CONCEPTIONS DU HASARD ET BIAIS PROBABILISTES CHEZ DES ENSEIGNANTS DU SECOND DEGRÉ : EFFET D'UNE FORMATION COURTE

Nicolas GAUVRIT[1]

TITLE

Effect of a 2-days training on teachers' (mis-)conceptions of randomness and probability

RÉSUMÉ

L'article présente une expérience en deux volets : l'un concerne les représentations du hasard chez des enseignants de mathématiques et de sciences et leur possible évolution à court terme lors d'une formation courte. L'autre porte sur quatre erreurs communes, leur importance, et l'effet possible à long terme d'une formation de deux jours sur ces biais. Nous concluons en donnant des pistes de remédiations.

Mots-clés : *probabilités subjectives, biais probabilistes, hasard.*

ABSTRACT

We present a two-fold study on in-service mathematics and science teachers. In a first experiment, conceptions of randomness is investigated, and the question of a possible effect of a short (2-days) training in probability is addressed. In a second pre-test post-test experiment, we analyse the long-term effect of the short training on four biases. We conclude by suggesting some possible remediations.

Keywords: *subjective probability, probabilistic bias, randomness.*

1 Introduction

Une des raisons pour lesquelles les probabilités et la statistique semblent être un domaine à part des mathématiques pour les élèves des collèges et lycées est la formation irrégulière des enseignants sur ces questions. Domaines spécifiques, un peu décalés parce que faisant partie des mathématiques appliquées dans une discipline où l'on valorise souvent le versant théorique, les probabilités et la statistique sont la bête noire de certains élèves, mais aussi de certains enseignants.

Des chercheurs en éducation s'inquiètent depuis des décennies et réfléchissent sur des biais et des conceptions trompeuses que les enseignants du secondaire semblent souvent tenir concernant le hasard ou le raisonnement probabiliste ou statistique (Lahanier-Reuter, 1999). Des dispositifs de formation des enseignants ciblés ont été testés parfois avec succès pour améliorer les difficultés rencontrées par les enseignants (Batanero, Godino & Roa, 2004).

Si ces formations longues semblent efficaces, la question reste ouverte de savoir si une formation courte (deux jours en l'occurrence) dans un cadre de médiation scientifique pourrait

[1] LDAR, Université Paris 7, ngauvrit@me.com

avoir un impact mesurable sur les conceptions du hasard ou sur les biais probabilistes présentés par les enseignants.

1.1 Conceptions du hasard

En ce qui concerne les conceptions du hasard, nous pouvons supposer que les enseignants de mathématiques disposent tout à la fois d'intuitions primaires, indépendantes de l'éducation, qui sont celles de tout un chacun, et d'intuitions secondaires développées au cours de leur formation. Cette distinction entre intuitions primaires et secondaires a d'abord été proposée par Fischbein (1987) pour séparer celles qui semblent provenir de notre perception naturelle du monde et qui seraient présentes même en l'absence d'éducation, et celles qui semblent naître de l'éducation. Par exemple, l'idée que toute variable aléatoire est nécessairement uniforme (voir plus bas) est peu présente chez les enfants ou les adultes n'ayant suivi aucune formation en probabilités. En revanche, elle est très fréquente chez les étudiants ayant suivi une formation de quelques années en probabilités. Cela fournit un argument pour dire que cette intuition est née de l'enseignement, ce qui caractérise les intuitions secondaires. À l'inverse, l'intuition que la suite de pile ou face PPPPPP est moins susceptible d'être obtenue avec une pièce équilibrée que PFFPFPP a tendance à diminuer avec les apprentissages probabilistes : il s'agit, selon la définition de Fischbein, d'une intuition primaire.

L'expérience 1 présentée plus bas a pour objectif de décrire quelques aspects de ces intuitions, et d'en discuter brièvement la pertinence mathématique. Nous verrons que pour ces conceptions, la formation courte de deux jours autour de laquelle a été conçue l'expérience ne semble pas avoir eu d'impact significatif.

Un nombre important de biais et d'heuristiques probabilistes ont été décrits en psychologie de l'éducation. Parmi ceux-ci, nous avons choisi d'étudier les quatre suivants : le biais d'équiprobabilité, l'heuristique d'invincibilité du hasard, la négligence des taux de base et le phénomène Falk. Nous les présentons ci-après.

1.2 Equiprobabilité

Le *biais d'équiprobabilité* (Lecoutre, 1985, 1992), ou *heuristique d'uniformité* (Falk & Lann, 2008), renvoie à l'idée que le hasard est équiprobable *par nature*. Ce biais évolue négativement avec l'éducation en probabilités, parce qu'il s'agit d'une intuition secondaire qui semble remplacer d'autres conceptions du hasard. Ce biais est communément mis en évidence dans le cadre des problèmes de type « Duc de Toscane », où il est demandé aux participants s'il est plus probable de trouver une somme de 11 ou de 12 lors du lancer simultané de deux dés équilibrés. La réponse fautive consiste à répondre que les deux alternatives sont équiprobables, alors que le 11 (un 5 et un 6) est deux fois plus probable que le double 6, mais il existe des tâches plus simples permettant de mettre en évidence des formes plus « sévères » de ce biais. Des élèves pensent par exemple qu'un échantillon de boules tirées d'une urne devrait avoir la même proportion de chaque couleur, même si l'urne ne contient pas des proportions égales (Zawojewski & Shaughnessy, 2000).

