

Nombres premiers et mouvement brownien
P. Billingsley, Chicago

Parce qu'il se factorise en un produit de nombres premiers, chaque entier contient en lui une sorte de trajectoire de mouvement brownien, et les mathématiques du mouvement brownien peuvent être utilisées pour en déduire des résultats sur la factorisation. Je suis conscient de la notion ancienne et lamentablement persistante selon laquelle un théorème énoncé en langage probabiliste est un peu moins vrai qu'il ne le serait autrement. Je vais néanmoins énoncer ces théorèmes en langage probabiliste et même en donner des preuves probabilistes, les rendant ainsi totalement faux aux yeux de ceux qui adhèrent à l'ancienne vision dont je parle.

Il y aura de toute façon peu de preuves, car pour l'essentiel je ne ferai qu'illustrer des théorèmes généraux par des exemples et des cas particuliers. Pour cela, j'ai l'autorité de William Feller. Lorsque j'étais son étudiant à Princeton, il avait l'habitude de dire que le meilleur en mathématiques, comme dans l'art, les lettres et tout le reste - que le meilleur consiste en le général incarné dans le concret. Bien que j'aie longtemps pris cela pour un sentiment antilittéraire, j'ai fini par le comprendre comme le principe intellectuel et esthétique qu'il entendait, et j'ai essayé depuis de le garder quelque part près du devant de mon esprit.

1. Mouvement brownien

Imaginez une particule en suspension dans un fluide, bombardée par des molécules se déplaçant à leur manière habituelle. La particule effectuera ce mouvement régulier et apparemment aléatoire décrit pour la première fois par Robert Brown en 1828. Comme nous ne nous intéresserons qu'à une seule composante de ce mouvement, imaginons-le projeté sur un axe vertical : à chaque instant t , nous notons la hauteur $x(t)$ de la particule au-dessus d'un plan horizontal fixe. Sur une unité de temps, le mouvement de la particule, que nous supposons partir de 0, est décrit par les positions $x(t)$ pour $0 \leq t \leq 1$ - c'est-à-dire par une fonction continue x sur $[0, 1]$ avec $x(0) = 0$. Cela nous conduit à considérer la collection $C = C_0[0, 1]$ de telles fonctions x .

Pour des raisons techniques, nous faisons de C un espace métrique en prenant la distance entre deux de ses éléments comme la distance verticale maximale entre leurs graphes. Cette topologie, appelée topologie uniforme, n'est pas directement notre propos ici ; elle est introduite principalement comme preuve que la discussion qui suit a bien une base mathématique.

Conférence Rouse Ball de l'année 1972.

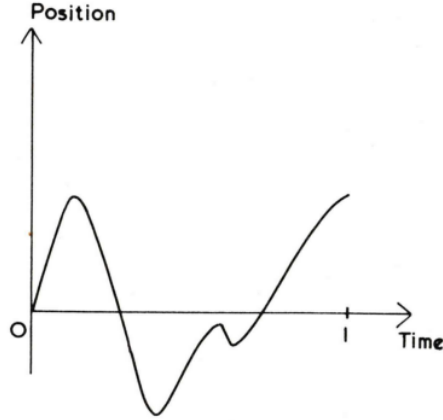


Figure 1

Le mouvement aléatoire de la particule est décrit par une attribution de probabilités $P(A)$ aux sous-ensembles A de C ; $P(A)$ représente la chance que la trajectoire tracée par la particule se trouve dans A , ou soit représentée par une fonction x qui se trouve dans A .

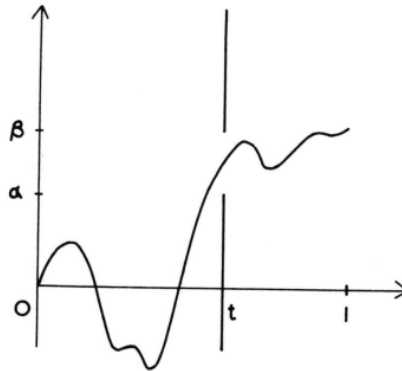


Figure 2

L'ensemble $[x : \alpha \leq x(t) \leq \beta]$, constitué des trajectoires qui passent par la porte de la Figure 2, représente l'événement qu'à l'instant t , la particule se situera entre α et β ; on lui attribue la probabilité

$$P[x : \alpha \leq x(t) \leq \beta] = \frac{1}{\sqrt{2\pi t}} \int_{\alpha}^{\beta} e^{-u^2/2t} du. \quad (1)$$

Ainsi, la distribution de la position au temps t suit la courbe gaussienne de moyenne 0 et de variance t . Que la moyenne soit 0 reflète le fait que la particule est aussi susceptible de monter que de descendre : il n'y a pas de dérive. La variance t croît linéairement : cela reflète le fait que la particule tend à s'éloigner de son point de départ et, ce faisant, ne subit aucune force tendant à la ramener à ce point de départ. L'équation (1) peut être étendue pour décrire la distribution de l'accroissement sur n'importe quel intervalle de temps.

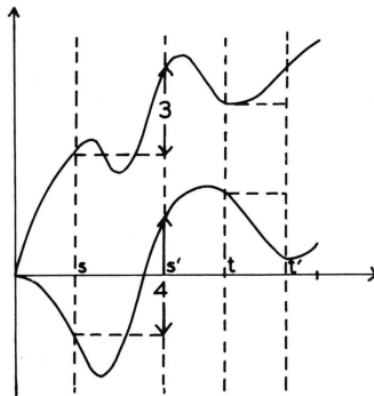


Figure 3

L'autre propriété importante du mouvement brownien est celle-ci : supposons que $s' < t < t'$ et considérons l'événement $A = [x : x(s') - x(s) \geq 3]$, qui est qu'à l'instant t , la particule subit un déplacement vers le haut d'au moins 3 unités pendant l'intervalle de temps $[s, s']$, ainsi que l'événement $B = [x : x(t') - x(t) < 0]$ que la particule subit un déplacement vers le bas pendant l'intervalle de temps $[t, t']$. Par exemple, la trajectoire supérieure de la Figure 3 se trouve dans A mais pas dans B , et la trajectoire inférieure se trouve à la fois dans A et dans B . Les probabilités de A et B et de leur intersection $A \cap B$ sont liées par

$$P(A \cap B) = P(A)P(B). \quad (2)$$

À l'instant t , A et B satisfont la définition de l'indépendance, ce qui reflète le fait que le déplacement subi par la particule pendant $[s, s']$ n'influence en rien le déplacement qu'elle subit pendant $[t, t']$. Cela implique une sorte d'absence de mémoire : bien que le comportement futur de la particule dépende de sa position actuelle, il ne dépend pas de la façon dont la particule y est arrivée. L'équation (2) a une forme plus générale montrant que les accroissements sur un nombre quelconque d'intervalles disjoints sont indépendants.

Les équations (1) et (2), ainsi que leurs versions généralisées, déterminent toutes les probabilités $P(A)$. (Cela ignore un point technique : $P(A)$ ne peut pas être défini pour tout sous-ensemble A de C , mais il peut l'être pour tout ensemble borélien A , c'est-à-dire pour tout A dans la tribu engendrée par les ensembles ouverts dans la topologie uniforme.) Ce fut l'une des réalisations de Norbert Wiener de prouver qu'il existe bien une attribution de probabilités satisfaisant ces règles, et P (la mesure correspondante sur les boréliens) est appelée en conséquence mesure de Wiener.

La mesure de Wiener, établie pour refléter par (1) et (2) les caractéristiques du mouvement brownien, a des propriétés surprenantes. Si A est l'ensemble des trajectoires à variation non bornée, alors $P(A) = 1$. À ce stade, les physiciens perdent tout intérêt, en raison de leur souci obsessionnel de la réalité, ainsi que du fait qu'une trajectoire à variation non bornée représente le mouvement d'une particule qui, dans ses errances d'avant en arrière, parcourt une distance infinie. Mais le fait est mathématiquement intéressant, et il en va de même du fait que $P(A) = 1$ si A est l'ensemble des x dans C qui ne sont dérivables nulle part. Construire une fonction continue nulle part dérivable est difficile, mais tirer un élément de C au hasard selon P produit une telle fonction de probabilité 1.

Dans ce qui suit, nous nous intéresserons à des ensembles qui correspondent de plus près à la réalité. Supposons $\alpha \geq 0$ et considérons l'événement $[x : \max x(t) \geq \alpha]$ que la particule atteigne la hauteur α à un moment t avec $0 \leq t \leq 1$. D'abord,

$$P[x : \max x(t) \geq \alpha] = P[x : \max x(t) \geq \alpha \& x(1) \geq \alpha] + P[x : \max x(t) \geq \alpha \& x(1) \leq \alpha].$$

Les deux probabilités de droite ici peuvent être prouvées égales, grossièrement parce qu'une fois que la particule atteint la hauteur α , elle est aussi susceptible de s'égarer vers le haut et de finir au-dessus de α au temps 1 que de s'égarer vers le bas et de finir en dessous de α . Ainsi

$$P[x : \max x(t) \geq \alpha] = 2P[x : \max x(t) \geq \alpha \& x(1) \geq \alpha].$$

Comme la condition $\max x(t) \geq \alpha$ est superflue en présence de la condition $x(1) \geq \alpha$, le membre de droite ici est $2P[x : x(1) \geq \alpha]$, et (1) avec $t = 1$, $\alpha \geq 0$, et $\beta = \infty$ implique maintenant

$$P[x : \max x(t) \geq \alpha] = \frac{2}{\sqrt{2\pi}} \int_{\alpha}^{\infty} e^{-u^2/2} du. \quad (3)$$

Bien que sa rigueur demande un certain effort, cette dérivation de (3) est intuitive. Le résultat suivant sera énoncé sans aucune preuve. Comme beaucoup d'assertions *ex cathedra*, il va à l'encontre de l'intuition. Considérons l'ensemble $\{t : x(t) > 0\}$ des instants $t, 0 \leq t \leq 1$, pour lesquels la particule est au-dessus de 0. Cet ensemble est une réunion d'intervalles (infiniment nombreux, contrairement à la Figure 4).

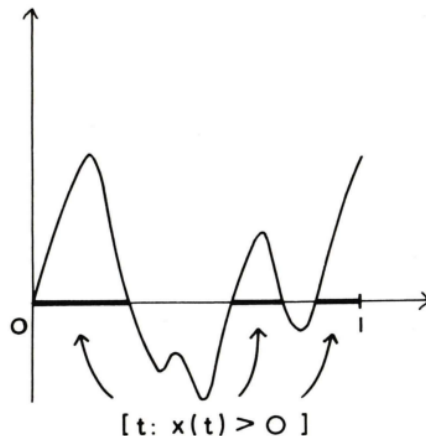


Figure 4

Désignons par des barres la mesure de Lebesgue de cet ensemble, la somme des longueurs des intervalles constitutifs : $||[t : x(t) > 0]||$. La distribution de cette quantité, le temps total passé au-dessus de 0, est donnée par

$$P[x : \alpha \leq ||[t : x(t) > 0]|| \leq \beta] = \frac{1}{\pi} \int_{\alpha}^{\beta} \frac{du}{\sqrt{u(1-u)}} \quad (4)$$

$\alpha \leq \beta \leq 1$. C'est la loi de l'arc sinus de Paul Lévy, ainsi appelée parce que l'intégration fait intervenir la fonction arc sinus.

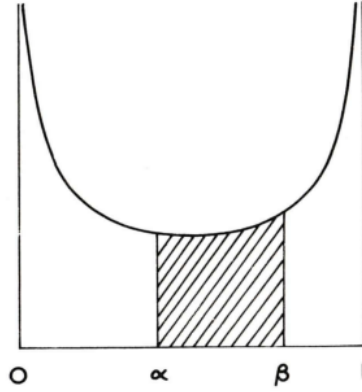


Figure 5

La Figure 5 montre la forme de la densité, l'aire de la région ombrée représentant le membre de droite de (4). La courbe est en forme de U, de sorte que si la longueur $\beta - \alpha$ de l'intervalle est fixée, la probabilité augmente à mesure que l'intervalle se rapproche de 0 ou de 1, étant minimale lorsque l'intervalle est centré sur $1/2$. C'est étrange car le temps passé au-dessus de 0 a une moyenne de $1/2$ par symétrie, et ordinairement les valeurs proches de la moyenne d'une quantité aléatoire sont plus susceptibles de se produire que les valeurs éloignées de la moyenne, alors qu'ici c'est juste le contraire.

2. Marche aléatoire

Imaginez une particule se déplaçant au hasard sur les nœuds d'un réseau cubique. La particule peut se déplacer dans l'une des six directions (Nord, Sud, Est, Ouest, Haut, Bas) vers un nœud adjacent. La direction est déterminée par le lancer d'un dé équilibré, la particule se déplace vers le nœud suivant, et le dé est relancé une fois de plus pour déterminer la direction du mouvement suivant, et ainsi de suite.

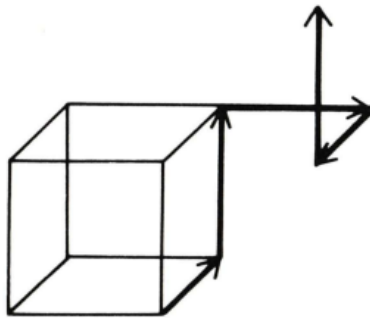


Figure 6

La Figure 6 montre cinq étapes d'une telle marche aléatoire, ainsi qu'une des cellules du réseau cubique. La figure est dans l'esprit d'un vénérable livre d'analyse vectorielle qui commençait une preuve du théorème de Gauss en enjoignant au lecteur de considérer "un élément infinitésimal de

volume de dimensions dx , dy et dz ". Cette injonction était accompagnée d'un diagramme soigneusement étiqueté comme la Figure 7, dont on disait qu'il montrait un tel élément infinitésimal de volume "beaucoup agrandi". Eh bien, la Figure 6 est elle aussi beaucoup agrandie, et si les cubes du réseau sont vraiment très petits et que la particule se déplace très rapidement d'un nœud à l'autre, il est naturel de s'attendre à ce que le mouvement approche le mouvement brownien.

Nous allons explorer une version unidimensionnelle de cette idée. Considérons un axe vertical avec les points entiers $0, \pm 1, \pm 2, \dots$ marqués. Nous partons de 0, lançons une pièce de monnaie, et montons d'une unité si la pièce tombe sur face et descendons d'une unité si elle tombe sur pile. Dans la nouvelle position ($+1$ ou -1), nous lançons à nouveau la pièce et montons ou descendons d'une unité selon qu'elle tombe sur face ou pile, et nous continuons ainsi pendant k étapes. Si nous prenons une unité de temps pour exécuter chaque étape de la marche aléatoire et procédons à un rythme uniforme d'un nœud au suivant, notre progrès est décrit par une fonction comme celle de la Figure 8, un chemin polygonal dont la hauteur au-dessus de i est la position à i - c'est-à-dire la position après la i -ème étape.

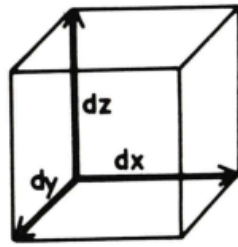


Figure 7

Ce chemin peut aussi être vu comme décrivant les fluctuations de la fortune d'un joueur. La position sur l'axe vertical représente la fortune du joueur (par rapport à son capital initial, de sorte qu'il commence conventionnellement à 0), et elle monte ou descend d'une unité - disons une livre - selon qu'il gagne ou perd le prochain coup.

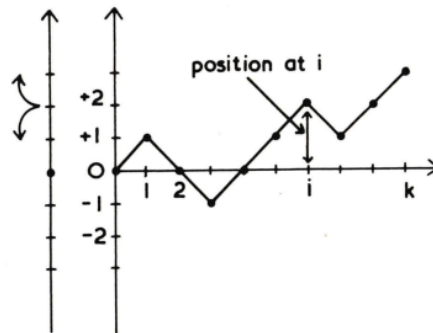


Figure 8

Le chemin de la marche aléatoire a certaines des propriétés du mouvement brownien. En premier lieu, si $i < i' < j < j'$, les déplacements subis sur les intervalles de temps $[i, i']$ et $[j, j']$ sont indépendants parce qu'ils dépendent de différents ensembles de lancers et que les lancers sont supposés indépendants (la pièce n'a pas de mémoire). Ainsi le chemin a essentiellement des accroissements indépendants (pour les intervalles dont les extrémités ne sont pas entières, les accroissements sont supposés indépendants). La distance parcourue en une étape a pour moyenne

$$\frac{1}{2}(+1) + \frac{1}{2}(-1) = 0 \quad (5)$$

et pour variance

$$\frac{1}{2}(+1)^2 + \frac{1}{2}(-1)^2 = 1, \quad (6)$$

et donc la position en i a pour moyenne 0 et, par indépendance, pour variance i , une autre propriété du mouvement brownien (voir équation (1)). Bien que le caractère polygonal de la trajectoire ne soit pas partagé par le mouvement brownien, une contraction appropriée des échelles de temps fera disparaître les segments de droite de la Figure 8 à la limite lorsque $k \rightarrow \infty$.

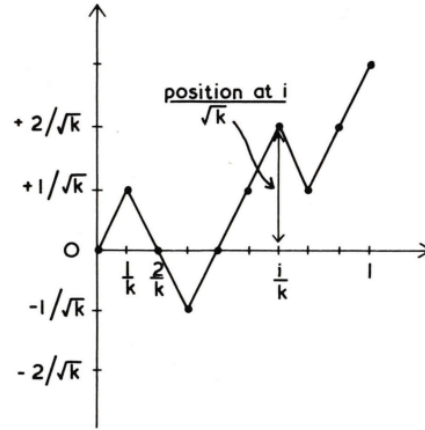


Figure 9

Supposons que nous contractions l'échelle de temps d'un facteur k et l'échelle verticale d'un facteur $\sqrt{k}$, passant de la Figure 8 à la Figure 9. Cela préserve l'indépendance essentielle des accroissements. La hauteur de la nouvelle trajectoire au-dessus du point i/k , étant $1/\sqrt{k}$ fois la hauteur de l'ancienne position après l'étape i , a pour moyenne 0 ; sa variance est $(1/\sqrt{k})^2$ fois l'ancienne variance i et vaut donc i/k . Contracter le temps d'un facteur k donne une trajectoire définie sur $[0, 1]$, en d'autres termes un élément de C . Le re-dimensionnement vertical par $1/\sqrt{k}$ fait que les variances s'ajustent correctement.

