

Par Pierre Livet, professeur émérite de philosophie à Aix-Marseille-Université, directeur de programme à IMÉRA.

Les propositions des Agences étatiques ou internationales qui annoncent qu'« il n'existe pas dans l'état actuel des recherches de preuve scientifique que » telle substance- pesticide par exemple- soit nocive pour la santé suscitent souvent des réactions soupçonneuses voire indignées, du genre : « pas de preuve, alors que c'est évidemment nocif », « s'il fallait attendre leurs preuves, on serait tous empoisonnés ». On peut s'interroger sur les éléments qui induisent ces réactions. Il est clair que ces réactions sont liées à des émotions ; les propositions faites ne parviennent pas à entamer des croyances, selon lesquelles tel pesticide, par exemple, est cancérigène. Et quand un non scientifique tente de lire un article scientifique qui fait la revue de l'état des connaissances et expériences, il voit qu'on parle d'incertitudes et de probabilité, et que les résultats sont jugés « significatifs » (le plus souvent, si $p < 0,05$). Si on combine « significatif » avec « incertitude », ou probabilité, on peut se demander dans quelle mesure on peut considérer les résultats comme des données stables et décisives. Et c'est encore pire quand on dit dans ces articles qu'on a pu observer des effets toxiques mais qui n'atteignent pas le seuil de significativité, bien que plusieurs expériences donnent des résultats qui vont dans un sens contraire. Puisqu'il y a des effets nocifs, - même non prouvés signifiants- pourquoi attendre pour interdire le produit ? Bref, incertitude, croyances et émotions se combinent pour créer soit défiance soit indignation envers les annonces pourtant murement réfléchies des Agences. Tentons de voir comment croyances et émotions se combinent, et pourquoi le traitement scientifique de l'incertitude est malaisé à interpréter.

Prenons le terme de « croyance » dans un sens plus logique que religieux. Une croyance est alors une attitude d'une personne envers une proposition (ex : le glyphosate est cancérigène) dont cette personne pense qu'elle est vraie ou a de bonnes chances de l'être. Comment une telle attitude a pu se mettre en place dans son esprit ? Il semble difficile qu'elle puisse décider consciemment d'avoir telle croyance- par exemple parce que cela irait dans le sens de ses désirs. Si elle sait que c'est seulement pour cette raison qu'elle a cette attitude, elle va avoir des soupçons et se douter que la croyance pourrait bien être fausse. Quoiqu'il en soit, cela va plutôt affaiblir sa croyance.

Puisqu'on ne peut pas adhérer fermement à une croyance simplement parce qu'on le veut, il faut donc que cette croyance s'installe « involontairement » dans l'esprit de celui qui y croit. Un beau jour, il découvre qu'il adhère à cette croyance – alors que peut-être jusque-là il n'avait pas d'attitude claire à son égard.

Si nous étions des esprits complètement lucides, les croyances ne s'installeraient en nous que lorsqu'elles seraient basées sur des raisons solides. Mais même alors, ce ne serait pas parce que nous avons pour politique de préférer des raisons que nous pourrions défendre que nous nous trouvons croire telle croyance. Elle apparaît à un moment faire corps avec toutes ces raisons. On pourrait dire alors que le ciment de la croyance a pris. Mais il en est comme le ciment. Une fois qu'il a pris, il est dur de défaire les liens qu'il a établis. Une fois que nous avons adhéré à une croyance, nous avons tendance à résister aux critiques qui lui sont faites, y compris quand cette croyance est beaucoup moins fondée que nous ne le croyions.

Les émotions participent-elles à cette résistance ? Pour simplifier, je considérerai seulement deux grandes classes d'émotion, les émotions d'inquiétude et les émotions d'espoir. Frijda disait que les émotions sont des potentiels de tendance à l'action. L'inquiétude, la peur, sont des émotions d'alerte, qui se déclenchent en nous rendant sensible à un danger et en nous motivant pour soit le fuir, soit y parer. Inversement les espoirs sont des émotions qui nous motivent à engager de nouvelles actions et à les poursuivre pour atteindre notre objectif. Une espèce qui est sensible aux dangers et qui est aussi motivée par des espoirs est plutôt mieux armée qu'une autre moins sensible et moins motivée pour survivre et se développer dans l'évolution. Si on s'en tient à une conception évolutionnaire de la rationalité, il sera rationnel évolutionnairement d'être un peu plus inquiet devant des dangers et inversement, un peu plus motivé par des espoirs, qu'un simple calcul qui combine les degrés d'incertitude concernant les faits connus ne le conseillerait. Cela nous donne plus de chances d'échapper à des dangers qui se révèlent plus graves que prévu, et plus de chances de saisir des opportunités qui sont plus à notre portée que prévu.

