IV

LE CALCUL DES PROBABILITÉS ET LA MÉTHODE DES MAJORITÉS

Nous donnerons, d'une manière générale, le nom de *méthode des majorités* au procédé qui consiste à regarder comme pratiquement valable l'avis exprimé par la majorité, après dépouillement des avis ou des opinions d'un nombre plus ou moins grand de personnes. Nous ne nous attarderons pas, pour l'instant, à énumérer et discuter les formes diverses que peut prendre la méthode, ni à essayer de classer ses nombreuses applications, réalisées ou possibles. La question qui nous occupera principalement est la suivante : la connaissance du calcul des probabilités peut-elle être de quelque utilité à ceux qui emploient ou qui critiquent la méthode des majorités?

I

Joseph Bertrand a rassemblé dans son *Calcul des Probabilités* (XLIII à XLIX et 319 à 327) les arguments qui opposent, pour ainsi dire, la question préalable à toute introduction du calcul des probabilités en ces matières. Quelques citations feront comprendre le point de vue auquel il se place.

L'application du calcul aux décisions judiciaires est, dit Stuart Mill, le scandale des Mathématiques. L'accusation est injuste. On peut peser du cuivre et le donner pour or, la balance reste sans reproche. Dans leurs travaux sur la théorie des jugements, Condorcet, Laplace et Poisson n'ont pesé que du cuivre.

.... « Condorcet a pris possession de l'univers moral pour le soumettre au calcul. » C'est la louange qu'on lui a donnée; on s'est demandé si c'est après l'avoir lu. Dans son livre sur *la probabilité des jugements*, il se propose d'abord deux problèmes. Premièrement : Quel est, pour chaque jugement et pour chaque juge, la probabilité de rencontrer juste? En second lieu, quelle est la probabilité d'erreur à laquelle la société peut se résigner sans alarmes?

La première question lui semble facile.

« Je suppose, dit Condorcet, que l'on ait choisi un nombre d'hommes véritablement éclairés et qu'ils prononcent sur la vérité ou la fausseté de la décision. Si, parmi les décisions de ce tribunal d'examen, on n'a égard qu'à celles qui ont obtenu une certaine pluralité, il est aisé de voir qu'on peut, sans erreur sensible, les regarder comme certaines. »

C'est un concile infaillible, tout simplement, qu'il définit et prétend convoquer. Sans douter, il hésite; non que les hommes véritablement éclairés soient rares, gardons-nous de le croire, mais leur temps est précieux. Pour l'épargner, Condorcet propose une seconde méthode dont Poisson, plus tard, n'a pas aperçu l'illusion. La probabilité d'erreur étant supposée pour un juré, on peut, sans augmenter leur nombre, la diminuer sans limite pour l'ensemble.

.... Si, comme le demande très sérieusement Cournot, on invitait le greffier à noter, après chaque jugement, l'opinion de chacun des juges, pour appliquer, quand les chiffres seront nombreux, la formule qui donne leur mérite, la perspicacité de chacun étant contrôlée par celles de ses deux collègues, le juge le mieux noté de France serait celui qui, sans discuter ni réfléchir, voterait toujours comme son président : s'il faut en croire la formule, un tel juge ne se trompe jamais.

Ni Cournot ni Poisson n'ont commis la plus petite faute comme géomètres; ils traduisent rigoureusement leurs hypothèses. Mais les hypothèses n'ont pas le moindre rapport avec la situation d'un accusé devant les juges.

Ils ont aperçu les différences et croient, en les signalant, acquérir le droit de n'en pas tenir compte.

Poisson, qui, comme Condorcet, a consacré à la théorie des jugements un volume entier rempli des plus savants calculs, croit atténuer les objections qu'il ne pouvait manquer d'apercevoir, en altérant, dans ses énoncés, la signification du mot *coupable*. On rendrait, dit-il, le langage plus exact en substituant le mot *condamnable*, qui avait besoin d'explications et que nous continuerons à employer pour nous conformer à l'usage.

L'innocent accablé sous des indices trompeurs ou victime de machinations trop habiles pour qu'aucun juge puisse les soupçonner est un accusé *condamnable*. Poisson, *pour se conformer à l'usage*, le classe parmi les coupables. L'erreur unanime des juges devient alors une preuve de sagacité dont l'algèbre leur tient compte en évaluant leur mérite avec une infaillible précision. Dans cette suite de calculs stériles qui resteront, comme l'a dit Stuart Mill, le scandale des mathématiques, Condorcet seul a donné un sage conseil : celui de choisir, pour composer les assemblées, des hommes *véritablement éclairés*.

J'ai tenu à faire ces citations, où l'ironie et le paradoxe cachent le plus souvent une pensée très juste; il y a beaucoup

à retenir des appréciations de Bertrand, et il était utile de les rappeler. Il serait intéressant de rechercher si Condorcet, Laplace, Poisson, Cournot furent vraiment aussi naïfs que Bertrand semble le croire; mais je n'entrerai pas dans cette discussion historique, assez inutile pour notre but; il ne s'agit pas de savoir si les affirmations de Condorcet sont justes, ou si Bertrand a raison de les critiquer; mais bien de déterminer sous quelle forme on peut actuellement songer à appliquer le calcul des probabilités à la méthode des majorités. Dans cette étude, les travaux antérieurs nous seront naturellement utiles; mais les citer et les discuter à chaque instant ne pourrait qu'alourdir sans profit l'exposition.

II

Le premier point qui est de toute évidence, c'est que la majorité ne peut pas posséder des lumières qui font complètement défaut à chacune des individualités qui la composent : un jury d'aveugles ne saurait juger des couleurs. L'impossibilité, qui tient ici à la nature du jury, peut, dans d'autres cas, être inhérente à la question posée : *connaissant la hauteur du grand mât, trouver l'âge du capitaine*; mille personnes ne résoudront pas cette question mieux qu'une seule.

Ces remarques paraîtront peut-être superflues, tellement elles sont évidentes; on peut cependant en tirer une conséquence qui est fondamentale dans notre étude : *la valeur du jugement de la majorité peut dans certains cas être plus grande que la valeur du jugement individuel, mais* ELLE NE PEUT PAS ÊTRE DE QUALITÉ DIFFÉRENTE.

Un exemple très simple fera bien comprendre le sens de cet énoncé : Paul joue à l'écarté, et est très inexpérimenté; il se trouve avoir la dame, l'as et le neuf de cœur, qui est atout, le roi de trèfle et le roi de pique. Il se demande s'il doit jouer d'autorité, et consulte des amis plus éclairés : ceux-ci l'y encouragent, mais son adversaire a le roi, le valet et le dix d'atout et deux petits carreaux; Paul perd la partie. A-t-il eu tort de suivre les conseils qu'on lui a donnés? Ou, plus précisément, doit-on dire que ces conseils étaient mauvais? Assurément non; ces conseils étaient aussi bons que possible, du moment que les conseilleurs ne connaissaient pas le jeu de l'adversaire de Paul: seul, un compère habile et peu honnête aurait pu, ayant entrevu ce jeu, donner à Paul un conseil destiné à mieux réussir dans la circonstance très particulière où il se trouvait.

Nous pouvons résumer ceci en disant que les conseils donnés à Paul par ses amis constituent la *vérité relative*, dans les conditions où se trouve Paul, c'est-à-dire dans l'ignorance où il se trouve du jeu de son adversaire; mais il peut arriver que cette *vérité relative* soit contradictoire à la *vérité absolue*, c'est-à-dire à la manière de jouer qui serait la meilleure pour celui qui connaîtrait les deux jeux. Il est d'ailleurs bon d'ajouter que cette vérité, que nous appelons *absolue*, est seulement *moins relative*; il peut se faire que la perte de cette partie, influant sur la disposition des cartes pour la partie suivante, soit, en fin de compte, avantageuse à Paul, de sorte que le conseilleur qui connaîtrait, non seulement le jeu de son adversaire, mais la manière dont les cartes seront ramassées et battues pour le coup suivant, pourrait donner un conseil différent et pratiquement meilleur. Il est donc bien entendu que le mot *absolu* ne signifie que *moins relatif*; il ne peut d'ailleurs avoir d'autre sens, dans toute étude portant sur des réalités.