Les sujets qui commettent le biais d'équiprobabilité justifient généralement leur intuition soit par l'évocation d'une loi du hasard, du type « les probabilités sont égales puisque c'est le hasard » (Callaert, 2004), soit par l'application à des combinaisons d'événements d'une équiprobabilité qui est valable sur les événements de base, comme « puisque chacun des deux dés est équilibré, il n'y pas plus de chance d'avoir 11 ou 12 » (Pratt, 2000). Il est certain que ce biais est lié à la conception du hasard et pas seulement à une difficulté d'ordre logique abstraite, puisqu'on peut le réduire significativement en masquant le caractère aléatoire de la situation (Lecoutre, 1992).

1.3 Invincibilité du hasard

Il est une intuition secondaire que l'éducation nous met très tôt en tête lorsque nous abordons les probabilités : *on ne peut pas gagner contre le hasard*. Enoncée dans cette version simplifiée, cette loi mathématique devient une heuristique, une conception fautive, parce qu'elle n'est vraie que pour les espérances, et non pour les probabilités de gagner.

Considérons en effet le cas d'un jeu équilibré purement aléatoire où, par exemple, on a la même probabilité de gagner et de perdre à chaque partie. Chaque partie fonctionne sur le principe du quitte ou double. On mise une somme x à sa convenance. Si l'on perd, la mise est perdue. Si l'on gagne, on récupère sa mise plus x (on double donc sa mise). On peut jouer autant de parties qu'on le souhaite, et s'arrêter quand on le souhaite. Une martingale classique fonctionne de la manière suivante (Delahaye, 1998) : on veut repartir en ayant au total gagné 1€. On commence par parier 1€. En cas de gain, on arrête et on repart donc avec un gain global de 1€. En cas de perte, on a perdu 1€ et il faut donc gagner 2€ à la partie suivante pour repartir avec un gain total de 1€. On parie donc 2€. En cas de gain, on arrête. En cas de perte, il nous faudra parier (1 + 2) + 1 = 4€, etc. En pariant selon cette méthode, soit on finira ruiné (i.e., avec trop peu pour parier à nouveau selon cette martingale), soit on repartira avec 1€ de plus. Si l'on est assez riche pour jouer n fois, la probabilité de ruine est $1/2^n$, qui tend rapidement vers 0. Cette martingale permet donc, pourvu que l'on ait suffisamment de réserves, de gagner avec une probabilité aussi élevée qu'on le souhaite. Bien entendu, on gagnera très probablement un peu, mais on risque de perdre énormément : l'espérance du gain reste nulle.

L'intuition que le hasard est invincible pourrait bien nous conduire à considérer que les martingales sont inefficaces même en termes de probabilité de gagner. Cette hypothétique heuristique d'invincibilité du hasard n'a à notre connaissance pas été étudiée.

1.4 Négligence du taux de base

La négligence des taux de base est un biais qui consiste à ne pas utiliser les probabilités *a priori* dans un raisonnement de type bayésien (Tversky & Kahneman, 1974 ; Bar-Hillel, 1980). Ce biais a été énormément discuté dans la littérature, parce qu'il conduit à des erreurs de jugement chez les médecins (Casscells, Schoenberger, & Graboys, 1978) ou les jurés (Thompson & Schumann, 1987).

Il est mis en évidence par des problèmes du type du « problème du taxi » (pour une revue, voir Birnbaum, 1983) :

« *A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data: (i) 85% of the cabs in the city are Green and 15% are Blue. (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.*

Question: What is the probability that the cab involved in the accident was Blue rather than Green? »

La réponse la plus fréquente est 80%, alors qu'un calcul de probabilité conditionnelle indique que la réponse correcte est de 41% – à condition de supposer (1) que les taxis verts et bleus ont *a priori* les mêmes probabilités d'être impliqués dans des accidents et (2) que les taux d'identifications correctes dans le test soient identiques pour les deux couleurs, ce que les auteurs pensent être des hypothèses *par défaut* naturelles. On interprète cet écart comme le signe d'une absence de prise en compte du taux de base (proportions de taxis bleus et verts dans la ville).

1.5 Phénomène Falk

Falk (1979, 1989) a posé la question suivante à des étudiants de psychologie – voir Batanero & Sanchez (2005, pp. 251-253) pour une présentation synthétique – :

Problem 11. An urn contains two white balls and two red balls. We pick up two balls at random, one after the other without replacement.

(a) What is the probability that the second ball is red, given that the first ball is also red?

(b) What is the probability that the first ball is red, given that the second ball is also red?

Alors que les questions (a) et (b) sont mathématiquement équivalentes, et que les étudiants répondent correctement à la question (a), beaucoup échouent à répondre correctement à la question (b), en ne prenant pas en compte le deuxième tirage dans l'estimation des probabilités. La justification la plus fréquente donnée par les participants est que le deuxième tirage ne peut pas influencer le premier puisqu'il arrive après. On appelle ce biais *fallacy of the axis time* (biais de l'axe du temps) ou « phénomène Falk ».

Gras et Totohasina (1995) ont identifié, à partir de l'observation d'étudiants de 17-18 ans, trois grandes conceptions des probabilités conditionnelles, dont les deux premières peuvent expliquer le biais décrit par Falk (1979) :

- la *conception chronologique*, selon laquelle $P_B(A)$ n'a de sens que si l'événement A a lieu après l'événement B ;
- la *conception causale*, selon laquelle $P_B(A)$ est compris comme la probabilité de A « sachant que B est une cause de A » ;
- enfin, la *conception cardinale*, qui utilise la définition abstraite $card(A \cap B)/card(B)$ (valable uniquement dans le cas fini équiprobable).

