

Comment **évaluer** la **scientificité** ?

**« L'habit ne fait pas le moine »,
un impératif de la démarche
scientifique**

Comment évaluer la scientificité ?

2e édition

2023

Titre de la 1ere édition : « L'évaluation de la scientificité et le mythe des pseudosciences »

Auteurs

Jessy Loranger

Jonathan Loranger

Tous droits réservés

© 2023, Communication Scientifique Ulysse S.E.N.C.

Saint-Hubert, (Qc) Canada

BIOGRAPHIE DES AUTEURS

Jessy Loranger

Après une Maîtrise en écologie terminée avec honneur en 2012 à l'Université de Sherbrooke (Québec, Canada) et en collaboration avec l'Université de Jena (Allemagne), Jessy Loranger a effectué un doctorat en cotutelle entre l'Université de Sherbrooke et l'Université de Montpellier (France) portant sur l'assemblage des communautés de plantes herbacées. En janvier 2016, il a ainsi obtenu un double titre de Docteur (PhD) en écologie des plantes et en écosystèmes, avec mention d'excellence. Au cours de ses études, il a produit une dizaine d'articles scientifiques publiés avec révision par les pairs et a agi en tant que réviseur pour une dizaine de revues ou organismes subventionnaires. Les impacts du système de publication scientifique l'intéressent depuis plusieurs années, encore plus à la suite à ses propres expériences de recherche, et après le partage des expériences des dizaines de scientifiques internationalement reconnus avec lesquels il a collaboré. Il est depuis 2018 un planificateur de permaculture certifié et travaille depuis 2023 en tant que coach de vie pour ceux qui veulent mieux gérer et dépasser leurs problématiques professionnelles, personnelles ou de santé physique et psychologique.

Jonathan Loranger

Docteur en psychologie depuis 2020, ses travaux de recherche ont principalement porté sur les avancées de la synergologie quant aux connaissances sur le langage corporel universel et sur leur application pratique, notamment en psychothérapie. Jonathan Loranger est aussi psychologue de profession, spécialisé en

violence conjugale et dans les facteurs de changement. Très impliqué dans le milieu des ressources communautaires tout au long de ses études en psychologie, il contribue aujourd'hui à la lutte contre la violence conjugale en supervisant des intervenants communautaires qui répondent aux demandes d'aide des gens qui veulent prévenir ou cesser la violence conjugale. Dès le début de ses études en psychologie, il étudie l'histoire et les fondements des différentes approches en psychologie. Depuis toujours passionné d'épistémologie, il est stimulé par les débats entourant l'évaluation de la scientificité de la psychologie. Il développe ainsi un regard critique sur l'accessibilité, le traitement et l'interprétation des résultats de recherche dans la société. C'est grâce aux échanges animés avec son frère Jessy Loranger que le projet du présent livre prit forme. Tous les deux étant bien conscients des croyances populaires sur la science, ils savaient qu'un livre s'imposait pour les démystifier.

TABLE DES MATIÈRES

BIOGRAPHIE DES AUTEURS 6

TABLE DES MATIÈRES 8

PRÉFACE 12

Le débat scientifique 14

LA DÉMARCHE SCIENTIFIQUE « DÉMYSTIFIÉE » 17

Les étapes de la démarche scientifique 17

**Se prononcer sur ce qu’est la science ou la
pseudoscience 23**

Le piège des faux sceptiques 30

Méthodes des faux sceptiques 33

Application de la démarche scientifique 38

*Une démarche scientifique, plusieurs méthodologies de
recherche 44*

L’étude de cas 48

Un exemple : les pensées 50

**L’INFLUENCE DU SYSTÈME DE PUBLICATION SUR LA
COMMUNICATION SCIENTIFIQUE 54**

**Les rôles du système de publication des revues
scientifiques 56**

Privatisation de la publication scientifique 57

Pressions des organismes subventionnaires	59
Obligations carriéristes des scientifiques	64
<i>Publie ou périclisse</i>	<i>67</i>
<i>La course aux citations</i>	<i>73</i>
<i>Le prestige des revues</i>	<i>76</i>
Une atmosphère de compétition insoutenable	80
<i>Inconduites professionnelles</i>	<i>84</i>
<i>L'invention de faits à travers les citations</i>	<i>90</i>
Biais contre les résultats non-significatifs	93
<i>Biais contre les résultats non-significatifs : la valeur-p et son influence</i>	<i>97</i>
Biais contre la reproduction	114
<i>Démonstrations empiriques</i>	<i>117</i>
<i>Reproductibilité protocolaire</i>	<i>118</i>
<i>Réplication des résultats</i>	<i>120</i>
<i>Mise en garde et pistes pour l'avenir</i>	<i>121</i>
Inconduites et biais de publication : les effets sur la science	128
<i>Les méta-analyses</i>	<i>128</i>
<i>Le gaspillage des ressources</i>	<i>132</i>
<i>La sélection de la mauvaise science</i>	<i>133</i>
<i>Les rétractions</i>	<i>137</i>
<i>L'aspect social</i>	<i>140</i>
Bilan et perspectives	142
<i>Des solutions concrètes</i>	<i>144</i>
LA RÉVISION PAR LES PAIRS	152
Fonctionnement et appréciation	154
Contrôle de qualité	157
<i>Rejet d'articles médiocres</i>	<i>157</i>