Cependant, si cela nous permet d'avoir plus de chances de survie, cela peut nous amener à surévaluer soit les dangers, soit les espoirs. Nos croyances seront alors moins en accord avec les calculs rationnels idéaux – ceux que nous aurions faits si nous étions certains de nos estimations des probabilités et si nous raisonnions de manière impeccable. Dans l'inquiétude ou dans l'espoir, nos propres croyances seront un peu trop hâtives.

Cela peut être gênant pour un individu, mais collectivement, une société a plus de chances de survivre et de se développer si elle compte à la fois des gens plus inquiets, qui sonnent plus facilement l'alerte, et par ailleurs des gens plus optimistes, qui peuvent donner de l'espoir.

Les espoirs – et parfois les craintes – peuvent motiver des recherches scientifiques, mais les scientifiques ne présentent pas les résultats de leurs travaux en indiquant leur aura émotionnelle. L'idéal de la science que nous avons tous, scientifiques et non scientifiques, est celui d'une science dont les résultats sont prouvés. Le parangon de la preuve est la preuve logique – ou mathématique si on marginalise, comme il est raisonnable de le faire, les résultats des logiciens concernant l'existence d'énoncés réflexifs indécidables à l'égard de l'arithmétique, dès qu'on y introduit la notion d'égalité. Quand on a donné une « vraie » preuve d'un théorème – preuve mise à l'épreuve par les autres mathématiciens – aucune mise à l'épreuve de cette démonstration ne pourront la disqualifier. Tout au plus les extensions des mathématiques pourront elles relativiser cette preuve au cadre axiomatique dans lequel elle se développe. Ainsi, quand en mathématiques on a prouvé une proposition, on a aussi prouvé qu'on l'a prouvée (les logiciens vont jusqu'à vouloir prouver la validité des méthodes pour prouver la preuve). Une preuve nous fait donc passer de la croyance au savoir. Une croyance est supposée vraie mais peut se révéler fautive, le savoir ne peut être faux – il est simplement limité.

Qu'en est-il dans le domaine des recherches expérimentales, et tout particulièrement de celles qui portent sur des domaines comme celui du vivant où les phénomènes et surtout leurs mesures peuvent révéler tout un éventail de variations ? On cherche à intégrer ensemble ces variations tout en tenant compte de leur diversité, et on travaille alors sur des distributions de valeur de mesure, ou plus généralement des distributions de données. Cela donne lieu à des statistiques. On recherche des corrélations entre les variations de différentes données. On tente de projeter les corrélations enregistrées sur d'autres

expériences – on utilise alors des probabilités- et on dispose de méthodes pour évaluer la validité de ces projections.

Quand les hypothèses sur ces projections sont validées à un certain degré, dont on a calculé qu'il minimise la possibilité d'erreurs, peut-on considérer que ces hypothèses sont prouvées ? L'étude des méthodes statistiques et probabilistes, dans leur formalisme, donne bien lieu à des théorèmes mathématiques, et donc à des preuves au sens classique. Mais en est-il de même du mode de validation des hypothèses, qui portent sur telle corrélation – sans parler de causalité- entre tel phénomène et tel autre phénomène ? Cette validation n'a pas le même statut que les preuves mathématiques. Il vaudrait mieux parler de « mise à l'épreuve » plutôt que de « preuve » en ce sens paradigmatique. En revanche, on a des preuves formelles des capacités – et des limites- des méthodes de mise à l'épreuve.

Considérons une de ces méthodes, le raisonnement bayésien. C'est un raisonnement qui nous permet des apprentissages en situation d'incertitude. Il nous dit comment évaluer la probabilité qu'aura notre hypothèse une fois confrontée à l'expérience, et cela en mettant en relation et en comparant différentes incertitudes¹.