Les deux premières conceptions amènent au phénomène Falk.

Dans la suite de cet article, nous présentons les résultats de deux expériences couplées portant pour l'une sur les conceptions du hasard et pour l'autre sur les quatre biais que nous venons de présenter. Ces expériences ont été réalisées auprès d'enseignants de mathématiques et de sciences des collèges et lycées (établissements accueillant les élèves, en France, de leur sixième à leur douzième année d'études).

2 Méthode

53 enseignants (33 femmes, 20 hommes) de mathématiques et de sciences des collèges et lycées de la région parisienne ont participé à l'expérience. Ils étaient répartis en deux groupes (de 28 et 25 personnes).

L'expérience s'est déroulée autour d'une formation courte de 2 jours organisée par Universcience et le Palais de la Découverte à Paris en 2012. Cette formation consistait en une suite de conférences autour du thème du hasard et des probabilités, et avait pour but de favoriser une meilleure compréhension des utilisations possibles de ces thèmes par les enseignants et pour l'enseignement. Les thèmes abordés par les conférenciers comprenaient l'histoire des probabilités, les grands paradoxes mathématiques liés au hasard, nos représentations du hasard (psychologie), le hasard en physique, en biologie, les algorithmes de productions pseudo-aléatoires, et l'utilisation trompeuse possible des résultats de sondages, par exemple par l'occultation de la variabilité des résultats. Il ne s'agit donc pas d'une formation à la théorie des probabilités, mais plutôt d'une mise en perspective des notions de hasard et de probabilité telles qu'elles sont utilisées dans les différents domaines de la connaissance.

L'étude proprement dite comporte deux volets sous forme pré-test-post-test. L'un concerne les conceptions du hasard, l'autre quatre erreurs communes en probabilités. Nous traiterons dans la suite chacun de ces deux thèmes séparément pour faciliter la lecture.

3 Expérience 1 : conceptions du hasard

3.1 Méthode

Les conceptions du hasard ont été évaluées au moyen de 14 questions (7 avant la formation, 7 juste après la dernière intervention du stage) de « jugement de représentativité » : les sujets devaient, pour une série d'événements, estimer s'ils étaient plus ou moins représentatifs du hasard en cotant selon une échelle de Lickert en 4 points (1 = pas représentatif du tout du hasard ; 4 = très représentatif du hasard). Il est bien évident qu'il n'y a pas de réponse « correcte » ou « fausse » à ces questions et les participants en étaient informés. Ce type de questionnaire est classiquement utilisé pour déterminer la structure des concepts, même si notre but était ici moins ambitieux. L'objectif est d'identifier ce qui est spontanément considéré comme plus ou moins « représentatif du hasard » par les sujets, indépendamment de toute référence théorique qui fournirait une réponse normative à cette question. Nous ne supposons pas que les participants ont *a priori* tous les mêmes conceptions du hasard, ni que la question telle qu'elle est posée devrait évoquer chez eux les mêmes images. La question est donc volontairement large et ne précise pas de quel hasard il s'agit, puisque c'est bien à l'image spontanée, et non aux connaissances que les participants seraient capables de convoquer, que nous nous intéressons ici.

Nous avons utilisé les deux groupes afin de contrebalancer l'ordre de passation des questions d'évaluation de représentativité des événements aléatoires : le groupe « direct » a répondu avant la formation aux questions Q1.x, et à la fin de la formation aux questions Q2.x. Le groupe « inverse » a d'abord répondu aux questions Q2.x.

3.2 Résultats

Les réponses aux 14 questions d'évaluation de représentativité des événements aléatoires ne montrent aucune évolution entre le pré-test et le post-test (après correction pour le nombre de tests, c'est-à-dire en calculant une significativité qui tient compte du fait qu'on a réalisé 14 tests, les différences ne sont pas significatives[2]). Il n'est donc pas nécessaire de distinguer les deux groupes. Le reste de l'analyse statistique porte sur l'ensemble des données disponibles sans distinction de groupe.

Le tableau 1 donne le résultat moyen et écart-type de la moyenne (ETM) pour chaque question.

TABLEAU 1 – *Moyenne et écart-type de la moyenne (ETM) de la note de représentativité (de 1 à 4) accordée aux événements par les sujets. Les événements sont rangés par ordre croissant de représentativité (les moins représentatifs en premier). Les couleurs indiquent les groupes d'événements qui seront analysés ensemble.*

Evénement	Moyenne ± ETM
(1.4) Un feu tricolore est réglé pour rester 50 secondes vert, 4 secondes orange, et 10 secondes rouge à chaque cycle. Il est vert lorsque vous arrivez au niveau du feu.	2,59 ± 0,14
(2.1) Dans les décimales de Pi (3,14159…), on trouve une série de six « 9 » consécutifs à partir de la 762ème décimale.	2,65 ± 0,14
(1.7) Une installation électrique tombe en panne à Nancy le 4 mars 2011.	2,70 ± 0,16
(2.2) Un dé truqué tombe sur 6 avec une probabilité de 50%, et sur chaque autre face avec une probabilité de 10%. On le lance, et il tombe sur 6.	2,75 ± 0,15
(1.6) On choisit 8 cartes de suite dans un jeu de 54 cartes ordinaires, et on considère la couleur des cartes (R : rouge ou N : noir). On obtient dans l'ordre RNRNRNRN.	2,91 ± 0,16
(2.7) Pendant un orage, une installation électrique tombe en panne à Nancy.	2,94 ± 0,16
(1.5) Lors d'une pluie de météorites qui touche les côtes de l'Europe, une petite météorite tombe dans l'océan à 1000km des côtes portugaises.	2,95 ± 0,17
(1.2) Un dé truqué tombe sur 6 avec une probabilité de 50%, et sur chaque autre face avec une probabilité de 10%. On le lance, et il tombe sur 1.	3,00 ± 0,19
(2.3) On lance 8 fois de suite une pièce. On obtient PFPFPFPF (P signifiant « pile » et F « face ») dans cet ordre.	3,04 ± 0,16
(1.1) La fonction ALEA() des tableurs courants permet, au moyen d'un calcul, de générer des nombres « suivant une loi uniforme » entre 0 et 1. Après avoir allumé un ordinateur, on tape « =ALEA() », et 0,384 s'affiche.	3,06 ± 0,15