La trajectoire aléatoire de la Figure 9 possède les propriétés essentielles du mouvement brownien. Comme lorsque $k \rightarrow \infty$ les liens du polygone disparaissent, on peut espérer avoir un mouvement brownien à la limite. Et en effet

$$\text{Prob} [\text{trajectoire} \in A] \rightarrow P(A) \quad (k \rightarrow \infty) \quad (7)$$

pour des sous-ensembles A de l'espace C , où $P(A)$ est la mesure de Wiener. Il y a 2^k chemins comme celui de la Figure 9, et $\text{Prob}[\text{chemin} \in A]$ est 2^{-k} fois le nombre d'entre eux qui se trouvent

dans A .

Supposons que A dans (7) soit l'ensemble $[x : \alpha \leq x(1) \leq \beta]$ des chemins dans C qui au point $t = 1$ ont une hauteur comprise entre α et β . Puisque la hauteur au-dessus de $t = 1$ dans la Figure 9 est $1/\sqrt{k}$ fois la position à k dans la marche aléatoire originale, (7) et (1) impliquent ensemble

$$\text{Prob} \left[\alpha \leq \frac{\text{position à } k}{\sqrt{k}} \leq \beta \right] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du. \quad (8)$$

Ainsi, le théorème central limite de De Moivre-Laplace pour les essais de Bernoulli décrit la position après un grand nombre d'étapes dans une marche aléatoire, ou la fortune du joueur à la fin d'une soirée de jeu de k étapes. Si $\alpha = \beta = 0.9$, l'intégrale dans (8) est d'environ 0.6. Si $k = 100$, le joueur a donc une probabilité d'environ 0.66 de terminer la soirée avec $9 \times \sqrt{100} = 9$ livres de son capital initial.

Supposons maintenant que A soit l'ensemble dans (3), l'ensemble des chemins dans C ayant quelque part une hauteur de α ou plus (ici $\alpha \geq 0$). Le chemin de la Figure 9 se trouve dans A si à un moment donné pendant la soirée de jeu la fortune du joueur est d'au moins $\alpha\sqrt{k}$ livres au-dessus de son capital initial, et par (7) cela converge vers l'intégrale dans (3). L'intégrale est d'environ 0.1 pour $\alpha = 1.7$. Avec $k = 100$, cela donne une probabilité approximative de 0.1 que le joueur ait été au moins 17 livres en avance au moment où il aurait dû s'arrêter.

Enfin, supposons que A soit l'ensemble $[x : \alpha \leq |[t : x(t) > 0]| \leq \beta]$ dans (4). Pendant la soirée, le joueur est en avance une certaine fraction du temps ; si la courbe de la Figure 9 représente l'histoire de sa fortune, elle est dans l'ensemble A si cette fraction se situe entre α et β . La chance de cet événement est donnée par (7) et (4) à peu près égale à l'aire de la région ombrée de la Figure 5. La chance que le joueur soit en avance plus de 90 % du temps s'avère être d'environ 0.2, tandis que la chance qu'il soit en avance entre 45 % et 55 % du temps n'est que d'environ 0.06. Une soirée sur cinq, le joueur sera en avance plus de 90 % de cette soirée de jeu. Par symétrie, une soirée sur cinq, le joueur sera en avance moins de 10 % de cette soirée de jeu. Pour le convaincre dans le premier [second] cas que son expérience est due simplement au hasard et non au fait qu'il soit le favori de la Fortune [le fou de la Fortune] sera difficile [impossible].

Nous avons appliqué (7) à trois ensembles intéressants A . Si A est l'ensemble des fonctions continues à variation non bornée, alors $P(A) = 1$, comme expliqué ci-dessus, tandis que $\text{Prob}[\text{chemin} \in A] = 0$ parce que la courbe de la Figure 9 est visiblement à variation bornée. Ainsi (7) échoue pour certains sous-ensembles A de C . Le fait mathématique est que (7) est valable pour tout ensemble (borélien) A dont la frontière ∂A (frontière au sens de la topologie uniforme) satisfait $P(\partial A) = 0$, une condition qui est vérifiée dans nos trois applications. La preuve du théorème utilise une combinaison de théorie des probabilités et d'analyse fonctionnelle. (Les détails peuvent être trouvés dans mon livre *Convergence of Probability Measures*).

Après avoir utilisé le mouvement brownien dans certains problèmes de marche aléatoire et de jeu, nous allons maintenant utiliser le mouvement brownien, la marche aléatoire et le jeu dans certains problèmes d'arithmétique multiplicative.

Diviseurs premiers

Pour un entier positif n , soit $f(n)$ le nombre de facteurs premiers distincts qu'il contient. Nous ne comptons pas la multiplicité : $f(3^4 \cdot 5^2)$ vaut 2, et non 6. Le tableau montre quelques valeurs de la fonction f . Elle croît lentement.

n	2	3	4	5	6	7	...	29	30	31	...	209	210	211	...
$f(n)$	1	1	1	1	2	1	...	1	3	1	...	2	4	1	...

les plus petits n ayant respectivement pour f -valeurs 2, 3 et 4 sont $2 \cdot 3 = 6$, $2 \cdot 3 \cdot 5 = 30$ et $2 \cdot 3 \cdot 5 \cdot 7 = 210$. Le fait qu'il existe une infinité de nombres premiers implique que f prend des valeurs arbitrairement grandes ; il implique aussi que f retombe infiniment souvent à 1, puisque $f(p) = 1$ pour p premier.

Comme f varie de façon irrégulière, il est naturel de s'interroger sur son comportement moyen. Par exemple, on peut montrer que

$$\frac{1}{N} \sum_{n=1}^N f(n) \sim \log \log N. \quad (9)$$

Comme $\log \log 10^{70} \approx 5$, l'entier typique inférieur à 10^{70} n'a que cinq diviseurs premiers. Des questions plus délicates concernent la distribution de f . Si S est un ensemble d'entiers positifs, soit $P_N(S)$ la fraction parmi les entiers $1, 2, \dots, N$ qui appartiennent à S :

$$P_N(S) = \frac{1}{N} \times \#\{n : 1 \leq n \leq N \text{ \& } n \in S\}. \quad (10)$$

Le problème est maintenant d'obtenir des informations sur des quantités comme

$$P_N[n : a \leq f(n) \leq b].$$

Or (10) peut être considéré comme une probabilité : nous tirons un entier n au hasard dans l'intervalle $1 \leq n \leq N$, et $P_N(S)$ est la probabilité qu'il appartienne à S . Que $P_N[n : a \leq f(n) \leq b]$ puisse être considéré comme une probabilité ne garantit pas, croyez-moi, que la théorie des probabilités aidera à l'évaluation. En fait, elle aide parce que la notion d'indépendance peut être invoquée. Si $\delta_p(n)$ vaut 1 ou 0 selon que le nombre premier p divise n ou non, alors $f(n) = \sum_p \delta_p(n)$. Nous pouvons comprendre la distribution de $f(n)$ si nous comprenons le comportement conjoint des $\delta_p(n)$ en tant que quantités aléatoires.

Le nombre de multiples de p jusqu'à N est la partie entière $[N/p]$ de N/p . La probabilité que $\delta_p(n) = 1$, ou que $p \mid n$, est donc

$$P_N[n : p \mid n] = \frac{1}{N} \left[\frac{N}{p} \right] \approx \frac{1}{p}, \quad (11)$$

l'approximation étant valable pour N grand. Cela reflète le fait que p divise un entier sur p , et cela ne nécessite pas que p soit premier.

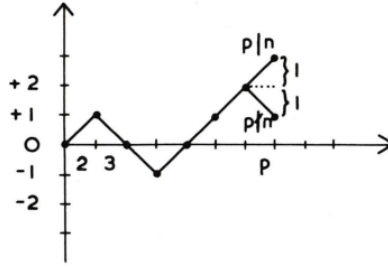


Figure 10

Des nombres premiers distincts p et q divisent individuellement n si et seulement si leur produit pq divise n , de sorte que

$$P_N[n : p \mid n \ \& \ q \mid n] = P_N[n : pq \mid n] \approx \frac{1}{pq} = \frac{1}{p} \cdot \frac{1}{q} \approx P_N[n : p \mid n] \cdot P_N[n : q \mid n]. \quad (12)$$

Ainsi les événements $[n : p \mid n]$ et $[n : q \mid n]$ satisfont approximativement la définition de l'indépendance si n est aléatoire, $1 \leq n \leq N$, avec N grand.

Nous pouvons utiliser ce fait pour construire une sorte de trajectoire de marche aléatoire contenant des informations sur la factorisation en nombres premiers de n et en particulier sur $f(n)$. Sur un axe vertical avec les points entiers marqués, nous partons de 0 et montons d'une unité si $2 \mid n$ et descendons d'une unité si $2 \nmid n$. De notre nouvelle position (+1 ou -1), nous montons d'une unité si $3 \mid n$ et descendons d'une unité si $3 \nmid n$. Nous procédons ainsi, en examinant chaque nombre premier successivement. La Figure 10 décrit cette marche aléatoire de factorisation de la même manière que la Figure 8 décrit la marche aléatoire de lancer de pièce. Chaque nombre sur l'axe du temps est le nombre premier correspondant à cette étape dans la marche aléatoire. Nous considérerons ultérieurement combien de temps continuer la marche.

Puisque n est aléatoire, ce chemin est aléatoire. Mais comme tout le caractère aléatoire réside dans le tirage de n avant que la marche ne commence, la marche aléatoire de factorisation peut sembler moins aléatoire que la marche aléatoire de lancer de pièce. C'est une illusion. Nous pouvons imaginer lancer la pièce k fois avant la marche, enregistrer la séquence de faces et de piles, et seulement ensuite effectuer la marche correspondante. La marche serait ennuyeuse, puisque nous verrions toute son histoire enregistrée avant de partir. Imaginons donc un ami qui lance la pièce k fois, enregistre les résultats à l'avance, et nous révèle les résultats un par un au fur et à mesure que nous exécutons la marche. Cela restaure le suspense. Pour la marche aléatoire de factorisation, nous pouvons imaginer un ami qui tire n au hasard, $1 \leq n \leq N$, factorise n en nombres premiers, et à chaque étape de la marche nous dit si le p correspondant divise n ou non.

L'accroissement du chemin aléatoire de la Figure 10 sur un intervalle dépend du nombre de nombres premiers dans l'ensemble correspondant qui divisent n . Les accroissements sur des intervalles disjoints dépendent d'ensembles disjoints de nombres premiers et donc, par (12), les accroissements seront approximativement indépendants si N est grand. Contrairement au mouvement brownien, cependant, la marche aléatoire de factorisation a une forte dérive vers le bas. Par (11), la chance de descendre est d'environ $1 - 1/p$, ce qui est presque 1 pour p grand. Le remède est de monter d'une

distance $1 - 1/p$ si $p \mid n$ et de descendre d'une distance $1/p$ si $p \nmid n$. La distance espérée parcourue est maintenant approximativement

$$\frac{1}{p} \left(1 - \frac{1}{p}\right) + \left(1 - \frac{1}{p}\right) \left(-\frac{1}{p}\right) = 0.$$

Cela se compare à (5), une équation qui montre que la marche aléatoire de lancer de pièce n'a pas de dérive.

La variance de la distance parcourue est maintenant approximativement

$$\frac{1}{p} \left(1 - \frac{1}{p}\right)^2 + \left(1 - \frac{1}{p}\right) \left(-\frac{1}{p}\right)^2 = \frac{1}{p} \left(1 - \frac{1}{p}\right) \approx \frac{1}{p}.$$

La distance parcourue à l'étape correspondant à p tend donc à être petite pour p grand, contrairement à la marche aléatoire de lancer de pièce, qui, par (6), procède avec une vigueur jamais diminuée. Le remède cette fois est de ne consacrer qu'une quantité $1/p$ de temps à exécuter l'étape correspondant à p . Avec ces deux modifications, le chemin est comme dans la Figure 11. Pour récapituler, sur un intervalle de longueur $1/p$, le chemin monte linéairement d'une quantité $\delta_p(n) - 1/p$ - c'est-à-dire monte de $1 - 1/p$ si $p \mid n$ (approximativement $1/p$) et atteint $0 - 1/p$ si $p \nmid n$ (approximativement probabilité $1 - 1/p$).

Le point T sur la figure (l'extrémité droite de l'intervalle correspondant à p) est $\sum_{q \leq p} 1/q$ (somme sur les nombres premiers q n'excédant pas p), qui peut être montré comme étant très proche de $\log \log p$ (voir (9)).

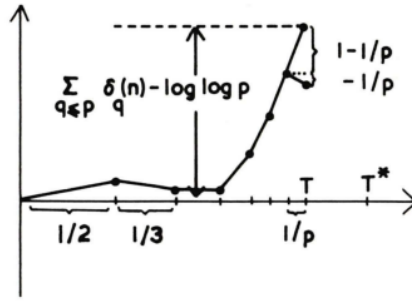


Figure 11

La hauteur de la courbe au-dessus de T est $\sum_{q \leq p} (\delta_q(n) - 1/q) \approx \sum_{q \leq p} \delta_q(n) - \log \log p$. Nous continuons la marche jusqu'à ce que chaque $p \ll N$ ait été traité, et le point correspondant sur l'axe du temps est $T^* = \sum_{p \leq N} 1/p \approx \log \log N$.

Le chemin aléatoire ressemble maintenant à un chemin de lancer de pièce en ce que les accroissements sont presque indépendants (pour N grand), il n'y a essentiellement pas de dérive, et les variances sont à peu près correctes. Comme dans le cas du lancer de pièce, une remise à l'échelle conduira à la limite ($N \rightarrow \infty$) au mouvement brownien. Pour envoyer T^* au point $t = 1$, nous contractons l'échelle horizontale d'un facteur $\log \log N$, et, comme dans le cas du lancer de pièce, nous contractons l'échelle verticale par la racine carrée de ce facteur. Le point T de la Figure 11 va à $\log \log p / \log \log N$, et le chemin est celui montré dans la Figure 12.

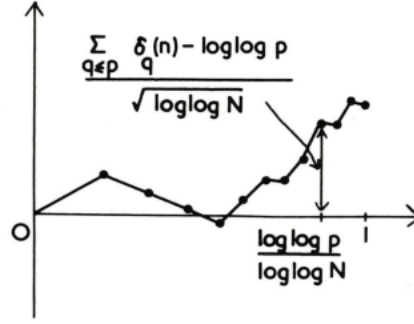


Figure 12

Puisque le chemin dépend de n et N , notons-le $\text{chemin}_N(n)$. Comme n est aléatoire, le chemin l'est aussi, et la chance qu'il se trouve dans un sous-ensemble donné A de l'espace C est $P_N[n : \text{chemin}_N(n) \in A]$. Le théorème reliant les nombres premiers au mouvement brownien est le suivant : Si A est un sous-ensemble (borélien) de C satisfaisant $P(\partial A) = 0$, alors

$$P_N[n : \text{chemin}_N(n) \in A] \rightarrow P(A) \quad (N \rightarrow \infty) \quad (13)$$

où P est la mesure de Wiener. La preuve de (13) utilise une combinaison de théorie des probabilités, d'analyse fonctionnelle et de théorie des nombres. (Le théorème est donné implicitement dans le livre de Kubilius *Probabilistic Methods in the Theory of Numbers*, p. 122, explicitement dans l'une des versions manuscrites de mon *Convergence of Probability Measures*, et sous une forme plus générale dans l'article à paraître de Philipp, *An invariance principle for additive number-theoretic functions*.)

Considérons les trois ensembles A auxquels nous avons appliqué le résultat analogue (7). Que le chemin de la Figure 12 se trouve ou non dans l'ensemble $A = [x : \alpha \leq x(1) \leq \beta]$ dépend de sa hauteur au-dessus de $t = 1$; cette hauteur est la quantité indiquée dans le diagramme avec N à la place de p . Puisque $\sum_{q \leq N} \delta_q(n)$ est le nombre total $f(n)$ de diviseurs premiers distincts de n , (13) ainsi que (1) donne

$$P_N[n : \alpha \leq \frac{f(n) - \log \log N}{\sqrt{\log \log N}} \leq \beta] \rightarrow \frac{1}{\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-u^2/2} du. \quad (14)$$

C'est le théorème central limite d'Erdős-Kac pour f . (Pour un traitement élémentaire de (14), voir *American Mathematical Monthly*, 76 (1969), 132.)

Pour $-\alpha = \beta = 0.9$, l'intégrale dans (14) est d'environ 0.6, et si $N = 10^{70}$, de sorte que $\log \log N \approx 5$, la double inégalité dans (14) est approximativement la même que $0.9 \leq (f(n) - 5)/\sqrt{5} \leq 0.9$, ce qui à son tour est approximativement la même que $3 \leq f(n) \leq 7$. Ainsi environ 60% des entiers inférieurs à 10^{70} ont de 3 à 7 diviseurs premiers.

En appliquant (13) à l'ensemble dans (3), on obtient la distribution approximative du maximum de la courbe de la Figure 12. Pour chaque p ($p \leq N$), nous regardons le nombre $\sum_{q \leq p} \delta_q(n)$ de diviseurs premiers de n qui ne dépassent pas p , le comparons à sa valeur $\log \log p$ pour un n "typique", et calculons la divergence (la différence). L'intégrale dans (3) étant d'environ 0.1 si $\alpha = 1.7$, pour

environ 10% des entiers inférieurs à 10^{70} , le maximum de cette divergence dépasse $1.7 \times \sqrt{5} \approx 3.8$.

Et considérons une fois de plus l'ensemble dans (4). Que la courbe de la Figure 12 se trouve ou non dans cet ensemble dépend du temps qu'elle passe au-dessus de 0. Le segment polygonal correspondant à p a une longueur $p^{-1}/\log \log N$ lorsqu'il est projeté sur l'axe horizontal, et donc le temps passé au-dessus de 0 est essentiellement

$$\Sigma \left[\frac{1}{p \log \log N} : p \leq N \text{ \& } \Sigma_{q \leq p} \delta_q(n) > \log \log p \right], \quad (15)$$

la somme s'étendant sur les p ($p \leq N$) pour lesquels n contient excessivement de facteurs premiers à gauche de p . Pour N grand, la distribution de (15) suit approximativement la courbe de densité de la Figure 5. Pour environ 20% des entiers inférieurs à N , elle dépasse 0.9, et pour environ 6%, elle se situe entre 0.45 et 0.55.

