études

les avatars de l'axiome du choix. d'où vient-il? où se cache-t-il? à quoi nous sert-il?

par Michel Bonnard

Avatars : 1. Métamorphoses de Vishnou
2. Mésaventures (pop.)

RÉSUMÉ

A ses débuts l'axiome du choix fut très controversé. Il joue aujourd'hui un rôle important, multiforme et souvent méconnu dans la plupart des branches des mathématiques. En fait, les mathématiciens ne sauraient guère s'en passer puisqu'ils l'utilisent très souvent sans le voir. C'est ce qu'on essaie de montrer ici.

AVERTISSEMENT

Le présent article est écrit par un utilisateur pour des utilisateurs. Que le lecteur n'attende pas un exposé dogmatique et axiomatique de la théorie des ensembles. D'ailleurs, il existe plusieurs systèmes d'axiomes qui ne sont pas tout à fait équivalents entre eux, bien qu'ils puissent tous servir à "fonder" (entre autres choses) les mathématiques "ordinaires". Ce dont nous parlons, c'est de la place de l'axiome du choix dans ces mathématiques ordinaires.

Nous esquissons seulement quelques démonstrations parmi les moins difficiles ; c'est pour donner une idée de la façon dont certaines questions sont liées entre elles.

Les quelques considérations historiques de cet article doivent beaucoup à Michel Guillemot que je remercie ici.

1. Les constantes auxiliaires

Une *démonstration* mathématique (idéale) est une suite finie d'énoncés constitués chacun d'un nombre fini de mots ou de signes. Il y a d'abord (généralement sous-entendus) les *axiomes* (jugés évidents) et les théorèmes *déjà connus* (et démontrés ailleurs) ; puis il y a les *hypothèses* ("soit n un entier naturel", "soit $f : \mathbf{R} \to \mathbf{R}$ une fonction continue", ...). Les lettres introduites par les hypothèses sont appelées les *constantes* de la démonstration ; leur signification est fixée pour toute la suite. Ensuite le *corps de la démonstration* est constitué par d'autres énoncés, agencés de telle sorte que chacun se déduit de certains de ceux qui le précèdent au moyen des règles de raisonnement (syllogisme, raisonnement par l'absurde, ...). Parmi ces règles, il y a la *méthode de la constante auxiliaire*.

Si, à un certain moment, on arrive à un énoncé M de la forme "il y a au moins un objet vérifiant la propriété P", on adjoint provisoirement aux hypothèses l'énoncé (appelons-le Q) qui dit "soit a un objet vérifiant P" ; on continue la démonstration, et si on arrive à un énoncé R où n'intervient ni a, ni rien qui dépende de a, c'est que R résulte des seules hypothèses de départ (sans Q). L'objet a, qui disparaît à la fin, est appelé *constante auxiliaire*.

Pour légitimer cette méthode on pourrait "montrer" qu'il est possible de réécrire tous les énoncés figurant entre M et R, en usant de "périphrases" de façon à ne jamais *nommer* la constante a.

Souvent la propriété P est l'appartenance à un certain ensemble A ; dans ce cas, M peut donc s'écrire "A est non vide", et Q peut alors s'écrire "soit a un élément de A".

A titre d'exemple, démontrons l'énoncé *"toute injection admet une rétraction"*. Pour cela, partons de l'hypothèse "soit $f : E \to F$ une injection" (en sous-entendant que E et F sont des ensembles non vides). Comme il y a des objets appartenant à E, introduisons la constante auxiliaire $a \in E$. Nous définissons une application $g : F \to E$ en disant que :

si $y \in f(E)$, $g(y)$ est l'unique élément de $f^{-1}(\{y\})$,
si $y \notin f(E)$, $g(y) = a$.

Nous vérifions alors immédiatement que l'on a $g \circ f = \mathrm{id}_E$, c'est-à-dire que g est une rétraction de f. Nous concluons "f admet une rétraction" (énoncé où a n'intervient plus — on pourrait aussi bien dire "f est rétractable"). •

On est naturellement amené à utiliser parfois plusieurs constantes auxiliaires en nombre fini. On a des énoncés :

"il y a des objets vérifiant P_1"
"il y a des objets vérifiant P_2"... ;

on introduit les constantes auxiliaires a_1 vérifiant P_1, a_2 vérifiant P_2... ; on arrive à R où ne figure ni a_1, ni a_2... ; on conclut que R résulte des seules hypothèses de départ. Cette méthode se légitimerait en montrant qu'on peut réécrire n fois la démonstration en "éliminant" à chaque fois une constante auxiliaire de plus.

Mais la question se pose de savoir si on peut utiliser *une infinité de constantes auxiliaires.* Ce n'est pas qu'on écrira une infinité d'énoncés (une démonstration n'en comporte, par principe, qu'un nombre fini — on doit pouvoir l'écrire effectivement) ; c'est qu'on aura une famille de constantes auxiliaires indexée par un ensemble infini.

A titre d'exemple, cherchons à démontrer l'énoncé *"toute surjection admet une section".* Par hypothèse, nous nous donnons une surjection $g : F \to E$; pour chaque $x \in E$, l'ensemble $g^{-1}(\{x\})$ est donc non vide. Si E comprend par exemple, 37 éléments, soient $a_1, a_2, ..., a_{37}$ nous fabriquons une section f de g en introduisant 37 constantes auxiliaires, $b_1 \in g^{-1}(\{a_1\}), ..., b_{37} \in g^{-1}(\{a_{37}\})$. Mais que faire si E est infini ? Nous dirons que nous définissons $f : E \to F$ en posant que, pour chaque $x \in E$, l'image $f(x)$ est un élément b_x que nous avons *choisi* arbitrairement dans l'ensemble non vide $g^{-1}(\{x\})$. Si nous *"avons le droit"* de faire ainsi, nous obtenons f qui vérifie $g \circ f = id_E$, c'est-à-dire que nous obtenons une section de g . Nous concluons donc que g est "sectionnable" en ayant utilisé une infinité de constantes auxiliaires, savoir les b_x pour x décrivant E.