La première incertitude est celle que présente notre hypothèse avant que nous l'ayons confrontée aux expériences. Si notre hypothèse est très innovante, elle est sans doute plus risquée, et son incertitude a priori est alors plus élevée. Si elle est déjà très plausible étant données ces connaissances, elle est plus probable – mais sans doute aussi moins innovante.

Nous la mettons ensuite à l'épreuve selon un protocole expérimental. Certaines données obtenues dans les expériences sont en accord avec notre hypothèse, et quelques autres moins proches des résultats attendus. Les proportions entre ces données favorables et ces données défavorables permettent de définir dans quelle mesure notre hypothèse a été renforcée par l'expérience.

Il nous faut encore évaluer cette probabilité relativement à celle de l'ensemble des hypothèses, qui ne comprend pas que la nôtre, mais aussi d'autres hypothèses concurrentes. Nous pouvons alors définir la probabilité de notre hypothèse une fois mise à l'épreuve des expériences - la probabilité dite « a posteriori ».

On retrouve alors le « degré de significativité » des résultats si omniprésent dans les articles scientifiques. Supposons que l'hypothèse envisage un effet (l'effet possible d'un pesticide sur une maladie), ou plutôt, en fait, une corrélation positive entre usage du pesticide et maladie. La mise à l'épreuve par l'expérience permet de mettre cette hypothèse en concurrence avec ce qu'on appelle « l'hypothèse nulle », celle selon laquelle il n'y a pas de corrélation, pas d'effet – hypothèse qui ne propose aucune explication, par ailleurs. La probabilité, après expérience, de l'hypothèse nulle deviendra pour nous l'indicateur du degré de significativité de l'hypothèse de la corrélation que nous voulons tester. Ce degré est d'autant plus élevé que la probabilité de l'hypothèse nulle est basse.

Par précaution, on souhaite que cette probabilité soit très basse (généralement, égale ou inférieure à 0,05 ou 5%). Dans certains cas, ce seuil n'est même pas assez bas, par exemple si la maladie qu'on envisage est rare ou très rare.

Il en est un peu différemment quand on compare l'hypothèse de telle corrélation avec les quelques hypothèses concurrentes qui proposent des explications du phénomène. Dans ce cas, une hypothèse risquée – et donc peu probable a priori- pourrait l'emporter par

¹ Voir en annexe l'interprétation de la formule de Bayes qui sert de base à ce raisonnement

comparaison, si elle est fortement renforcée par l'expérience et si les autres hypothèses ne le sont pas.

Tout cela peut paraître bien compliqué – et c'est une version simplifiée ! Mais les conséquences, pour ce qui concerne en quoi des dispositifs expérimentaux, en situation d'incertitude, fournissent des preuves d'hypothèses, sont plus claires.

Si l'hypothèse nulle a une très faible probabilité après expérience, notre hypothèse a une significativité élevée. Mais cela ne permet pas de dire qu'elle est « prouvée », au sens d'un théorème mathématique. Ce qu'on a prouvé, ce n'est pas l'hypothèse, c'est seulement « que l'hypothèse a été mise à l'épreuve de l'expérience et a un seuil de significativité élevé ». Pour faire plus simple, **on a prouvé que l'hypothèse a passé l'épreuve.**

Cela ne prouve pas que les hypothèses concurrentes ont été éliminées – même l'hypothèse nulle a encore une probabilité, bien que très faible.

Inversement, on ne peut pas prouver l'absence de nocivité d'un produit – autrement dit, l'absence d'effet nocif et donc l'hypothèse nulle- parce que pour cela, il faudrait considérer toutes les hypothèses nulles concernant tous les types et modes de nocivité, alors que la liste des nocivités possibles que l'on peut faire ne peut pas prétendre être exhaustive.

Il faut aussi se rappeler, ce qui va dans le même sens, que lorsque l'on considère plusieurs hypothèses différentes (et non pas simplement l'hypothèse d'un effet et l'hypothèse nulle), elles peuvent avoir des degrés de significativité dont le total dépasse 1. Ce qui veut dire qu'une hypothèse ne consomme pas une part de significativité comme on consommerait une part de gâteau (dont la mesure est de 1), en ne laissant pour les autres qu'un gâteau réduit d'autant.