Les facteurs premiers présentent à cet égard le même comportement étrange que les pièces de monnaie. D'une certaine manière, ils sont encore plus étranges. Une quantité peut-être plus naturelle à considérer que (15) est

$$\frac{1}{\pi(N)} \times \#[p : p \leq N \text{ \& } \Sigma_{q \leq p} \delta_q(n) > \log \log p], \quad (16)$$

le nombre de p pour lesquels n contient excessivement de nombres premiers à gauche de p , normalisé par division par $\pi(N)$, le nombre total de nombres premiers impliqués. Pour N grand, parmi les points de rupture du polygone de la Figure 12, la grande majorité sont très proches de 1, avec pour résultat qu'à la limite la distribution de (16) consiste en une masse de 0.5 en 0 et une masse de 0.5 en 1. Si $\epsilon > 0$ et N dépasse un certain N_ϵ , alors (16) est inférieur à ϵ avec une probabilité comprise entre $0.5 - \epsilon$ et $0.5 + \epsilon$ et est supérieur à $1 - \epsilon$ avec une probabilité comprise dans la même plage.

Courbes parallèles

D. R. Woodall

Le concept de parallélisme est généralement réservé aux droites. Mon objectif ici est d'étendre sa définition aux courbes. Mon idée de départ était que des courbes parallèles devraient ressembler à des voies ferrées, ayant partout une normale commune et étant partout à la même distance l'une de l'autre mesurée le long de la normale ; mais d'autres définitions sont possibles. En particulier, on dit souvent que deux courbes sont parallèles si elles peuvent être superposées par une simple translation, et j'examinerai ce cas plus tard. L'autre propriété intrinsèquement différente des droites parallèles semble être celle utilisée par Euclide comme définition, à savoir qu'elles ne se rencontrent pas, quelle que soit la distance à laquelle on les prolonge ; mais cela ne conduit à aucune généralisation intéressante.

Considérons donc deux segments de courbe simples et réguliers AB et CD (Fig. 1). Nous disons que deux segments sont parallèles si toute normale à l'un est aussi une normale à l'autre. On peut montrer aisément que dans ce cas les deux courbes ont le même centre de courbure le long d'une telle normale, et que deux courbes sont parallèles si et seulement si l'une d'elles est le lieu des points situés à une distance fixe de l'autre mesurée le long des normales à celle-ci : c'est-à-dire que les courbes sont à la même distance l'une de l'autre le long de toutes les normales. C'est le seul morceau de vraies mathématiques dans tout l'article, aussi je suis la loi bien connue et établie de le laisser en exercice au lecteur !

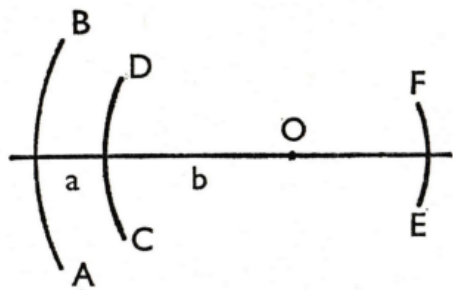


FIG. 1

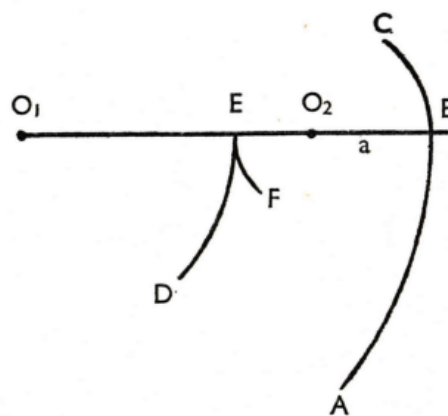


FIG. 2

Comme le montre la Fig. 1, on constatera parfois (mais pas sur les chemins de fer) que le centre de courbure se trouve entre les courbes. Il est commode de distinguer ce cas et de l'appeler parallélisme négatif, le parallélisme non négatif étant appelé parallélisme positif. Il est également commode de considérer une courbe comme décrite dans une direction spécifique, ce qui fixe automatiquement la direction de toute courbe parallèle, et de définir la distance d'une courbe parallèle à celle-ci comme positive si elle se trouve à gauche et négative si elle se trouve à droite. Par exemple, dans la Fig. 1, la courbe AB est à la distance $+a$ de la courbe CD , et la courbe BA est à la distance $-a$ de la courbe DC . Le signe du parallélisme et le signe de la distance sont indépendants, l'un étant une propriété intrinsèque des courbes, l'autre dépendant de la direction de description des courbes et de laquelle est considérée comme parallèle à l'autre. Si la direction de description des courbes est

fixée, le parallélisme a le signe opposé à celui du produit des distances des courbes l'une à l'autre. Ainsi, dans la Fig. 1, les courbes CD et FE sont négativement parallèles, à la distance $+b$ l'une de l'autre. Clairement, en général, si la courbe MN est positivement parallèle, à la distance x , à la courbe PQ et que PQ est positivement parallèle, à la distance y , à RS , alors MN est positivement parallèle, à la distance $x + y$, à RS ; et si MN est parallèle (positivement ou négativement) à PQ et que PQ est parallèle à RS , alors MN est parallèle à RS . Ainsi, le parallélisme et le parallélisme positif sont des relations d'équivalence.

Nous sommes maintenant en mesure de considérer deux situations qui peuvent survenir lorsque de tels segments sont joints pour former une courbe continue et dérivable. La première est illustrée par la Fig. 2. En passant par B , le rayon de courbure de la courbe ABC devient inférieur à la distance de DEF à celle-ci, avec pour résultat que le signe du parallélisme change, produisant une pointe en E . Cette pointe donne lieu au phénomène suivant : lorsque la courbe DEF est décrite, le changement de direction en E provoque un changement soudain du signe de la distance de ABC à celle-ci. Car AB est à la distance $-a$ de DE , mais BC est à la distance $+a$ de EF . Ainsi, en un point où le signe du parallélisme change, il y a un changement correspondant du signe de la distance d'une courbe à l'autre. (Le signe de la distance n'a bien sûr pas de sens aux points tels que E où la direction de la courbe n'est pas définie, et le signe du parallélisme n'a pas de sens aux points tels que B et E où le rayon de courbure n'est pas défini.) Nous disons que deux courbes sont proprement parallèles si la distance de chacune à l'autre, au sens défini ici, reste constante lorsque chacune des courbes est décrite. Cela implique que le signe du parallélisme ne change pas, et qu'il n'y a pas de pointes.

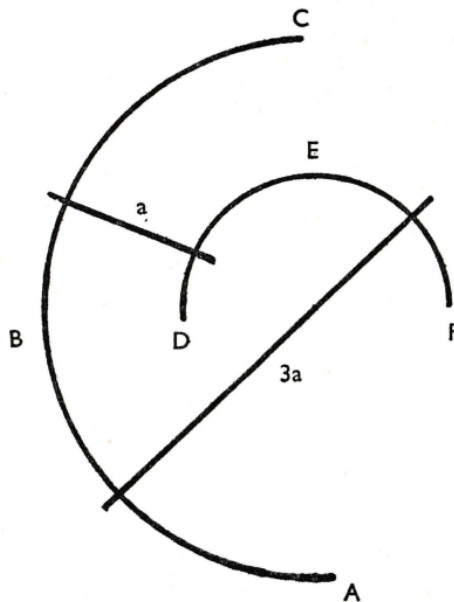


FIG. 3

La seconde situation est illustrée par la Fig. 3. Ici, le segment AB est parallèle à EF , et BC à DE , de sorte qu'il y a clairement un sens dans lequel ABC est parallèle à DEF , mais il y a une discontinuité assez violente en B et E . Nous appelons ce parallélisme discontinu. On dit que deux courbes sont continûment parallèles si à tout segment continu de l'une correspond un segment

continu parallèle de l'autre. Notons que le parallélisme propre est continu, mais que la réciproque n'est pas nécessairement vraie (la Fig. 2 en fournit un contre-exemple).

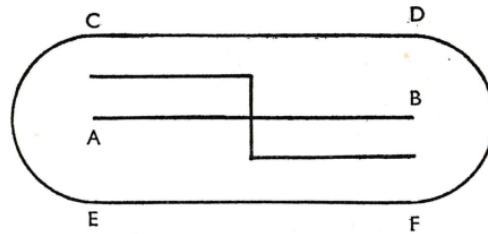


FIG. 4

Il y a essentiellement 3 façons dont des paires de segments parallèles peuvent être assemblées pour former des courbes “parallèles”. Premièrement (Fig. 4), les “courbes” résultantes peuvent consister en une collection de telles paires agencées au hasard sans continuité particulière, et ayant des extrémités. Ce cas n'est intéressant que dans la mesure où il faut se donner beaucoup de mal pour spécifier exactement en quoi consiste chaque courbe, en particulier aux extrémités.

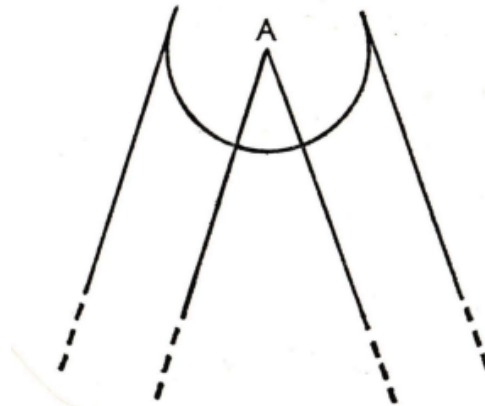


FIG. 5

Ainsi, dans la Fig. 4, par exemple, la droite AB est-elle parallèle à la paire de droites CD , EF ? Ou à la courbe $CDFE$? Je préférerais dire que la droite AB décrite deux fois était (continûment) parallèle à la paire de droites CD , EF , et que la courbe formée par la droite AB , le demi-cercle ponctuel en B décrit dans le sens horaire de la direction BD à BF , la droite BA , et le demi-cercle ponctuel en A décrit dans le sens horaire de AE à AC était (proprement) parallèle à la courbe $CDFE$. Mais c'est peut-être être pédant.

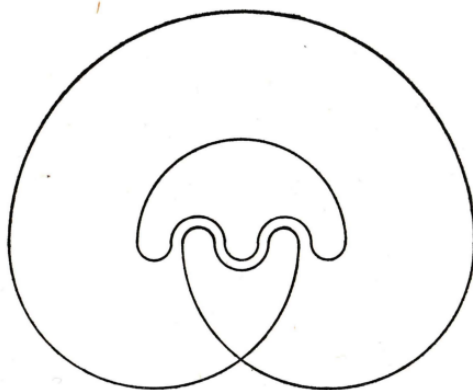


FIG. 6

Deuxièmement (Fig. 5), il peut ne pas y avoir d'extrémités, mais il peut y avoir des points (tels que A) où la courbe n'est pas dérivable. Des pérégrinations similaires sont nécessaires ici.

Troisièmement, les courbes résultantes peuvent être partout continues et dérivables. Il peut y avoir des pointes, mais il n'y a pas d'extrémités ni de points où il y a un changement de pente. Le parallélisme peut être discontinu et impropre (Fig. 6), continu et impropre (Fig. 7), ou continu et propre. Deux cercles concentriques, le cercle extérieur de rayon a , l'intérieur de rayon b fournit un exemple extrême de ce cas, car le cercle de rayon a peut être considéré comme étant proprement et continuellement parallèle

1. positivement, à distance $\pm(a - b)$, au cercle de rayon b ,
2. négativement, à distance $\pm(a + b)$, au cercle de rayon b ,
3. positivement, à distance $\pm a$, au cercle-point, son centre,
4. négativement, à distance $\pm 2a$, à lui-même.

Les cercles et les courbes formés à partir d'ensembles de cercles complets sont, je crois, les seules courbes différentiables connexes qui sont proprement et non trivialement (c'est-à-dire à distance $\neq 0$) parallèles à elles-mêmes ; mais il existe d'autres courbes qui sont continûment parallèles à elles-mêmes, celle de la Fig. 8 par exemple.

J'ai mentionné au début qu'il y avait d'autres définitions possibles du parallélisme. Par exemple, on pourrait dire que deux courbes sont parallèles par rapport à un ensemble donné de droites si chacune d'elles rencontre chacune des droites sous le même angle θ . Ainsi, les spires successives d'une spirale équiangulaire (ou logarithmique) (Fig. 9) peuvent être dites parallèles à l'angle θ à l'ensemble des droites passant par l'origine. On pourrait appeler cela le parallélisme radial, puisque l'ensemble des droites prises est l'ensemble des "vecteurs radiaux" issus de l'origine. Si l'on prend plutôt l'ensemble des normales aux courbes, ce type de parallélisme se réduit à celui considéré initialement, et on pourrait donc raisonnablement l'appeler parallélisme normal. Deux cercles concentriques sont à la fois normalement et radialement parallèles. Mais dans le cas du parallélisme non normal, il n'est pas possible de définir la distance entre les deux courbes de manière à ce qu'elle reste constante au fur et à mesure que les courbes sont parcourues. J'ai également mentionné que deux courbes sont souvent dites parallèles si elles peuvent être amenées à coïncider par une simple translation, ou, en fait, par translation et grossissement, c'est-à-dire par projection. Ainsi, du pantographe, un instrument pour agrandir des dessins, on dit souvent que lorsque le pointeur parcourt la courbe à agrandir, le crayon

trace une courbe parallèle, en ce sens qu'il se déplace dans la même direction. On pourrait étendre la définition pour couvrir cela en disant que deux courbes sont pseudo-parallèles ou semi-parallèles par rapport à un ensemble donné de droites si une droite quelconque rencontre les deux courbes sous le même angle, mais que des droites différentes rencontrent la courbe sous des angles différents. Ainsi, les figures translatées sont pseudo-parallèles par rapport à l'ensemble des droites parallèles à la direction de translation, tandis que les figures projetées sont pseudo-parallèles par rapport aux droites passant par le point de projection fixe. Là encore, aucune distance invariante ne peut être définie, sauf dans le cas de la translation pure, où la distance parcourue, ou distance de décalage, est invariante.

Il faut également mentionner qu'il n'y a aucune raison de ne pas définir une sorte de parallélisme curviligne, par lequel deux courbes sont parallèles par rapport à un ensemble donné de courbes si elles coupent chacune des courbes sous le même angle. Ainsi, l'affirmation selon laquelle deux cercles concentriques sont parallèles par rapport à l'ensemble des vecteurs radiaux passant par leur centre se généralise à l'affirmation selon laquelle deux ellipses confocales sont parallèles par rapport à l'ensemble orthogonal d'hyperboles confocales. Mais cela ouvre tout un nouveau champ de la géométrie, et bien que ce soit un sujet fascinant, je ne pense pas que la relation entre ces courbes soit suffisamment proche de l'idée intuitive de parallélisme pour justifier vraiment ce nom.

Il est possible d'aller encore plus loin et d'appliquer les définitions de base en 3 dimensions ou plus. Ainsi, en 3 dimensions, par exemple, on dirait que deux surfaces sont parallèles si une normale à l'une est aussi une normale à l'autre, et que deux courbes sont parallèles si en tout point de l'une il existe une normale (c'est-à-dire une droite dans le plan normal) qui est aussi une normale à l'autre. On verra que la majeure partie de la théorie ci-dessus s'étend à ces cas, bien que des complications supplémentaires surgissent.

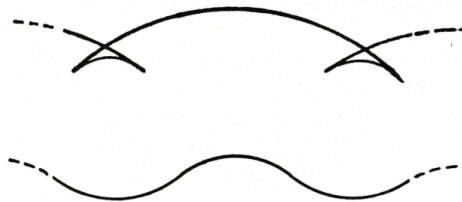


FIG. 7

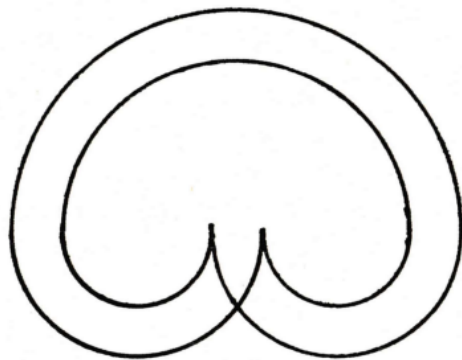


FIG. 8

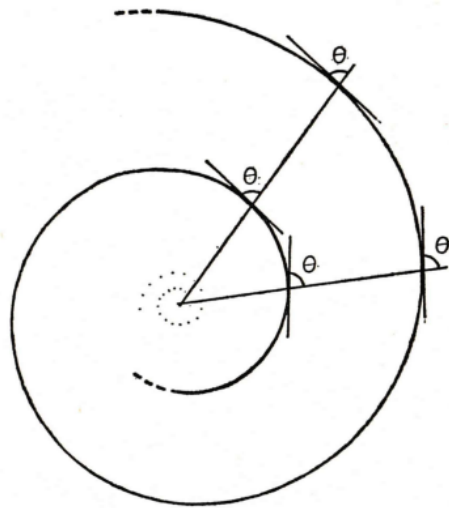


FIG. 9

Pentaplexité

Une classe de pavages non périodiques du plan

par le Professeur R. Penrose

Certains lecteurs connaissent peut-être un article de Martin Gardner paru dans le numéro de janvier 1977 de *Scientific American*. Il y décrivait une paire de formes planes que j'avais trouvées, appelées “cerfs-volants” et “fléchettes”, qui, lorsqu'elles sont assemblées selon certaines règles simples, pouvaient paver tout le plan, mais uniquement de manière non périodique. Les pavages présentent un certain nombre de propriétés remarquables, dont certaines ont été décrites dans l'article de Gardner. Je vais donner ici un bref compte rendu expliquant comment ces tuiles sont apparues et indiquant certaines de leurs propriétés.