<i>Acceptation d'articles importants</i>	160
Suppression de l'innovation	162
Biais de révision	165
Inconduites professionnelles et biais de révision	168
<i>Les conflits d'intérêts</i>	168
<i>Fabrication de réviseurs et rudesse</i>	171
Bilan	173
Solutions et alternatives au système habituel	179
<i>Formation à la révision et différentes motivations</i>	179
<i>Modifications du système lui-même</i>	181
<i>Les banques d'archivage</i>	185
 DIFFÉRENTS MILIEUX DE RECHERCHE ET LEUR FAÇON	
D'ABORDER LA COMMUNICATION SCIENTIFIQUE	189
 Recherche fondamentale versus recherche appliquée	190
Valorisation de la recherche : une question de financement	192
<i>Recherche fondamentale financée par le public</i>	193
<i>Recherche fondamentale financée par le privé</i>	196
<i>Recherche appliquée financée par le public</i>	199
<i>Recherche appliquée financée par le privé</i>	203
Bilan	204
 REMARQUES FINALES	206
 Alternatives et perspectives par rapport aux systèmes actuels	207
<i>Ébauche d'un plan d'action</i>	208
<i>De grands défis à surmonter</i>	211

<i>Vivre avec le paradoxe</i>	217
Le mythe des pseudosciences	221
BIBLIOGRAPHIE	226
REMERCIEMENTS	239

PRÉFACE

« L'habit ne fait pas le moine. » Voilà un proverbe que nous avons tous déjà entendu à un moment ou un autre de notre vie. Nous pourrions le mettre en relation avec une autre phrase autant connue : « Le contenu est plus important que le contenant. » Ce sont deux morceaux de sagesse populaire qui nous serviront de fils conducteurs tout au long de cet ouvrage et qui nous aideront à ne pas perdre de vue l'objectif que nous poursuivons. Ces proverbes nous indiquent que peu importe notre impression initiale d'une chose (un texte ou une personne, par exemple), peu importe l'apparence extérieure ou la provenance, seule une inspection minutieuse et détaillée du contenu peut nous permettre d'en développer une opinion réelle et informée. Encore faut-il avoir le désir de développer une opinion réelle et la motivation d'y investir temps et énergie. C'est d'ailleurs la raison pour laquelle nous ne pouvons pas développer d'opinion informée sur chaque personne, sur chaque idée croisant notre route. Tout est une question d'intérêt et de ressources : si nous sommes intéressés par la nouvelle situation, nous serons plus à même d'investir ce qu'il faut pour nous informer.

Paradoxalement, dès que nous nous trouvons face à une nouvelle personne ou à une nouvelle idée, notre cerveau a tendance à faire automatiquement des liens avec nos expériences et nos connaissances passées, avec nos représentations mentales, afin de forger une opinion le plus rapidement possible. Ceci est une fonction de survie importante, nous permettant de réagir à une nouvelle situation rapidement et de la façon la plus conservatrice possible grâce à l'information que nous possédons déjà. Faisant fi du fait que l'habit ne fait pas le moins, nous aurons

alors tendance à juger rapidement les nouvelles personnes que nous rencontrons par rapport à comment elles nous font penser aux personnes que nous avons rencontrées dans le passé, et les nouvelles idées que nous entendons par rapport aux convictions que nous avons déjà. Ceci peut nous éviter des souffrances et des pertes de temps inutiles.

Cependant, cette fonction de jugement rapide avec de l'information limitée n'est pas destinée à constituer une fin en soi. Il s'agit plutôt de nous inciter à poser les bonnes questions, à observer prudemment, à colliger de l'information, afin d'ultimement confirmer notre idée d'origine, ou l'infirmier et apprendre. Ceci est d'autant plus important – et même attendu – au sein de l'activité scientifique. Dans le monde scientifique, cette fonction est nommée « sain scepticisme » et représente une qualité nécessaire. Pareillement, ce scepticisme n'est pas supposé être une fin en soi, mais bien la base pour une plus ample investigation.

Nous le répétons encore une fois : en science ou ailleurs, aller au-delà de notre première impression n'est possible que si l'intérêt pour la nouvelle situation et les ressources pour l'étudier sont présentes, car nous ne pouvons pas nous investir pour nous informer sur tout ce avec quoi nous entrons en contact ; nous sommes souvent obligés de prendre des décisions avant de pouvoir se permettre une investigation assez élaborée pour apporter un éclairage définitif. Autrement, nous n'en finirions plus et nous ne rentrions jamais en action; nous serions paralysés par l'incertitude. Par contre, sachant que notre cerveau développera inmanquablement une opinion sur chaque nouvelle situation, même pour lesquelles notre intérêt n'est pas suffisant pour creuser plus loin, la rigueur veut que nous prenions conscience

que cette opinion n'est alors pas une opinion informée. En d'autres mots, nous ne devons jamais perdre de vue qu'il ne s'agit que d'hypothèses et nous nous devons de les traiter ainsi.

L'origine et le rôle du sain scepticisme, ainsi que le concept de l'habit qui ne fait pas le moine – ou du contenu plus important que le contenant – sont deux points essentiels sur lesquels nous reviendrons régulièrement au travers de ce livre. Comme nous le verrons, ce sont deux concepts cruciaux à intégrer afin d'évaluer la scientificité avec succès, mais aussi afin de s'apercevoir quand d'autres ne le font pas correctement.