2. L'axiome du choix

Cet axiome affirme qu'il est légitime d'utiliser une famille de constantes auxiliaires indexées par un ensemble éventuellement infini. Plus précisément, il dit que si on a une famille (X_i) d'ensembles non vides, indexée par un ensemble non vide, *"on peut choisir arbitrairement"* un élément dans chacun des X_i ; c'est-à-dire sous forme plus mathématique :

Soit I un ensemble non vide et soit $(X_i)_{i\in I}$ une famille d'ensembles non vides ; alors il existe (au moins) *une famille $(a_i)_{i\in I}$ telle que, pour chaque i , on ait $a_i \in X_i$.*

Mais une famille (a_i) telle que, pour chaque i , on ait $a_i \in X_i$, c'est par définition un élément du produit cartésien ΠX_i ; l'axiome du choix équivaut donc à l'énoncé suivant (dit *axiome multiplicatif*) :

Le produit d'une famille non vide d'ensembles non vides est non vide.

Cet énoncé est parfois donné sous la forme apparemment plus faible mais en fait équivalente :

Le produit d'une famille non vide d'ensembles non vides, *deux à deux disjoints,* est non vide.

Pour démontrer le sens non évident de l'équivalence de ces deux énoncés, on utilise la *somme ensembliste.* Si $(X_i)_{i\in I}$ est une famille d'ensembles, et si X est leur réunion, leur somme est par définition :

$$\sum_{i\in I} X_i = \{(x,i) \in X \times I \ ; \ x \in X_i\} .$$

Cette somme est réunion d'ensembles X'_i, qui sont 2 à 2 disjoints, chacun d'entre eux étant en bijection avec le X_i correspondant. Alors le fait que ΠX_i soit non vide équivaut à ce que $\Pi X'_i$ soit non vide.

Cela étant dit, on sent bien que la légitimité de l'axiome du choix ne va pas de soi. Certes, *on ne voit pas ce qui empêcherait* qu'il existe des éléments de ΠX_i dès que tous les X_i sont non vides ; mais dire que l'on choisit arbitrairement un élément dans chacun des X_i est en quelque sorte une *affirmation gratuite* quand on sait très peu de chose sur les X_i.

Insistons sur le fait que, ce qui pose problème, c'est la légitimité d'une *infinité de choix arbitraires*. Pour un nombre fini de choix, on peut utiliser des constantes auxiliaires. Par ailleurs il n'y a pas de problème non plus si, dans chacun des X_i, il y a un procédé particulier qui donne un élément bien défini ; par exemple, si les X_i sont tous des segments de $\mathbf{R}$, nous pouvons définir $(a_i)\in\Pi X_i$ sans utiliser l'axiome du choix, en disant que, pour chaque i, a_i est le milieu de X_i.

Nous parlerons plus loin des utilisations des conséquences de l'axiome du choix (bon ordre, lemme de Zorn, ...). Mais les *utilisations directes* sont nombreuses et moins remarquées (et même, le plus souvent, inconscientes). Nous en avons vu une au **§ 1** (toute surjection admet une section) ; en voici une autre :

Soit $(X_n)_{n\in\mathbf{N}^*}$ une famille dénombrable d'ensembles dénombrables. Pour simplifier, nous supposons que les X_n sont 2 à 2 disjoints. Nous voulons montrer que $\bigcup_n X_n$ est dénombrable. Pour chaque n nous faisons correspondre aux éléments (dénombrables) de X_n les puissances successives du n-ième nombre premier. Nous fabriquons ainsi une injection de $\bigcup X_n$ dans $\mathbf{N}$, ce qui prouve que $\bigcup X_n$ est dénombrable. Mais en faisant tout cela, nous avons utilisé l'axiome du choix ; en effet, pour faire correspondre aux éléments de X_n les puissances du n-ième nombre premier, nous avons subrepticement *choisi pour chaque* n une injection de X_n dans $\mathbf{N}$; nous avons donc bien fait une infinité de choix arbitraires.

En fait, pour démontrer que *toute* union dénombrable d'ensembles dénombrables est dénombrable, on est *obligé* d'utiliser, d'une manière ou d'une autre, un axiome qui ressemble à l'axiome du choix. Par contre, si par exemple, pour chaque n, X_n est l'ensemble des valeurs d'une suite *donnée*, on sait fabriquer une injection particulière bien définie de X_n dans $\mathbf{N}$ et on montre alors que $\bigcup_n X_n$ est dénombrable sans utiliser l'axiome du choix.

Tout cela doit montrer assez que, pour faire des mathématiques en étant *certain* de ne jamais utiliser l'axiome du choix sans l'avoir vu, il y a lieu d'avoir subi un entraînement spécial.

Une *classe propre* est une collection d'objets mathématiques, "trop grosse" pour être un ensemble (par exemple : la classe de tous les cardinaux, la classe de tous les ensembles, ...). Ce dont nous parlons dans le présent article c'est l'axiome du choix *pour les ensembles* — c'est celui de "tout le monde". Mais il existe aussi (plus fort mais rarement utilisé) l'*axiome du choix pour les classes* qui dit que, si on a une classe propre d'ensembles non vides, on peut choisir un élément dans chacun d'eux. Cet axiome entraîne notamment que toute classe propre peut être bien ordonnée (alors que l'axiome du choix pour les ensembles entraîne seulement que tout ensemble peut être bien ordonné).

