:

  $$f(A) = \bigcup_{I \in \mathcal{P}_F(\mathbf{N})} f(I)$$

  (où $\mathcal{P}_F(\mathbf{N})$ désigne l'ensemble des parties finies de $\mathbf{N}$).

Tout ceci est encore très abstrait, mais nous allons voir dans la section suivante des applications informatiques plus concrètes du paradoxe du menteur.

Vers la calculabilité

Restons pour l'instant dans des considérations de points fixes et considérons un langage de programmation très simple[5] constitué d'expressions e qui sont construites de la manière suivante :

- si x est une variable, alors x est une expression,
- si e est une expression et x une variable, alors :

  $$\texttt{fun}\ x \to e$$

  est une expression qui représente la fonction qui à x associe l'objet décrit par l'expression e,
- si e et e' sont deux expressions, alors :

  $$e(e')$$

  est une expression qui représente l'application de e à e' (en général, e sera de la forme $\texttt{fun}\ x \to e''$ et dans ce cas la valeur de $e(e')$ est celle de e'' où l'on a remplacé tous les x par e).

Par exemple les expressions suivantes :

$$\texttt{fun}\ x \to x$$
$$\texttt{fun}\ f \to (\texttt{fun}\ x \to f(x))$$

désignent respectivement l'identité et la fonction qui applique son premier argument au deuxième. Dans ce langage de programmation, on peut appliquer n'importe quelle expression à n'importe quelle autre, en particulier à elle-même. On peut ainsi réaliser les mêmes constructions que celles qui nous ont permis d'obtenir des points fixes. On vérifie sans peine que si e est une expression, alors :

$$(\texttt{fun}\ x \to e(x(x)))(\texttt{fun}\ x \to e(x(x)))$$

est un point fixe pour e. Par suite si on considère l'expression :

$$\texttt{fun}\ f \to (\texttt{fun}\ x \to f(x(x)))(\texttt{fun}\ x \to f(x(x)))$$

notée Y, alors quelle que soit l'expression e que l'on considère, $Y(e)$ est un point fixe pour e, *i.e.* :

$$Y(e) = e(Y(e)).$$

Ainsi dans ce langage il y a un moyen uniforme de construire des points fixes. Nous allons voir que ceci permet en particulier de définir des fonctions par récurrence. Considérons par exemple la fonction factorielle fact : $\mathbf{N} \to \mathbf{N}$ définie par :

$$\begin{cases} \mathrm{fact}(0) = 1, \\ \mathrm{fact}(n) = n \times \mathrm{fact}(n-1) \text{ si } n \geq 1. \end{cases}$$

La factorielle peut alors être considérée comme le point fixe de la fonctionnelle :

$$\begin{array}{rccc} T: & (\mathbf{N} \to \mathbf{N}) & \to & (\mathbf{N} \to \mathbf{N}) \\ & f & \mapsto & \begin{bmatrix} \mathbf{N} & \to & \mathbf{N} \\ 0 & \mapsto & 1 \\ n \geq 1 & \mapsto & n \times f(n-1) \end{bmatrix} \end{array}$$

Dans notre langage de programmation, en supposant qu'on dispose des entiers et de l'instruction conditionnelle if ... then ... else ..., T se définirait par :

```
T = fun f →
      fun n →
         if(n = 0)
            then 1
            else n × f(n − 1)
```

[4] On pourra se reporter à [B1] ou [L] pour plus de détails sur le λ-calcul

[5] Il pourrait s'agir d'un sous-langage de Lisp

et compte tenu du fait que Y permet d'obtenir des points fixes, la factorielle est définie par :

$$\text{fact} = Y(T).$$

On voit donc ici que l'autoapplication donne une grande puissance au langage en permettant la construction de fonctions définies par récurrence. La contrepartie à cette puissance est qu'on ne peut pas en général assurer que tous les programmes terminent (considérer par exemple le programme :

$$Y(\texttt{fun}\ f \to \texttt{fun}\ x \to f(x))).$$

Il est d'ailleurs amusant de constater que la théorie des types qui est peu employée dans les mathématiques en général, trouve ici sa pleine application : en effet un moyen classique d'empêcher que des programmes puissent s'appliquer à eux mêmes est d'utiliser une notion de typage, qui doit tout de même rester suffisamment souple pour garder des programmes expressifs.