[2] On a simplement divisé le risque à partir duquel on considère qu'un résultat est significatif (5 %) par le nombre de tests (14). Il aurait ici fallu que l'un des tests donne un résultat significatif à 0,36 % pour pouvoir conclure.

(1.3) On lance 8 fois de suite une pièce. On obtient PFFFPFPP (P signifiant « pile » et F « face ») dans cet ordre.	3,11 ± 0,14
(2.4) Un feu tricolore est réglé pour rester 30 secondes vert, 4 secondes orange, et 30 secondes rouge à chaque cycle. Il est orange lorsque vous arrivez au niveau du feu.	3,13 ± 0,14
(2.6) On choisit 8 cartes de suite dans un jeu de 54 cartes ordinaires, et on considère la couleur des cartes (R : rouge ou N : noir). On obtient dans l'ordre RRNRNNNR.	3,29 ± 0,14
(2.5) Une petite météorite tombe dans l'océan à 1000km des côtes portugaises.	3,35 ± 0,16

Les items 1.4 et 2.4 portent tous deux sur des événements similaires, mais de probabilités différentes. L'application du test de Student établit qu'il s'agit de résultats significativement différents : la valeur de la statistique de Student est t = 2,911, à laquelle est associée la significativité p = 0,006. Les items 1.2 et 2.2 entretiennent le même rapport que les questions 1.4 et 2.4. Les résultats observés sont conformes à ceux du cas précédent (t = 2,084, p = 0,044). Les tailles d'effet sont de 0,47 pour la comparaison 1.4-2.4, et de 0,33 pour la comparaison 1.2-2.2. Il s'agit donc dans les deux cas d'effet faible à moyen selon les conventions proposées par Cohen (1992)[3].

Les couples 1.6-2.6 et 1.3-2.3 entretiennent le même rapport : une situation équivalente donne une suite régulièrement alternée, ou au contraire sans régularité apparente. Dans les deux cas, le pattern irrégulier est jugé plus représentatif du hasard, mais les tailles d'effet sont différentes : un effet moyen-faible de 0,28 pour les couleurs de cartes (1.6-2.6), contre un effet très faible, de seulement 0,04 pour le pile-ou-face (1.3-2.3).

Les questions 2.1 et 1.1 concernent des événements déterministes et dont on aurait pu s'attendre à ce qu'ils soient perçus comme peu représentatifs du hasard. L'un (l'algorithme ALEA()) est généralement considéré comme produisant un pseudo-hasard « raisonnable », contrairement à l'autre (les décimales de pi). La différence entre les deux événements correspond à un effet moyen-faible (d = 0,28).

Les questions 1.5 et 2.5 se distinguent seulement par l'ajout, pour la question 1.5, d'un contexte explicatif pour l'événement (une pluie de météorites). L'événement hors contexte est jugé plus significatif du hasard, avec une taille d'effet moyenne de 0,25. Le même pattern avec les questions 1.7 et 2.7 donne des résultats différents : c'est l'événement en contexte qui est jugé plus significatif, avec cependant un effet très faible (d = 0,05). Cela peut être dû à l'effet de la précision de date dans l'item 1.7.

3.3 Discussion

Ces différentes estimations de représentativité permettent de jeter un premier regard sur la force et la réalité d'un ensemble de présupposés sur le hasard (Nickerson, 2002). Certains sont partagés par les probabilistes, d'autres non. Commençons par remarquer que la note la

[3] Dans le cas de la comparaison de deux moyennes, la taille d'effet vaut $d = |m - m'|/s$ où m et m' sont les moyennes observées, et s l'écart-type de la différence. Cohen (1992) propose de considérer qu'une valeur pour d de 0,2 est faible, 0,5 moyenne et 0,8 forte. Il ne s'agit là que d'une définition subjective, mais largement utilisée en sciences humaines.

plus basse est de 2,59, une valeur supérieure au centre de l'échelle de Lickert (2,5). Globalement, il y a donc une tendance à considérer les événements, fussent-ils choisis pour leur caractère déterministe, comme des exemples d'événements « dus au hasard ». Il est difficile d'interpréter cette valeur, du fait que sept sujets ont répondu « très représentatif » à *tous* les items. On peut bien sûr considérer que le hasard est en quelque sorte un mot vide. Dans ce cas, la question de la représentativité n'a pas de sens. Il reste étonnant qu'une telle idée s'exprime par l'affirmation que tous les événements sont « très représentatifs » du hasard, et non sous forme de réponses plus modérées.