3. Cardinaux, injections et surjections

En pratique, pour "définir" les *cardinaux* il suffit de dire que l'on a $\mathrm{card}\,E = \mathrm{card}\,F$ si et seulement si il existe une bijection entre les ensembles (finis ou infinis) E et F (si on veut un cardinal est une *classe d'équivalence* d'ensembles modulo la relation "être en bijection" — mais, pour les puristes, il existe aussi de "vraies" méthodes ensemblistes pour "fabriquer" les cardinaux).

On définit la *relation d'ordre* entre les cardinaux en disant que l'on a $\mathrm{card}\,E \leqslant \mathrm{card}\,F$ s'il existe une injection $E \to F$ (cela définit bien une relation entre cardinaux car, s'il existe les bijections $E \to E'$ et $F \to F'$, et si de plus il existe une injection $E \to F$, il existe aussi une injection $E' \to F'$). On voit (facilement) que cette relation est réflexive et transitive et on peut montrer (difficilement, mais *sans* utiliser l'axiome du choix) qu'elle est antisymétrique, c'est-à-dire que, s'il existe à la fois une injection $E \to F$ et une injection $F \to E$, il existe une bijection entre E et F (théorème de Bernstein).

On appelle *principe de trichotomie* l'énoncé, naturel a priori, qui dit que les cardinaux sont totalement ordonnés, c'est-à-dire que, si on se donne deux ensembles, il y en a un qui peut être envoyé injectivement dans l'autre. Cet énoncé a été jadis jugé évident par des mathématiciens qui, par ailleurs, refusaient l'axiome du choix. Pourtant, il a été démontré depuis que ces deux énoncés sont en fait équivalents (cf. § 5).

Par ailleurs, une autre idée naturelle est de caractériser l'ordre entre cardinaux au moyen de surjections. S'il existe une injection $E \to F$ on sait montrer *sans* axiome du choix qu'il existe aussi une surjection $F \to E$ (il suffit de prendre une rétraction de l'injection donnée — cf. § 1). Le *principe de partition* est l'énoncé affirmant que la réciproque est vraie : s'il existe une surjection $F \to E$, il existe aussi une injection $E \to F$. Cet énoncé équivaut à dire que l'on a $\mathrm{card}\,E \leqslant \mathrm{card}\,F$ si et seulement si il existe une surjection $F \to E$.

Si l'axiome du choix est vrai, le principe de partition est vrai, en effet, toute surjection admet alors une section, laquelle est une injection (cf. § 1). Mais on ignore si l'axiome du choix est *équivalent* au principe de partition. En fait on peut montrer (ce n'est pas très difficile) que l'axiome du choix est équivalent au fait que toute surjection admet une section ; mais si le principe de partition est vrai et si on se donne une surjection $g : F \to E$, on sait qu'il y a aussi une injection $f : E \to F$, mais rien ne dit que f soit une section de g.

Le nom du principe de partition vient de ce que cet énoncé équivaut au suivant :

Soit $(A_i)_{i \in I}$ une partition d'un ensemble A (en parties non vides, 2 à 2 disjointes), alors on a $\operatorname{card} I \leqslant \operatorname{card} A$.
(L'équivalence est facile à montrer ; le point crucial est que, si on se donne une surjection quelconque, les images réciproques des éléments de l'espace d'arrivée forment une partition de l'espace de départ).

Toujours à propos d'injections et de surjections, signalons que l'axiome du choix (pour les ensembles) équivaut à *chacun* des énoncés :

Tout ensemble peut être envoyé injectivement dans toute classe propre ;

Toute classe propre peut être envoyée surjectivement sur tout ensemble.

4. Le bon ordre

Un ensemble ordonné est dit *bien ordonné* (ou muni d'un *bon ordre*) si toute partie comprend un plus petit élément (je dis bien "plus petit élément", et pas seulement "borne inférieure"). Les exemples les plus simples sont **N** muni de l'ordre habituel et **N** × **N** muni de l'ordre lexicographique (il y a aussi, bien sûr, des exemples beaucoup plus compliqués). Par contre, ni **Z** ni aucun intervalle non réduit à 1 point de **R** ou **Q** ne sont bien ordonnés pour l'ordre habituel.

Le *problème du bon ordre* c'est de savoir si tout ensemble peut être bien ordonné. S'il en est ainsi, l'axiome du choix est vrai ; en effet, si (X_i) est une famille (infinie) non vide d'ensembles non vides, nous munissons l'ensemble $X = \bigcup_i X_i$ d'un bon ordre et nous obtenons $(a_i) \in \prod X_i$ en disant que, pour chaque i, l'élément a_i est le plus petit élément de la partie X_i de X. En faisant ainsi nous avons fait le choix arbitraire d'un bon ordre sur X. Si, suivant la première idée qui vient à l'esprit, nous avions muni *chacun* des X_i d'un bon ordre (au lieu d'aller chercher la réunion X), nous aurions fait une infinité de choix arbitraires, ce qui n'aurait pas été acceptable pour *démontrer* l'axiome du choix à partir d'autre chose.

Réciproquement, on peut montrer (difficilement) que, si l'axiome de choix est vrai, tout ensemble peut être bien ordonné ; c'est le *théorème de Zermelo* (ou du bon ordre) dont la publication fut à l'origine de violente polémique (cf. § 7).

Les ensembles bien ordonnés servent à faire des démonstrations par *récursion transfinie*. Montrons ce que c'est par un exemple.

Précisons d'abord que, dans un ensemble bien ordonné, tout élément autre que le plus grand (s'il y en a un) admet un *successeur* (plus petit majorant strict) ; mais un élément peut avoir ou non un *prédécesseur* (plus grand minorant strict).