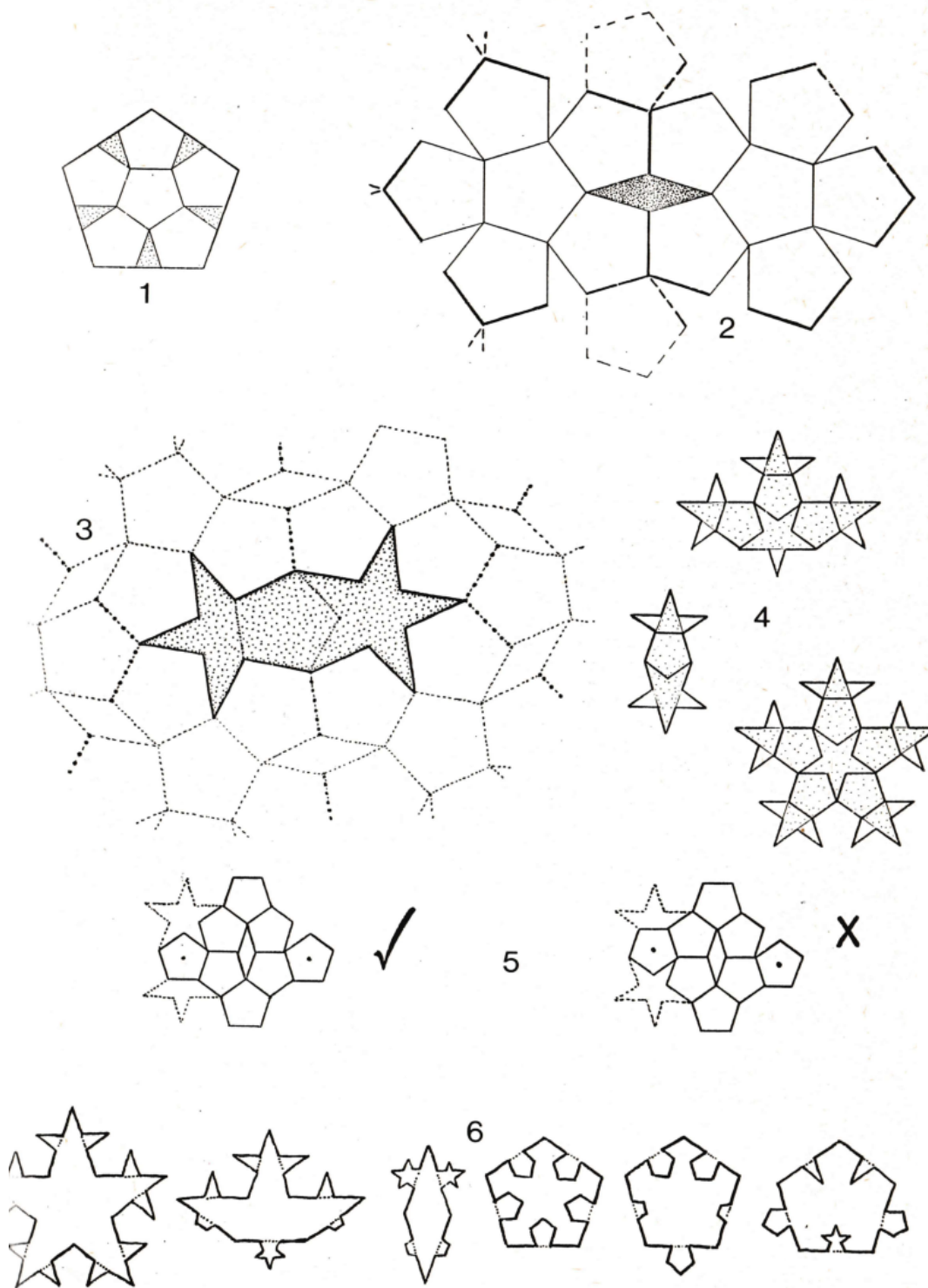
Le point de départ était l'observation qu'un pentagone régulier peut être subdivisé en six plus petits, ne laissant que cinq fines lacunes triangulaires¹. (Voir Fig. 1 ; c'est un élément familier du “patron” habituel qui se replie en un dodécaèdre régulier). Imaginez maintenant que ce processus soit répété un grand nombre de fois, où à chaque étape les pentagones de la figure sont subdivisés selon le schéma de la Fig. 1. Il y aura alors des lacunes de formes variées qui apparaîtront et nous souhaitons voir comment les combler au mieux. À la deuxième étape de subdivision, des lacunes en forme de losange apparaissent entre les pentagones (Fig. 2). À la troisième, ces losanges développent des “pointes”, mais il est possible de trouver de la place, à l'intérieur de chacun de ces “losanges pointus”, pour un autre pentagone, de sorte que la lacune se sépare en une étoile (pentagramme) et un “bateau en papier” (ou chapeau de bouffon ?) (Fig. 3). À l'étape suivante, l'étoile et le bateau développent également des “pointes”, et, de même, nous pouvons trouver de la place pour de nouveaux pentagones en leur sein, les lacunes restantes étant de nouvelles étoiles et de nouveaux bateaux (comme précédemment). Ces subdivisions sont représentées sur la Fig. 4.

Puisqu'aucune nouvelle forme n'est introduite aux étapes suivantes, nous pouvons envisager que ce processus de subdivision se poursuive indéfiniment. À chaque étape, l'échelle des formes peut être agrandie vers l'extérieur afin que les nouveaux pentagones qui apparaissent deviennent de la même taille que ceux de l'étape précédente. Cependant, en l'état, cette procédure autorise une ambiguïté que nous aimerions supprimer. La subdivision d'un “losange pointu” peut être réalisée de deux manières, car il y a deux positions alternatives pour le pentagone. Insistons sur une seule de ces possibilités, la règle étant celle donnée sur la Fig. 5. (Lorsque nous examinons la disposition des pentagones environnants, nous constatons nécessairement qu'ils sont agencés selon le type de configuration montré sur la Fig. 5). Il convient de mentionner que si la règle opposée avait été adoptée pour subdiviser un “losange pointu”, une contradiction apparaîtrait à l'étape de subdivision suivante, mais cela ne se produit jamais avec la règle de la Fig. 5.

Cette procédure, poursuivie jusqu'à la limite, conduit à un pavage de tout le plan avec des pentagones, des losanges, des bateaux et des étoiles. Mais il existe de nombreux pavages “incorrects” avec ces mêmes formes, n'étant pas construits selon la prescription ci-dessus. En fait, la “correction” peut être imposée en adoptant des règles d'appariement appropriées. La manière la plus claire de représenter ces règles est de modifier les formes pour en faire une sorte de puzzle infini, où une telle modification suggérée est donnée sur la Fig. 6. Il n'est pas trop difficile de montrer que tout

1. [Vraiment contente d'avoir lu cette motivation initiale.](#)

pavage avec ces six formes est forcé d'avoir une structure hiérarchique du type qui vient d'être décrit.



De plus, la nature hiérarchique forcée de ce motif implique que le pavage possède un certain nombre de propriétés très remarquables. En premier lieu, il est nécessairement non périodique (c'est-à-dire sans aucun parallélogramme de période). Nous y reviendrons plus loin. Deuxièmement, bien que le

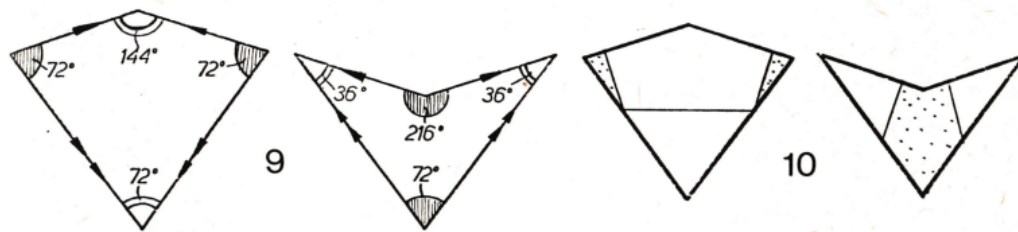
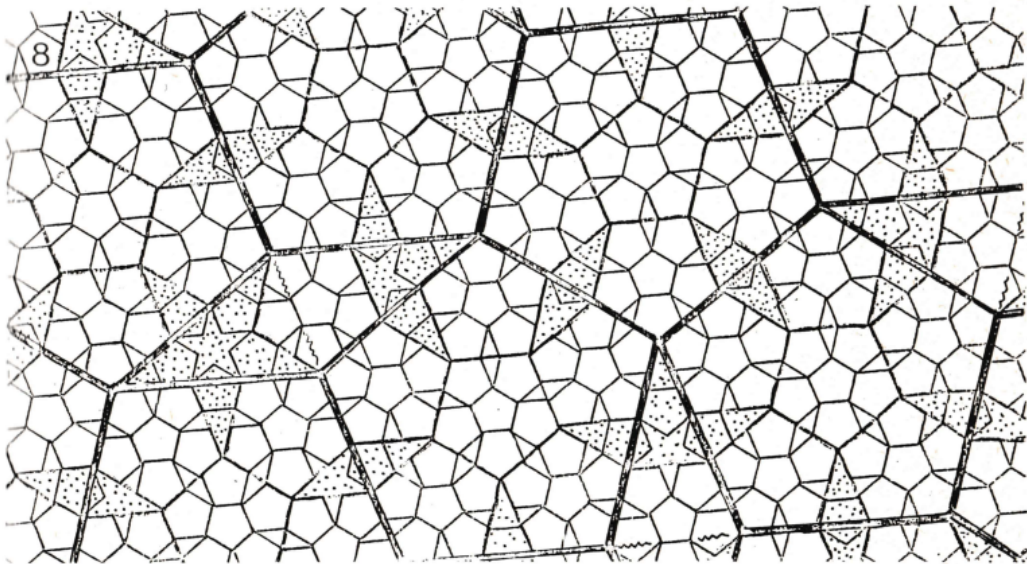
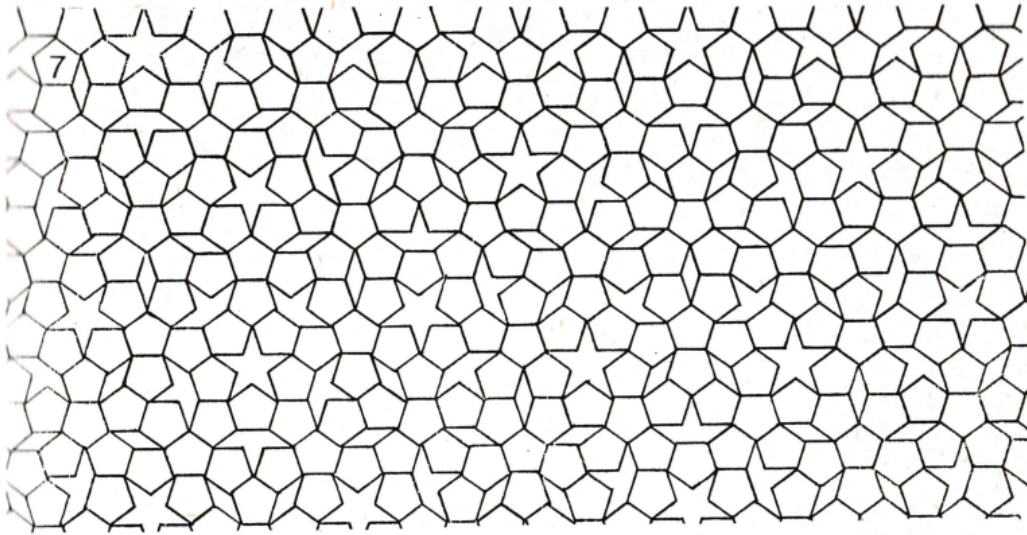
motif complet ne soit pas déterminé de manière unique - car il existe 2^{n_0} arrangements différents - ces différents arrangements sont, dans un certain sens “fini”, tous indiscernables les uns des autres ! Ainsi, quelle que soit la taille d’une portion finie sélectionnée dans l’un de ces motifs, cette portion finie apparaîtra quelque part dans tout autre motif complet (une infinité de fois, en fait). Troisièmement, il existe de nombreuses caractéristiques inattendues et esthétiquement plaisantes que ces motifs présentent (voir Fig. 7). Par exemple, on voit apparaître de nombreux décagones réguliers, qui tendent à se chevaucher par endroits. Chaque décagone est entouré d’un anneau de douze pentagones, et il existe également des anneaux plus grands de divers types. Notez que chaque segment de droite du motif s’étend vers l’extérieur indéfiniment, pour contenir un nombre infini d’autres segments de droite de la figure. L’arrangement hiérarchique de la Fig. 7 est mis en évidence sur la Fig. 8.

Après que j’ai trouvé cet ensemble de six tuiles qui force la non-périodicité, il m’a été fait remarquer (par Simon Kochen) que Raphael Robinson avait, plusieurs années auparavant, également trouvé un ensemble (tout à fait différent) de six tuiles qui force la non-périodicité. Mais il m’est venu à l’esprit qu’avec mes tuiles, on pouvait faire mieux. Si, par exemple, la troisième forme “pentagone” est éliminée en étant jointe en deux endroits au “losange” et en un endroit au bas du “bateau”, alors on obtient un ensemble de cinq tuiles qui force la non-périodicité. Il n’a pas été difficile de réduire encore ce nombre à quatre. Et ensuite, avec un peu de découpage et de recollage, à deux !

Les deux tuiles ainsi obtenues sont les “cerfs-volants” et les “fléchettes” mentionnés au début ². Les formes précises sont illustrées sur la Fig. 9. Les règles d’appariement sont également montrées, où les sommets de même couleur doivent être placés les uns contre les autres. Il existe de nombreuses façons alternatives de colorer ou d’ombrer ces tuiles pour forcer les arrangements corrects. Une façon qui met en évidence la relation avec les pavages pentagone-losange-bateau-étoile est montrée sur la Fig. 10. Une portion de tuiles assemblées (partiellement ombrée de cette manière) est montrée sur la Fig. 11 (ci-contre).

La nature hiérarchique des pavages cerf-volant-fléchette peut être vue directement, et est illustrée sur la Fig. 12. Prenez un tel pavage et coupez chaque fléchette symétriquement par un segment de droite. Les demi-fléchettes et les cerfs-volants résultants peuvent alors être rassemblés pour former des fléchettes et des cerfs-volants à une échelle légèrement plus grande : deux demi-fléchettes et un cerf-volant forment une grande fléchette, deux demi-fléchettes et deux cerfs-volants forment un grand cerf-volant. Il n’est pas difficile de se convaincre que tout pavage cerf-volant-fléchette correctement apparié est assemblé de cette manière. Cette propriété d’“inflation” sert également à prouver la non-périodicité. Car supposons qu’il y ait un parallélogramme de période. Les cerfs-volants et fléchettes gonflés correspondants devraient également avoir le même parallélogramme de période. Répétez le processus d’inflation de nombreuses fois, jusqu’à ce que la taille des cerfs-volants et fléchettes gonflés résultants soit supérieure à celle du parallélogramme de période supposé. Cela donne une contradiction.

2. Ces noms ont été suggérés par John Conway.

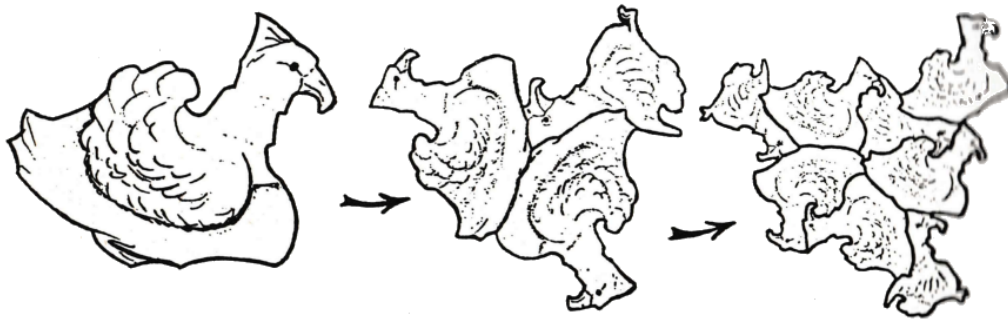


La contradiction avec la périodicité se manifeste d'une autre manière frappante. Considérons une très grande surface contenant d fléchettes et k cerfs-volants, obtenue en se référant au processus d'inflation un grand nombre de fois. Plus la surface est grande, plus le rapport $x = k/d$ des cerfs-volants aux fléchettes sera proche de satisfaire la relation de récurrence $x = (1 + 2x)/(1 + x)$ (puisque, par inflation, une fléchette et deux cerfs-volants forment un plus grand cerf-volant, tandis qu'une fléchette et un cerf-volant forment une plus grande fléchette). Cela donne, à la limite d'un

motif infiniment grand, $x = \frac{1 + \sqrt{5}}{2} = T$, le nombre d'or ! Ainsi, nous obtenons une densité relative irrationnelle³ de cerfs-volants par rapport aux fléchettes - ce qui est impossible pour un pavage périodique.

Il existe une autre paire de quadrilatères qui, avec des règles d'appariement appropriées, ne pave le plan que de manière non périodique. C'est une paire de losanges représentée sur la Fig. 13. Sur la Fig. 14, un ombrage approprié est suggéré où les bords ombrés de manière similaire doivent être appariés les uns aux autres. Sur la Fig. 15, la relation hiérarchique avec les cerfs-volants et les fléchettes est illustrée. Les losanges apparaissent à mi-chemin entre un niveau cerf-volant-fléchette et le niveau cerf-volant-fléchette gonflé suivant.

De nombreuses versions différentes de puzzles de la paire cerf-volant-fléchette ou de la paire de losanges peuvent manifestement être données. Une suggestion pour des cerfs-volants et fléchettes modifiés, sous la forme de deux oiseaux, est illustrée sur la Fig. 16. Le processus d'inflation (en sens inverse) est illustré sur la Fig. 17.