En admettant cette manière de parler, on peut distinguer, au point de vue pratique, trois catégories principales dans les applications du calcul des probabilités à la méthode des majorités.

1° La *vérité relative* que l'on atteint a une signification intéressante par elle-même, au point qu'il est légitime de la prendre comme but absolu de la recherche.

2° La vérité relative que l'on atteint n'a qu'un rapport éloigné et inconnu avec la vérité absolue qui seule intéresse.

3° La question est posée de telle manière que l'on doute s'il y a une vérité relative; on doit donc se poser d'abord la question de savoir si cette vérité relative existe; on pourra ensuite se proposer d'en déterminer la nature.

Ces distinctions, je le répète, ne peuvent avoir qu'une valeur *pratique*; au point de vue théorique, tous les cas intermédiaires sont possibles; au point de vue abstrait, tous les cas sont semblables en ce que leur traduction mathématique est la même; mais la signification utile de ces mêmes formules est fort différente suivant les cas : la même balance, dirait Joseph Bertrand, pèse tantôt du cuivre et tantôt de l'or.

III

Occupons-nous d'abord des cas où la vérité relative, non seulement est intéressante par elle-même, mais doit être regardée comme ayant une valeur absolue : ce sont les cas où le fait crée le droit.

L'un des meilleurs exemples que l'on puisse en donner est celui des questions de langage[1]; l'on se propose, par exemple, de savoir si telle locution est actuellement usitée ou comprise en un lieu donné : le moyen le plus sûr est d'interroger un grand nombre des habitants de ce lieu; si on pouvait les consulter tous, individuellement et indépendamment l'un de l'autre, l'ensemble de leurs réponses, supposées sincères et exactement dénombrées, fournirait évidemment, par définition même, la réponse la plus satisfaisante possible à la question posée. Mais il est clair qu'une telle consultation est pratiquement impossible; la méthode des majorités consistera donc, en consultant un petit nombre seulement de personnes, à chercher à prévoir le résultat que donnerait le dénombrement général. Par exemple, pour savoir si les Parisiens de 1908 disent *un* automobile ou *une* automobile, on consultera vingt personnes, et on prendra l'avis de la majorité. C'est ce que font souvent les journaux lorsqu'ils instituent une « enquête ». Que vaut la méthode? Sans entrer dans les détails, il est nécessaire, pour que le calcul puisse être correctement appliqué à une question, que le groupe total par rapport auquel cette question est posée, soit sensiblement homogène; cette homogénéité n'est jamais absolue, puisque deux individus quelconques ne sont pas identiques; il faut donc de plus que le groupe partiel que l'on choisit comme représentant du groupe total ait sensiblement la même hétérogénéité. Par exemple, il serait absolument incorrect de prendre les cinquante voyageurs qui sont dans un tramway déterminé comme représentants de l'ensemble des Parisiens; il serait plus correct de prendre comme représentants de l'ensemble des conscrits parisiens nés en 1887, ceux d'entre eux qui sont nés, par exemple, le 14 mai de cette année; ce choix serait meilleur que le choix d'un quartier déterminé, ou même qu'un choix alphabétique; en supposant

1. Voir mon article : *Un paradoxe économique; le calcul des probabilités et les vérités statistiques* (*Revue du Mois* du 10 décembre 1907; t. IV, p. 688).

les noms inscrits par lettre alphabétique et choisissant, comme on le fait souvent pour des examens ou concours, une portion de la liste désignée par le sort, on risquerait de tomber sur des portions trop homogènes, les personnes dont les noms commencent par *La* ou *Le*, ou bien par W ou Z pouvant avoir en commun certains caractères ethniques les différenciant des autres.

Si les conditions d'homogénéité sont réalisées, c'est-à-dire si le groupe restreint sur lequel porte l'expérience est vraiment une image fidèle du groupe total, quelles conclusions peut-on tirer de l'étude de ce groupe partiel? Par exemple, sur 36 500 conscrits d'une région donnée nés en 1887, il y en a 100 qui sont nés le 14 mai; on constate que, sur ces 100, la majorité absolue, soit 51, a une taille supérieure ou égale à 1 m. 65, les 49 autres ayant une taille inférieure; on constate, d'autre part, sur la majorité des 100, telle ignorance, ou telle opinion, ou tel jugement particulier; que peut-on induire de là pour l'ensemble total dont ils ont été extraits? Ce n'est point ici le lieu de traiter mathématiquement ces questions; il suffit d'avoir indiqué qu'il est légitime de se les poser et de signaler la forme de la réponse qu'y fournit le calcul. Cette forme est nécessairement un coefficient de probabilité. Par exemple, on pourra, des faits observés, conclure ceci : la probabilité pour que la majorité des faits non observés soit conforme à la majorité des faits observés est de 0,999, la probabilité opposée étant par suite 0,001. En d'autres termes, on peut parier 999 contre 1 que l'on ne se trompe pas, mais on n'a pas la certitude de ne pas se tromper. Cette forme particulière d'affirmation est commune à toutes les questions où intervient le calcul des probabilités; certains esprits se refusent à la comprendre et préfèrent déclarer qu'ils veulent ignorer un calcul qui conduit à des résultats aussi incertains; ils ne se rendent pas compte que cette incertitude est commune à toutes nos affirmations, et n'est pas moins dangereuse lorsqu'elle est masquée sous des apparences dogmatiques[1]. En réalité, au point de vue pratique, une probabilité suffisamment voisine de l'unité doit, dans l'action, être confondue avec la certitude.

Indiquons enfin que le calcul des probabilités permet, de l'étude des chiffres observés, de tirer des conclusions sur

1. Voir mon article : *Sur la valeur pratique du calcul des probabilités* (*Revue du Mois* du 10 avril 1906; t. I, p. 424).

l'homogénéité du groupe étudié et, par suite, sur la valeur que l'on peut attribuer aux observations faites sur ce groupe. Je suppose, par exemple, que les 10 premiers conscrits examinés au conseil de revision de la Seine aient mesuré, cinq d'entre eux 1 m. 55 et cinq d'entre eux 1 m. 75; la simple constatation du fait qu'aucun d'eux n'a une taille voisine de la moyenne 1 m. 65 suffit pour inspirer les doutes les plus sérieux sur l'exactitude des conclusions que l'on obtiendrait en supposant que ce groupe partiel est l'image exacte du groupe total dont il est extrait : on doit, au contraire, être certain que des circonstances particulières ont influé sur le choix des dix individus composant ce groupe.

IV

Je serai très bref sur le cas où la *vérité relative*, que la méthode des majorités permet d'atteindre, est sans rapport réel, ou du moins sans rapport connu, avec la vérité absolue qu'il serait utile de connaître. Ces cas ne sont pas intéressants, on doit les signaler seulement, à titre d'indication, afin d'éviter les erreurs auxquelles leur étude pourrait conduire. On peut se demander si toutes les applications du calcul aux décisions judiciaires rentrent dans cette catégorie, c'est-à-dire si la distinction faite par Poisson entre *coupable* et *condamnable* ne suffit pas pour justifier les sévérités de Joseph Bertrand. Il semble que celui-ci ait posé la question sous une forme trop abstraite. Pour obtenir des jugements équitables, il est clair qu'il faut avant tout choisir des juges éclairés et dépourvus de passion : mais ce n'est pas ainsi que la question est posée dans la pratique, et il est vraiment trop simple de proposer cette solution. La question n'est pas de réaliser un idéal impossible à atteindre : c'est d'arriver au moindre mal en se servant des institutions imparfaites et des hommes faillibles dont on dispose, jusqu'au jour, que nous souhaitons tous avec Joseph Bertrand, mais sans y compter plus que lui, où les institutions seront parfaites et les hommes infaillibles.