La comparaison des items 1.4 et 2.4 d'une part, 1.2 et 2.2 d'autre part, suggère qu'un événement est plus représentatif du hasard s'il est de probabilité plus faible. Cette conception du hasard est en phase avec la théorie de l'information : l'information associée à un événement de probabilité p est en effet $-\log_2(p)$. Elle est donc maximale si p est minimal. À l'inverse, l'information moyenne (ou entropie) d'un *système* aléatoire sera maximisée par une loi de probabilité uniforme. Les deux comparaisons faites ici donnent des résultats différents. Dans le cas du dé (questions 1.2 et 2.2), l'effet de la probabilité de l'événement est plus faible. Cela peut s'expliquer par l'apparition d'un prototype (le dé) dont le caractère emblématique dans le cadre du hasard pourrait agir comme un modulateur de l'effet d'information.

Les couples 1.6-2.6 et 1.3-2.3 permettent d'aborder la question de la régularité ou de la complexité des suites aléatoires. Ces deux couples suggèrent une prédominance des patterns irréguliers ou complexes, en parfaite concordance avec ce que prévoit la théorie de la complexité algorithmique (Delahaye, 1999) et son application aux suites courtes (Delahaye & Zenil, 2012). Dans le cas prototypique du pile-ou-face pourtant, l'effet de complexité semble presque neutralisé. Ce phénomène peut s'expliquer par un effet de l'éducation, qui peut produire une intuition secondaire : l'éducation en probabilité engage les sujets à activer la représentation équiprobable (correcte ici) selon laquelle toute suite de pile-ou-face est aussi probable qu'une autre. Cet « effet secondaire de l'enseignement » (Morsanyi, Primi, Chiesi, & Handley, 2009) ne semble pas être transposé à des cas moins conventionnels comme celui des feux tricolores.

La comparaison des items 1.1 et 2.1, qui concernent deux événements déterministes, montre un décalage intéressant : si les décimales de pi sont considérées comme peu aléatoires (avec tout de même une moyenne de représentativité supérieure à la valeur centrale 2,5), la fonction ALEA() est considérée comme un bon producteur de hasard. Cela laisse présager encore une fois une double conception : une intuition primaire que le hasard n'est pas déterministe, modulée par une intuition secondaire provenant de l'éducation selon laquelle le pseudo-aléatoire de l'ordinateur est un bon prototype de machine aléatoire « malgré tout ».

Les items 1.5 et 2.5 illustrent l'effet possible d'une phrase explicative : le simple fait d'avoir ajouté une expression contextualisant la chute de la météorite (« pendant une pluie de météorites ») rend l'événement moins représentatif du hasard. Entre les items 1.7 et 2.7, on peut voir une double modification : d'une part l'ajout en 2.7 d'une expression contextualisante (l'orage), mais d'un autre côté l'ajout d'une précision (la date) en 1.7 qui semble compenser l'effet de contextualisation.

Au final, on peut supposer à partir de ces observations l'existence d'un certain nombre de présupposés, d'intuitions primaires ou secondaires concernant les événements représentatifs du hasard. Ils sont résumés dans le tableau 2. Certains (e.g., complexité et équiprobabilité) sont en conflit, et l'éducation pourrait modifier l'équilibre, favorisant éventuellement un type de conception sur un autre, comme cela a déjà été avancé pour le couple biais d'équiprobabilité/ biais de représentativité (Morsanyi, Handley, & Serpell, 2012).

TABLEAU 2 – *Quelques hypothèses concernant les intuitions et leur type (primaires ou secondaires) concernant des événements « dus au hasard » (EA)*

	Type	Explication
Information	primaire	Un EA a une probabilité faible
Complexité	primaire	Un EA est complexe
Equiprobabilité	secondaire	Tous les patterns possibles sont aussi représentatifs du hasard
Déterminisme 1	primaire	Un EA est non-déterministe
Déterminisme 2	secondaire	Certains algorithmes engendrent des EA
Inexpliqué	primaire	Un EA n'a pas d'explication (de cause)

4 Expérience 2 : quatre erreurs fréquentes

4.1 Méthode

Les 53 sujets ont effectué une série de quatre tâches présentées dans le tableau 3, au début de la formation sur le hasard. Comme le montre le tableau, chacune de ces tâches renvoie à un des biais que nous avons choisi d'étudier. Ces exercices ont été réalisés sur papier, en dix minutes, dans la salle où avait lieu la formation.

Les participants ont été contactés à nouveau 6 mois après la formation par mail pour réaliser à nouveau ces mêmes exercices. Les questionnaires étaient remplis sur Internet. 18 participants ont réalisé cette seconde série (soit un taux d'acceptation de 34%, ce qui est assez élevé), tous en moins de 8 minutes (temps de connexion).

A chaque question, les participants devaient cocher une et une seule réponse.