Cela dit, soit E un espace vectoriel (de dimension finie ou non, sur un corps commutatif quelconque) et supposons E muni d'un bon ordre noté $\leqslant$ (sans aucune relation avec les lois de composition). Nous allons fabriquer dans E des parties libres B_x de plus en plus grosses, indexées par x décrivant l'ensemble ordonné $E \setminus \{0\}$. D'abord, si x_0 est le plus petit élément de $E \setminus \{0\}$, on pose $B_{x_0} = \{x_0\}$. Ensuite les B_x se fabriquent de proche en proche par "récursion transfinie" :

• Si x n'a pas de prédécesseur, on pose
$$B_x = \bigcup \{B_z \,;\, z \in E \setminus \{0\},\ z < x\} \,;$$

• Si x a un prédécesseur y et si x n'appartient pas au sous-espace vectoriel engendré par B_y, on pose
$$B_x = B_y \cup \{x\} \,;$$

• Si x a un prédécesseur y et si x appartient au sous-espace vectoriel engendré par B_y, on pose $B_x = B_y$.

Ayant ainsi fabriqué des B_x pour tous les $x \in E \setminus \{0\}$, nous appelons B la réunion de tous ces B_x. Nous pouvons alors montrer (ce n'est pas très difficile) que les B_x sont libres et que B est libre maximale, c'est-à-dire que B est une base (algébrique). Nous avons donc démontré que tout espace vectoriel qui peut être bien ordonné admet une base.

Notons bien au passage qu'on n'a pas besoin d'axiome du choix pour "avoir le droit" de faire des récursions transfinies, dès lors qu'on a un bon ordre sous la main. Là où intervient l'axiome du choix, c'est pour pouvoir bien ordonner tel ensemble sur lequel on travaille.

Revenant à notre exemple, signalons son intérêt historique. Cauchy avait montré que, dans $\mathbb{R}$, tout endomorphisme de groupe qui est continu est une homothétie. La question se posait donc de chercher des endomorphismes de groupe (discontinus) qui ne soient pas des homothéties. En 1905, Hamel répondit à la question (en admettant l'axiome du choix) : le théorème de Zermelo dit que $\mathbb{R}$ peut être bien ordonné, et admet donc une base en tant qu'espace vectoriel sur $\mathbb{Q}$; au moyen de cette base, on peut alors fabriquer des applications de $\mathbb{R}$ dans $\mathbb{R}$ qui sont $\mathbb{Q}$-linéaires (ce sont donc des homomorphismes de groupe) et ne sont pas $\mathbb{R}$-linéaires (ce ne sont donc pas des homothéties).

5. Les principes de maximum

Dans les mathématiques "ordinaires", la récursion transfinie sert surtout à fabriquer, dans un ensemble, des parties de plus en plus grosses jusqu'à en attraper une qui soit maximale parmi celles à qui on s'intéresse. A cette méthode, jugée trop technique, on préfère habituellement aujourd'hui l'emploi de principes de maximum. Ces principes sont, en pratique, les conséquences les plus efficaces de l'axiome du choix ; il s'agit d'énoncés affirmant *l'existence d'éléments maximaux* dans certains ensembles ordonnés. Signalons d'abord quelques-uns de ces énoncés moins connus que le lemme de Zorn mais ayant un intérêt par eux-mêmes :

• Tout ensemble partiellement ordonné contient une chaîne maximale pour l'inclusion (Hausdorff — une *chaîne* est une partie totalement ordonnée pour l'ordre induit) ;

• Tout ensemble ordonné contient une chaîne n'ayant pas de majorant strict (Tukey) ;

• Dans tout ensemble, toute famille de parties de caractère fini comprend un élément maximal pour l'inclusion (Hausdorff, Bourbaki, Tukey, Teichmuller — énoncé habituellement nommé "lemme de Tukey" — une famille $\mathcal{E}$ de parties d'un ensemble X est dite de caractère fini si une partie A de X est élément de $\mathcal{E}$ dès que toute partie finie de A est l'élément de $\mathcal{E}$; par exemple, dans un espace vectoriel, la famille constituée par les parties libres est de caractère fini ; l'énoncé ci-dessus permet donc de retrouver l'existence de bases en dimension infinie).

L'utilisation d'un principe de maximum est déjà sous-jacente dans la démonstration par Zermelo du théorème qui porte son nom (1904). Elle est explicite chez Hausdorff (1909). Mais c'est seulement depuis Zorn (1935) qu'elle s'est popularisée, avec l'emploi général de l'énoncé suivant :

Soit $\mathcal{E}$ un ensemble partiellement ordonné, non vide, dans lequel toute chaîne admet un majorant. Alors tout élément de $\mathcal{E}$ est majoré par un élément maximal.

On peut obtenir $2^3 - 1 = 7$ variantes de cet énoncé en remplaçant, à volonté, "ensemble ordonné" par "famille de parties d'un ensemble donné", "chaîne" par "partie bien ordonnée", "majorant" par "borne supérieure" — toutes ces variantes sont en fait équivalentes à l'énoncé donné ci-dessus.

Cet énoncé est habituellement appelé *lemme de Zorn*, du nom de celui qui l'a effectivement démontré en 1935 sans savoir que d'autres (Hausdorff, Kuratowski, Bochner, R.L. Moore) avaient déjà établi, indépendamment les uns des autres, des résultats très voisins, restés peu connus. Une raison (entre autres) du fait que tant de mathématiciens aient ignoré les résultats obtenus par d'autres est sans doute que les principes de maximum qu'ils utilisaient n'étaient pas au centre de leurs préoccupations, mais leur apparaissaient comme de simples outils techniques.