Lorsque l'on quitte ainsi le terrain de la justice idéale pour se placer sur celui des faits, il est clair qu'il peut exister des circonstances où tel accusé innocent doive être fatalement condamné par la quasi-unanimité des juges ou jurés appelés à prononcer sur son sort. On doit naturellement déplorer qu'il en soit ainsi et tâcher, en améliorant les garanties que la procédure donne aux accusés, en améliorant surtout la mentalité des juges, de rendre un tel cas de plus en plus improbable, mais le désir d'améliorer la réalité ne doit pas empêcher de la constater. Dès lors, il est parfaitement légitime de se poser une question telle que la suivante : en augmentant le nombre des jurés et en modifiant la majorité requise pour la condamnation, augmente-t-on ou diminue-t-on les chances pour qu'un accusé soit condamné, lorsque son innocence non seulement est réelle, mais encore n'est pas masquée par des apparences mensongères et des machinations habiles.

Lorsque la question est ainsi posée, il ne semble pas douteux que les calculs de Laplace et de Poisson n'y fournissent la réponse correcte [1]. Seulement, pour que cette réponse reste correcte, il ne faut pas en modifier le sens en faisant abstraction des différences concrètes qui existent entre les divers cas. Par exemple, il résulte aisément d'un calcul simple qu'il est préférable pour un accusé d'être jugé par un tribunal de 7 juges, où 5 voix sont nécessaires pour la condamnation, que par un tribunal de 12 juges où 7 voix suffisent à condamner. Mais il serait évidemment abusif d'en conclure qu'il est préférable pour un soldat accusé d'être traduit devant un conseil de guerre que devant un jury civil. Car, indépendamment de toute autre considération, le seul fait que le vote du jury a lieu au scrutin secret est un élément dont il a été impossible de tenir compte dans les calculs.

Parmi les cas où il n'y a *rien* à trouver, et où par conséquent la méthode des majorités et toute méthode statistique perd ses droits, il n'est pas inutile de mentionner les jeux de hasard, sans trop compter toutefois être cru par les joueurs, gens à préjugés tenaces et plus disposés à écouter les prometteurs de martingales que les conseils de l'algèbre. Il est fâcheux de voir des chercheurs sérieux perdre à des recherches aussi vaines un temps qui pourrait être mieux employé. Je fais allusion ici à un travail récent de M. Charles Henry, où, sous couvert de distinction entre la vérité psychologique et la vérité mathématique, il réédite des sophismes que l'on pouvait croire définitivement percés à jour [1]. La question de savoir si l'on peut gagner de l'argent à coup sûr en jouant à la roulette est une question essentiellement pratique et concrète; à ce titre, elle appartient entièrement à la science et, quelque opinion que l'on professe sur l'intérêt de spéculations métaphysiques sur le psychologisme, on doit admettre qu'elles n'ont rien à voir avec la solution pratique de cette question, solution qui a été donnée depuis longtemps par Bernoulli et confirmée par tous les mathématiciens : un joueur qui joue indéfiniment à un jeu équitable arrive forcément à la ruine, au bout d'un temps plus ou moins long; ce temps devient très court si le jeu n'est pas équitable, ce qui est toujours le cas dans la pratique.

1. Il est cependant un point, que Bertrand paraît avoir négligé et dont l'importance peut parfois être grande : la connaissance des conditions requises pour la condamnation ne peut-elle influer sur l'opinion des juges? Par exemple, parmi les conventionnels ayant voté la mort de Louis XVI, ne peut-on pas regarder comme très vraisemblable qu'il en est au moins un qui se fût prononcé pour une condamnation moins sévère si, juge unique ou membre d'un jury très peu nombreux, son vote lui eût paru plus décisif? Cette question est distincte de celle du vote secret des juges, qui fait partie des dispositions pouvant être modifiées par la loi et les mœurs.

V

Nous arrivons maintenant au cas le plus intéressant pour les psychologues, car il se présente souvent dans leurs recherches, bien que la question ne soit pas toujours posée par eux sous la forme que nous allons lui donner. Un ensemble de phénomènes étant observés par des individus divers, puis classés de telle manière qu'une majorité s'en dégage, le premier problème qui se pose est le suivant :

L'opinion de cette majorité correspond-elle à quelque réalité?

Déterminer quelle est cette réalité et si, en particulier, elle est adéquate ou non à ce que l'on aurait pu espérer, en instituant les expériences, c'est un second problème, dont je ne méconnais nullement l'importance, mais qu'il est essentiel de distinguer du premier, car c'est seulement lorsque le premier sera résolu que le second pourra être abordé avec fruit.

Précisons cette distinction par un exemple; supposons que l'on ait lu à haute voix une même page dans un grand nombre d'écoles primaires et demandé à chaque enfant d'indiquer par écrit quel est le mot qui revient le plus souvent dans cette page [2]. Si les trois quarts des réponses s'accordent pour dési-

1. CHARLES HENRY. — La loi des petits nombres. — Recherches sur le sens de l'écart probable dans les chances simples à la roulette, au trente-et-quarante, etc., suivies d'une instruction pratique pour le joueur. Paris, 1908 (Laboratoire d'Energétique d'Ernest Solvay).

2. Une expérience analogue a été faite récemment par M. Binet; on

gner le même mot, on peut conclure avec certitude que cette coïncidence n'est pas fortuite, mais a une raison. Cette raison peut être le fait que le mot désigné est effectivement celui qui figure le plus souvent dans la page lue; elle peut aussi être tout autre; mais c'est l'existence même d'une raison qui est le résultat essentiel fourni par la méthode des majorités dans cet exemple particulier.

De même si, faisant comparer à de nombreux expérimentateurs les poids de deux objets A et B dont les formes et dimensions sont très différentes, on obtient des réponses dont les trois quarts s'accordent pour attribuer un poids plus élevé à l'objet A, on peut en conclure qu'il y a une raison à cela, mais cette raison n'est peut-être pas que A soit effectivement plus lourd que B. Au contraire si, sur 1 000 expériences, 520 seulement déclarent A plus lourd et 480 B plus lourd (je suppose qu'on laisse de côté les réponses douteuses), l'écart observé est de ceux qui se produiraient fréquemment dans une série de 1000 réponses tirées à pile ou face; on ne doit donc rien en conclure, sinon une forte présomption pour qu'une nouvelle série de 1 000 expériences fournisse un résultat aussi incertain.

Avant d'arriver à la discussion d'expériences à résultats numériques, je voudrais dire un mot d'une expérience fort intéressante dans laquelle le résultat se traduisait par un dessin. Il s'agit d'observations de la surface d'une planète, faites en un même instant, par des observateurs éloignés les uns des autres, dont chacun communiquait un dessin de son observation. Ces dessins ont été réunis et comparés par M. Jean Mascart, de l'Observatoire de Paris, à la très intéressante étude duquel je renvoie[1]. L'expérience suggérée par M. Nicolas Poutiata a été organisée par la Société Astronomique de France, sur la proposition de M. Camille Flammarion; le programme en avait été arrêté avec soin et fut exécuté avec le concours de 36 observateurs, disposant de moyens d'action fort divers; elle a duré 19 jours (du 2 au 20 janvier 1906), le mauvais temps ayant naturellement empêché certains jours le concours de certains observateurs; mais le nombre des observations faites un même jour à la même heure dépasse souvent 10 et s'élève jusqu'à 17. Nous ne pouvons entrer ici dans la discussion détaillée à laquelle se livre M. Jean Mascart, ni dans la description des divers procédés photographiques employés pour *prendre la moyenne* des dessins divers (signalons toutefois l'ingénieuse idée de proportionner les temps d'action photographique, soit au diamètre, soit à la surface de l'objectif employé dans l'observation particulière utilisée). Contentons-nous de retenir l'impression qui résulte de la seule inspection des dessins juxtaposés : ces dessins exécutés le même jour à la même heure, sont très différents les uns des autres et fort peu de détails se retrouvent sur plusieurs d'entre eux. Cette simple constatation doit rendre très prudent dans l'interprétation de dessins de cette nature; il est manifeste que des observateurs dont la bonne foi n'est pas douteuse, et qui savent de plus qu'ils seront soumis à un contrôle, croient voir des apparences ne correspondant à aucune réalité. Ces apparences sont-elles dues aux conditions atmosphériques ou à l'imagination de l'observateur, c'est ce que de nouvelles expériences pourraient peut-être déterminer. En tous cas, l'expérience mériterait d'être reprise, car elle n'est pas moins intéressante pour le psychologue qu'elle ne promet d'être profitable pour les progrès de l'astronomie planétaire; dans ces nouvelles expériences, on tirera grand profit de maintes remarques judicieuses de M. Jean Mascart.