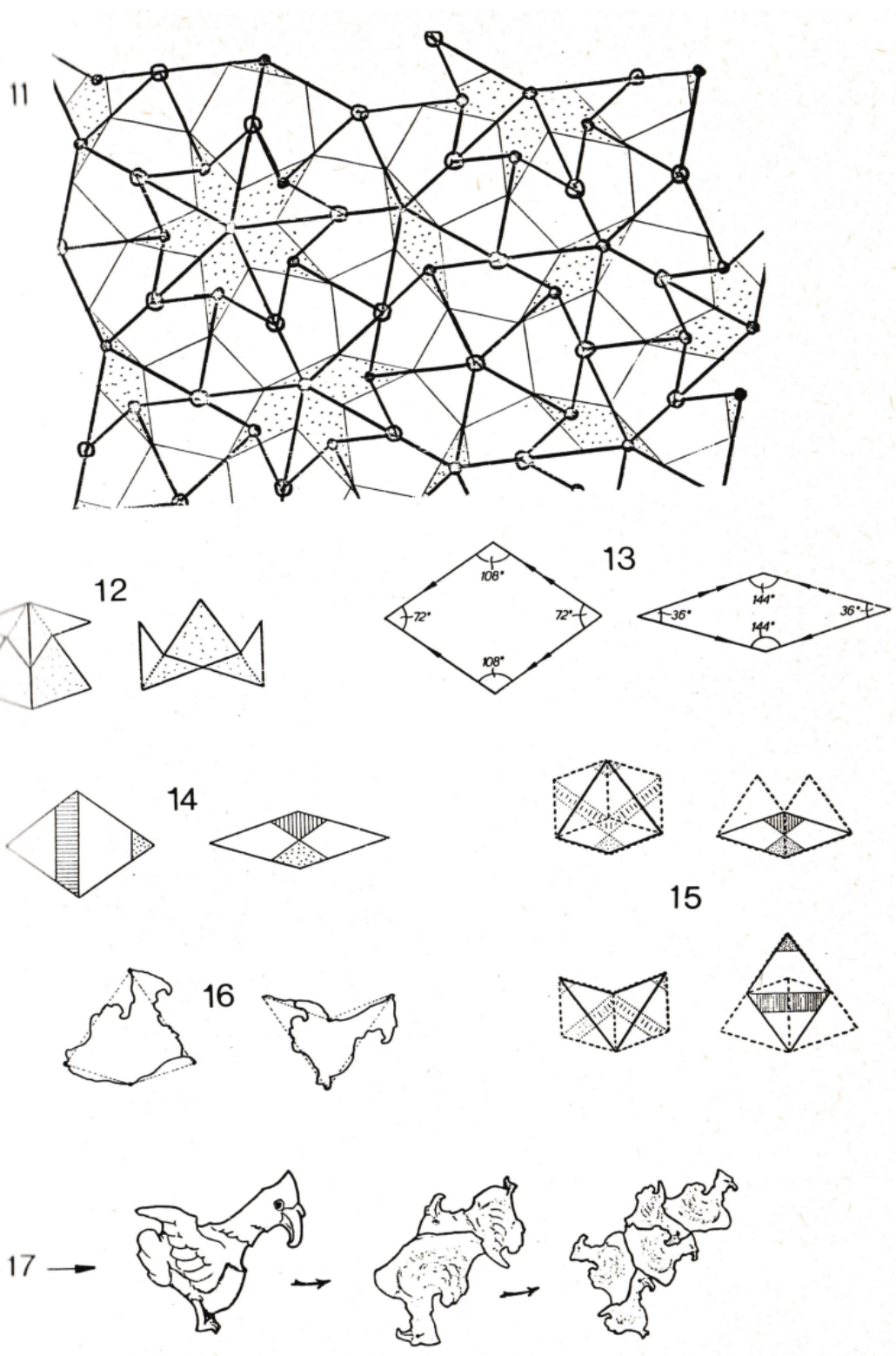
D'autres modifications sont également possibles, telles que des règles d'appariement alternatives, suggérées par Robert Ammann (voir Fig. 18) qui forcent la moitié des tuiles à être inversées.

De nombreuses caractéristiques intrigantes des pavages n'ont pas été mentionnées ici, telles que les anneaux à symétrie pentagonale que produisent les bandes de la Fig. 14, la classification par Conway des "trous" dans les motifs cerf-volant-fléchette (c'est-à-dire les régions entourées de pavages "légaux" mais qui ne peuvent elles-mêmes être légalement comblées), la version tridimensionnelle des losanges d'Ammann (quatre solides qui apparemment ne remplissent l'espace que de manière non périodique), l'analyse par Ammann et Conway des "empires" (le système infini de tuiles partiellement déconnectées dont les positions sont forcées par un ensemble donné de tuiles).

On ne sait pas s'il existe une forme unique qui puisse paver le plan euclidien uniquement de manière non périodique. Pour le plan hyperbolique (de Lobatchevski), une forme unique peut être fournie qui, dans un certain sens, ne pave que de manière non périodique (voir Fig. 19) - mais dans un

3. C'est la densité numérique. Le cerf-volant a τ fois l'aire de la fléchette, donc l'aire totale couverte par les cerfs-volants est $\tau^2 (= 1 + \tau)$ fois celle couverte par les fléchettes.

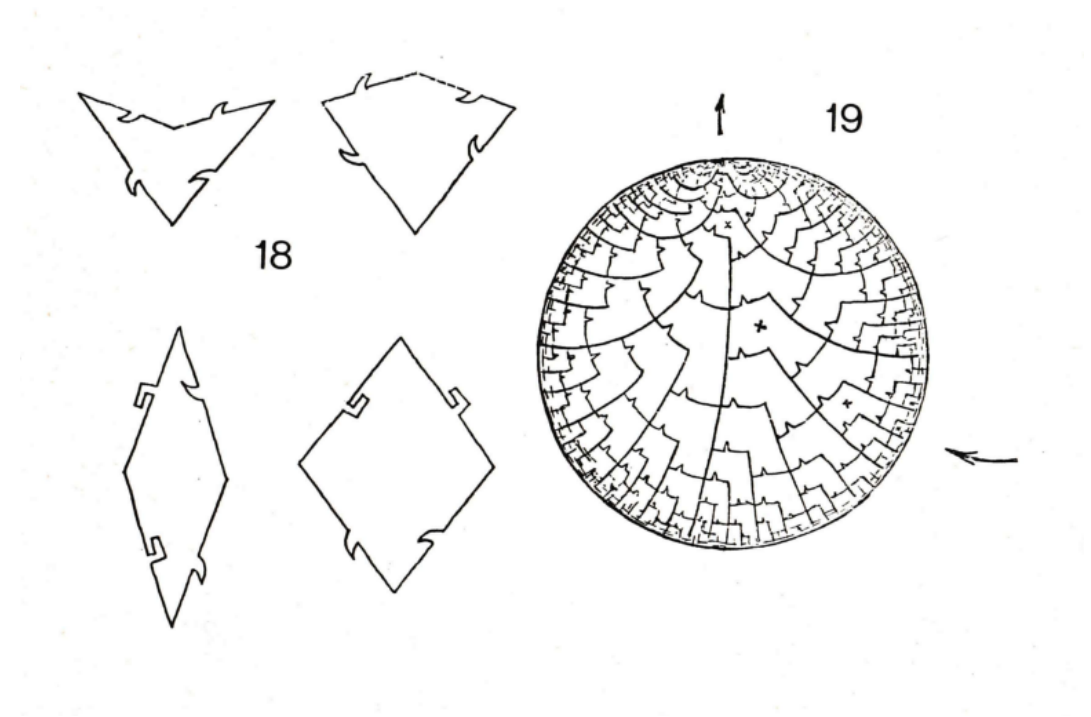
autre sens, une périodicité (dans une seule direction) peut se produire.



3. Références

[1] Gardner, M., *Scientific American* janvier 1977, p. 110-121.

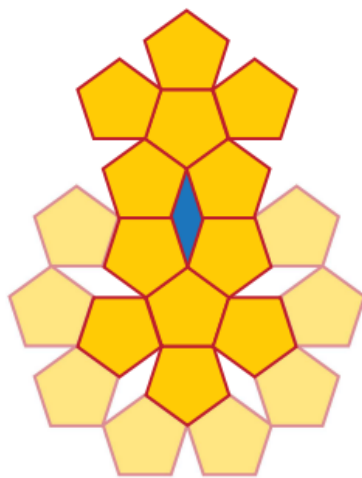
[2] Penrose, R., *Bull. Inst. Maths. and its Applications*. 10 No 7/8 (1974) p. 266-271.



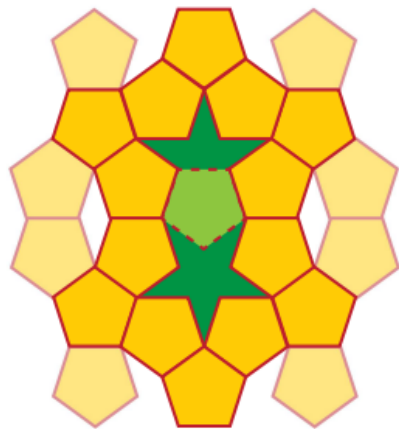
Figures colorées dans la réédition de l'article



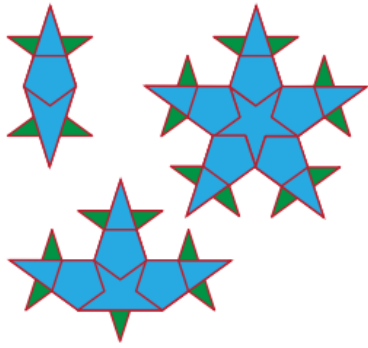
▲ Figure 1



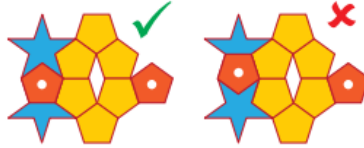
▲ Figure 2



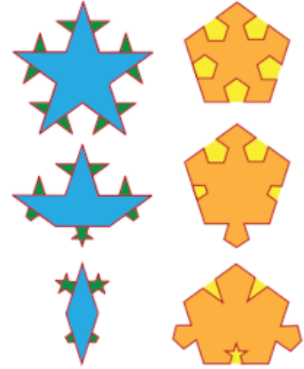
▲ Figure 3



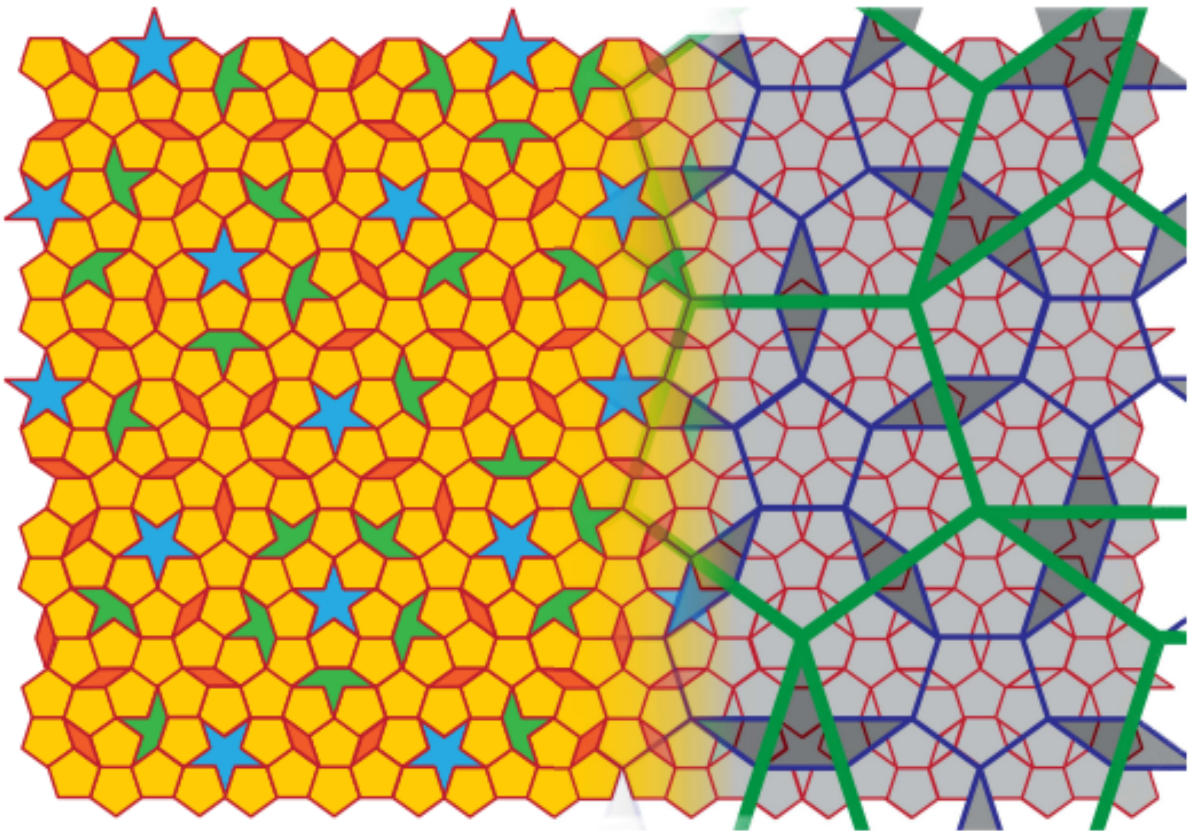
▲ Figure 4



▲ Figure 5

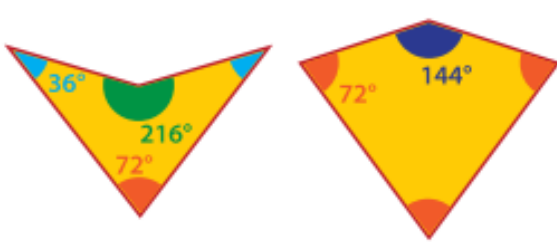


▲ Figure 6

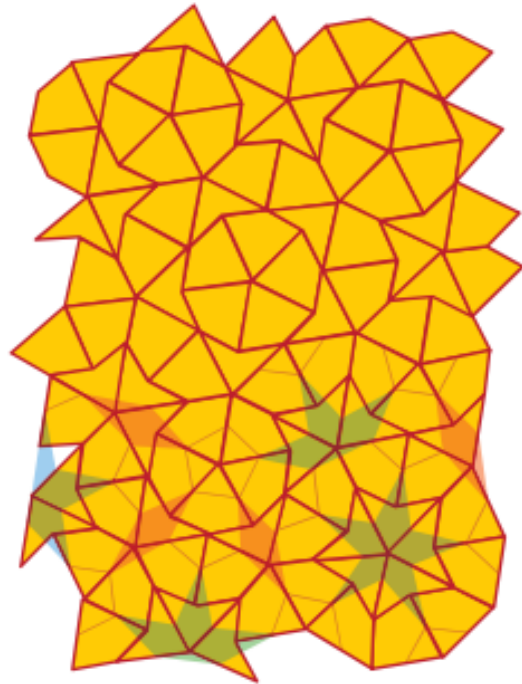


▲ Figure 7

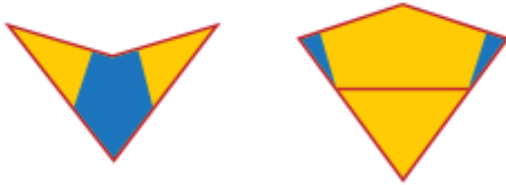
▲ Figure 8



▲ Figure 9



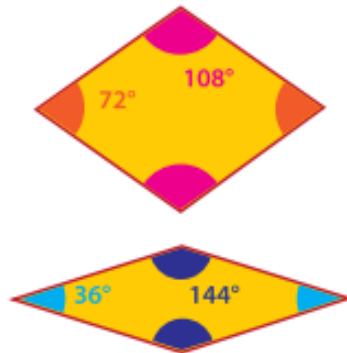
▲ Figure 11



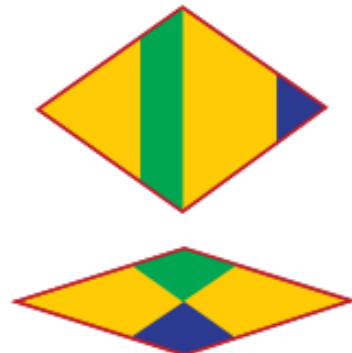
▲ Figure 10



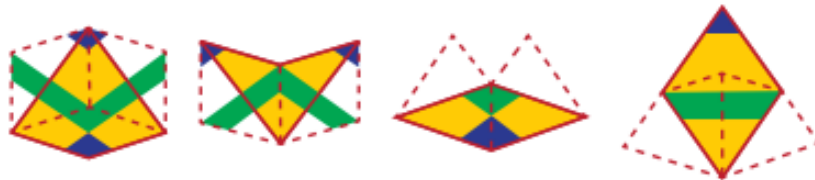
▲ Figure 12



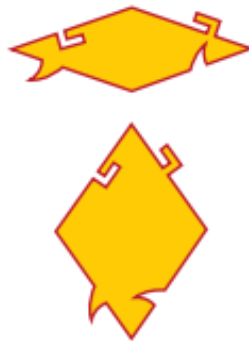
▲ Figure 13



▲ Figure 14

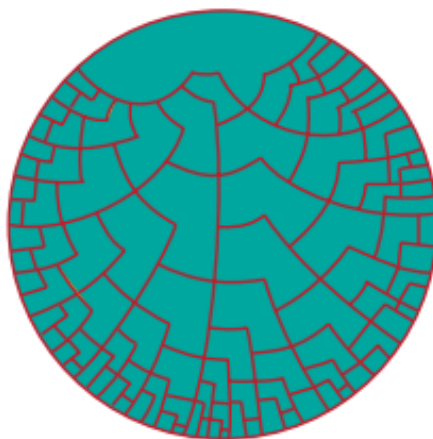


▲ Figure 15



▲ Figure 16

▲ Figure 17



▲ Figure 18

Petites Flèches

par Dr P. T. Johnstone

L'une des caractéristiques saillantes des mathématiques du vingtième siècle a été la prolifération de différents types de structures étudiées par les mathématiciens, et la spécialisation croissante de la recherche mathématique qui en a résulté. Une exception notable à cette tendance, au cours des trente dernières années, a été l'essor de la théorie des catégories, qui trouve son origine dans les travaux conjoints de Samuel Eilenberg et Saunders Mac Lane vers 1945 [1]. Dans cet article, je vais tenter de décrire les objectifs de la théorie des catégories et certains de ses domaines d'application actuels.

La théorie des catégories commence avec l'idée que si vous voulez étudier un type particulier de structure mathématique, vous devez étudier non seulement les structures de ce type, mais aussi les applications qui préservent les structures entre elles. Ainsi, si vous vous intéressez aux espaces topologiques, il est inutile d'étudier les espaces par eux-mêmes, sans le concept d'application continue qui les lie. De même, l'étude des groupes, des anneaux ou des espaces vectoriels ne peut aller loin sans la notion appropriée d'homomorphisme.