Pour "voir" les liens entre ces questions, nous allons esquisser deux tronçons faciles de la démonstration de l'équivalence entre l'axiome du choix, le lemme de Zorn et le principe de trichotomie. Pour cela, précisons d'abord qu'un ensemble ordonné dans lequel toute chaîne admet un majorant est appelé un ensemble *inductif*.

Montrons que, si le lemme de Zorn est vrai, toute surjection admet une section. Soit $g : F \to E$ une surjection. Trouver une section de g revient à trouver une partie A_0 de F telle que $g(A) = E$ et que $g\|_{A_0}$ soit injective ; en effet, la réciproque de la bijection $g\|_{A_0}$ sera alors une section de g. Pour cela, notons $\mathcal{E}$ l'ensemble des parties A de F telles que $g\|_{A_0}$ soit injective. L'ensemble $\mathcal{E}$ est non vide (il comprend les singletons) et, ordonné par inclusion, il est inductif (si (A_i) est une famille totalement ordonnée d'éléments de $\mathcal{E}$, on peut vérifier que $\bigcup_i A_i$ est élément de $\mathcal{E}$). Il y a donc dans $\mathcal{E}$ un élément maximal A_0 ; et de la maximalité de A_0 on déduit que $g(A_0) = E$.

Montrons maintenant que le lemme de Zorn entraîne le principe de trichotomie. Soient E et F deux ensembles et soient p et q les projections canoniques de $E \times F$ sur E et F respectivement. Appelons $\mathcal{E}$ l'ensemble constitué par les parties A de $E \times F$ qui sont telles que $p|_A$ et $q|_A$ soient toutes deux injectives. Là encore, on vérifie que $\mathcal{E}$ est non vide et inductif et admet donc un élément maximal A_0. De la maximalité de A_0 on déduit que l'on a $p(A_0) = E$ ou $q(A_0) = F$. Dans le premier cas, on a une bijection $A_0 \to E$ et une injection $A_0 \to F$, d'où on déduit une injection $E \to F$. Dans le deuxième cas, on a une injection $A_0 \to E$ et une bijection $A_0 \to F$, d'où une injection $F \to E$.

Dans ces deux esquisses de démonstration, et suivant en cela l'usage établi, nous n'avons pas vérifié en détail que l'ensemble $\mathcal{E}$ était inductif. L'usage d'escamoter ces vérification (fastidieuses mais pas énormément difficiles) vient de ce qu'elles se ressemblent toutes (on a l'impression de faire toujours la même chose). Mais l'étudiant moyen ne sait guère les faire seul, et, habitué à nous voir les escamoter, il trouvera facilement des ensembles inductifs là où il n'y en a pas. C'est pour cela que votre serviteur a proposé une formulation qui permet d'éviter ces vérifications (elles sont, en quelque sorte, faites une fois pour toutes).

Soit E un ensemble, soit $n \geqslant 1$ un entier et soit $P(A, x_1, \ldots, x_n)$ une propriété concernant A partie de E et $x_1, \ldots, x_n$ éléments de E. Nous dirons que P est une *propriété inductrice* de type n si

$$[P(A, x_1, \ldots, x_n) \text{ et } (A \subseteq A')] \Rightarrow P(A', x_1, \ldots, x_n).$$

Dans ce cas, nous dirons que $A \subset E$ est *compatible* avec une telle P si, quels que soient $x_1, x_2, \ldots, x_n$ éléments de A, la propriété $P(A, x_1, \ldots, x_n)$ est vraie. Avec ce vocabulaire, on peut démontrer (au moyen du lemme de Zorn) l'énoncé suivant :

Soit E *un ensemble ; soit* $(P_i)_{i \in I}$ *une famille* (éventuellement infinie) *de propriétés inductrices sur* E. *Soit* A *une partie de* E *compatible avec toutes les* P_i. *Alors* A *est incluse dans une certaine* A_0, *maximale* (pour l'inclusion) *parmi les parties qui sont compatibles avec toutes les* P_i.

Par exemple, si E est un groupe et si a est un élément de E autre que l'élément neutre e, introduisons les propriétés :

$$P_1(A,x) \underset{\text{def}}{\Longleftrightarrow} x \neq a$$

$$P_2(A,x) \underset{\text{def}}{\Longleftrightarrow} (x^{-1}) \in A$$

$$P_3(A,x_1,x_2) \underset{\text{def}}{\Longleftrightarrow} (x_1.x_2) \in A.$$

On vérifie immédiatement que P_1, P_2, P_3 sont inductrices (P_1 ne fait intervenir A qu'en apparence). Alors dire que A est compatible avec P_1 c'est dire "pour tout $x \in A$, on a $x \neq a$", autrement dit "$a \notin A$". Dire que A est compatible avec P_2 c'est dire que, pour tout $x \in A$, on a $x^{-1} \in A$. Dire que A est compatible avec P_3 c'est dire que, quels que soient $x_1, x_2 \in A$, on a $(x_1.x_2) \in A$. Finalement, dire que A est compatible *à la fois* avec P_1, P_2 et P_3, c'est dire que A est un sous-groupe de E ne comprenant pas a. Notre énoncé nous permet donc de conclure :

Tout sous-groupe A ne comprenant pas a est inclus dans un certain sous-groupe A_0, maximal parmi les sous-groupes ne comprenant pas a.

Si nous avions voulu déduire ce résultat directement du lemme de Zorn, nous aurions dû vérifier que l'ensemble des sous-groupes ne comprenant pas a est inductif (ce qui demande — théoriquement — pas mal de détails techniques). Au contraire, vérifier que P_1, P_2, P_3 sont inductrices est immédiat.

Nous ne ferons pas ici un exposé détaillé sur les techniques des propriétés inductrices. Signalons cependant que ces techniques permettent de traiter toutes les applications courantes du lemme de Zorn. Plusieurs exemples ont été exposés (oralement) au Colloque International de Logique à Orsay (juillet 1985) ; un article (avec ces exemples et d'autres) devrait être publié prochainement.