lisait à des enfants 100 substantifs et on leur demandait d'écrire la liste des mots dont ils se souvenaient. Je n'ai pas eu les éléments suffisants pour la traiter par le calcul; je pourrai avoir l'occasion d'y revenir.

1. Observations simultanées de la surface de Jupiter, réunies par M. Jean Mascart (*Bulletin de la Société astronomique de France*, 1907).

VI

Parmi les observations psychologiques qui ont été l'occasion des développements mathématiques se rattachant au calcul des probabilités, on doit mentionner particulièrement la méthode bizarrement dénommée méthode *des cas vrais et des cas faux* (*Richtig und falsch*; *Right and wrong*). Cette méthode a donné lieu à de nombreux calculs, dans lesquels interviennent l'intégrale de Gauss et des formules qui s'y rattachent; ce n'est pas ici le lieu d'en faire un exposé critique, qui exigerait un appareil mathématique relativement considérable; j'aurai peut-être l'occasion d'y revenir ailleurs; je me contenterai de signaler ici que le point délicat est de trouver une expression mathématique de l'erreur élémentaire satisfaisant également bien aux conditions de continuité et de discontinuité qui sont dans la nature de la question; les tentatives faites dans ce sens ne me paraissent pas entièrement satisfaisantes au point de vue théorique et, d'autre part, conduisent rapidement, au point de vue pratique, à des complications analytiques qu'il serait préférable d'éviter, en des questions où sont intéressées beaucoup de personnes dont la culture mathématique est forcément limitée.

Je voudrais surtout indiquer comment l'emploi systématique de la méthode des majorités permet de poser la question sous une forme assez différente de la forme usuelle : au lieu de s'intéresser, comme il semble qu'on l'ait fait surtout jusqu'ici, à l'étude de la sensation individuelle, qu'on cherchait à dégager de l'ensemble des expériences, c'est la précision de la « sensation collective » qu'il me paraîtrait intéressant de dégager de cet ensemble. Prenons un exemple précis. L'un des buts que se sont proposés les expérimentateurs en psychologie, c'est, par exemple, de déterminer quel était l'accroissement minimum de poids que pouvait apprécier un observateur moyen; étant donné un poids étalon de 1000 grammes, on constate que c'est seulement lorsque le poids à juger est inférieur à 950 grammes ou supérieur à 1050 que la différence est assez généralement perçue[1]; on dira donc que 50 grammes représentent la différence minimum perceptible par rapport à 1000 grammes; c'est de la détermination de cette différence minimum qu'on s'est beaucoup occupé depuis la loi de Weber. Mais on peut poser autrement la question ; au lieu de se proposer d'étudier la loi d'*une* expérience, on peut rechercher quelle précision pourraient donner un grand nombre d'expériences, et il est visible que cette précision serait plus grande, en même temps qu'il est évident qu'elle ne serait pas indéfinie. Supposons, par exemple, que l'on fasse comparer à un grand nombre d'observateurs deux poids, l'un de 1000 grammes, l'autre de 1010 grammes[2]; si, sur 1000 observateurs, il y en a 600 qui donnent la solution juste, il sera légitime de dire que cet

1. Je laisse ici entièrement de côté la discussion des erreurs systématiques qui s'introduisent dans les expériences de ce genre; je suppose, par exemple, que l'on élimine l'erreur relative à l'ordre des comparaisons en faisant chaque comparaison deux fois; je laisse aussi de côté la définition précise de l'écart *généralement perçu*; si l'on admet que les écarts perçus suivent la loi de Gauss, on peut adopter, soit l'écart probable, soit l'écart médian, assez voisins l'un de l'autre, comme on sait.

2. Bien entendu, je suppose toujours que les erreurs systématiques soient évitées.

ensemble de 1000 observateurs a la perception collective de la différence de 10 grammes bien que, si l'on recommençait plusieurs fois l'expérience, il puisse arriver que chacun des 1000 observateurs se trompe quelquefois, de sorte qu'aucun d'entre eux ne peut être regardé comme ayant la perception individuelle de cette différence. J'ai supposé ici que l'on exigeait une réponse affirmative ou négative, en excluant les réponses douteuses; il serait aisé de poser le problème d'une manière analogue dans le cas où l'on admet ces réponses douteuses. Supposons, par exemple, que l'on ait une échelle de poids de comparaison, marqués en chiffres, et variant de gramme en gramme depuis 900 jusqu'à 1100 grammes; on donne à l'observateur un poids P de valeur inconnue et on lui demande quels sont ceux des poids connus qui lui paraissent égaux, lesquels supérieurs, lesquels inférieurs au poids à déterminer. Passant encore ici sur les détails de l'expérience, non que j'en méconnaisse la grande importance, mais je dois me borner, je suppose que l'on déduise de son résultat la valeur la plus probable du poids P[1]. Quelle confiance peut-on accorder à une telle détermination individuelle? Et quelle confiance peut-on attribuer à la moyenne de 1 000 déterminations analogues, faites par des observateurs différents? Si l'on acceptait d'une manière aveugle la loi de Gauss, ou toute autre loi mathématique précise, pour exprimer les erreurs individuelles, on serait forcément conduit à la conclusion que la précision croît comme la racine carrée du nombre des observations et peut, par suite, être rendue aussi grande que l'on veut. Cette méthode d'estimation en arriverait ainsi à pouvoir dépasser, si on l'employait avec soin, la précision des meilleures mesures faites au moyen d'instruments, ce qui est en apparence paradoxal. Il serait intéressant de déterminer par l'expérience la précision collective qui peut ainsi être atteinte par des mesures estimatives. C'est là un problème qui ne me paraît pas indigne de retenir l'attention des observateurs; on trouverait d'ailleurs dans les publications déjà faites, et surtout sans doute dans les cahiers d'expériences, des éléments pour en commencer l'étude[2].

1. Les méthodes mathématiques pour atteindre ce but sont très simples si l'on suppose l'intervalle assez faible pour que l'on puisse y considérer le minimum perceptible comme constant; elles se compliquent si l'on tient compte de sa variation suivant la loi de Weber.

2. D'après des renseignements qui m'ont été fournis par M. Ch.-Éd. Guillaume, directeur-adjoint du Bureau International des Poids et Mesures, on est arrivé, en multipliant les lectures faites par des observateurs

Pour donner une idée de la marche que l'on pourrait suivre, j'emprunte les chiffres suivants au *Manuel de psychologie expérimentale* de Titchener[1]. Un poids S de 1 071 grammes était comparé à divers poids C, au nombre de 7, dont les valeurs croissaient de 921 à 1 221 grammes par intervalles de 50 grammes. Le tableau suivant donne les résultats de 100 observations pour chaque poids C, soit 700 observations en tout; on trouve dans les trois colonnes les nombres respectifs des observations où C fut jugé inférieur, égal, ou supérieur à S.