En fait, savoir si un programme termine[6] ou non sur une entrée est indécidable, *i.e.* il n'existe pas de programme m à deux arguments tel que $(m(e))(e')$ nous indique si le programme e termine ou non sur l'entrée e'. Nous allons montrer ce résultat car c'est encore un bel exemple d'utilisation du paradoxe du menteur faisant intervenir l'autoapplication. Pour savoir si un programme e termine sur l'entrée e' il suffirait que l'on dispose d'un programme m possédant la propriété suivante :

> $(m(e))(e')$ termine si et seulement si $e(e')$ ne termine pas

(en effet, il suffirait alors de lancer les deux calculs $(m(e))(e')$ et $e(e')$ en parallèle et on serait assuré que l'un des deux termine).

Supposons donc que l'on dispose d'un tel programme et considérons alors le programme m' définit par :

$$m' = \texttt{fun}\ x \to (m(x))(x)$$

et on regarde ce qui se passe pour $m'(m')$:
- soit $m'(m')$ termine, or $m'(m') = (m(m'))(m')$ qui par définition de m termine si et seulement si $m'(m')$ ne termine pas, ce qui est contradictoire,
- soit $m'(m')$ ne termine pas ce qui par définition de m entraîne que $m'(m')$ termine et on obtient là aussi une contradiction.

On en déduit qu'un tel programme m ne peut pas exister. On va voir dans la section suivante que ces considérations d'aspect pratique possèdent des conséquences mathématiques capitales.

Vers les théorèmes d'incomplétude

Nous allons ici montrer comment à partir d'un résultat d'indécidabilité on peut obtenir le premier théorème d'incomplétude de Gödel. Ce théorème montre que dans toute théorie mathématique raisonnable, il existe des énoncés qui ne sont ni prouvables ni réfutables. La présentation que l'on va donner est en partie inspirée de [S$_3$] et [CL].

Une théorie mathématique T peut être vue comme la donnée :
- d'un ensemble $\mathcal{E}$ dont les éléments sont appelés les énoncés de la théorie T, muni d'une involution calculable $\neg$ (*i.e.* une application $\neg : \mathcal{E} \to \mathcal{E}$ telle que $\neg\neg e = e$ pour tout énoncé e, $\neg e$ sera appelé la négation de e),
- une notion de prouvabilité qui est représentée par une partie $\mathcal{P}$ de $\mathcal{E}$. On dira qu'un énoncé e est prouvable (respectivement réfutable) dans T si et seulement si $e \in \mathcal{P}$ (respectivement $\neg e \in \mathcal{P}$).

Une telle théorie est dite :
- cohérente si, et seulement si, il n'existe pas d'énoncé qui soit à la fois prouvable et réfutable,
- complète si, et seulement si, tout énoncé est soit prouvable soit réfutable,
- énumérable si, et seulement si, il existe un programme qui consiste à afficher la liste (éventuellement infinie) des énoncés prouvables,
- décidable si, et seulement si, il existe un programme qui pour chaque énoncé e indique si il est prouvable ou non dans la théorie.

Le résultat clé est le suivant :

Théorème. — ***si une théorie est cohérente, complète et enumérable, alors elle est décidable.***

Pour le montrer, construisons (informellement) un programme qui prend en entrée un énoncé e et qui nous dit si e est prouvable ou non. La première chose à faire est de lancer le programme qui énumère tous

[6] Il faudrait définir formellement les notions d'« exécution » et de « terminaison » pour un programme. On préférera se contenter ici d'une notion intuitive. Le lecteur pourra bien entendu se reporter à [B$_1$] ou [L] pour une définition formelle ou à [B$_3$] pour une simple introduction

les énoncés prouvables. Comme la théorie est complète, on sait que l'on verra apparaître soit e soit $\neg e$ au bout d'un temps fini. Deux cas se présentent :

– soit on voit apparaître e, dans ce cas e est prouvable,
– soit on voit apparaître $\neg e$, dans ce cas e est réfutable et comme la théorie est cohérente, il n'est pas prouvable.

Dans chacun des cas on peut répondre, soit « e est prouvable », soit « il ne l'est pas ».