6. Efficacité de l'axiome du choix

Donnons un choix arbitraire fini de résultats qui sont liés à cet axiome.

D'abord, parmi les énoncés qui lui sont *équivalents*, en plus des trois formulations du **§ 2**, du théorème de Zermelo et des cinq principes de maximum de **§ 5**, il y a aussi :

- toute surjection admet une section ;
- toute relation d'équivalence sur un ensemble admet un système complet de représentants ;
- les cardinaux sont totalement ordonnés (principe de trichotomie) ;
- tout ensemble infini A peut être mis en bijection avec $A \times A$;
- tout ensemble peut être envoyé injectivement dans toute classe propre ;
- toute classe propre peut être envoyée surjectivement sur tout ensemble ;
- tout treillis unitaire non trivial comprend un élément maximal ;
- tout anneau commutatif dans lequel $1 \neq 0$ contient un idéal maximal (théorème de Krull) ;
- tout espace vectoriel de dimension infinie admet une base algébrique ;
- dans tout espace vectoriel, tout sous-espace admet un supplémentaire ;
- tout produit d'espaces quasi-compacts est quasi-compact (théorème Tychonoff).

D'autres résultats sont liés à l'axiome du choix en ce sens qu'on ne peut pas (ou que l'on ne sait pas) les démontrer sans utiliser au moins une forme affaiblie de l'axiome du choix. Par exemple :

- tout ensemble fini-au-sens-de-Dedekind est fini (un ensemble est dit fini au sens de Dedekind s'il ne peut être mis en bijection avec aucune de ses parties propres) ;
- tout ensemble infini contient un sous-ensemble infini dénombrable ;
- toute réunion dénombrable d'ensembles dénombrables est dénombrable ;
- s'il existe une surjection $F \to E$, il existe aussi une injection $E \to F$ (principe de partition) ;
- la continuité des fonctions $\mathbf{R} \to \mathbf{R}$ peut se caractériser au moyen des suites ;
- la mesure de Lebesgue est dénombrablement additive ;
- l'intégrale d'une limite simple de fonctions est (dans les bons cas) la limite de leurs intégrales (théorèmes de Beppo Levi et de Lebesgue) ;
- toute solution locale d'une équation différentielle peut se prolonger en une solution maximale ;
- toute forme linéaire continue sur un sous-espace d'un espace normé peut se prolonger à l'espace entier (théorème de Hahn-Banach) ;
- en topologie, la définition des limites "par les filtres" est équivalente à la définition "par les suites généralisées" ;
- tout filtre sur un ensemble est majoré par un ultrafiltre ;
- toute fonction continue sur un fermé d'un espace topologique normal se prolonge à l'espace entier (théorème d'Urysohn) ;
- etc.

Il faut cependant remarquer que, si l'axiome du choix (éventuellement affaibli) est nécessaire pour que les énoncés ci-dessus soient vrais dans *toute* leur généralité, cet axiome n'est pas toujours indispensable dans les cas "concrets" courants (par exemple, le théorème de Krull est équivalent à l'axiome du choix, mais les idéaux maximaux de **Z** ou de **R**[X] se fabriquent "à la main").

7. Légitimité de l'axiome du choix ?

Jusqu'ici nous avons parlé de ce qu'on sait faire en admettant que l'axiome du choix est vrai. Mais est-il "vrai" ? La question s'est posée bizarrement car les mathématiciens ont d'abord utilisé cet axiome sans y prendre garde.

Dès 1821, Cauchy emploie des infinités de choix arbitraires dans une démonstration du théorème de la valeur intermédiaire. Il est vrai que cet emploi est une facilité et peut être évité en recourant, par exemple, à des bornes supérieures. Par contre, la caractérisation bien connue de la continuité au moyen des suites, publiée en 1872 par Heine (mais semblant due à Cantor), utilise de façon incontournable une infinité dénombrable de choix arbitraires. Il y aurait bien d'autres exemples, mais ces deux là sont volontairement choisis dans la bonne vieille analyse réelle. Avant la date fatidique de 1904, les infinités de choix arbitraires ne semblent guère inquiéter les mathématiciens, hormis quelques Italiens (dont Peano et Beppo Levi — ce qui n'est quand même pas négligeable).

Les problèmes graves sortiront de l'affaire du bon ordre. En 1883, Cantor avait admis comme évident que tout ensemble peut être bien ordonné. Devant le scepticisme de la plupart des mathématiciens, il finit par reconnaître le besoin d'une démonstration. En 1904, J. König fait une communication (qui se révèlera erronée par la suite) où il pense avoir démontré que **R** *ne peut pas* être bien ordonné. En réponse, Zermelo démontre le théorème qui porte son nom (si l'axiome du choix est vrai, tout ensemble peut être bien ordonné — et réciproquement). C'est à cette occasion que l'axiome du choix est enfin explicité (avec le statut d'axiome).

L'article de Zermelo provoqua de violentes polémiques. Beaucoup de mathématiciens semblent s'être mis subitement à refuser les infinités de choix arbitraires parce que cela permettait de démontrer que **R** pouvait être bien ordonné, possibilité à laquelle ils ne croyaient pas. De nombreux adversaires déclarés de l'axiome du choix l'avaient largement utilisé sans le voir avant 1904 ; et certains continueront après (toujours sans le voir). D'autres, refusant l'axiome du choix (jugé faux) proposeront à sa place des énoncés (jugés évidents) qui se révèleront plus tard être des équivalents de l'axiome refusé. Certains dépenseront beaucoup d'énergie et d'imagination à essayer (en vain) de démontrer la vérité (ou au contraire la fausseté) de l'axiome du choix. Ces polémiques dureront de nombreuses années, mais l'axiome du choix s'imposera à la longue, grâce à son efficacité et au fait que personne ne sera parvenu à le réfuter.