TABLEAU 3 – *Les quatre items de l'expérience 2. Les réponses en italiques sont les erreurs attendues (correspondant au biais étudié). Les réponses soulignées sont correctes.*

Item	Réponses possibles	Biais étudié
On lance trois dés équilibrés en même temps. En rangeant les trois valeurs obtenues par ordre décroissant, on forme un nombre de trois chiffres, par exemple 421 (si on a tiré un 4, un 1 et un 2) ou 422 (si on a tiré un 4 et deux 2).	*(1) probabilités égales* ; <u>(2) 421 plus probable</u> ; (3) 422 plus probable ; (4) on ne peut pas savoir.	Équiprobabilité
Un jeu de pile ou face fonctionne de la manière suivante : on parie une mise X (au choix) sur le résultat du tirage (pile ou face). Si on a deviné juste, on empoche le double de sa mise. Dans le cas contraire, on perd la mise. On peut parier autant de fois qu'on le souhaite. On dira qu'on a	*(1) aucune méthode ne peut donner une probabilité de gain de plus de 50%* ; <u>(2) il existe une méthode donnant une probabilité de gain de plus de 90% ;</u> (3) il existe une méthode donnant une probabilité de gain dépassant 50%, mais pas 90% ;	Heuristique d'invincibilité (moyenne/proba)

« gagné » si on a plus d'argent à la fin du jeu qu'au début.	(4) je ne sais pas.	
Une maladie M touche 1 personne sur 100. Il existe un test fiable à 90% pour détecter la maladie : le test sera positif pour 90% des malades, et négatif pour 90% des personnes n'ayant pas la maladie[4]. On choisit une personne au hasard dans la population générale et on lui fait passer le test. Il est positif. La probabilité que cette personne soit atteinte de la maladie M est…	(1) comprise entre 0 et 0,5 ;(2) comprise entre 0,5 et 0,9 ; (3) supérieure ou égale à 0,9.	Négligence du taux de base
Une urne contient 4 boules rouges, 2 blanches et une noire. On tire une première boule de l'urne. Sans la replacer, on tire une seconde boule. Cette seconde boule est blanche. La probabilité que la première boule soit blanche est	(1) inférieure à la probablité qu'elle soit noire ; (2) égale à la probabilité qu'elle soit noire ; *(3) supérieure à la probabilité qu'elle soit noire* ; (4) on ne peut pas savoir.	Phénomène Falk

4.2 Résultats

La figure 1 représente les données obtenues. On peut observer des taux de réponses incorrectes conformes au biais étudié (colonne 1) similaires aux résultats des études antérieures : 58,5% pour l'équiprobabilité ; 37,7% pour l'invincibilité ; 60,4% (en cumulant les deux types d'erreur) pour la négligence du taux de base ; et 56,6% pour le phénomène Falk.

[4] Cette hypothèse, qui n'est peut-être pas réaliste d'un point de vue médical, est régulièrement utilisée dans ce type d'expériences. Elle rend la question plus simple à traiter, et réduit ainsi les risques d'erreurs qui seraient dues à une accumulation trop importante de données à retenir. Contrairement au cas du problème historique des taxis bleus et verts, l'égalité des probabilités est ici explicite, et il n'est donc pas besoin d'attendre que cette égalité soit supposée par les participants.