VALEUR DE C	$C < S$	$C = S$	$C > S$
921	1	8	91
971	6	11	83
1 021	10	37	53
1 071	30	37	33
1 121	55	34	11
1 171	76	20	4
1 221	90	9	1

Si l'on examine ce tableau au point de vue auquel nous nous plaçons, c'est-à-dire si l'on cherche à déduire de l'ensemble des observations la relation réelle supposée inconnue entre C et S, on aperçoit immédiatement que cette relation apparaît sans ambiguité; la lecture des trois dernières colonnes indique sans qu'il soit nécessaire de regarder la première si C est inférieur, égal, ou supérieur à S. Bien entendu, l'égalité rigoureuse ne peut pas être établie, mais simplement présumée d'après le résultat des expériences. Du fait que sur 100 expériences, C fut jugé 30 fois inférieur, 37 fois égal et 33 fois supérieur, on peut simplement conclure ceci : il n'y a aucune raison[2], d'après ces expériences, pour supposer $C > S$ plutôt

exercés, à atteindre dans les moyennes une précision *beaucoup plus* grande que la précision des observations individuelles; par exemple, au moyen de lectures faites à la loupe sur une échelle millimétrique, on a pu atteindre avec certitude *le centième de millimètre*, avec un nombre suffisant d'observations. Je me propose de faire l'étude de ces résultats, en grande partie inédits, au point de vue développé ici.

1. Titchener, *Experimental Psychology. Students' Manual-Quantitative*, p. 107; j'ai fait la somme des nombres correspondant aux deux séries d'expériences (Time Order I, S first et Time Order II, S second).

2. En réalité, la petite différence entre 30 et 33 donnerait une faible raison de supposer $C > S$; c'est là un problème de probabilité des causes, qui peut être traité, suivant la méthode classique, à l'aide d'une hypothèse supplémentaire sur la probabilité *a priori*. Mais il n'est pas besoin de

que $C < S$; on doit donc, si l'on n'a pas d'autres éléments d'information, supposer $C = S$. De même, la majorité distingue à coup sûr un poids C de 1 021 grammes du poids S de 1 071 grammes, bien que chacun des observateurs puisse se tromper individuellement; car, en laissant de côté les 37 réponses affirmant l'égalité ou le doute, il reste 53 réponses exactes contre 10 inexactes; une telle répartition ne peut pas être l'effet du hasard; si un jeu de pile ou face, sur 63 épreuves, donne 53 fois pile et 10 fois seulement face, la certitude est presque absolue que le jeu était truqué.

Mais on peut aller plus loin et se demander ce qui se passerait si C avait une valeur intermédiaire entre 1 021 grammes et 1 071 grammes; vu la régularité de la variation des nombres inscrits dans les trois colonnes, on commet une erreur assez faible en procédant par interpolation linéaire, c'est-à-dire en supposant que les nombres varient en progression arithmétique, ce qui donne les résultats suivants[1] :

C	$C < S$	$C = S$	$C > S$
1 021	10	37	53
1 031	14	37	49
1 041	18	37	45
1 051	22	37	41
1 061	26	37	37
1 071	30	37	33

Bien entendu, si l'expérience était réellement faite, les nombres ne seraient pas exactement ceux que nous inscrivons à cause des erreurs fortuites; mais le calcul des probabilités nous indique quelles sont les erreurs fortuites qui peuvent être regardées comme probables et celles qui sont hautement improbables. Par exemple, sur 63 jugements (ce qui est le cas ici, puisqu'il y en a 37 neutres), on doit regarder, s'ils sont l'effet du hasard, comme peu probable qu'il y en ait plus de 40 dans un même sens, 23 seulement étant de sens opposé[2].

calcul, pour peu qu'on soit familier avec ces questions, pour être certain que la probabilité pour que l'on ait $C > S$ dépasse très peu 0,5, et est par suite très voisine de la probabilité pour que $C < S$.

1. En tenant compte des différences secondes, on obtient des nombres assez voisins qui conduisent aux mêmes conclusions finales; j'ai préféré présenter le calcul sous sa forme la plus intuitive.

2. La probabilité est $1 - \Theta\left(\frac{9}{\sqrt{32}}\right) = 1 - \Theta(1,6) = 0,024$ (Θ désignant l'intégrale de Gauss.

L'ensemble des observations faites dans le cas où $C = 1051$ permet donc de conclure avec beaucoup de vraisemblance que l'on a $C < S$, c'est-à-dire dans le sens de la vérité. De l'ensemble des nombres de notre premier tableau, nous tirons donc la conclusion que l'ensemble des 100 observateurs considérés, lorsque l'on considère leur jugement collectif, arrive à distinguer une différence de poids de 20 grammes. Or, par des calculs assez longs et très corrects, Titchener trouve (*loc. cit.*, p. 109 et p. 113), pour les valeurs de DL[1] déduites de ce même tableau les divers nombres suivants[2] : 43, 36, 30, 53, 38, 34, dont la moyenne 39 est à peu près le double du nombre 20 que nous trouvons. Cette différence des résultats déduits des mêmes expériences montre bien que les problèmes sont distincts : Titchener détermine la sensibilité individuelle tandis que nous déterminons la sensibilité collective de l'ensemble des observateurs. Il serait d'ailleurs possible de donner plus de précision à notre détermination, mais ce serait aux dépens de la rigueur, vu la petitesse des nombres employés[3]. Il vaut mieux s'en tenir à cette constatation nette, que la différence de 20 grammes est perceptible presque à coup sûr par la méthode des majorités appliquée à l'ensemble des 100 observations; cette méthode donne donc, comme on pourrait s'y attendre, une précision bien plus grande que l'observation individuelle.

Je me contenterai de cet exemple, espérant qu'il suffira pour montrer quel intérêt il pourrait y avoir à appliquer la méthode à des expériences combinées spécialement en vue de cette application. Il n'est pas douteux qu'on serait ainsi conduit à des résultats importants à divers titres.

1. C'est-à-dire du minimum d'accroissement moyen perceptible pour l'observateur individuel.
2. Les différences entre ces nombres sont dues, d'une part à la distinction entre les deux « time order », d'autre part à la distinction entre les accroissements positifs et négatifs, et, enfin, à des différences entre les méthodes suivies; ces différences entre les résultats montrent combien il est illusoire d'appliquer des méthodes de calcul trop précises à des nombres forcément approchés.
3. Si l'on suivait une méthode parallèle à celle qui est généralement employée dans ces questions quand on use de l'intégrale de Gauss, on rechercherait quelle valeur doit être donnée à C pour que la probabilité de l'écart observé soit inférieure à $\frac{1}{2}$; on déduirait ainsi des nombres donnés que la valeur de DL, pour l'ensemble des 100 observateurs, considérés collectivement, est d'environ 8 grammes; mais le nombre des observations est trop faible pour que ce résultat puisse être regardé comme précis.

VII

Pour terminer, je voudrais discuter quelques applications de la méthode des majorités à des questions qualitatives, car c'est peut-être là surtout que cette méthode fournit des résultats qui ne pourraient être obtenus par aucun autre moyen.

La première idée de ce genre d'application m'a été suggérée par une enquête que j'ai faite sur des textes qui avaient servi à des expériences graphologiques de M. Binet[1]. Il s'agissait de distinguer, sur 12 textes très courts, reproduits en typographie, lesquels des 12 scripteurs étaient des hommes supérieurs et lesquels des hommes médiocres[2]. Les réponses individuelles furent intéressantes, mais le classement déduit du dépouillement du scrutin le fut plus encore; les six hommes supérieurs furent déclarés tels par la majorité absolue; un seul homme médiocre eut aussi la majorité absolue et c'est, si l'on peut dire, le moins médiocre des hommes « médiocres » (ou « simplement intelligents ») choisis par M. Binet, qui en dit : « C'est un publiciste scientifique qui ne manque pas de mérite et qui, s'il se reconnaît, ne m'en voudra pas de lui avoir préféré Claude Bernard ». Somme toute, l'opinion de la majorité était conforme à la vérité pour tous les hommes supérieurs et l'erreur unique qu'elle commettait en classant comme supérieur un des hommes médiocres pouvait se justifier par les renseignements mêmes donnés sur lui. Je n'insisterai pas plus longtemps sur cette expérience, mais j'ai tenu à la rappeler, car ce fut à son occasion que je fus amené à correspondre avec M. Binet au sujet de cette méthode des majorités (c'est lui qui employa le premier cette expression), et qu'il voulut bien m'engager à écrire cette petite étude pour son *Année*.