Le débat scientifique

En science, le débat scientifique est un outil utile pour s'assurer que le sain scepticisme puisse faire son travail, mais aussi pour s'assurer qu'il reste sain. Le débat scientifique est utilisé pour comparer des points de vue et pour pouvoir arriver à une conclusion satisfaisante au sein de la communauté scientifique. Pouvoir débattre de la rigueur scientifique d'un travail, de l'interprétation des résultats ainsi que des théories qui y sont rattachées permet d'évaluer les travaux le plus objectivement possible et de tirer les conclusions les plus logiques. Dépendamment de la quantité et de la qualité de l'information disponible aux scientifiques avant et après le début d'un débat, la durée de ces débats peut varier énormément. Pour que de tels débats puissent vraiment remplir leur mission, il est important que chaque scientifique suive et comprenne bien la démarche scientifique, puisque c'est le respect de celle-ci qui déterminera si un travail est scientifique ou non.

Ceci paraît parfaitement raisonnable et même évident, mais malheureusement la démarche scientifique semble demeurer un mystère pour bien des gens, malgré sa nature simple et logique (Fig. 1). Même les scientifiques en milieu universitaire ne sont étonnamment pas à l'abri d'une méprise ou d'une confusion quant à l'évaluation de la rigueur scientifique d'une recherche^{29,48}. Ainsi, l'évaluation de la scientificité d'un travail donné est souvent faite de façon inadéquate pour diverses raisons, phénomène qui, à son tour, compromet les débats scientifiques et, ultimement, la crédibilité de la science elle-même.

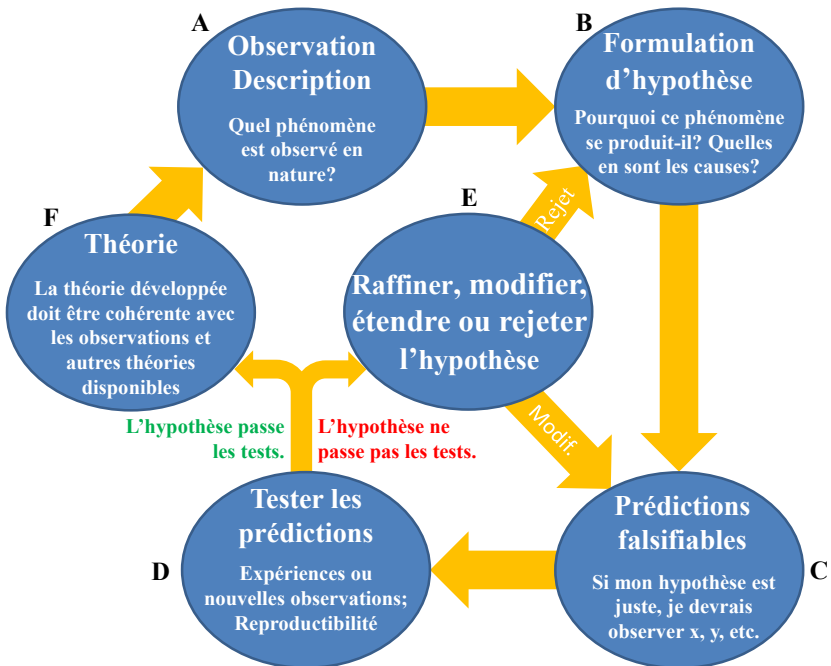


Figure 1. Les différentes étapes de la démarche scientifique Tout travail suivant ces étapes peut être considéré comme scientifique, peu importe l'objet d'étude ou le contenant, donc l'« habit » de présentation. Figure tirée et modifiée d'après Garland⁷⁵ et Wolfs²⁰⁹.

En utilisant les fondements de la démarche scientifique comme point de référence, nous viserons, au fil du texte, à expliciter ce qu'une personne a besoin de connaître pour évaluer la scientificité d'un travail ainsi que les pièges les plus communs lors de cette évaluation.

Premièrement, la personne doit être au clair avec la démarche scientifique et avec les principales mésinterprétations sur la science. Deuxièmement, elle doit bien comprendre l'influence du système actuel de publication sur la communication scientifique, puisque ce système est aujourd'hui intimement lié à la scientificité dans l'esprit de bien des gens, incluant les scientifiques. Troisièmement, elle doit être à même de réfléchir sur ce qu'implique réellement le système actuel de révision par les pairs étant donné le rôle clé que cette révision joue aujourd'hui dans la crédibilité scientifique. Finalement, le point reliant toutes ces idées est, encore une fois, qu'il ne faut pas oublier que ce n'est pas l'habit qui fait le moine. Ceci peut paraître trivial à première vue, mais trop souvent nous sommes confrontés à la situation inverse, où le contenant est mis de l'avant jusqu'à parfois même devenir le focus général et à faire oublier l'importance du contenu.

Une telle situation peut se produire autant dans la population en générale qu'au sein de l'activité scientifique et peut être causée autant par des erreurs honnêtes que par des intentions malhonnêtes. Pour nous, le but principal de l'écriture de ce livre est donc de dissiper la confusion régnant autour de ces situations, en expliquant les causes et les enjeux. Tout cela nous paraît nécessaire pour une évaluation la plus juste possible de la scientificité de tout travail.