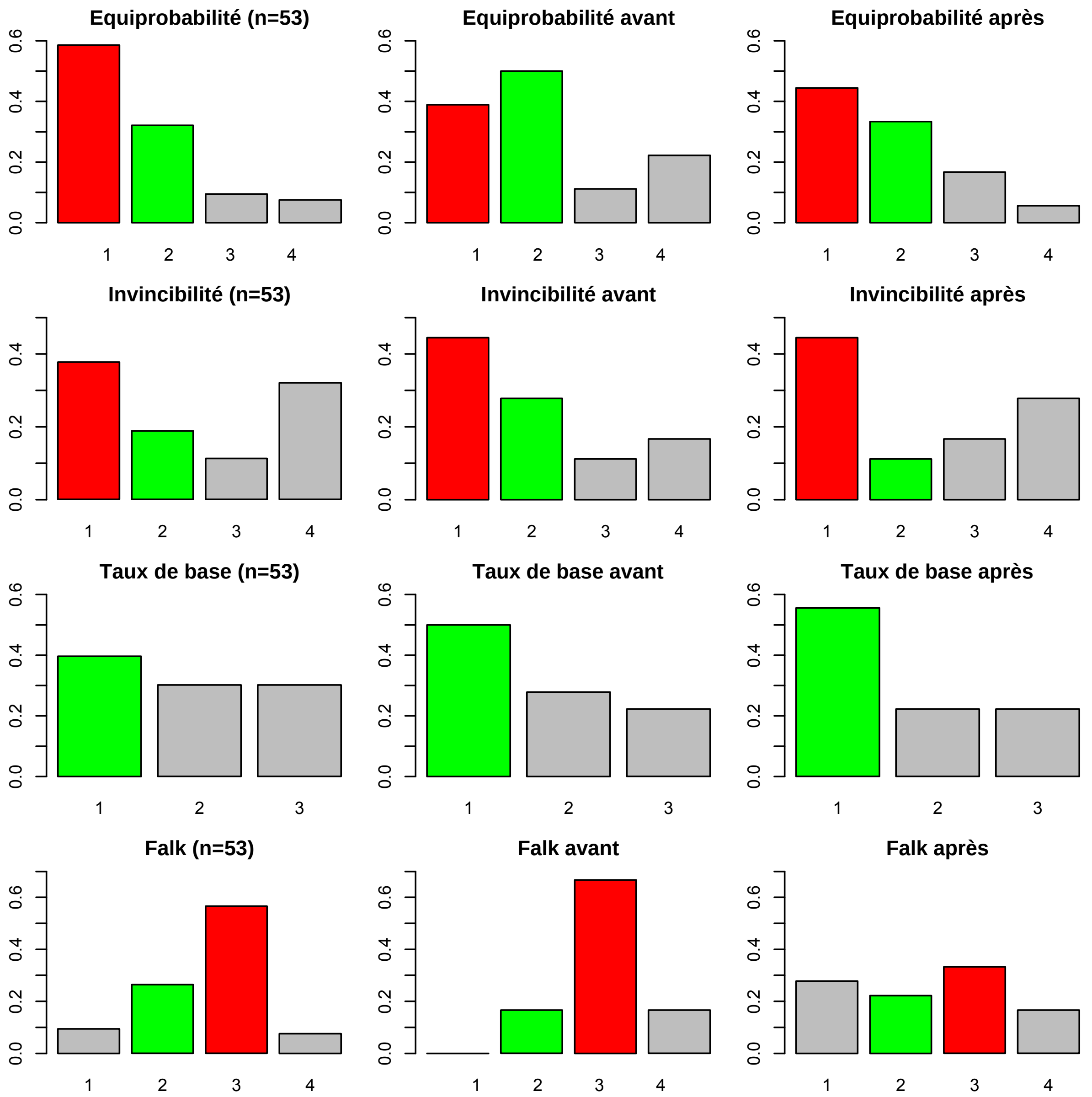

FIGURE 1 – *Distribution des réponses des participants. Chaque ligne correspond à un item (ou biais) indiqué en titre de graphique. Les graphes de la première colonne donnent les résultats des 53 participants avant la formation. Ceux de la seconde colonne représentent les réponses* avant la formation *données par les 18 sujets qui ont répondu au post-test. Ceux de la troisième colonne, à comparer avec les précédentes pour visualiser l'évolution, représentent les 18 réponses fournies 6 mois après la formation. Les barres en vert représentent les bonnes réponses, celles en rouge les erreurs attendues selon le biais étudié. Les barres en gris représentent les autres modalités, « je ne sais pas / on ne peut pas savoir » ou les réponses incorrectes intermédiaires.*

4.3 Discussion

Commençons par signaler que les réponses d'ignorance sont tantôt libellées « je ne sais pas » et tantôt « on ne peut pas savoir », ce qui est évidemment bien différent. Cela limite les comparaisons possibles dans les taux de réponses d'ignorance d'un item à l'autre, mais n'empêche pas, en revanche, d'étudier les évolutions, pour un même item, des réponses. Le temps de réponse des sujets n'était pas limité, mais on leur demandait pourtant d'être assez rapides. Cette condition expérimentale a pu augmenter le taux de réponses d'ignorance.

Les biais étudiés ici sont largement représentés chez les enseignants du secondaire, tant avant qu'après la formation.

En ce qui concerne l'évolution attribuable à la formation, il faut comparer les colonnes 2 et 3 de la figure 1, qui portent sur les mêmes individus.

Des études antérieures avaient déjà montré que le biais d'équiprobabilité a tendance, dès le début du niveau universitaire, à augmenter avec l'éducation en probabilités. Cela est confirmé ici, même sur une formation de deux jours seulement et à 6 mois d'intervalle : le taux de réponses incorrectes équiprobables augmente, tandis que celui de bonnes réponses diminue.

En ce qui concerne l'heuristique d'invincibilité du hasard, on constate un tableau paradoxal : alors que les réponses correctes diminuent en nombre, cela n'augmente pas pour autant le nombre d'erreurs du type attendu (qui stagne). Les réponses correctes « manquantes » sont remplacées par des réponses prudentes : soit « je ne sais pas », soit la réponse incorrecte intermédiaire qu'il existe des méthodes permettant de battre le hasard avec une probabilité comprise en 0,5 et 0,9 seulement. S'il s'agit bien d'un effet de la formation, on peut supposer que ce tableau est le résultat de l'apprentissage d'un rapport plus prudent aux probabilités, et non d'un contenu probabiliste spécifique.

En ce qui concerne la négligence du taux de base et le phénomène Falk, l'effet de la formation semble positif, avec une augmentation des bonnes réponses et une baisse des réponses incorrectes du type attendu.

5 Conclusion

Même les enseignants en fonction, formés en probabilités, continuent à tenir des intuitions ou à utiliser des heuristiques trompeuses, qui mènent parfois à des erreurs. L'échec – au moins pour certains biais – de l'enseignement, l'effet parfois délétère même de la formation (notamment sur l'heuristique d'équiprobabilité) suggère que ces formations ne font qu'adosser aux intuitions primaires trompeuses des intuitions secondaires qui peuvent se révéler trompeuses elles aussi, mais qui apportent surtout la confusion en créant des contradictions internes.

L'éducation aux probabilités dispensée aux futurs enseignants, telle qu'elle est envisagée aujourd'hui, ne remplit pas complètement son rôle. La formation de deux jours étudiée ici n'avait pas vocation à résoudre ces problèmes, et d'autres formes de stages plus ciblées devront être testées pour savoir si un tel dispositif est susceptible de corriger les biais que nous avons évoqués.

Une piste possible est donnée par des résultats suggestifs selon lesquelles une formation attaquant frontalement les biais est finalement efficace – il s'agit de formation où les biais sont nommés et démontrés faux. Des stages axés sur le formalisme mathématique et la modélisation (un des points clés dans les problèmes de probabilités) existent. Des études ultérieures devront trancher la question de leur efficacité.

Une autre piste est celle d'un enseignement sur le *hasard*. La formation des enseignants n'évoque jamais la question de la définition du hasard. La théorie de la complexité algorithmique, qui donne une assise à l'intuition qu'une suite aléatoire doit être complexe, ou encore la théorie de l'information qui permet de justifier qu'un événement probable n'est pas « aussi aléatoire » qu'un autre moins probable, ne sont pas abordées. Cette approche théorique pourrait pourtant peut-être fournir aux futurs enseignants une base d'analyse réflexive de leurs intuitions.