*
* *

Je discuterai en détail une expérience intéressante faite par Mme Rousson sur la lecture d'une physionomie d'enfant[3].

1. Voir la *Revue du Mois* d'août et de septembre 1906; t. II, p. 244 et 366.
2. La classification était due à M. Binet; voir son livre : *Les Révélations de l'écriture soumises à un contrôle scientifique*.
3. *Bulletin de la Société libre pour l'étude psychologique de l'enfant*, juin 1906 (6e année, p. 76).

Rappelons-en brièvement les conditions, en laissant de côté les détails. On soumit séparément à 20 observateurs 40 photographies d'enfants (chacune double; face et profil), dont 23 étaient arriérés et 17 normaux; chaque observateur déclarait l'enfant arriéré, ou normal, ou refusait de répondre (il y eut seulement 18 refus de réponse sur 800 observations [1]). Les pourcentages de réponses justes varièrent suivant les observateurs, entre 67 p. 100 et 92 p. 100; leur moyenne fut de 78 p. 100. On peut conclure de là qu'il y avait vraiment dans les photographies [2], ou du moins dans certaines d'entre elles, des indices d'intelligence. Mais cette conclusion générale peut être de beaucoup précisée par l'application de la méthode des majorités. Je voudrais montrer en détail sur cet exemple quels principes simples on doit utiliser dans cette application, en tâchant de faire appel au minimum de connaissances mathématiques.

Le nombre des votants était 20; chacun d'eux devait voter : *normal* (N) ou *arriéré* (A), de sorte que l'ensemble des votes relatifs à une même photographie était représenté par une suite telle que la suivante [3] (photographie n° 7) :

N N A A A N A N A A A N A A A A A A A A

Combien peut-il exister de telles suites de 20 lettres N ou A? La première lettre peut être N ou A, ce qui fait deux possibilités; pour chacune de ces 2 possibilités, la seconde lettre peut être N ou A, ce qui fait 2×2 possibilités (NN, NA, AN, AA); pour chacune de ces 2×2 possibilités, la troisième lettre peut être N ou A, ce qui fait $2 \times 2 \times 2$ possibilités

1. Ces refus sont assez peu nombreux pour pouvoir être négligés Mme Rousson n'en tient pas compte dans le pourcentage des réponses justes, c'est-à-dire que si un observateur refuse de répondre 2 fois et se trompe 3 fois, sur 40 enfants, elle compte 35 réponses justes sur 38. Ce procédé augmente légèrement les pourcentages de Mme Rousson; nous les avons reproduits tels qu'elle les donne.

2. Je dis les *photographies*, et non pas seulement les *physionomies*, car un des observateurs a fait remarquer qu'il eût mieux valu « écarter toutes les différences extérieures qui peuvent influencer le jugement ». Cette remarque est sans importance pour nos déductions; elle doit intervenir dans les conclusions spéciales à la physionomie que l'on pourrait chercher à en tirer.

3. Je laisse de côté les quelques votes *nuls* (au nombre de 18 sur 800); pour donner plus de certitude à mes conclusions, je les compterai avec la *minorité*; c'est le procédé le plus désavantageux et avec toute autre convention, les conclusions seraient vraies *a fortiori*. Vu leur petit nombre, la chose est d'ailleurs de peu d'importance.

(NNN, NNA, NAN, NAA, ANN, ANA, AAN, AAA), etc. En répétant le même raisonnement jusqu'à la vingtième lettre, on trouve un nombre de possibilités égal au produit de 20 facteurs égaux à 2, c'est-à-dire à 2^{20}, ou à 1 048 576. Tel est le nombre des diverses possibilités du scrutin considéré. Si chaque votant, au lieu de délibérer son vote, le jouait à pile ou face, chacune de ces possibilités serait également probable, et il y aurait par suite exactement 1 chance sur 1 048 576 pour que l'une quelconque d'entre elles se produisît (par exemple, celle que nous avons écrite plus haut).

Il faut maintenant déterminer combien, parmi les diverses possibilités du scrutin, il y en a qui donnent pour N un certain nombre de voix; il est bien clair, par exemple, que N peut avoir une voix contre 19 à A de 20 manières différentes, car l'unique voix N peut-être celle du premier votant, ou celle du second,..... ou celle du vingtième. La solution complète de cette question est fournie par l'analyse combinatoire et le calcul numérique se fait simplement par l'algorithme dit « triangle arithmétique de Pascal »; je me contenterai d'en indiquer le résultat.

On peut obtenir	D'un nombre de manières égal à	Somme des nombres de la colonne précédente
—	—	—
20 N et 0 A	1	21
19 N et 1 A	20	211
18 N et 2 A	190	1 351
17 N et 3 A	1 140	6 196
16 N et 4 A	4 845	21 700
15 N et 5 A	15 504	60 460
14 N et 6 A	38 760	137 980
13 N et 7 A	77 520	263 950
12 N et 8 A	125 970	
11 N et 9 A	167 960	
10 N et 10 A	184 756	
9 N et 11 A	167 960	
8 N et 12 A	125 970	

Il a été inutile d'écrire la fin du tableau, qui reproduit évidemment le commencement en ordre inverse. Dans la troisième colonne, nous avons inscrit la somme des nombres de la seconde (21 700, par exemple, est la somme $1 + 20 + 190 + 1\,140 + 4\,845 + 15\,504$); on verra tout à l'heure quelle en est l'utilité.

Le tableau précédent est l'élément essentiel de la discussion de la méthode des majorités; lorsque le nombre des votants est petit on l'obtient sans peine à l'aide du triangle arithmétique de Pascal[1]; lorsqu'il est grand, les calculs seraient inextricables et l'on utilise des formules approchées que l'on établit par le calcul intégral et qui conduisent à considérer la célèbre intégrale de Gauss; telle est la signification de cette intégrale, qui n'a rien de mystérieux; mais, dans l'exemple actuel, il m'a paru préférable d'en éviter l'emploi.

Le nombre total des possibilités est 1 048 576; si les votes étaient joués à pile ou face, ce qui rendrait toutes ces possibilités également probables, la probabilité pour que se produise l'une des 38 760 combinaisons qui donne 14 N et 6 A serait égale au quotient de 38 760 par 1 048 576, c'est-à-dire, en chiffres ronds, à 4 p. 100. On peut ainsi déduire du tableau précédent les résultats probables que donneraient des votes émis au hasard; les voici, en arrondissant les chiffres :

On obtiendrait	20 N et 0 A.	1	fois sur	1 000 000
—	19 N et 1 A.	1	—	50 000
—	18 N et 2 A.	1	—	2 500
—	17 N et 3 A.	1	—	1 000
—	16 N et 4 A.	1	—	200
—	15 N et 5 A.	1,5	—	100
—	14 N et 6 A.	4	—	—
—	13 N et 7 A.	8	—	—
—	12 N et 8 A.	12	—	—
—	11 N et 9 A.	17	—	—
—	10 N et 10 A.	18	—	—

Il en résulte que, sur 40 scrutins, on devrait s'attendre, en gros, aux résultats suivants :

1. Ce triangle a la disposition suivante :

1								
1	1							
1	2	1						
1	3	3	1					
1	4	6	4	1				
1	5	10	10	5	1			
1	6	15	20	15	6	1		
1	7	21	35	35	21	7	1	
1	8	28	56	70	56	28	8	1

Le nombre 56, par exemple, inscrit le quatrième dans la dernière ligne, s'obtenant en additionnant le nombre 35 inscrit au-dessus de lui au nombre 21 qui précède ce nombre 35; les nombres de notre tableau formeraient la 21[e] ligne du triangle, commençant par 1, 20, suivis de 190, 1140, etc.