Cependant, si une autre hypothèse n'a pas atteint le seuil de significativité (c'est-à-dire si la probabilité de son hypothèse nulle n'est pas égale ou inférieure à 5%), on alors bien prouvé quelque chose : on a **prouvé** que l'hypothèse en question **n'a pas surmonté l'épreuve.** Et c'est ce que veulent dire les phrases du genre « l'existence d'un effet de tel pesticide sur telle maladie n'a pas été prouvée ». Cela du moins si on a fait des expériences, car si on n'en a pas faites, l'hypothèse de cet effet n'aura effectivement pas été mise à l'épreuve, mais on n'aura évidemment pas prouvé non plus qu'elle n'a pas surmonté l'épreuve.

Revenons sur les modes de réception de ces types de résultats par les publics dont nous faisons tous partie, modes de réception qui font jouer les relations entre émotions et croyances. Quand nous avons été alertés sur une possible corrélation entre tel pesticide et telle maladie – ou tel effet sur les abeilles, ou la décroissance de la biodiversité- et que des scientifiques nous répondent « on n'a pas prouvé que... » nous avons tendance à comprendre : « vos croyances dans cet effet nocif, et vos émotions d'alerte, vous n'avez qu'à les refouler », le rouleau compresseur scientifique est passé par là.

La disqualification d'émotions d'alerte, quand elle est fondée sur une justification qui reste obscure, et qu'on ne saisit pas le travail de mise à l'épreuve qui a eu lieu, peut déclencher des indignations, qui tiennent à la fois à ce qu'on peut prendre pour un mépris des alertes et pour le recours à une langue de bois (qui disqualifie ceux qui ne la décodent pas).

Ces émotions motivent la recherche d'arguments qui puissent rétablir une certaine symétrie entre les lanceurs d'alerte et les scientifiques. Puisque les formulations des chercheurs étaient disqualificatrices, ces arguments vont viser à disqualifier les chercheurs. Les rares cas de fraude scientifique seront montrés du doigt, et surtout on soupçonnera très vite des

conflits d'intérêts entre les chercheurs et des entreprises et multinationales qui visent le profit avant la santé publique.

Or quand le chercheur veut répliquer, il a un problème de taille : montrer qu'une relation avec une entreprise qui finance un laboratoire et un projet n'a pas influencé chez le chercheur les hypothèses et l'évaluation des résultats est à peu près aussi difficile que prouver l'hypothèse nulle ! Imaginons que les contre-arguments fournis puissent renforcer cette défense, il reste toujours un soupçon. Ce qui veut dire que, si élevé que puisse être le degré de significativité de l'hypothèse : « cette relation n'a pas empêché le chercheur de mener sa recherche en toute indépendance, et ne l'a pas amené à biaiser ses résultats », ce degré sera toujours insuffisant pour éliminer le soupçon. En un certain sens, l'opinion publique utilise à l'envers les précautions que prennent les scientifiques pour leurs mises à l'épreuve - mais sans fixer la méthode de cette mise à l'épreuve !

Il n'empêche que dans une discussion argumentée, mettre l'adversaire dans la position où il devrait prouver l'hypothèse nulle est une stratégie payante. Les mises à l'épreuve scientifiques, de leur côté, ne peuvent consister qu'à diminuer la probabilité de l'hypothèse nulle en renforçant l'hypothèse d'un effet. Or, dans une argumentation qui ressort du simple débat, et non de la mise à l'épreuve expérimentale, on ne peut mesurer le degré auquel un argument renforce une position par rapport à la position opposée. Les arguments peuvent faire chacun appel à des valeurs différentes, qu'il est difficile de rendre commensurables, voire même simplement d'ordonner les unes par rapport aux autres. La vista des débatteurs persuadés qu'un effet nocif est scandaleusement négligé a dans ce champ du débat tout autant de force de disqualification que les mises à l'épreuve scientifique.

Ce conflit entre disqualifications réciproques est un conflit bloqué. Pour en sortir, il faut que les deux parties puissent sortir de ces postures de disqualification. Cela demande, de la part des scientifiques, l'abandon d'une posture surplombante. En particulier, les exigences de significativité ou de puissance d'une expérimentation sont des exigences limites. Entre ces limites, entre la probabilité très faible de l'hypothèse nulle et la forte significativité de l'hypothèse d'un effet, bien des éléments de probabilité intermédiaire existent, qui ne sont pas pour autant « insignifiants ». Différents éléments peuvent être d'utiles indicateurs sans permettre de conclure. Un de ces éléments peut donner quelque pertinence à telle hypothèse, et un autre à telle autre. Ces hypothèses peuvent être partiellement compatibles. Les mises à l'épreuve peuvent donc aussi révéler des zones de compatibilités entre des croyances que les scientifiques et le public pensaient au départ en conflit.