LA DÉMARCHE SCIENTIFIQUE

« DÉMYSTIFIÉE »

Les étapes de la démarche scientifique

Étant donné que la confusion régne souvent sur la démarche scientifique et que la compréhension de cette démarche est cruciale pour faire une évaluation de la scientificité, nous en étayons les détails dans cette section. Le rôle de la démarche scientifique est de s'affranchir le plus possible de la subjectivité de l'observateur^{Note 1}. C'est un processus partiellement itératif qui se divise en plusieurs étapes (Fig. 1).

La première étape consiste à observer et décrire une nouvelle situation, un phénomène inexpliqué qui attise la curiosité du scientifique¹⁵⁶ (Fig. 1A). Notez que cette première étape est donc ancrée dans une lacune des connaissances sur le sujet¹⁶⁶, ainsi que sur un intérêt d'en savoir plus, de s'informer.

Le scientifique a toutefois accès à une certaine quantité d'information dans la littérature scientifique. Lorsqu'il vérifie si un autre scientifique a déjà répondu à la question qu'il se pose, il est déjà dans la deuxième étape de la démarche scientifique qui est de s'interroger sur les causes du phénomène observé et de formuler une hypothèse tentant de l'expliquer (Fig. 1B). Le

¹ Dans ce livre, l'emploi du masculin pour désigner des personnes est utilisé comme générique avec comme unique fonction d'alléger le texte, sans aucune intention discriminatoire. Nous considérons que le féminin peut être aussi bien utilisé comme générique que le masculin, mais comme tous les auteurs sont des hommes, nous nous sommes décidés pour le masculin puisque c'est ce que nous utilisons naturellement.

scientifique utilise donc sa perspicacité, ses connaissances, ses observations et l'information pertinente à sa question dans la littérature scientifique afin de formuler une hypothèse ou un modèle niché dans un cadre théorique raisonnable pour expliquer le phénomène d'intérêt¹⁵⁶.

À partir de cette hypothèse, le scientifique doit ensuite pouvoir établir des prédictions concernant le phénomène, c'est-à-dire que si l'hypothèse est juste, elle doit permettre de prédire en tout ou en partie l'occurrence du phénomène observé ou de prédire les résultats de futures observations^{156,209} (Fig. 1C). L'hypothèse consiste typiquement en une cause possible du phénomène ou en une variable pouvant l'influencer. Par exemple, lorsque les feuilles d'une plante d'intérieur commencent à flétrir, une hypothèse possible pour ce phénomène est le manque d'eau, surtout si on sait que la plante n'a pas été arrosée depuis plusieurs jours. La prédiction est que lorsque cette cause ou cette variable est présente, le phénomène en question sera observé ou influencé de façon prévisible. Par exemple, d'après l'hypothèse mentionnée, on prédit que si on arrose la plante, les feuilles devraient retrouver leur turgescence et qu'elles flétriront à nouveau si elles manquent d'eau dans l'avenir.

Les hypothèses et les prédictions formulées doivent être falsifiables, suivant la logique hypothéticodéductive^{154,156}. Ceci implique qu'il doit être possible de démontrer que l'hypothèse est fausse si c'est le cas, ce qui nous amène à la quatrième étape de la démarche scientifique : l'établissement d'une méthodologie de recherche appropriée pour tester la justesse et la validité des prédictions (Fig. 1D). Cette méthodologie consistera normalement à amasser et analyser un ensemble de données permettant ce test. Si nous revenons à la plante, il se peut que

cela nous suffise de tout simplement l'arroser, et dans la vie de tous les jours, ça s'arrête normalement là si la plante redevient saine et que la prédiction semble ainsi validée. Cependant, si nous voulons vraiment suivre la démarche scientifique, il faudrait prévoir plusieurs répétitions avec par exemple différents temps entre les arrosages, plusieurs individus de la même espèce de plante dans différents pots et un groupe de plantes qui ne sont pas arrosées (ce qu'on appelle le groupe contrôle) pour vérifier si seules les plantes arrosées retrouvent leur santé.

Il est essentiel ici de comprendre qu'il n'y a pas qu'une seule méthodologie de recherche qui soit adéquate et acceptable au sein de la démarche scientifique (Fig. 2) : ceci dépend du phénomène à l'étude, de la nature de l'information colligée, de la question et de l'hypothèse qui sont posées, de l'intérêt à trouver une réponse juste et aussi des ressources à la disposition du scientifique. Ce point sera discuté plus en détail dans la section « Application de la démarche scientifique ».

Suite à l'analyse des résultats, trois possibilités peuvent se dégager : les résultats corroborent l'hypothèse de recherche, ils la contredisent ou ils sont non concluants (Fig. 1D-F). Par exemple, une fois les plantes arrosées, les feuilles peuvent retrouver leur turgescence, elles peuvent rester flétries, ou encore une partie seulement retrouvent leur turgescence. Évidemment, ceci est une formulation simplifiée et il se peut que de nombreuses expériences conduites par de nombreux scientifiques soient nécessaires afin de corroborer ou rejeter une hypothèse donnée.

Dans le cas où l'hypothèse ne passe pas les tests, le scientifique doit alors la modifier, la raffiner ou encore la rejeter et en formuler une toute nouvelle et revenir à l'étape de

l'établissement de prédictions testables (Fig. 1E, B, C). Il se peut par exemple qu'une carence en nutriments ou une maladie soit la cause du flétrissement des feuilles, ou que l'eau était effectivement un facteur, mais que d'autres éléments soient en cause, comme une hausse importante de la température ambiante.