Aujourd'hui, l'usage est d'employer librement l'axiome du choix dans presque toutes les branches des mathématiques. On ne peut d'ailleurs guère faire autrement : quand on utilise un résultat démontré par un autre, lequel utilise des résultats démontrés par plusieurs autres, chacun desquels utilisant..., on *ne peut pas* remonter toute la filière pour savoir si l'axiome du choix est utilisé quelque part (surtout qu'il est parfois bien caché — cf. **§ 2**). Pour être certain de ne *jamais* utiliser l'axiome du choix, il faudrait réécrire *toutes* les mathématiques (ce que certains Italiens avaient un moment envisagé).

Ce sont les polémiques sur l'axiome du choix qui ont amené Zermelo à axiomatiser la théorie des ensembles. Avant 1904, suivant l'idée initiale de Cantor, la théorie des ensembles était purement naïve. Mais, pour arriver à *prouver* quelque chose au sujet de la légitimité de l'axiome du choix, la naïveté ne suffit plus ; il faut fixer des limites *extrêmement précises* à ce qu'il est permis de faire avec les ensembles. Finalement un assez large consensus s'est fait autour du système dit "ZF" (amélioration par Fraenkel d'un système dû à Zermelo). *Très* grossièrement, les axiomes de ZF disent qu'on a "le droit" de faire les opérations usuelles sur les ensembles (produit, réunion, ensemble des parties, partie définie par une propriété...). Si on adjoint l'axiome du choix à ce système (qui ne le comprend pas au départ), on obtient le système dit "ZFC". Les mathématiques "courantes" qui s'écrivent aujourd'hui peuvent être raisonnablement considérées comme basées sur ZFC.

Parmi d'autres systèmes, il y a aussi (moins populaire chez les mathématiciens "ordinaires") celui de Von Neumann-Bernays-Gödel (dit NBG). En fait, les différences entre ZF et NGB n'affectent guère les mathématiques courantes.

Pour en revenir à la légitimité de l'axiome du choix, les polémiques étaient bien calmées quand la question fut résolue, essentiellement par Gödel (1940) et Cohen (1963), qui démontrèrent (dans NBG) qu'on ne peut pas démontrer que l'axiome du choix est faux ("consistance relative") mais qu'on ne peut pas démontrer non plus qu'il est vrai ("indépendance") — il est *indécidable*. Il ne peut donc avoir que le statut d'axiome supplémentaire que l'on peut accepter ou refuser.

Mais c'est là que se pose la "vraie" question. Les mathématiques seront-elles "meilleures" avec l'axiome du choix ou sans lui (et avec éventuellement autre chose à la place) ? Pour répondre à cette question, il faudrait d'abord lui donner un sens précis. Entre autres choses, il y a sans doute lieu de considérer l'aptitude des mathématiques à décrire la réalité concrète. Par exemple, les nombres réels servent à mesurer les grandeurs physiques ; alors, le fait qu'il existe ou non des parties non mesurables dans **R** (ce qui dépend de l'axiome du choix — cf. plus loin), influe peut-être sur la pertinence avec laquelle on décrit les phénomènes physiques.

En attendant, on peut énumérer quelques arguments pour ou contre l'axiome du choix.

En faveur de cet axiome, il y a son efficacité et le fait que, de toute façon, on l'utilise souvent sans s'en apercevoir (cf. **§ 2**) — et puis, étant donné une famille d'ensembles non vides, on ne voit pas ce qui empêcherait de prendre un élément dans chacun d'eux.

L'argument contre cet axiome c'est qu'il suscite des êtres "insaisissables", peu intuitifs ou même vraiment bizarres. Par exemple, on se demande vraiment à quoi peut ressembler ce fameux bon ordre sur **R** dont l'existence est affirmée. Par ailleurs, la projection canonique **R**→**R**/**Q** admet une section dont l'image est une partie A de **R** ayant un et un seul élément dans chaque classe modulo **Q** ; non seulement on ne voit pas du tout comment est faite A, mais en plus A n'est pas mesurable, ce qui est bien gênant (ce serait si agréable si toute partie de **R** était mesurable...).

Mais il y a pire, notamment le *paradoxe de Hausdorff* : au moyen de l'axiome du choix, on démontre que la sphère S de $\mathbf{R}^3$ est union disjointe de 4 ensembles A,B,C,D tels que D soit dénombrable et que les ensembles A,B,C et B∪C soient tous isométriques entre eux (donc A est à la fois le tiers et la moitié de S \ D...). La publication de cet exemple par Hausdorff en 1914 fut prise comme argument contre l'axiome du choix — ce qui n'était d'ailleurs pas du tout l'intention de l'auteur.

Si on refuse l'axiome du choix sans le remplacer par autre chose, on renonce à beaucoup trop de résultats (cf. §6) ; il a donc été proposé de le remplacer par tel ou tel énoncé plus faible, notamment :

- *L'axiome des choix dénombrables* qui dit que tout produit dénombrable d'ensembles non vides est non vide (i.e. on peut faire des infinités dénombrables de choix arbitraires) ;

- *L'axiome des choix dépendants* (plus fort que le précédent) qui dit qu'on peut construire des suites par récurrence en faisant un choix arbitraire à chaque pas (cet axiome permet de conserver l'essentiel de l'analyse réelle classique) ;

- *L'axiome de la détermination des jeux* (dont l'énoncé n'apporterait rien ici) qui permet, lui aussi, de conserver l'analyse classique, mais entraîne en plus que toute partie de **R** est mesurable (on ne sait pas encore si cet axiome est compatible avec ZF — les avis sont partagés ; mais il entraîne par exemple, que la projection canonique **R**→**R**/**Q** n'a pas de section, ce qui peut être jugé bizarre...).