Une catégorie consiste donc en trois choses : une collection d'objets, une collection de flèches ou de morphismes, chacun étant associé à une paire d'objets appelés sa source et son but (nous utilisons la notation diagrammatique " $f : A \rightarrow B$ " pour " f est une flèche de source A et de but B "), et une loi de composition qui assigne à chaque "paire composable" de flèches ($f : A \rightarrow B$, $g : B \rightarrow C$) une troisième flèche $gf : A \rightarrow C$. De plus, nous exigeons que la composition soit associative (c'est-à-dire que $h(gf) = (hg)f$ chaque fois que les composés sont définis), et que pour chaque objet A il existe une flèche identité $1A : A \rightarrow A$ telle que $f1A = f$ et $1Ag = g$ chaque fois que les composés sont définis.

Les exemples familiers incluent la catégorie Set des ensembles et des fonctions, les catégories algébriques Gp, Rng, Lat, ... des groupes, anneaux, treillis, ... et les homomorphismes appropriés, la catégorie Top des espaces topologiques et des applications continues, la catégorie Ban des espaces de Banach et des contractions linéaires, et ainsi de suite. Ce sont toutes des catégories concrètes ; c'est-à-dire que leurs objets sont des ensembles avec un certain type de structure, leurs flèches sont des applications préservant la structure, et leur loi de composition est la composition habituelle des fonctions. Mais il existe aussi de nombreuses catégories qui ne sont pas concrètes ; par exemple, nous pouvons faire de tout ensemble partiellement ordonné P une catégorie dont les objets sont les éléments de P , et dont les flèches sont des instances de la relation d'ordre (c'est-à-dire qu'il y a exactement une flèche $p \rightarrow q$ si $p \leq q$, et aucune si $p \not\leq q$). (La loi de composition dans ce cas est évidente, puisqu'il y a toujours au plus une flèche avec une source et un but donnés.)

La lectrice astucieuse aura remarqué que j'ai utilisé le mot vague "collection" plutôt que le mot précis "ensemble" dans la définition d'une catégorie, et le dernier paragraphe en donne la raison : d'après le paradoxe de Russell, nous savons que la collection de tous les ensembles ne peut pas elle-même être un ensemble. La relation entre la théorie des catégories et la théorie des ensembles est souvent quelque peu tendue, précisément parce que le métier de la première est d'étudier des choses telles que la totalité de tous les ensembles ou de tous les groupes, qui sont des objets inadmissibles pour la seconde. Il existe des modifications de la théorie standard des ensembles qui fournissent un fondement "respectable" à la théorie des catégories [2] ; mais en fait la théorie des

catégories, en raison de sa nature fondamentale, peut elle-même être utilisée comme base pour le reste des mathématiques, y compris la théorie des ensembles ! Nous reviendrons sur ce point plus tard ; pour l’instant, introduisons simplement le terme *petite catégorie* (aussi appelée *kittygory* [3] !) pour décrire une catégorie dont les objets et les flèches sont les éléments d’un certain ensemble. Une notion plus faible, mais tout aussi utile, est celle de *catégorie localement petite* ; nous disons que C est localement petite si, pour chaque paire d’objets (A, B) , les flèches de C de source A et de but B sont les éléments d’un ensemble $C(A, B)$. Notez qu’une catégorie concrète est automatiquement localement petite.

Nous pouvons maintenant prendre n’importe quel concept ou théorème mathématique qui peut être exprimé uniquement en termes d’objets, de flèches et de composition, et l’appliquer à n’importe laquelle des catégories mentionnées précédemment. Un exemple évident est le concept d’isomorphisme ; nous disons que deux objets A et B sont isomorphes s’il existe $f : A \rightarrow B$ et $g : B \rightarrow A$ tels que $gf = 1_A$ et $fg = 1_B$. Plus intéressant, la notion de produit cartésien est catégorique. Normalement, nous définissons le produit cartésien de deux ensembles en spécifiant ses éléments, mais nous pouvons aussi le définir (au moins à isomorphisme près) en spécifiant les flèches vers celui-ci : un produit de deux objets A et B est un objet $A \times B$ muni de flèches $p : A \times B \rightarrow A$ et $q : A \times B \rightarrow B$ telles que, pour toute paire de flèches $(f : C \rightarrow A, g : C \rightarrow B)$, il existe un unique $h : C \rightarrow A \times B$ avec $ph = f$ et $qh = g$.

Cette définition coïncide avec la définition habituelle dans toutes les catégories concrètes mentionnées ci-dessus, mais comme elle ne fait pas intervenir d’éléments, nous pouvons aussi l’appliquer à des catégories non concrètes ; dans un ensemble partiellement ordonné, elle coïncide avec la notion de borne inférieure (plus grand élément inférieur). Mais il y a une autre possibilité : en inversant toutes les flèches dans une définition ou un théorème catégorique, nous obtenons un nouveau concept “dual” qui peut être tout aussi utile que l’ancien. Ainsi, nous avons la notion duale de coproduit : un coproduit de A et B consiste en un objet $A+B$ et des flèches $A \rightarrow A+B$, $B \rightarrow A+B$ tels que ... En exercice, essayez de déterminer à quoi ressemblent les coproduits dans Set , Top , Gp , Ab (groupes abéliens) et CRng (anneaux commutatifs) ; la réponse dans chaque cas est une construction familière dans le sujet approprié.

Jusqu’ici, cependant, nous n’avons pas pris à cœur le dicton fondamental de la théorie des catégories : “Regardez toujours vers les morphismes”. La notion appropriée de morphisme entre catégories est appelée un foncteur : un foncteur $C \rightarrow D$ consiste en deux fonctions envoyant les objets de C vers les objets de D , les flèches de C vers les flèches de D , et préservant la composition et les identités.

(Ainsi, les (petites) catégories et les foncteurs entre elles forment elles-mêmes une catégorie Cat). On l’appelle parfois un foncteur covariant, pour le distinguer de la notion de foncteur contravariant qui est similaire, mais inverse la direction des flèches et l’ordre de composition. Parmi les exemples les plus simples de foncteurs (covariants) figurent les foncteurs d’oubli $\text{Gp} \rightarrow \text{Set}$, $\text{Top} \rightarrow \text{Set}$, etc., qui oublient simplement la structure supplémentaire sur les objets et les flèches d’une catégorie concrète. De nombreuses constructions familières en algèbre, telles que “sous-groupe des commutateurs” : $\text{Gp} \rightarrow \text{Gp}$, “groupe des unités” : $\text{Rng} \rightarrow \text{Gp}$, et “anneau des matrices $n \times n$ ” : $\text{Rng} \rightarrow \text{Rng}$, sont en fait des foncteurs ; bien que nous les considérions normalement comme définis uniquement sur les objets de la catégorie appropriée, en fait chacun d’eux a une définition naturelle sur les

homomorphismes également. Comme exemple de foncteur contravariant, nous donnons le foncteur d'ensemble des parties $P : \text{Set} \rightarrow \text{Set}$, qui envoie A vers l'ensemble de tous les sous-ensembles de A et $f : A \rightarrow B$ vers l'application $(B' \mapsto f^{-1}(B'))$.

Mais il y a en fait un troisième niveau de structure ici, puisque nous pouvons aussi parler d'applications entre foncteurs. Si S et T sont deux foncteurs de C vers D , une transformation naturelle $\alpha : S \rightarrow T$ est une fonction assignant à chaque objet A de C une flèche $\alpha_A : SA \rightarrow TA$ dans D , telle que pour chaque $f : A \rightarrow B$ dans C les composés $\alpha_B \circ Sf$ et $Tf \circ \alpha_A : SA \rightarrow TB$ soient égaux dans D . L'exemple le plus connu de transformation naturelle est l'"isomorphisme naturel" familier entre un espace vectoriel de dimension finie et son double dual ; en fait, on pourrait presque dire que la théorie des catégories a été inventée afin de donner un sens précis au mot "naturel" dans ce contexte.

Pour voir comment cela fonctionne, soit $R\text{-Mod}$ la catégorie des modules sur un anneau commutatif R . Pour tout module A , l'ensemble des applications linéaires $A \rightarrow R$ forme un module A^* , le dual de A , et l'assignation $A \mapsto A^*$ est un foncteur contravariant $R\text{-Mod} \rightarrow R\text{-Mod}$ (puisque si l'on se donne $f : A \rightarrow B$, la composition avec f induit une application linéaire $B^* \rightarrow A^*$). Le composé de ce foncteur avec lui-même, cependant, est covariant ; et il existe une transformation naturelle α du foncteur identité $1_{R\text{-Mod}}$ vers $\star\star$, telle que $\alpha_A(a)$ soit l'application linéaire "évaluer en a " : $A^{**} \rightarrow R$. Si A est un module libre de type fini, alors α_A est un isomorphisme ; c'est l'isomorphisme naturel que nous connaissons et aimons tous.

La notion de transformation naturelle nous permet de faire des foncteurs de C vers D les objets d'une catégorie D^C . Une partie importante de la théorie des catégories consiste à étudier ces catégories de foncteurs, en particulier celles de la forme Set^C où C est localement petite. L'importance de ces dernières est liée à l'existence, pour chaque objet A de C , d'un foncteur $h^A : C \rightarrow \text{Set}$ qui envoie B vers $C(A, B)$ et $f : B \rightarrow B'$ vers la fonction "composer avec f ". (De même, nous avons un foncteur contravariant $h_A : C \rightarrow \text{Set}$ qui envoie B vers $C(B, A)$).

Le lemme de Yoneda [4] dit que, pour tout foncteur $T : C \rightarrow \text{Set}$, les flèches $h^A \rightarrow T$ dans Set^C sont en correspondance bijective naturelle avec les éléments de l'ensemble TA ; explicitement, une transformation naturelle $\alpha : h^A \rightarrow T$ est complètement spécifiée par l'élément $\alpha_A(1_A)$ de TA . Cette observation apparemment anodine a des conséquences profondes. Par exemple, nous en déduisons que les flèches $h^A \rightarrow h^B$ dans Set_C^C correspondent à des éléments de $h^B(A)$, c'est-à-dire à des flèches $B \rightarrow A$ dans C ; ainsi l'assignation $A \mapsto h^A$ est non seulement un foncteur contravariant $C \rightarrow \text{Set}_C^C$, mais en fait un plongement plein (c'est-à-dire qu'il est bijectif sur les flèches entre une paire d'objets donnée). Nous disons qu'un foncteur $T : C \rightarrow \text{Set}$ est représentable s'il est isomorphe à un certain h^A , et définissons une représentation de T comme une paire (A, a) avec A un objet de C et $a \in TA$, telle que la flèche induite $h^A \rightarrow T$ soit un isomorphisme. Une autre conséquence du lemme de Yoneda est que la représentation de T , si elle existe, est unique à isomorphisme canonique près.

Presque tous les concepts importants de la théorie des catégories peuvent être décrits, au moins pour les catégories localement petites, comme des représentations de certains foncteurs. Par exemple, un produit de A et B est une représentation du foncteur (contravariant) qui envoie C vers l'ensemble des paires $(f : C \rightarrow A, g : C \rightarrow B)$. Le foncteur contravariant d'ensemble des parties est représenté par $(2, \{1\})$ où 2 est l'ensemble à deux éléments $\{0, 1\}$, et le foncteur d'oubli $U : \text{Gp} \rightarrow \text{Set}$ est re-

présenté par $(Z, 1)$ où Z est le groupe additif des entiers. Plus généralement, pour chaque ensemble A , le composé $h^A.U$ est représenté par (FA, n_A) où FA est le groupe libre engendré par A et $n_A : A \rightarrow UFA$ est l'insertion des générateurs.

Ce dernier est un exemple d'adjonction ; si $F : C \rightarrow D$ et $U : D \rightarrow C$ sont des foncteurs, nous disons que F est adjoint à gauche de U si l'on se donne une bijection, naturelle en A et B , entre les ensembles $D(FA, B)$ et $C(A, UB)$. Un théorème remarquable affirme que toute telle bijection provient d'une paire de transformations naturelles $\eta : 1_C \rightarrow UF$, $\epsilon : FU \rightarrow 1_D$ (l'unité et la coïunité de l'adjonction) telles que $\epsilon_{FA} \cdot F n_A = 1_{FA}$ et $U \epsilon_B \cdot n_{UB} = 1_{UB}$ pour tous A et B ; ainsi la notion d'adjonction est une notion "élémentaire" qui peut être définie sans référence à la petitesse locale. Un grand nombre de constructions clés en mathématiques s'avèrent être des adjoints à gauche ou à droite de certains foncteurs familiers ; nous venons de mentionner l'exemple des groupes libres, et la construction des anneaux de groupes intégraux est un foncteur qui est adjoint à gauche du foncteur "groupe des unités" mentionné plus tôt. De même, la notion catégorique de produit peut être décrite comme un adjoint ; si chaque paire d'objets de C a un produit, alors l'assignation $(A, B) \mapsto A \times B$ définit un foncteur du produit cartésien $C \times C$ vers C qui est adjoint à droite du "foncteur diagonal" $A \mapsto (A, A)$.

Nous disons qu'une catégorie C est cartésienne close si elle a des produits comme ci-dessus et, pour chaque objet A , le foncteur $(-) \times A : C \rightarrow C$ a un adjoint à droite, le foncteur exponentiel $(-) ^A$. Set est cartésienne close, avec $BA = Set(A, B)$, car il est bien connu que les fonctions $C \rightarrow Set(A, B)$ correspondent aux fonctions $C \times A \rightarrow B$. De même, Cat est cartésienne close, les exponentielles étant les catégories de foncteurs définies plus tôt. L'importance de la clôture cartésienne est que nous pouvons considérer les exponentielles B^A comme des équivalents "internes" des hom-ensembles $C(A, B)$, et (en reformulant convenablement des notions telles que la représentabilité) nous pouvons supprimer la dépendance de ce que nous avons fait à l'égard de notions ensemblistes comme la petitesse locale - c'est-à-dire que nous pouvons donner un développement véritablement autonome de la théorie des catégories.

Cette ligne de développement a été inaugurée par William Lawvere, qui, en 1963 [5], a donné un certain nombre d'axiomes purement catégoriques caractérisant les propriétés catégoriques de Set ; plus récemment, Lawvere et Myles Tierney ont introduit la notion de topos, qui a été à l'origine de certains des développements catégoriques les plus passionnants de ces dernières années. Un topos est simplement une catégorie cartésienne fermée qui satisfait une version "autonome" de la représentabilité du foncteur ensemble-des-parties (comme Lawvere l'a fait remarquer, cette dernière condition pour Set n'est que la version catégorique de l'axiome de compréhension des théoriciens des ensembles !) ; pour le théoricien des catégories, il sert d'"univers de discours" au sein duquel il est capable d'effectuer catégoriquement toutes sortes de constructions précédemment comprises uniquement pour la catégorie des ensembles.

Le plus frappant dans ce développement est que le concept de topos avait été introduit plus tôt (certes avec une définition différente, qui dépendait de la théorie des ensembles) par l'école française de géomètres algébriques dirigée par Alexander Grothendieck [6] ; et en effet des exemples de topos surgissent naturellement en géométrie algébrique, en topologie et ailleurs. Ainsi, par la théorie des topos, il est possible pour le théoricien des catégories d'apporter des idées de logique (telles que

le théorème de complétude) directement à la géométrie et à l'algèbre, ou inversement d'appliquer des notions topologiques telles que la localisation à des problèmes de logique. De cette manière, on peut voir que la théorie des catégories réalise effectivement son ambition de rassembler différents domaines de la recherche mathématique et d'exposer leur unité fondamentale.

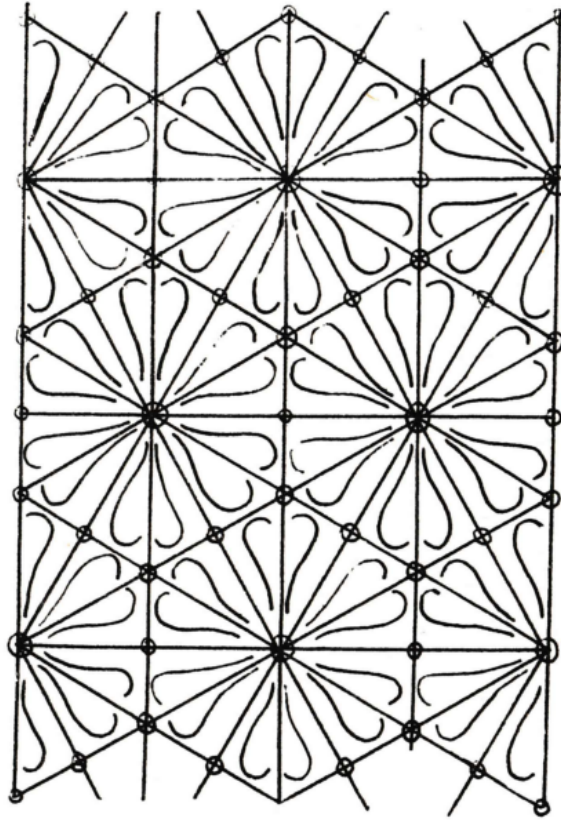
Je termine avec quelques suggestions de lectures complémentaires. [7] est la bible du théoricien des catégories, écrite par l'un des fondateurs du sujet, qui est aussi un excellent exposant ; mais elle peut être assez difficile pour l'étudiant débutant. [8] est une tentative louable, et pas totalement infructueuse, d'écrire un texte de théorie des catégories pour ceux qui sont beaucoup moins familiers avec les mathématiques abstraites. [9] est une introduction courte et lisible aux merveilles de la théorie des topos.

Références

- [1] S. Eilenberg et S. Mac Lane, *Trans. A.M.S.* 58 (1945), 231-294.
- [2] par ex. S. Mac Lane, *Springer Lecture Notes* 106 (1969), 192-200.
- [3] Comme presque tous les calembours catégoriques, celui-ci est de Peter Freyd. Il provient de son délicieux livre *Abelian Categories* (Harper & Row, 1964), malheureusement épuisé aujourd'hui.
- [4] Attribué à N. Yoneda, *J. Fac. Sci. Tokyo* 7 (1954), 193-227. En fait le lemme n'y apparaît pas explicitement, et Lawvere a fait remarquer que l'idée du lemme est plus ancienne que la théorie des catégories elle-même - elle remonte au théorème de représentation de Cayley pour les groupes.
- [5] F. W. Lawvere, *Proc. Nat. Acad. Sci.* 52 (1964), 1506-1511.
- [6] J. Giraud, *Séminaire Bourbaki* 256 (1963).
- [7] S. Mac Lane, *Categories for the Working Mathematician*, Springer Graduate Texts no. 5, 1971.
- [8] M. Arbib and E. G. Manes, *Arrows, Structures and Functors*, Academic Press 1975.
- [9] G. C. Wraith, *Bull. Inst. Math. and its Applications* 12 (1976), 115-119.