Considérons maintenant une théorie dans laquelle il est possible de « faire » des mathématiques. En particulier, cette théorie doit satisfaire un certain nombre de conditions :

– elle est enumérable puisque les énoncés prouvables sont obtenus à partir d'axiomes (qui sont énumérables) en appliquant des règles de déduction qui sont elles aussi énumérables,
– elle doit être suffisamment puissante, en particulier elle doit contenir l'arithmétique et, ce faisant, elle ne peut pas être décidable[7].

En appliquant le résultat précédent, on voit donc qu'une telle théorie ne peut pas être à la fois cohérente et complète. Par conséquent, si on suppose que la théorie que l'on utilise est cohérente, il existe des énoncés qui ne sont ni prouvable ni réfutables. Ce résultat constitue le premier théorème d'incomplétude de Gödel. Le second théorème consiste à expliciter un tel énoncé. On construit pour cela un énoncé noté coh_T qui exprime la cohérence de la théorie T en ce sens que si coh_T est réfutable dans T, alors T n'est pas cohérente. On montre ensuite que si la théorie considérée est cohérente, coh_T n'est pas prouvable (et puisque T est cohérente, il n'est pas réfutable non plus).

Quelques remarques à propos de ces résultats :

– le premier théorème nous dit que si on note $\mathcal{P}$ et $\mathcal{R}$ les énoncés prouvables et réfutables, alors $(\mathcal{P}, \mathcal{R})$ n'est jamais une partition de l'ensemble des énoncés. En effet, soit la théorie est incohérente et $\mathcal{P} \cap \mathcal{R} \neq \varnothing$, soit elle est incomplète et $\mathcal{P} \cup \mathcal{R} \neq \mathcal{E}$,
– par ailleurs, à supposer que la théorie soit cohérente, il est impossible de la compléter afin d'obtenir une théorie qui reste énumérable, cohérente et qui soit complète,
– le résultat sur la cohérence est assez amusant : soit la théorie est cohérente et elle ne peut alors pas prouver sa propre cohérence, soit la théorie est incohérente et alors elle peut prouver sa cohérence (puisqu'une théorie incohérente contient une contradiction et à partir d'une telle contradiction, on peut prouver n'importe quel résultat).

Pour conclure, voici un petit exercice (dont je ne connais pas la réponse) : dans la théorie des ensembles de Zermelo-Fraenkel, qui est la théorie communément utilisée, rien n'interdit *a priori* de considérer l'ensemble :

$$\Omega = \{x \in A \mid x \notin x\}$$

dès que l'on s'est donné un ensemble A. Cette construction (ou tout autre construction faisant intervenir $x \notin x$) est-elle utilisée dans les mathématiques de « tous les jours » ?

Pour en savoir plus...

[B$_1$] Hendrik Peter Barendregt, *The lambda calculus, its syntax and semantics*, North-Holland, 1981.

[B$_2$] John Barwise (ed.), *Handbook of Mathematical Logic*, North-Holland, 1977.

[B$_3$] Alexandre Boisseau, *Lambda calcul*, Hubris (1998), no. 3, 7–12.

[B$_4$] Raphaël Bomboy, *L'axiome du choix*, Hubris (1997), no. 0, 2–7.

[B$_5$] Nicolas Bourbaki, *Éléments d'histoire des mathématiques*, Masson, 1974.

[CL] René Cori et Daniel Lascar, *Logique mathématique (tomes 1 et 2)*, Masson, 1994.

[D] Gilles Dowek, *La logique*, Domino, Flammarion, 1995.

[G] Martin Gardner, *La magie des paradoxes*, Bibliothèque pour la science, Belin, 1980.

[L] René Lalement, *Logique, réduction, résolution*, Masson, 1990.

[S$_1$] Raymond Smullyan, *Le livre qui rend fou*, Dunod, 1984.

[S$_2$] Raymond Smullyan, *Ça y est, je suis fou*, Dunod, 1993.

[S$_3$] Raymond Smullyan, *Les théorèmes d'incomplétude de Gödel*, Masson, 1993.

[7] Il n'est pas surprenant (du moins de nos jours) que l'arithmétique permette de coder de nombreux problèmes indécidables, on pourra consulter l'article de M. Davis dans [B$_2$] à ce sujet