Si l'hypothèse est corroborée, elle peut alors être développée davantage pour produire des prédictions dans des cadres spatiaux ou temporels plus larges ou encore pour formuler d'autres hypothèses¹⁵⁶. Nous pourrions par exemple inclure des groupes de plantes soumises à différentes variations de température pour vérifier à quel point l'eau seule est un facteur déterminant, ou encore un groupe arrosé avec de l'eau distillée (du moins à courts termes) pour vérifier que c'est bien l'eau le facteur limitant et non une carence (ou un surplus) en nutriments qui serait influencée par l'eau non-distillée.

Lorsqu'une hypothèse a passé tous les tests – et dépendamment de sa nature – elle peut potentiellement servir au développement de théories générales ou même de lois, celles-ci devant être en accord avec la majorité sinon toutes les données disponibles ainsi qu'avec les théories déjà existantes et qui ne sont pas détrônées par la nouvelle théorie (Fig. 1F). Si nous suivons cette idée par rapport à notre exemple des plantes et de l'eau et que nous poussons (sans jeu de mot) un peu (beaucoup), nous pourrions éventuellement en arriver aux fonctions vitales de l'eau pour la respiration, la photosynthèse et l'intégrité des cellules des êtres vivants.

Un dernier point que nous tenons à souligner est qu'au sein de la démarche scientifique, les tests (Fig. 1D) et les

développements théoriques permis par les résultats (Fig. 1F) doivent avoir le potentiel d'être reproduits. On distingue trois types de reproductibilité⁷⁹, soient la reproductibilité du protocole utilisé, la reproductibilité des résultats et la reproductibilité des inférences qualitatives d'une étude. Étant donné le caractère absolument essentiel de la reproductibilité pour la démarche scientifique, nous donnerons ici au moins quelques lignes d'explication pour chacun de ces types.

La reproductibilité du protocole utilisé dans une étude signifie qu'on y fournit une description détaillée et surtout complète des méthodes de collecte et d'analyse des données, des outils et des programmes utilisés, des données récoltées et des résultats suite à l'analyse. Ceci peut prendre une forme assez différentes d'un domaine de recherche à l'autre⁷⁹, mais ça doit être suffisamment complet et détaillé pour permettre à d'autres scientifiques de reproduire exactement les mêmes analyses et résultats à partir des mêmes données, ou de refaire l'étude au complet en suivant exactement le même protocole (si possible). Notez qu'une bonne reproductibilité du protocole ne dit absolument rien quant à la validité des résultats originaux ; ça dit seulement que les méthodes ont été décrites correctement et qu'il n'y a pas eu d'erreurs d'analyse.

La reproductibilité des résultats fait référence à la réplication d'une étude. Une réplication est lorsque des scientifiques suivant le même protocole ou un protocole similaire répliquent l'étude originale de façon indépendante de celle-ci pour voir s'ils peuvent arriver aux mêmes résultats. Il peut s'agir des mêmes scientifiques que lors de la première étude s'ils veulent vérifier eux-mêmes la validité de leurs résultats, ou d'autres scientifiques qui veulent peut-être vérifier des résultats

qui sont surprenants ou d'importance particulière dans le cadre de leur propre recherche. Dans les sciences du vivant (biologie, biomédecine, écologie, psychologie, etc.), de bonnes réplifications sont relativement difficiles à produire puisqu'il est attendu et naturel que beaucoup de facteurs de variation ne puissent pas être pris en compte ou contrôlés d'une étude à l'autre. Ceci rend donc délicate l'évaluation de la réplification des résultats¹⁵¹. Nous y reviendrons plus tard.

Finalement, nous avons la reproductibilité des inférences qualitatives d'une étude. Parmi les différents domaines de recherche, ce dernier type de reproductibilité est souvent le plus important^{12,79}, puisque tellement de facteurs n'ayant rien à voir avec la véracité d'une hypothèse peuvent influencer les deux premiers types de reproductibilité. De plus, il se peut que l'étude d'un phénomène particulier nécessite une méthodologie de recherche dont le protocole ou les résultats ne peuvent pas être exactement reproduits, que ce soit pour des raisons techniques ou éthiques. Nous avons en effet déjà mentionné comment la reproductibilité des résultats est parfois impossible à faire exactement étant donné de nombreux facteurs impossibles à contrôler, situation typique des systèmes vivants. Quant aux soucis éthiques, prenez par exemple un écologue étudiant les effets d'un feu de forêt ayant dévasté un écosystème entier, ou un psychologue étudiant les effets d'une prise d'otage sur un groupe de personnes. Dans un cas comme dans l'autre, il serait de toute façon complètement impossible de reproduire exactement l'expérience, mais, surtout, il serait éthiquement inacceptable de le tenter. Parallèlement, il serait éthiquement inadmissible de refuser d'étudier ces phénomènes sous prétexte qu'on ne peut pas les reproduire exactement.

En plus d'être parfois la seule méthode de reproductibilité possible, la reproduction des inférences qualitatives issues de la recherche est *toujours* possible. Ceci peut se faire en comparant ces inférences à la théorie existante et/ou aux résultats d'autres cas mettant en évidence un phénomène commun^{12,156}. Cette comparaison permet à son tour d'accumuler des connaissances et des évidences scientifiques et ainsi d'augmenter (ou d'amoindrir) la fiabilité des conclusions inférées d'après les premiers résultats. Ce type de reproduction est particulièrement approprié pour des phénomènes incluant beaucoup de facteurs incontrôlables⁷⁹ et peut être utilisée même dans les cas où les résultats ne sont pas exactement répliqués. Évidemment, le plus haut degré de confiance dans la généralisation des résultats est atteint lorsque les trois types de reproduction peuvent être réunis.