On Wallpaper

par Dr S. J. Patterson



Mon intérêt pour le papier peint date de mes premiers jours. La plupart des pièces de la maison de mon enfance étaient tapissées de motifs floraux, qui non seulement impliquaient les décalages et demi-décalages de l'art scolaire, mais qui pouvaient être tournés, ou réfléchis avec un changement de couleur, et ainsi de suite. Démêler ces motifs était l'une des façons de passer de longues heures au lit avec la rougeole ou une autre maladie infantile. Il s'avère que les motifs sont pleins de symétries ; mais il n'y a que dix-sept papiers peints essentiellement différents. Ainsi l'œil de l'adulte voit-il le monde de l'enfant.

Pourtant, ayant commencé, je vais poursuivre en décrivant, du point de vue du mathématicien, les motifs de papier peint. Mieux, nous allons examiner certains groupes d'isométries du plan. L'isométrie générale du plan a la forme $x \rightarrow Ax + B$ où $A \in \mathbb{O}(2)$, le groupe des matrices orthogonales 2×2 (qui inclut les réflexions), et $B \in \mathbb{R}^2$. Nous considérerons un groupe G de telles applications ; si $g \in G$ nous écrirons $g(x) = A(g)x + B(g)$. Si nous vérifions $g_1(g_2(x)) = (g_1g_2)(x)$ alors $A(g_1g_2) = A(g_1)A(g_2)$, et $B(g_1g_2) = B(g_1) + A(g_1)B(g_2)$. En particulier, $A : G \rightarrow \mathbb{O}(2)$ est un homomorphisme. Soit $G_0 = \ker(A)$. Si alors, $g_1, g_2 \in G_0$ nous avons que $B(g_1g_2) = B(g_1) + B(g_2)$, de sorte que $B : G_0 \rightarrow \mathbb{R}^2$ est aussi un homomorphisme.

Le moment est venu de poser l'hypothèse “papier peint” sur G . C'est que l'image de G sous B doit être un réseau, désormais noté L . (Un réseau est un sous-groupe de $\mathbb{R}^2$ de la forme $(m + nv : m, n \in \mathbb{Z})$, où $u, v \in \mathbb{R}^2$ sont linéairement indépendants.) C'est l'ensemble des translations qui est si caractéristique du papier peint.

Si maintenant $g \in G$, $h \in G_0$ alors $(ghg^{-1})(x) = x + A(g)B(h)$; en particulier, si $l \in L$ alors $A(g)(l) \in L$. Il s'ensuit (et ce n'est pas trop difficile) que $A(G)$ est un sous-groupe fini de $O(2)$. Il s'avère que le fait que $A(G)$ préserve le réseau L impose des restrictions très fortes sur ce que $A(G)$ et L peuvent être. Si $A(g) = (I)$, ou $\{I, -I\}$ alors il n'y a aucune restriction sur L . Si $A(g) = (I, R)$ ou $\{I, I + R\}$, où R est une réflexion, alors nous pouvons prendre R comme $R : (x, y) \rightarrow (x, -y)$ et L engendré par $(1, 0)$ et $(0, t)$ ou $(1, t)$ où $t > 0$. Sinon $A(G) \cap SO(2)$ est d'ordre 3, ou 4; alors le réseau L est engendré par $(1, 0)$ et $(1, \sqrt{3}/2)$ dans les deux premiers cas, et par $(1, 0)$ et $(0, 1)$ dans le troisième.

Maintenant, nous avons une bonne idée des $A(G)$ et L possibles; nous devons maintenant essayer de voir comment ceux-ci peuvent être assemblés pour former le groupe G . C'est plutôt plus subtil qu'on ne le suppose d'abord. Soit l'ordre de $A(G)$ égal à n , et soient $g_1, \dots, g_n \in G$ tels que les $A(g_j)$ ($1 < j < n$) constituent $A(G)$. Nous allons examiner $b = \frac{1}{n} \sum_{j=1}^n B(g_j)$. Notez que si $g \in G$ alors gg_j doit être de la forme $h(j, g) \in G_0^3$ et σ est une permutation de $\{1, 2, \dots, n\}$. Alors $A(g)b = \frac{1}{n} \sum_j A(g)B(g_j) = \frac{1}{n} \sum_j (B(gg_j) - B(g)) = \frac{1}{n} \sum_j B(h(j, g)gg_j) - B(g)$. $\Sigma = b - B(g) + \frac{1}{n} \sum_{\Sigma} -lB(h(j, g))$. Le dernier terme du membre de droite est de la forme $n^{-1}L(g)$, avec $1(g) \in L$. Ainsi $B(g) = b - A(g)b + n^{-1}L(g)$. Si nous remplaçons maintenant x par $x + b$ (en changeant l'origine dans le papier peint) nous voyons que nous pouvons supposer que $B(g)$ a la forme $n^{-1}L(g)$. En général, nous ne pouvons pas supposer que $B(g)$ est dans L .

Ceci nous permet essentiellement de réduire le problème à un problème fini. Pour le voir, nous considérerons le groupe quotient L/nL qui est fini, d'ordre n^2 . Soit $c \in A(G)$; alors $c = A(g)$ pour un certain $g \in G$ et nous posons $1_0(c) = 1(g)$; ceci, modulo nL , est bien défini. Ainsi les 1_0 satisfont $1_0(8_1c_2) = 1_0(c_1) + c_11_0(c_2) \pmod{nL}$. Un argument similaire à celui ci-dessus montre qu'il existe au plus un G qui donne lieu à un L_0 avec cette propriété. La question réciproque, à savoir si, étant donné $G, c \in O(2)$, L un réseau préservé par G , et 1_0 comme ci-dessus, il existe un groupe de papier peint G avec ce réseau, $A(G) = G_1$ et le 1_0 donné (à une notion d'équivalence près), est un peu plus difficile mais peut aussi être résolue.

Quand tout est dit et fait, on trouve 17 motifs essentiellement différents, qui sont reproduits ici. Dans ceux-ci, le motif répété est un simple $\mathcal{I}$; même avec cela on trouve, comme dans un kaléidoscope, des motifs attrayants. Les centres de rotation sont marqués par $\bullet$.

Dans l'art mauresque, où aucune représentation de la nature n'était autorisée (car il était blasphématoire de rivaliser avec le créateur), les dix-sept motifs étaient connus et (réputés) utilisés dans l'Alhambra à Grenade.

Enfin, selon l'éthique moderne, je dois ajouter que tout cela peut être généralisé à 3 (et plus) dimensions; les groupes sont alors appelés groupes cristallographiques. Ils jouent un rôle majeur en cristallographie, qui est l'un des principaux domaines d'application de la théorie des groupes. Il y

en a 219 en dimension 3, et 4783 en dimension 4. Au-delà, on en sait peu.

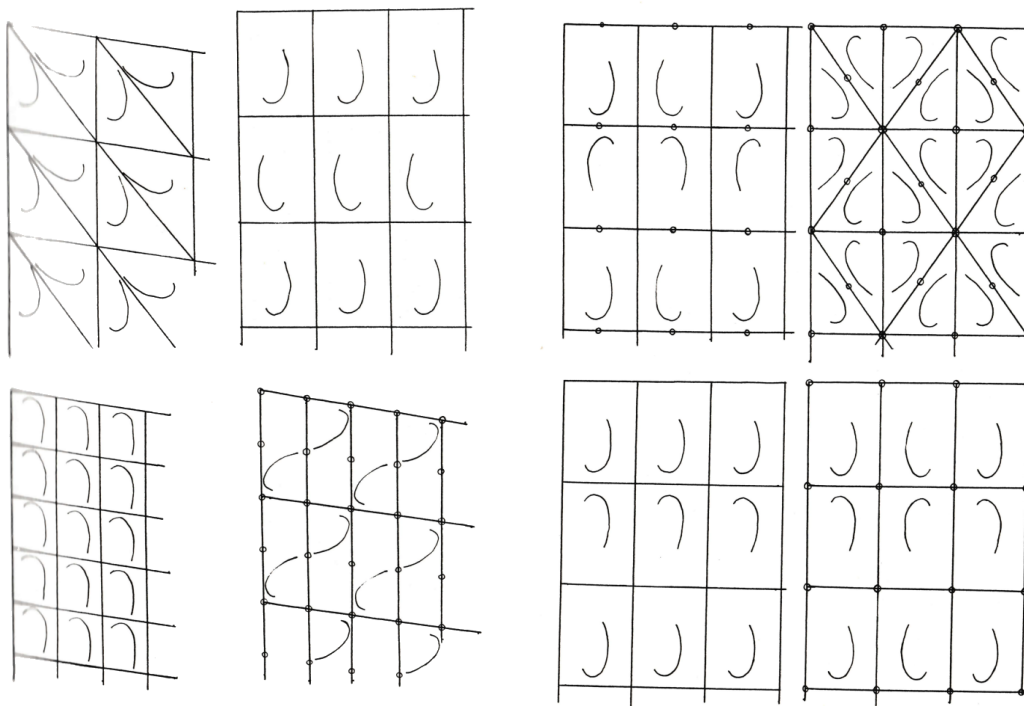
Le problème analogue peut être posé pour tout espace sur lequel agit un grand groupe de transformations. Ainsi, on considère les tessellations du plan hyperbolique, qui sont étroitement liées aux surfaces de Riemann, et des espaces de dimension supérieure, qui touchent à certains des problèmes les plus profonds de la théorie des nombres et des mathématiques en général.

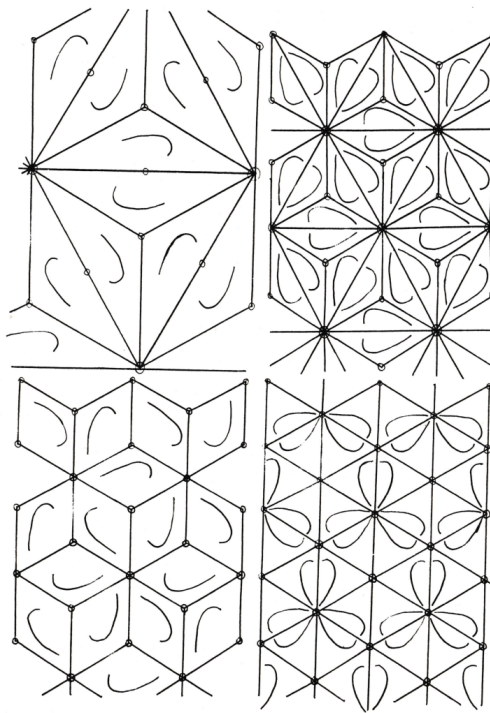
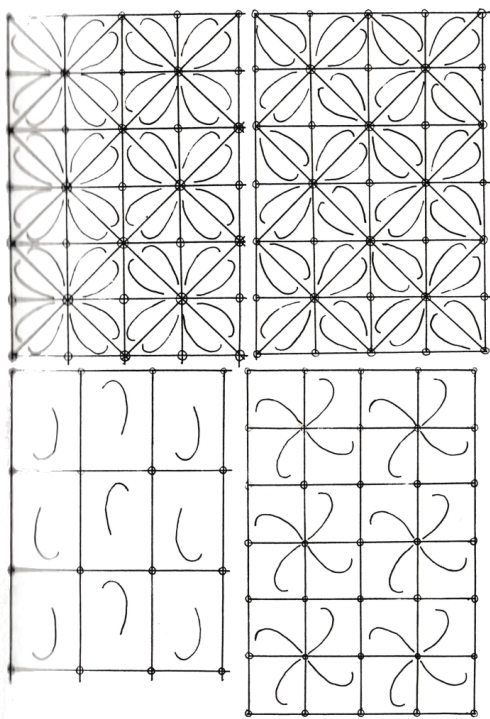
4. Littérature

L. Fjes-Toth. *Regular figures*. Pergamon, 1964.

R.L.E. Schwarzenberger. *Crystallography in spaces of arbitrary dimension*. Proc. Camb. Phil. Soc. 76 (1974) 23-32.

H. Weyl. *Symmetry*. Princeton, 1952.





Nous supposons que le soleil est infiniment plus massif que la planète. La première loi de Kepler stipule que l'orbite de la planète est une ellipse avec le soleil à un foyer. Sur la Fig. 1, F est le foyer occupé par le soleil, tandis que E est le foyer vide. IJ est un diamètre quelconque de l'ellipse et OC est le demi-petit axe. SK et HT sont les tangentes à l'ellipse en I et J respectivement. Le reste de la construction sur la figure est explicite.

$$v_I \cdot FG = v_J \cdot FH = v_J \cdot EK = v_C \cdot OC. \quad (1)$$

Il existe également une propriété importante de l'ellipse qui stipule que le produit des deux perpendiculaires focales à la tangente en tout point I est constant et égal au carré du demi-petit axe (cf. [1] et [2]),

$$FG \cdot EK = OC^2 \quad (2)$$

En éliminant OC entre (1) et (2),

$$v_I \cdot v_J = v_C^2 \quad (3)$$

Autrement dit, cela donne le théorème suivant :

Théorème 1 La vitesse d'une planète à l'extrémité du petit axe est égale à la moyenne géométrique des vitesses aux extrémités d'un diamètre quelconque.

Ce théorème a été démontré pour deux diamètres particuliers. Freeman [3] a démontré le théorème pour le cas particulier où le diamètre IJ est le grand axe de l'ellipse, tandis que Tan [4] démontre le même lorsque I et J sont aux extrémités des latus rectum de l'ellipse.

Il est également instructif de déterminer le rapport des vitesses aux extrémités d'un diamètre. À partir des triangles semblables FGS et FHT,

$$FH/FG = FT/FS.$$

Donc, d'après (1),

$$v_I/v_J = FH/FG = FT/FS.$$

Ainsi nous avons le second théorème :

Théorème 2 Les vitesses aux extrémités d'un diamètre sont inversement proportionnelles à la distance entre le foyer et les points où les tangentes à l'ellipse rencontrent le grand axe prolongé.

À partir de (3) et (4), nous pouvons déterminer les vitesses aux extrémités de tout diamètre en fonction de la vitesse à l'extrémité du petit axe,

$$v_I = (FT/FS)^{\frac{1}{2}} v_C \quad (5)$$

Les vitesses peuvent être calculées explicitement en fonction de v_C et de l'excentricité de l'orbite, e , pour des points spéciaux de l'ellipse orbitale tels que l'aphélie, le périhélie et les extrémités des latus rectum. Sur la Fig. 2, A désigne l'aphélie et P le périhélie de l'orbite. FL est le demi-latus rectum et EM le demi-latus rectum du foyer vide. Nous avons les relations bien connues

$$AF = a(1 + e), PF = a(1 - e),$$

$$OC = b = a(1 - e^2)^{\frac{1}{2}}, LF = EM = a(1 - e^2),$$

a et b étant respectivement les demi-grand et demi-petit axes. Clairement,

$$v_P/v_A = AF/PF = (1 + e)/(1 - e),$$

un résultat cité par Freeman [3]. Aussi,

$$v_P = (AF/PF)^{\frac{1}{2}} v_C = (1 + e)^{\frac{1}{2}}/(1 - e)^{\frac{1}{2}} v_C. \quad (7)$$

$$v_A = (PF/AF)^{\frac{1}{2}} v_C = (1 - e)^{\frac{1}{2}} / (1 + e)^{\frac{1}{2}} v_C. \quad (8)$$

(7) et (8) donnent les vitesses de la planète à son périhélie et à son aphélie respectivement.

De plus, nous savons que la pente de la tangente au *latus rectum* est égale à l'excentricité de l'ellipse (cf. [5]). Par conséquent, dans le triangle LSF, $FS = FL/e = a(1 - e^2)/e$. Aussi, $FT = FE + ET = FE + FS$. Après un peu d'algèbre, $FT = a(1 + e^2)/e$.

Par conséquent, $v_L/v_M = FT/FS = (1 + e^2)/(1 - e^2)$. Aussi, pour les vitesses aux extrémités des *latus rectum*, nous avons

$$v_L = (FT/FS)^{\frac{1}{2}} \cdot v_C = (1 + e^2)^{\frac{1}{2}} / (1 - e^2)^{\frac{1}{2}} \cdot v_C, \quad (9)$$

$$v_M = (FS/FT)^{\frac{1}{2}} \cdot v_C = (1 - e^2)^{\frac{1}{2}} / (1 + e^2)^{\frac{1}{2}} \cdot v_C. \quad (10)$$

Les équations (7) à (10) présentent des relations remarquablement symétriques entre les vitesses de la planète et ses positions dans son ellipse orbitale.

5. Références

- [1] G. Salmon, *A Treatise on Conic Sections*, p. 179-180, Chelsea, New York (1954).
- [2] E.H. Lockwood, *A Book of Curves*, p. 15, Cambridge University Press (1961).
- [3] I.M. Freeman, *Amer. J. Phys.*, 45, 585-586 (1977).
- [4] A. Tan, à paraître dans *Amer. J. Phys.*, 47, (1979).
- [5] P. Van De Kamp, *Elements of Astrodynamics*, p. 13-15, Freeman, San Francisco (1964).
- [6] *The Geometry Problem Solver*, p. 700, Research and Education Association, New York (1977).

Comment simplifier les mathématiques ? par P. Freyd

Un certain membre de l'American Mathematical Society publie un résumé tous les mois environ, prétendant démontrer l'incohérence de ZFC, c'est-à-dire que la théorie des ensembles de Zermelo-Fraenkel contredit l'axiome du choix. Le schéma est toujours le même : il énonce un "théorème" dans ZF puis construit un contre-exemple dans ZFC. Ce qui rend tout cela insatisfaisant, c'est que son "théorème" a toujours un contre-exemple évident qui vous saute aux yeux et n'a jamais besoin de l'axiome du choix.

Je propose de combler cette lacune, car cela est rendu très nécessaire par l'état actuel de la littérature.