8. Et Bourbaki, dans tout cela ?

Beaucoup de lecteurs français abordent la théorie des ensembles dans le fameux traité de Bourbaki et y cherchent en vain l'axiome du choix, lequel n'est pas mentionné dans le corps du traité, mais figure seulement dans la note historique et dans le fascicule de résultats. La raison de ce paradoxe est que, suivant une idée de Hilbert, Bourbaki utilise le symbole τ (appelé ϵ chez Hilbert) qui permet d'escamoter l'axiome du choix.

Etant donné une propriété R concernant l'objet mathématique arbitraire x, on introduit un certain objet noté $\tau_x(R)$. Cet objet *existe toujours* mais la *seule* chose que l'on sache à son sujet c'est que :

- s'il y a au moins un objet vérifiant R, alors $\tau_x(R)$ vérifie R ;
- s'il n'y a pas d'objet vérifiant R, alors $\tau_x(R)$ ne vérifie pas R.

La conception d'un tel objet est tout à fait contraire à l'intuition dans la pratique naïve du mathématicien ordinaire. En effet, soit, par exemple, E un ensemble non vide et prenons pour R la propriété "$x \in E$" ; comme il y a des objets qui appartiennent à E, l'objet $\tau_x(x \in E)$ est un élément "bien défini" de E, mais on n'a *aucun moyen* de savoir lequel. En particulier, on ne saura jamais si $\tau_x(x \in \{0,1\})$ est 0 ou 1.

Il y a cependant un cas où ce τ peut être utile au naïf : s'il y a *un et un seul* objet vérifiant R, alors $\tau_x(R)$ est cet unique objet (par exemple $\tau_x\ (\forall y,\ y \notin x)$ est l'unique ensemble ayant la propriété de ne pas avoir d'élément — c'est-à-dire que c'est ϕ).

Cela étant, l'emploi du τ, s'il est légitime, entraîne de façon évidente l'axiome du choix pour les classes : si (X_i) est une classe non vide d'ensembles non vides, on a un choix canonique d'un élément a_i de chaque X_i en écrivant $a_i = \tau_x(x \in A_i)$. En voyant cette "démonstration" on a bien l'impression que le τ sert à *nous empêcher de nous demander* s'il est légitime de faire des infinités de choix arbitraires.

En fait, l'axiomatique de Bourbaki est "essentiellement équivalente" à ZFC — en un sens que votre serviteur serait bien en peine de préciser complètement, mais qui veut dire en pratique qu'on peut aborder *avec profit* le livre de théorie des ensembles de Bourbaki directement par le chapitre 2 (en sautant au besoin les passages où sont utilisés des notations ou du vocabulaire "non courants" — il faudra un jour comprendre pourquoi on arrive à pratiquer une théorie sans connaître ce qui est *censé* en être le fondement — et Bourbaki est loin d'être le seul qui conduise à se poser cette question).

9. Pour en savoir plus

On trouve des démonstrations (plus ou moins détaillées) du lemme de Zorn à partir de l'axiome du choix dans un certain nombre de livres assez courants (niveau licence - maîtrise) notamment (choix arbitraire) :

P. DUBREIL et M.L. DUBREIL-JACOTIN, *Leçons d'algèbre moderne,* Paris, Dunod, 1964.

A.G. KUROSH, *Algèbre générale*, Paris, Dunod, 1967.

J.M. EXBRAYAT et P. MAZET, *Algèbre 1*, Paris, Hatier, 1971.

R. et A. DOUADY, *Algèbre et théories galoisiennes*, (vol. 1), Paris, Cedic - Fernand Nathan, 1977.

Pour une introduction rapide et efficace (en anglais) à l'axiomatique de ZF (avec démonstration des équivalents usuels de l'axiome du choix), on pourra lire le petit fascicule (très apprécié de votre serviteur) :

K.J. DEVLIN, *Fundamentals of contemporary set theory*, New York, Springer, 1979.

Une autre introduction à ZF, avec des ouvertures vers la logique, des applications à la topologie et une présentation de candidats au remplacement de l'axiome du choix se trouve dans :

F. GUENARD et G. LELIÈVRE, *Compléments d'analyse* (vol. 1), Fontenay-aux-Roses, les Cahiers de Fontenay, 1985.

Pour une liste très complète (avec des démonstrations) des équivalents de l'axiome du choix, on consultera :

H. et J. RUBIN, *Equivalents of the axiom of choice* (2e éd.), Amsterdam, North-Holland, 1985.

Pour l'**histoire** de l'axiome du choix et des polémiques y afférentes, deux ouvrages sont fondamentaux :

J. CASSINET et M. GUILLEMOT, *L'axiome du choix dans les mathématiques de Cauchy (1821) à Gödel (1940)*, Thèse à l'Université Paul Sabatier (n° 1094/95), Toulouse, 1983.

G.H. MOORE, *Zermelo's axiom of choice*, New York, Springer, 1982.

Pour rattacher tout cela à l'histoire générale de la logique et de la théorie des ensembles, on lira la note historique de :

N. BOURBAKI, *Théorie des ensembles*, Paris, Hermann, 1970.

Par contre, pour comprendre en quoi consiste la formalisation faite dans le chapitre I de ce dernier ouvrage (notamment l'emploi du τ), on consultera plutôt :

R.E. EDWARDS, *A formal background to mathematics* (I-a), New York, Springer, 1979.

Signalons enfin qu'il y a un article qui parle notamment du paradoxe de Hausdorff dans le numéro de février 1987 de la revue "Pour la Science".