Ce qui est certain, c'est que la reproductibilité, de quelque type que ce soit, n'est possible qu'avec une description détaillée et complète des résultats et de la manière dont les résultats ont été obtenus⁷⁹. Il est toutefois important de noter qu'un obstacle à l'accessibilité d'une description précise ne signifie pas que les résultats ne peuvent pas être généralisés, mais seulement que nous manquons de l'information nécessaire à l'évaluation de la généralisation potentielle des résultats.

Se prononcer sur ce qu'est la science ou la pseudoscience

Un piège très courant dans l'évaluation de la scientificité est de tenter de se prononcer sur ce qu'est la science. Cette mise en matière peut vous sembler paradoxale et embrouillante, mais

justement, la perception que l'on a du concept de science interfère avec l'évaluation de la scientificité, confusion qui est même présente dans les universités, parmi les scientifiques².

En effet, tenter de donner une définition à la science est le travail des philosophes de la science, contrairement à l'évaluation de la scientificité, qui doit être faite en référence à la démarche scientifique. En d'autres mots, l'évaluation de la scientificité ne se fait pas par rapport à une conceptualisation de ce que pourrait vouloir dire le mot « science », mais par rapport à la démarche scientifique, qui elle, ne souffre pas d'une pluralité de définitions.

De nombreux philosophes de la science tentent de définir ce qu'est la science. Le débat reste ouvert¹¹⁷, car ils tiennent compte de la dimension sociale de la science¹¹¹ ainsi que de sa dimension cognitive²⁸. En d'autres mots, ils tentent de répondre à des questions comme : « Qu'est-ce que représente la science pour l'être humain? » ; « Comment les scientifiques perçoivent-ils la science? » ou « Quelle est la valeur des connaissances apportées par la science? » Toutes ces questions restent donc matière à débat, notamment parce que pour y répondre, il faut faire intervenir d'autres facultés cognitives que le raisonnement hypothéticodéductif^{Note 2}. En d'autres mots, les réponses à ces questions sont de nature changeante et matière à interprétation

² Le raisonnement hypothéticodéductif fait partie intégrante de la démarche scientifique. Il consiste en la formulation d'une hypothèse (H1) suivant un raisonnement hypothétique en imaginant les conséquences si l'hypothèse était vraie, c'est-à-dire en faisant des prédictions : si H1 est vraie, alors on devrait observer tels ou tels résultats. Ensuite, on compare ces prédictions aux résultats d'expériences ou de nouvelles observations et si H1 tient toujours la route, nous en déduisons que nous pouvons la conserver jusqu'à preuve du contraire.

et la démarche scientifique ne permettra généralement pas d'y répondre de manière définitive puisqu'elle n'est pas fait pour cela.

Quant à elle, l'évaluation de la scientificité d'un travail est faite en comparant la démarche utilisée dans le travail évalué à la démarche scientifique, simplement et rigoureusement. Dans ce contexte, on pourrait se demander pourquoi on ne définit pas tout simplement la science comme ce qui est conforme à la démarche scientifique. La raison est fort simple : ce qu'on appelle « sciences » sont en fait des domaines de recherche. Par exemple, on dit que l'écologie est une science. Ensuite, on qualifie cette science par rapport à d'autres sciences. Cependant, en écologie comme dans toutes les sciences, des travaux sont faits avec une grande rigueur scientifique, tandis que d'autres comptent des entorses sévères à la démarche scientifique. Ainsi, par l'utilisation même que l'on fait du mot « science », « science » et « scientificité » ne peuvent pas être équivalents et c'est pourquoi il faut évaluer la scientificité de chaque travail fait en écologie pour en déterminer la contribution singulière à cette science.

Pareillement, si nous prenons la situation inverse, nous voyons bien pourquoi « science » et « scientificité » ne sont pas équivalents ; il serait par exemple tout à fait aberrant de dire qu'un article est scientifique uniquement parce que son sujet est l'écologie. Pourtant, l'erreur est souvent faite, certaines personnes croyant que certains sujets suffisent pour dire que les articles qui y sont associés ne sont pas scientifiques. Si vous y prêtez attention, vous constaterez ce phénomène à maintes reprises, même de la part de scientifiques.

Les auteurs qui parlent de pseudoscience ont le même problème que ceux qui tentent de définir la science, car la

pseudoscience n'est généralement définie que par rapport à la science ou par rapport à d'autres concepts, et non par rapport à la conformité avec la démarche scientifique. Une pseudoscience serait alors un savoir organisé de façon à paraître scientifique, mais qui ne l'est pas. Afonso et Gilbert² analysent les contributions de Popper, Kuhn, Lakatos et Laudan par rapport à la démarcation entre science et pseudoscience. Ils concluent qu'il ne semble pas y avoir de ligne de démarcation franche entre la science et la pseudoscience, malgré certains indicateurs pouvant servir de signaux d'alarme pour la présence de pseudoscience, tels que le refus de considérer des évidences ou des théories contrariant la théorie favorisée, l'utilisation d'un langage « obscur » pour décrire le phénomène, ou encore le manque de répliques. Ils placent donc la science et la pseudoscience sur un continuum. Ceci paraîtrait aberrant si nous n'avions pas déjà vu que la science et la démarche scientifique sont deux choses différentes, mais devient parfaitement intelligible une fois cette distinction faite. Afonso et Gilbert² concluent d'ailleurs qu'un facteur déterminant dans l'acceptation de croyances pseudoscientifiques est un manque de compréhension de ce qui définit la scientificité et la démarche scientifique.