Je suis tombé sur cette "preuve" particulière lorsque j'étais étudiant de premier cycle, et il a fallu à moi et à deux de mes professeurs la meilleure partie d'une journée pour trouver mon erreur. Depuis, la personne la plus rapide que j'aie trouvée a été P.T. Johnstone, qui a mis environ dix secondes. Mais je vous préviens - j'ai vu de solides mathématiciens réduits à l'état de babillards croyant à l'incohérence de leur propre profession. (Comme vous le remarquerez, vous n'avez pas besoin de connaître Zermelo-Fraenkel, juste la théorie naïve des ensembles. ZF peut être considéré comme une axiomatisation particulière de la théorie naïve des ensembles. Le problème est que Gödel a prouvé que l'incohérence de ZFC implique l'incohérence de ZF, ce qui implique à son tour que notre notion fondamentale de la théorie des ensembles est incohérente.)

Théorème

Le groupe circulaire, $\mathbb{R}/\mathbb{Z}$, n'a que deux automorphismes.

Preuve

Il suffit clairement de montrer que tous les automorphismes sont continus. Soit G un groupe compact quelconque et supposons que θ est un automorphisme discontinu. Soit G_2 le groupe compact obtenu en conservant le même ensemble sous-jacent et la même structure de groupe que G_1 mais topologisé de sorte que $C \subset G_2$ soit fermé si et seulement si $\theta(C)$ est fermé dans G_1 . Nous n'avons besoin de montrer que

Lemme

Un groupe possède au plus une topologie compacte.

Preuve

Étant donné G_1 et G_2 comme décrits ci-dessus, soit G_{12} le groupe topologique obtenu en conservant la même structure de groupe mais topologisé de sorte que $C \subset G_{12}$ soit fermé si C est fermé à la fois dans G_1 et G_2 . De toute évidence, G_{12} est T_1 (les points sont fermés) et tout T_1 groupe est T_2 (séparé). L'application identité donne des applications continues $G_i \rightarrow G_{12}$ pour $i = 1, 2$ et une application continue bijective d'un espace compact vers un espace séparé est un homéomorphisme. C.Q.F.D.

Avec l'axiome du choix, nous pouvons décomposer $\mathbb{R}/\mathbb{Z}$ (comme pour tout groupe abélien divisible) en somme directe de son sous-groupe de torsion, T , et d'un sous-groupe sans torsion divisible; utiliser la divisibilité de T pour obtenir une rétraction $T \rightarrow \mathbb{R}/\mathbb{Z} \twoheadrightarrow T$, puis $\mathbb{R}/\mathbb{Z} = T \oplus F$ où F est le noyau de $\mathbb{R}/\mathbb{Z} \twoheadrightarrow T$. Tout groupe abélien sans torsion divisible peut être considéré comme un espace vectoriel sur les rationnels. Nous pouvons choisir une base et il est clair que celle-ci doit être de l'ordre du continu (le sous-groupe de torsion est dénombrable). Toute permutation de la base induit un automorphisme de F , donc de $\mathbb{R}/\mathbb{Z}$. Le groupe d'automorphismes dans ZFC possède $2^{2^{\aleph_0}}$ éléments.

Cette "preuve" est satisfaisante en ce sens qu'il faut réellement une sorte d'axiome du choix pour construire un troisième automorphisme. Étant donné

un homomorphisme quelconque $f : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$; considérons l'application $g : \mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z} \xrightarrow{f} \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ des réels vers les nombres complexes, où $\mathbb{R} \rightarrow \mathbb{R}/\mathbb{Z}$ est le revêtement canonique et $\mathbb{R}/\mathbb{Z} \rightarrow \mathbb{C}$ est le plongement canonique (il envoie x sur $e^{2\pi i x}$). L'axiome de mesurabilité, connu comme étant cohérent avec ZF, dit que toute fonction de $\mathbb{R}$ vers $\mathbb{C}$ est mesurable (cela a un sens de l'intégrer). Pour tout $n \in \mathbb{Z}$ définissons $h(x) = g(x)e^{2\pi i n x}$. Parce que $h(x+y) = h(x)h(y)$ nous avons $\int_0^1 h(x+y)dx = h(y) \int_0^1 h(x)dx$. Mais h est périodique et nous avons $\int_0^1 h(x+y)dx = \int_y^{1+y} h(x)dx = \int_y^1 h(x)dx + \int_1^{1+y} h(x)dx = \int_y^1 h(x)dx + \int_0^y h(x)dx = \int_0^1 h(x)dx$.

De $h(y) \int_0^1 h(x)dx = \int_0^1 h(x)dx$ nous concluons que soit $h \equiv 1$ soit $\int_0^1 h(x)dx = 0$. Par conséquent, pour tout $n \in \mathbb{Z}$ soit $g(x) = e^{-2\pi i n x}$ pour tout x soit $\int_0^1 g(x)e^{2\pi i n x}dx = 0$. Mais le théorème fondamental des séries de Fourier dit que $\int_0^1 |g|^2 dx = \sum |\int_0^1 g(x)e^{2\pi i n x}dx|^2$. Puisque $|g| \equiv 1$ nous concluons que pour un certain $n \in \mathbb{Z}$, $g(x) = e^{-2\pi i n x}$ pour tout x et que l'application originale $f : \mathbb{R}/\mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ est juste la multiplication par $-n$. Il est cohérent avec ZF de supposer que le groupe de $\mathbb{R}/\mathbb{Z}$ a deux éléments.

Avec un peu de travail, nous pouvons remplacer le groupe circulaire par le corps des nombres complexes : l'axiome de déraisonnabilité implique que le groupe de Galois de $\mathbb{C}$ a deux éléments; l'axiome du choix fait qu'il a $2^{2^{\aleph_0}}$ éléments.

Une application génétique des coordonnées homogènes

A. W. F. Edwards

Dans le numéro d'Eureka de l'année dernière, le professeur C. A. B. Smith a donné un compte rendu de certains théorèmes dans l'étude mathématique de l'évolution (*Mathématiques et Évolution*, p. 5). Le plus simple est le théorème de Hardy-Weinberg, selon lequel les proportions des génotypes AA , Aa et aa parmi les zygotes issus d'une génération, après une reproduction aléatoire dans une génération parentale où les fréquences géniques sont pA et qa ($p + q = 1$), sont p^2 , $2pq$ et q^2 respectivement. En 1926, le professeur Bruno de Finetti, aujourd'hui célèbre pour ses contributions à la théorie des probabilités, a fait remarquer que si l'on représente une population avec des proportions génotypiques u, v et w , respectivement, par un point dans le plan, en utilisant des coordonnées homogènes et (puisque $u + v + w = 1$) un triangle équilatéral de hauteur unité comme triangle de référence (Figure 1), toute population de zygotes qui satisfait la loi de Hardy-Weinberg se situera sur la parabole $v^2 = 4uw$. Haldane et Moshinsky ont montré plus tard que si une population est consanguine à un degré F , auquel cas les proportions génotypiques sont $p^2 + Fpq$, $2pq(1 - F)$ et $q^2 + Fpq$, elle se situera sur la parabole $4uw - v^2 = F(2u + v)(2w + v)$. La Figure 2 montre cette parabole pour $F = \frac{1}{2}$ et pour $F = 0$, qui est le cas de Hardy-Weinberg.

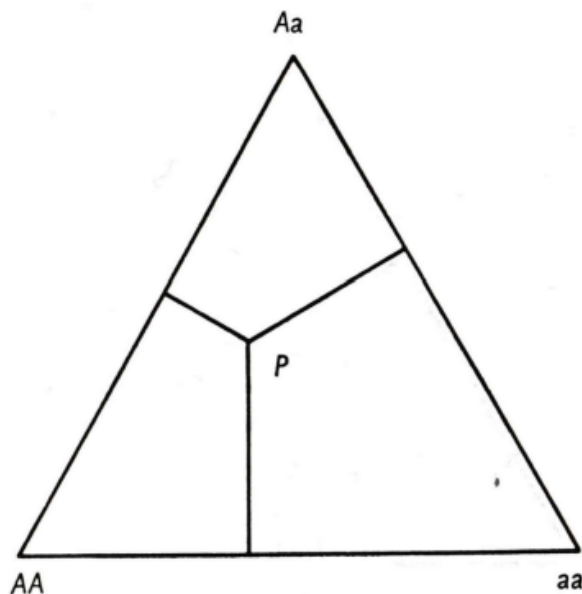


FIG. 1. Un diagramme de de Finetti pour un locus diallélique. Les longueurs des perpendiculaires du point P aux trois côtés du triangle équilatéral représentent les proportions des trois génotypes AA , Aa et aa dans la population.

Dans cet article, j'indique comment certains des théorèmes élémentaires de l'évolution peuvent être prouvés et éclairés en utilisant des coordonnées homogènes. Les résultats ont été obtenus par le Dr C. Cannings et moi-même, et ont été publiés sous le titre *Natural Selection and the de Finetti Diagram* dans les *Annals of Human Genetics*, 31, 421 (1968).

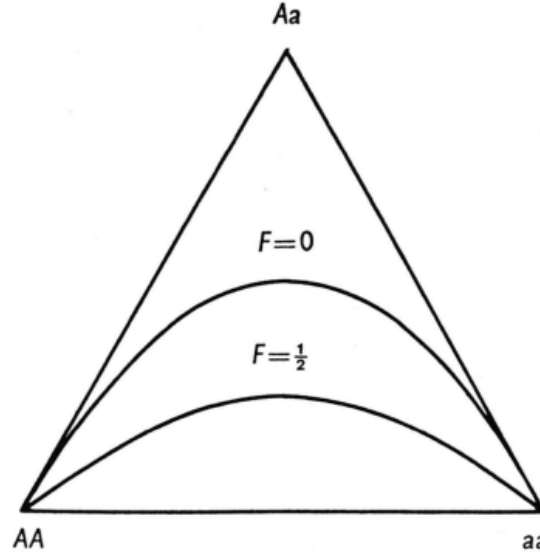


FIG. 2. Les lieux des populations avec des proportions génotypiques en équilibre de Hardy-Weinberg ($F = 0$) et en équilibre sous consanguinité de degré $F = 1/2$.

La “parabole de Hardy-Weinberg” passe par les sommets de la base du triangle de référence, et y est tangente aux côtés. Supposons maintenant que les zygotes de chaque génération soient différenciellement éliminés par la sélection naturelle avant qu’ils ne deviennent eux-mêmes parents, de sorte que les proportions génotypiques zygotiques u, v et w deviennent des proportions adultes au, hv et bw , pour les génotypes AA, Aa et aa respectivement.

Puisque, par le théorème de Hardy-Weinberg, $v^2 = 4uw$, après sélection le point représentant la population adulte satisfait $abv^2 = 4h^2uw$. Cette équation est évidemment celle d’une conique symétrique par rapport à l’axe vertical du diagramme, $u = w$, et dépendant d’une seule constante, $k = h^2/ab$. En considérant l’intersection de cette conique et de la droite à l’infini $u + v + w = 0$, il est apparent qu’elle prend les formes suivantes :

$$h^2 < ab, k < 1 : \text{ellipse},$$

$$h^2 = ab, k = 1 : \text{parabole},$$

$$h^2 > ab, k > 1 : \text{hyperbole}.$$

Dans le cas particulier $h^2 = ab$, nous disons qu’il n’y a pas de dominance dans l’épaisseur à une échelle multiplicative, car l’“épaisseur” h de l’hétérozygote Aa est alors la moyenne géométrique des épaisseurs des deux homozygotes AA et aa . La courbe “après sélection” coïncide alors avec la parabole de Hardy-Weinberg “avant sélection”.

Dans le cas général $k \neq 1$, la conique touche encore les côtés du triangle aux sommets de la base, puisque poser soit $u = 0$ soit $w = 0$ induit la racine double $v^2 = 0$. Il s’ensuit que les coniques de la famille ont un double contact entre elles, et qu’elles ne se croisent pas ailleurs. Si chaque point de la courbe avant sélection est joint à son point correspondant sur la courbe après sélection,

la sélection peut être visualisée graphiquement (Figure 3). En partant du point représentant les proportions zygotiques avant sélection (P), la population se déplace vers le point correspondant Q sous l'effet de la sélection. La reproduction aléatoire réassortit alors les gènes de sorte que les proportions de Hardy-Weinberg sont restaurées sans changer la fréquence génique, la population revenant ainsi verticalement à la parabole de Hardy-Weinberg en P'. Le cycle se répète alors, la population suivant un chemin en dents de scie vers son équilibre ultime.

Q1 : Prouver que les lignes verticales sont des lignes de fréquence génique constante, et qu'une telle ligne divise la base dans le rapport $p : q$.

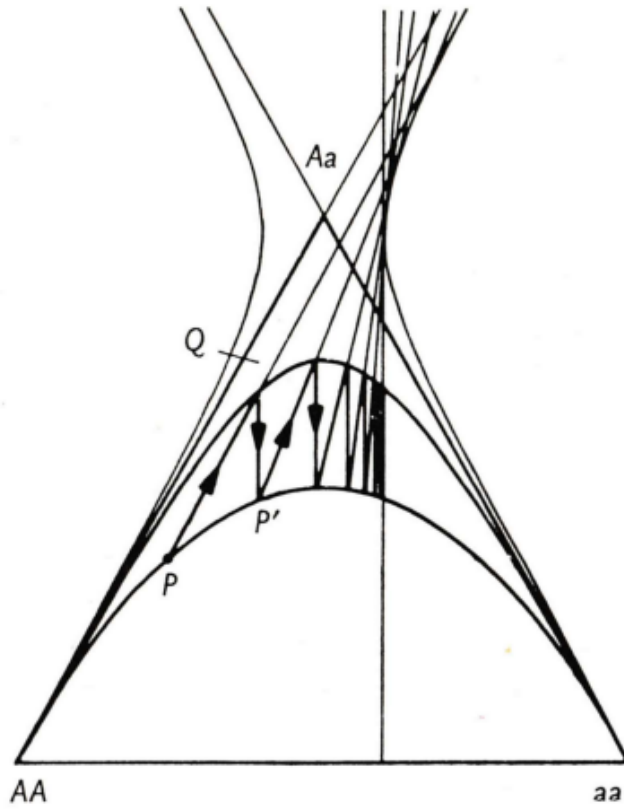


FIG. 3. La sélection vue graphiquement. La courbe passant par P et P' est la courbe avant sélection, et celle passant par Q la courbe après sélection. Les valeurs sélectives utilisées dans cet exemple sont $(1/4, 1, 1/2)$, la fréquence génique d'équilibre étant ainsi de 0,4 pour le gène A. Voir le texte pour plus d'explications.

Afin de construire les lignes joignant les points correspondants sur les courbes avant et après sélection, nous pouvons trouver leur enveloppe.

Q2 : Prouver que l'enveloppe des lignes joignant (x, y, z) à (ax, hy, bz) , où $y^2 = 4xz$, est $v^2 = 4kuw$, où

$$k = 4(h - a)(b - h)/(a - b)^2.$$

Ce résultat montre que les enveloppes sont aussi des membres de la famille de coniques que nous avons déjà considérée. Mais notez comment cette fois $k = 1$ lorsque $h = (a + b)/2$, ou lorsqu'il n'y a pas de dominance additive dans l'épaisseur. Les lignes de jonction sont alors tangentes à la

parabole de Hardy-Weinberg. Lorsque $h = a$ mais $h \neq b$, comme lorsque le gène A est complètement dominant sur a en épaisseur, les lignes forment un faisceau passant par le sommet “ aa ”.

Nous pouvons donc construire le diagramme pour toutes valeurs sélectives a, h et b , et tout point de départ, si nous avons à disposition un graphique de la famille de coniques $v^2 = 4kuw$, pour tout k . Celui-ci est donné à la Figure 4, qui a été tracé par ordinateur.

Q3 : Trouver une preuve simple que $v^2 = uw$ est un cercle.

Plusieurs des résultats bien connus de la théorie de la sélection sont maintenant facilement dérivés. Premièrement, il est clair que la sélection est toujours monotone en fréquence génique ; deuxièmement, il y aura équilibre chaque fois que la ligne joignant $(p^2, 2pq, q^2)$ et $(ap^2, 2hpq, bq^2)$, les points avant et après sélection, intercepte la base en $(p, 0, q)$, car alors la sélection laisse la fréquence génique inchangée.

Q4 : Prouver que la condition pour que la fréquence génique reste inchangée de cette manière est $p/q = (h - b)/(h - a)$.

Ce résultat établit la fréquence génique d’équilibre. Troisièmement, la stabilité de l’équilibre est facilement déterminée : car p/q se situe entre sa valeur dans la génération précédente et $(h - b)/(h - a)$, étant leur moyenne géométrique. Ainsi toutes les lignes passant par les points correspondants avant et après sélection s’inclinent vers la verticale passant par le point d’équilibre (non trivial), étant plus proches de lui au sommet du diagramme. Il s’ensuit que l’équilibre sera stable si la conique après sélection est au-dessus de la parabole de Hardy-Weinberg, et instable si elle est en dessous. Mais nous savons déjà que ces événements correspondent à $k > 1$, $h^2 > ab$ (hyperbole) et $k < 1$, $h^2 < ab$ (ellipse) respectivement, établissant ainsi les règles bien connues pour trouver la stabilité de l’équilibre.

Pour différentes valeurs de V , ceci est évidemment une famille de lignes parallèles sur le diagramme, et puisque deux valeurs différentes de V ne donnent pas la même ligne, elle change de façon monotone à mesure que les lignes sont parcourues. Pour que le point d’équilibre soit une valeur stationnaire de V , il suffit de montrer que la tangente à la parabole de Hardy-Weinberg au point d’équilibre est l’une des lignes de la famille de parallèles :

Q5 : Prouver que la tangente à $v^2 = 4uw$ en $(p^2, 2pq, q^2)$ où $p/q = (h - b)/(h - a)$, est l’une de la famille $au + hv + bw = V$.

Il est facile de montrer que l’épaisseur moyenne augmente plutôt que diminue à mesure que l’équilibre est approché.

Enfin, le diagramme nous permet de voir pourquoi, s’il n’y a soit pas de dominance additive, soit pas de dominance multiplicative dans l’épaisseur, l’épaisseur moyenne des zygotes d’une génération est égale à l’épaisseur moyenne de leurs parents dans la génération précédente : c’est-à-dire que la reproduction ne change pas l’épaisseur moyenne. Car dans le premier cas, il est facile de montrer que les lignes de V constant sont verticales, et dans le second cas, la courbe après sélection, ou

parentale, coïncide avec la parabole de Hardy-Weinberg. Dans les deux cas, le retour vertical à la parabole de Hardy-Weinberg ne change pas l'épaisseur moyenne.