7	fois chacune des combinaisons	11 N et 9 A,	10 N et 10 A,	9 N et 11 A
5	— —	12 N et 8 A,	8 N et 12 A	
3	— —	13 N et 7 A,	7 N et 13 A	
3	fois en tout les deux combinaisons	14 N et 6 A,	6 N et 14 A	

l'arrivée des autres combinaisons étant, non pas impossible, mais assez peu probable pour 15 N ou 16 N, très peu probable pour 17 N et 18 N, tout à fait improbable pour 19 N ou 20 N. En particulier, cette dernière combinaison (l'unanimité) ne devrait se produire que 1 fois sur 1 000 000, or, en réalité, elle s'est produite 5 fois et chaque fois dans le sens de la vérité[1]. Attribuer cet événement au hasard est aussi improbable qu'admettre le fait suivant : dans une ville comme Londres, renfermant environ 1 000 000 d'hommes adultes, 40 individus investis du pouvoir suprême sont forcés de démissionner pour remettre le pouvoir à 5 citoyens tirés au sort parmi le million, et le tirage au sort désigne précisément 5 des démissionnaires. Le peuple sera sûr qu'il y a fraude; et il aura raison. De même, nous devons être certains que ce n'est pas le hasard seul qui a produit le résultat du vote; il y a donc une cause, et il est assez naturel de penser que c'est bien la cause recherchée, c'est-à-dire l'inspection des photographies (ce pourrait être aussi une entente entre les votants, ou une suggestion; ces hypothèses paraissent devoir être écartées d'après ce qui est rapporté des conditions de l'expérience).

Mais entrons dans le détail des résultats; nous les grouperons en exacts, douteux et inexacts.

I. Résultats exacts[2].

5 à l'unanimité de 20 voix.
5 par 19 voix contre 1 (ou contre 1 abstention).
6 par 18 voix contre 2 (ou abstentions).
2 par 17 voix contre 3 (—).
4 par 16 voix contre 4 (—).
3 par 15 voix contre 5 (—).
2 par 14 voix contre 6 (—).

27 résultats exacts au total.

1. C'est parce qu'elle s'est produite dans le sens de la vérité que nous disons 1 fois sur 1 000 000 et non 2 fois sur 1 000 000, qui serait la probabilité pour que se produise l'unanimité dans un sens non indiqué d'avance (soit 20 N, soit 20 A).

2. Ces résultats exacts ou douteux se répartissent à peu près également

II. Résultats douteux.

2 à la majorité de 13 voix.
2 — 12 —
1 — 11 —
4 à l'égalité des voix (10 contre 10).
1 à la minorité de 9 voix (contre 10 et 1 abstention).
1 — 5 voix[1] (contre 12 et 3 abstentions).
11 douteux.

III. Résultats inexacts.

1 où la solution exacte a seulement 5 voix contre 15.
1 — — 3 — 17.
2 inexacts.

Les résultats classés comme douteux sont ceux pour lesquels *il n'est pas improbable que l'exactitude du résultat soit due au hasard*. Leur nombre est 11; si l'on se reporte à la troisième colonne du tableau de la page 143 on voit que sur 1 024 576 épreuves régies par le hasard il y en a 60 460 pour lesquelles la majorité atteint ou dépasse 14 et 137 980 pour lesquelles cette majorité atteint ou dépasse 13; donc, sur 11 épreuves régies par le hasard, le nombre probable de celles pour lesquelles la majorité dans un sens donné d'avance serait égale ou supérieure à 14 est inférieur à 1, tandis que le nombre de celles pour lesquelles la majorité atteint 13 est supérieure à 1. C'est pour cela que nous avons classé les 2 scrutins où la majorité atteint 14 parmi les résultats exacts et les 2 scrutins où la majorité atteint 13 parmi les douteux. Il y a évidemment un certain arbitraire dans cette démarcation absolue; il serait puéril de dissimuler cet arbitraire; mais il serait excessif d'en tirer argument contre les

entre les normaux et les arriérés; voici cette répartition :

I. Exacts.		II. Douteux.	
20 voix	4 A, 1 N	13 voix	1 A, 1 N
19 —	2 A, 3 N	12 —	1 A, 1 N
18 —	5 A, 1 N	11 —	0 A, 1 N
17 —	0 A, 2 N	10 —	3 A, 1 N
16 —	1 A, 3 N	9 —	0 A, 1 N
15 —	3 A, 0 N	5/8 —	1 A, 0 N
14 —	0 A, 2 N		

Les deux résultats inexacts concernent deux arriérés déclarés normaux par le vote.

1. Ce résultat aurait pu aussi être rangé parmi les inexacts; il est douteux si l'on ajoute les abstentions à la minorité.

conclusions dans ce qu'elles ont de général; on doit simplement ne pas oublier que ces conclusions comportent quelque jeu, c'est-à-dire qu'au lieu de 27 résultats exacts sur 40, on pourrait en compter 25 à 29, peut-être même 24 à 30; mais cette légère incertitude dans la précision numérique n'entache pas le principe même des déductions : il est dans la nature même de tout calcul basé sur des expériences de ne fournir que des résultats approchés. Ces réserves faites, nous adopterons comme exacts les nombres des pages 145 et 146.

On voit que, pour 27 photographies sur 40, c'est-à-dire pour *les deux tiers* environ, on est conduit à admettre que le résultat du scrutin a été déterminé par l'aspect des photographies; il est tout à fait vraisemblable que ces 27 photographies, soumises à d'autres observateurs de même nature (instituteurs et institutrices), auraient conduit aux mêmes résultats; il serait intéressant de tenter l'expérience.

Pour 13 photographies, c'est-à-dire pour *un tiers*, l'ensemble des réponses ne donne pas une exactitude supérieure à ce qu'aurait fourni le hasard. On doit, semble-t-il, en conclure que, pour ces photographies, les caractères sur lesquels se basaient en général les observateurs pour juger étaient absents ou partiellement contradictoires. En résumé, il semblerait que, pour un tiers des photographies, l'expérience doive être interprétée comme conduisant à la conclusion que les caractères de ces photographies *ne permettent pas* de déterminer à quelle catégorie appartient l'enfant. Telle est la conséquence logique de cette expérience; je m'empresse d'ajouter que cette conséquence pourrait être partiellement modifiée par une expérience plus étendue portant sur les mêmes photographies; je suis convaincu toutefois qu'elle ne serait pas entièrement infirmée.

Enfin, disons quelques mots des deux scrutins dont le résultat est inexact, auxquels on doit peut-être joindre l'un de ceux qui ont été classés comme douteux. Il s'agit de trois enfants arriérés qui ont été déclarés normaux par 17 voix contre 3, 15 contre 5 et 12 contre 5. Il y a d'assez sérieuses présomptions pour que deux tout au moins des photographies (sinon les trois) possèdent les caractères des enfants normaux, c'est-à-dire pour que l'erreur de la majorité corresponde à une vérité relative, bien que contredisant la vérité absolue (c'est-à-dire l'appréciation des personnes qui connaissent très bien les enfants). On pourrait se demander si ce n'est pas parce que ces photographies sont plus difficiles à juger qu'elles ont ainsi conduit à des conclusions inexactes, et si des juges plus habiles n'arriveraient pas, sur elles, à la vérité. Il ne le semble pas, car ceux des observateurs qui ont donné sur ces photographies une appréciation exacte ne paraissent pas se distinguer par une habileté particulière; les pourcentages respectifs de réponses exactes sont, pour les observateurs ayant bien répondu dans les trois scrutins à résultat inexact :

82 0/0, 75 0/0, 67 0/0, 78 0/0, 67 0/0
82 0/0, 67 0/0, 92 0/0
77 0/0, 76 0/0, 67 0/0, 82 0/0, 75 0/0

Ces pourcentages oscillent entre le minimum général et le maximum général des 20 observateurs (67 p. 100 et 92 p. 100); leur moyenne, 76 p. 100, est légèrement inférieure à la moyenne générale, 78 p. 100, des 20 observateurs. Ce ne sont donc pas les observateurs les plus habiles qui ont le mieux répondu dans ces cas-là.