Sortir de ces blocages demande aussi, de la part du public, d'utiliser aussi ses capacités d'alerte pour repérer les positions qui n'ont pas de fondements avérés et qui n'ont pas été mises à l'épreuve – ce que les scientifiques appellent les pseudo-sciences. Il est aisé de se rendre compte que telle théorie pourtant très répandue n'a simplement jamais passé l'épreuve des tests expérimentaux. Il est un peu plus compliqué, mais cela ne demande pas des capacités scientifiques élaborées, de montrer que telle citation d'une justification d'une décision d'agence est tronquée, et que la partie de la phrase qui est coupée lui donne un sens opposé à son utilisation dans le débat. Inversement, les scientifiques jouent aussi eux-mêmes le rôle de lanceurs d'alertes, et ils peuvent dans bien des cas collaborer avec les donneurs d'alertes publiques. Ils peuvent alors tenter de nous éclairer sur les liens entre les alertes qu'ils soutiennent et les mises à l'épreuve expérimentales. Cette coopération entre

scientifiques et public n'est possible que si l'éducation permet de mieux faire comprendre quelle est la spécificité d'une mise à l'épreuve scientifique. Cela implique de présenter non seulement les résultats scientifiques, mais peut-être davantage les démarches qui les font avancer, les exigences et les difficultés de ces démarches, comme les repères que ces démarches se donnent. On peut alors mieux comprendre ce que veut dire la significativité, pourquoi il faut répéter les expériences, pourquoi il faut faire des analyses de cohortes ; en quoi une recherche prospective, qui, en termes de mise à l'épreuve, est plus exigeante qu'une enquête rétrospective puisqu'on ne dispose pas des données pour y ajuster l'hypothèse, apporte un renfort plus important à l'hypothèse qui passe cette épreuve. On peut aussi mieux comprendre pourquoi le temps de la recherche scientifique n'est pas celui d'une démonstration éclair ou de la simple application directe d'un résultat, mais est plutôt le temps qu'il faut pour mettre et de remettre à l'épreuve, et non pas celui d'un débat le temps d'une soirée télévisée. C'est là un point commun entre les scientifiques et ceux des donneurs d'alerte engagés : ils travaillent sur du long terme, ce qui les oblige à voir leurs croyances mises à l'épreuve, mais les amène aussi à savoir garder espoir dans les difficultés et à rester en alerte quand d'autres préféreraient oublier leurs inquiétudes.

Annexe : Interprétation de la formule de Bayes

La formule de Bayes est la suivante :

$$P(H_j \text{ si } E) = \frac{P(H_j) \times P(E \text{ si } H_j)}{\text{Somme des } (P(H_i) \times P(E \text{ si } H_i))}$$

$P(H_j \text{ si } E)$ est dite « probabilité a posteriori », après l'expérience E , de l'hypothèse H_j .

$P(H_j)$ est la probabilité « a priori » de l'hypothèse H_j . Elle tient au rapport entre cette hypothèse et les connaissances les plus probables de nos connaissances au moment où nous la formulons.

$P(E \text{ si } H_j)$ est la probabilité que l'expérience E fournisse des données en accord avec H_j , et donc renforce H_j . Nous combinons donc la probabilité a priori de H_j avec ce renforcement.

Nous devons pondérer cette probabilité renforcée de H_j en la comparant aux probabilités a priori de toutes les hypothèses H_i (incluant H_j), du moins de celles qui sont compatibles avec les connaissances acquises, et de leurs degrés de renforcement par les données de l'expérience.

Une fois que nous aurons calculé les rapports entre toutes ces probabilités ou incertitudes, nous pourrons définir la probabilité « a posteriori » de notre hypothèse, celle que sa mise à l'épreuve par l'expérience aura révélée.

Pour atteindre une probabilité a posteriori très basse pour l'hypothèse nulle, il faut a contrario que la probabilité a posteriori de l'hypothèse H_j soit très haute. La formule nous montre que cela exige d'une part que la probabilité a priori de H_j soit très haute, et d'autre part que son renforcement par l'expérience ($E \text{ si } H_j$) soit très fort.