Dans un effort pour offrir des critères concrets pour différencier la science de la pseudoscience, Fanelli⁶³ développe quant à lui un modèle mathématique permettant de quantifier le niveau de savoir permis par une science à partir de variables telles que le pouvoir explicatif d'une théorie ou les coûts de méthodologie pour obtenir de l'information. Sa définition d'une pseudoscience est alors une discipline qui produit un savoir net négatif, c'est-à-dire que plus on s'y attarde, moins on en retire de connaissances par rapport à l'investissement qu'on y met. En d'autres mots, plus on se penche sur le sujet et plus on y investit

de ressources, moins les concepts deviennent clairs et pragmatiques.

Nous considérons ses efforts louables et très utiles pour évaluer la contribution d'investigations, d'articles, ou de théories. Cependant, nous nous permettons de remettre en question l'utilité de ce modèle pour démarquer une science d'une pseudoscience. En effet, établir les valeurs des variables nécessaires pour calculer le savoir permis par ce que nous appelons une science, par exemple l'écologie, semble aussi difficile et subjectif que de définir ce qu'est une science. En théorie, ce modèle pourrait fonctionner, mais il nous semble nécessiter de mettre des valeurs finies sur des notions impossibles à quantifier objectivement. Il s'agit néanmoins d'un modèle très intéressant pour de futures recherches sur le sujet. De plus, il explicite mathématiquement la notion selon laquelle les concepts de science et de pseudoscience se trouve en fait sur un continuum, car ce qui les sépare selon Fanelli se trouve dans une zone plus ou moins floue où le savoir net que l'on obtient passe du positif (science) au négatif (pseudoscience). Cette zone est floue, car les variables servant au calcul ne peuvent pas être exactement quantifiées et parce que cette difficulté est fortement dépendante du contexte de recherche.

Tout ceci témoigne du fait qu'introduire un concept extérieur à la démarche scientifique (comme celui de pseudoscience) ne permet pas l'évaluation de la scientificité. Ainsi, la thèse que nous supportons dans ce livre – stipulant que se concentrer sur la démarche scientifique permet mieux d'évaluer la scientificité que d'essayer de séparer les concepts de science et de pseudoscience – reste très pertinente et appropriée. En effet, l'échec à parvenir à un consensus sur une définition de la science

est inévitable, car on critique des ensembles de travaux et de theories, et on tente de qualifier le niveau de connaissance que permet la démarche scientifique par rapport aux autres types de connaissance. En d'autres mots, il s'agit de questions méta-scientifiques, c'est-à-dire des questions par rapport à une théorie de la science et de ce qu'elle représente. Les définitions de la science ne permettent donc pas à un scientifique de déterminer si un travail a rigoureusement suivi la démarche scientifique. Il doit évaluer lui-même la scientificité des travaux qui l'intéressent. Ainsi, quoique la capacité à philosopher sur la nature de la science puisse grandement aider un scientifique à poser les bonnes questions et à innover dans l'application de la démarche scientifique, c'est la capacité d'une personne à évaluer la scientificité des travaux qu'elle lit qui va l'aider à retenir les articles les plus rigoureux lors de sa revue de littérature et donc à se forger une opinion scientifique informée.

Ceci nous ramène à la notion que l'habit ne fait pas le moine, que le contenu est plus important que le contenant, notion introduite dans la préface. Il semble important ici d'élaborer sur les liens que nous voyons entre cette notion et celles que nous avons présentées dans les paragraphes précédents. Pour nous, se concentrer sur la personne qui porte un habit, de moine ou autre, est analogue à l'évaluation de la scientificité d'un article. Ceci implique une attention particulière à la démarche utilisée et aux implications des résultats, puis une comparaison avec la démarche scientifique. Évaluer la scientificité nécessite alors de se pencher sérieusement sur le contenu, sinon la tâche est impossible.

Il s'en suit que, même si nous ne croyons pas exactement que tenter de donner une définition à la science et à la

pseudoscience soit la même chose que de se concentrer sur l'habit ou le contenant, ces deux activités sont tout de même reliées sur certains points. Celui qui nous apparaît le plus important est que, malgré le fait que les définitions de science et de pseudoscience soient dynamiques, changeantes dans le temps suivant l'évolution de nos connaissances, elles représentent toutefois des catégories relativement fixes à un point donné dans le temps. Ainsi, une catégorisation d'un travail ou d'une idée en tant que science ou pseudoscience qui ne se base que sur le sujet d'étude met un premier filtre par rapport au contenant uniquement (la catégorie) avant de passer au contenu (la démarche utilisé). En effet, une personne ne s'intéressant que pour ce qui est scientifique pourra être tentée de ne pas pousser son investigation plus loin en apprenant qu'un travail porte sur un sujet ayant été positionné sous la catégorie de pseudoscience. Ce filtre peut être raisonnable et sauver du temps, mais il sera en réalité impossible de savoir si c'est le cas avant de s'être penché sérieusement sur le contenu, c'est-à-dire sur la scientificité.