Comme conclusion générale de cette application de la méthode des majorités à l'expérience de Mme Rousson, nous arrivons à la conclusion que les photographies renfermaient les éléments d'une solution exacte dans environ 65 p. 100 des cas [1], devaient laisser dans le doute dans environ 30 p. 100 des cas, et devaient conduire à une solution inexacte (arriéré pris pour normal) dans environ 5 p. 100 des cas.

*
* *

J'ai appliqué une critique analogue à la précédente à une expérience fort intéressante de M. Binet sur des photographies de mains [2]. Il s'agissait de déterminer, sur une photographie de main d'enfant (ou plutôt deux photographies : paume et dos) 1° quel était le sexe de l'enfant; 2° si l'enfant était intelligent ou bête. La réponse à cette seconde question posée sous cette forme par M. Binet devait être donnée sans ambiguïté. L'expérience a porté sur 20 photographies, qui ont été soumises à 20 observateurs; j'ai posé moi-même à 4 observa-

1. Ceci ne contredit pas le fait que le pourcentage général des réponses exactes a été de 78 p. 100, car, dans les cas douteux, il y a environ une chance sur deux pour que la réponse soit exacte; si tous les cas étaient douteux, le pourcentage général des réponses exactes serait de 50 p. 100.
2. Cette expérience est inédite : je remercie M. Binet de m'avoir communiqué son dossier.

teurs la question de M. Binet relative à l'intelligence, et j'ai joint dans ce qui suit ces 4 réponses aux 20 obtenues par M. Binet [1]; voici les résultats :

1° Question relative au sexe.

13 réponses exactes :

2 par 19 voix sur 20 votants, 1 par 18 voix, 6 par 17 voix, 1 par 16 voix, 1 par 15 voix, 1 par 14 voix.

5 réponses douteuses :

1 par 13 voix, 2 par 12, 1 par 11, 1 par 7.

2 réponses inexactes :

5 voix pour la vérité contre 15.

Les conclusions sont à peu près semblables à celles que nous avons tirées de l'expérience de Mme Rousson, sauf, toutefois, que les majorités dans les réponses exactes sont moins considérables; les photographies des mains dévoilent un peu moins bien le sexe que les photographies des physionomies ne dévoilent l'intelligence. A signaler que les deux réponses inexactes se rapportent à deux garçons pris pour des filles : ces deux garçons sont intelligents : l'un deux a été jugé tel par 18 voix sur 24 et l'autre a eu dans cette seconde enquête 12 voix sur 24.

2° Question relative à l'intelligence.

6 réponses exactes :

1 par 19 voix sur 24, 2 par 18, 1 par 17 et 1 par 16;

Ces réponses exactes concernent 2 enfants jugés intelligents et 4 jugés bêtes.

12 réponses douteuses :

La solution exacte ayant obtenu, sur 24 votants, 3 fois 14 voix, 4 fois 12 voix, 2 fois 11 voix, 2 fois 10 voix, 1 fois 9 voix.

2 réponses inexactes :

La solution exacte n'ayant obtenu que 7 voix sur 24.

Ces deux réponses inexactes se réfèrent respectivement aux deux catégories.

Ces résultats sont, comme il était naturel de s'y attendre,

1. Cette adjonction diminue d'une unité le nombre des *douteux*; le sens général du résultat reste le même. Dans le dossier que m'a communiqué M. Binet, j'ai utilisé, pour chaque observateur, les réponses de la première épreuve, sans tenir compte des modifications de jugement à la seconde épreuve, dont je dirai un mot tout à l'heure.

très inférieurs comme exactitude à ceux qu'a fournis l'étude des physionomies; la majorité n'est jamais aussi nette et, dans les deux tiers des cas, les écarts sont de l'ordre de ceux qu'on pourrait attendre du hasard. Il semble donc bien qu'il y ait environ 60 p. 100 de cas dans lesquels on ne peut rien tirer des photographies de mains. Ce qui corrobore cette impression c'est l'étude d'une seconde expérience, dans laquelle un certain nombre des mêmes observateurs ont varié dans leur jugement sur les mêmes photographies (dans cette seconde épreuve, ils connaissaient le sexe des enfants); ces variations de jugement sont au total de 62, dont 46 sur les 12 cas que nous avons classés comme douteux, 15 sur les 6 cas où la réponse fut exacte et 1 seulement sur les deux cas où la réponse fut inexacte. Cela donne environ 4 variations pour chacun des cas douteux, 2,5 seulement pour les réponses exactes et 0,5 pour les réponses inexactes. Il y a donc des cas où le jugement est très incertain, et d'autres cas où il est plus sûr, et où l'examen décèle « quelque chose d'objectif ». Ce « quelque chose d'objectif » s'est trouvé coïncider avec l'intelligence dans 6 cas sur 8; ces chiffres sont trop faibles pour que le calcul des probabilités permette de rien dire à leur sujet; il serait illusoire de faire des pourcentages ; en remplaçant 6 sur 8 par 75 p. 100, on commettrait une grave erreur; car il peut fort bien arriver que le hasard pur (c'est-à-dire une probabilité de 50 p. 100) donne 6 succès sur 8 expériences, tandis que 75 succès sur 100 expériences décèlent, avec une certitude pratiquement absolue, une cause.

C'est donc par une étude concrète que l'on peut chercher à déterminer la nature de ce « quelque chose d'objectif » que décèle l'expérience. L'examen des photographies a suggéré à M. Binet l'hypothèse suivante : lorsque les doigts sont courts et laids, on juge l'enfant bête; lorsqu'ils sont longs et effilés, on le juge intelligent. Pour vérifier cette hypothèse, M. Binet a demandé à deux nouveaux expérimentateurs de faire une classification en basant leur jugement uniquement sur le critérium précédent. Cette expérience figure dans le dossier qu'il m'a remis; pour ne pas me laisser influencer par elle, je ne l'ai dépouillée qu'après avoir rédigé tout ce qui précède : elle en apporte une confirmation remarquable, de sorte que l'hypothèse de M. Binet me paraît entièrement confirmée par mon étude arithmétique. En effet, dans les 6 cas où la réponse de la majorité a été exacte, et dans les 2 cas où elle a été

inexacte, *toutes* les réponses de ces deux derniers observateurs concordent exactement avec celles de la majorité[1]. Cette concordance parfaite, portant sur 16 opinions (2 observateurs jugeant chacun 8 photographies), est en quelque sorte la vérification expérimentale de la méthode des majorités, et cet exemple montre bien comment cette méthode devra être appliquée dans les cas, comme celui-ci, où l'on ne sait pas très exactement *a priori* sur quoi porte la recherche. La méthode des majorités nous apprend qu'il y a « quelque chose d'objectif »; l'observation directe des faits conduit ensuite à formuler des hypothèses, qui doivent être vérifiées par de nouvelles expériences. Le calcul n'a joué là, comme toujours dans les sciences expérimentales, qu'un rôle d'intermédiaire entre deux expériences, mais ce rôle de guide de l'expérience, pour être modeste, est fort souvent des plus utiles.

ÉMILE BOREL,
Professeur adjoint à la Sorbonne.

1. Une seule différence insignifiante a signaler; pour l'un des enfants jugé bête par la majorité (réponse exacte), l'un des deux observateurs a inscrit « moyen », l'autre ayant d'ailleurs inscrit « bête ».