Une troisième piste enfin, que suggèrent les résultats obtenus ici à l'item portant sur l'heuristique d'invincibilité, consisterait en un apprentissage de plus haut niveau, visant non à améliorer les intuitions, mais à améliorer le choix d'une procédure dans une approche de double voie. Les théories de la double voie supposent que deux modes de raisonnement sont possibles : le premier, rapide, se fonde sur des heuristiques et intuitions ; le second, lent, passe par un raisonnement mathématique.

L'expérience 2 présentée ici favorise l'utilisation de la voie rapide, parce qu'il était demandé aux participants de répondre en quelques minutes et que les dispositions matérielles ne prévoyaient pas de quoi travailler longuement à l'écrit. L'augmentation à l'item 2 du nombre de réponses « je ne sais pas » pourrait correspondre à une augmentation du sentiment que le problème ne doit pas être résolu par la voie rapide (sans que celle-ci soit mise en branle pour autant, à cause de la contrainte de temps). Si tel est le cas, on peut espérer que la formation a finalement eu sur cet exemple l'effet positif de *faire douter* les participants de leurs intuitions, une qualité connue pour être importante dans le développement de l'esprit critique (Cassotti & Moutier, 2010).

Références

[1] Bar-Hillel, M. (1980), The base-rate fallacy in probability judgments, *Acta Psychologica*, **44**(3) , 211-233.

[2] Batanero, C. and E. Sanchez (2005), What is the nature of high school students' conceptions and misconceptions about probability? Dans Jones, G. A., *Exploring probability in school: Challenges for teaching and learning* (Vol. 40, pp. 241-266), Mathematics Education Library.

[3] Batanero, C., J. D. Godino, and R. Roa (2004), Training teachers to teach probability. *Journal of Statistics Education*, **12**(1), *www.amstat.org/publications/jse/v12n1/batanero.html*.

[4] Birnbaum, M. H. (1983), Base rates in bayesian inference: Signal detection analysis of the cab problem, *The American Journal of Psychology*, **96**(1), 85-94.

[5] Callaert, H. (2004), In Search of the Specificity and the Identifiability of Stochastic Thinking and reasoning, *Proceedings of the Third Conference of the European Society for Research in Mathematics Education*, Center for Statistics.

[6] Casscells, W., A. B. Schoenberger, and T. Graboys (1978), Interpretation by Physicians of Clinical Laboratory Results, *New England Journal of Medecine*, **299**, 999-1001.

[7] Cassotti, M. and S. Moutier (2010), How to explain receptivity to conjunction-fallacy inhibition training: Evidence from the Iowa Gambling Task, *Brain and Cognition*, **72**, 378–384.

[8] Cohen, J. (1992), A power primer, *Psychological Bulletin,* **112**(1), 155-159.

[9] Delahaye, J.-P. (1999), *Information, complexité et hasard*, Éditions Hermès, Paris.

[10] Delahaye, J.-P. (1998), Les martingales et autres illusions, *Pour la Science,* **251**, 100-103.

[11] Delahaye, J.-P. and H. Zenil (2012), Numerical evaluation of algorithmic complexity for short strings: A glance into the innermost structure of randomness, *Applied Mathematics and Computation,* **219**(1), 63-77.

[12] Falk, R. and A. Lann (2008), The allure of equality: Uniformity in probabilistic and statistical judgment, *Cognitive Psychology,* **57**(4), 293-334.

[13] Fischbein, E. (1987), *Intuition in science and mathematics*, Reidel.

[14] Gras, R. and A. Totohasina (1995), Chronologie et causalité, conceptions sources d'obstacles épistémologiques a la notion de probabilité conditionnelle, *Recherches en Didactique des Mathématiques,* **15**(1), 49-95.

[15] Lahanier-Reuter, D. (1999), Conceptions du hasard et enseignement des probabilités et statistique, PUF, Paris.

[16] Lecoutre, M.-P. (1992), Cognitive models and problem spaces in “purely random” situations, *Educational Studies in Mathematics,* **23**(6), 557-568.

[17] Lecoutre, M.-P. (1985), Effet d'informations de nature combinatoire et de nature fréquentielle sur les jugements probabilistes, *Recherche en Didactique des Mathématiques*, **6**, 193-213.

[18] Morsanyi, K., S. J. Handley, and S. Serpell (2012), Making heads or tails of probability: An experiment with random generators, *British Journal of Educational Psychology*.

[19] Morsanyi, K., C. Primi, F. Chiesi, and S. J. Handley (2009), The Effects and Side-Effects of Statistics Education: Psychology Students' (Mis-)Conceptions of Probability, *Contemporary Educational Psychology,* **34**(3), 210-220.

[20] Nickerson, R. (2002), The production and perception of randomness, *Psychological Review,* **109**(2), 330-357.

[21] Pratt, D. (2000), Making sense of the total of two dice, *Journal for Research in Mathematics Education,* **31**(5), 602-625.

[22] Thompson, W. C. and E. L. Schumann (1987), Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy, *Law and Human Behavior,* **11**(3), 167-187.

[23] Tversky, A. and D. Kahneman (1974), Judgment under uncertainty: Heuristics and biases, *Science,* **185**, 1124–1131.

[24] Zawojewski, J. S. and J. M. Shaughnessy (2000), Data and chance. Dans Silver, E. A. and P. A. Kenney, *Results from the seventh mathematics assessment of the National Assessment of Educational Progress* (pp. 235-268), National Council of Teachers of Mathematics, Reston, VA.