Évidemment, il est probable qu'à ce point de la lecture, vous puissiez penser à plusieurs exemples de pseudosciences pour lesquelles ce genre de filtre serait tout à fait justifié et où le caractère pseudoscientifique de la discipline ne laisse presque aucun doute (en science, il y a toujours un doute). Nous avons ici deux points à préciser pour expliquer pourquoi nous considérons les explications des dernières pages utiles, même essentielles.

Premièrement, une étude scientifique peut se faire même sur un sujet étant reconnu comme une complète pseudoscience (voir par exemple McGrew et McFall¹²⁷) et il serait absurde de rejeter une étude scientifique rigoureuse à cause de son sujet

d'observation. Nous pourrions nous demander à quoi servirait de telles études si elles ne font que présenter des évidences supplémentaires qu'une discipline considérée comme une pseudoscience n'est effectivement pas supportée par des tests scientifiques. N'est-ce pas tout simplement une perte de temps et d'argent? Ceci dépend du sujet et du contexte social l'entourant. Par exemple, si une grande quantité de personnes se sentent reliées à une pseudoscience et y croient, ceci peut justifier d'y consacrer du temps et de l'argent pour réaliser des études scientifiques de qualité sur ce sujet, car cette discipline a un réel impact sur la société.

Le deuxième point encore plus important est que ce ne sont pas toutes les catégorisations de pseudoscience qui sont claires et limpides ; certaines restent matière à débat. En effet, des sujets étiquetés comme « pseudosciences » le sont parfois avant qu'une investigation scientifique rigoureuse ait été réalisée, avant que le débat scientifique ne soit terminé, ou même avant qu'il ait eu lieu. Ce premier filtre par rapport au contenant au détriment du contenu peut alors potentiellement être dommageable pour l'avancement de nos connaissances en science, surtout si la catégorisation de pseudoscience est faite par incompétence ou par malhonnêteté. Nous explorons ce dernier point dans la prochaine section.

Le piège des faux sceptiques

Dans un débat scientifique, le mot « pseudoscience » est parfois utilisé comme un argument en soi (voir par exemple Denault et coauteurs⁴⁵) afin de décrédibiliser une idée, une

personne ou une théorie. Ceci est parfois justifiable, parfois non. En effet, comme nous l'avons mentionné dans la section précédente, cet argument pose un problème majeur par rapport à l'évaluation de la scientificité. De fait, l'argument qui dit « c'est une pseudoscience » vient souvent subtilement remplacer l'évaluation de la scientificité, c'est-à-dire que le contenant devient plus important que le contenu, que l'habit prime sur la personne qui le porte. Sans cette évaluation pourtant, le constat de science ou de pseudoscience n'est qu'une croyance, une catégorisation sans intérêt d'un point de vue scientifique. Lorsque ce n'est pas justifier, ce type de comportement peut retarder les avancées scientifiques et créer ainsi du dommage si ces avancées auraient permis à des applications importantes de voir le jour plus tôt.

Tout scientifique est effectivement supposé savoir que c'est une grave erreur de refuser sans fondement de considérer une idée, un traitement, ou une méthode qui n'a pas encore été testée¹¹⁸. C'est pourquoi nous voulons introduire ici le concept du faux sceptique. Nous appelons « faux sceptique » un auteur qui se prononce sur la scientificité sans l'évaluer et sans démontrer les failles méthodologiques. Un faux sceptique est donc une personne qui tente de faire oublier aux autres la personne qui porte l'habit du moine.

Le phénomène du faux scepticisme peut avoir plusieurs causes telles que l'ignorance bornée, l'incompétence, ou encore un intérêt malhonnête (personnel ou monétaire) à discréditer un savoir, un laboratoire, une personne ou un nom. D'après cette description, vous vous rendez bien compte que le faux scepticisme est bien différent du sain scepticisme que nous avons présenté dans la préface et qui est quant à lui causé par un esprit

scientifique curieux et rigoureux, élément essentiel pour poser les bonnes questions et trouver les bonnes réponses face à une situation nouvelle, essentiel aussi pour ne pas être berné par ses propres préconceptions, quelles qu'elles soient.

Ainsi donc, le faux scepticisme peut être le résultat de plusieurs causes différentes, toutes indésirables pour l'activité scientifique. Cependant, il se manifeste généralement de façon relativement invariable et *nécessite une participation active de la part de la personne qui le pratique*. En effet, si une personne refuse de considérer une idée ou un travail sans avoir d'arguments scientifiques pour ce refus, mais ne se prononce pas professionnellement sur la scientificité de la chose, on parlera alors d'un manque d'intérêt, ce qui est tout à fait normal et ne pose pas de problème. Non, le faux scepticisme requiert une action visant le plus souvent à empêcher les autres de s'intéresser à un sujet particulier, avec un raisonnement ou des méthodes qui n'ont pas de fondements scientifiques. Déjà, comme il serait fâcheux de reproduire les résultats d'une étude que l'on souhaite discréditer, les faux sceptiques utilisent une supercherie en deux étapes pour s'en sortir sans arguments scientifiques.

Premièrement, ils ne se penchent jamais sur la méthodologie des scientifiques qu'ils veulent discréditer, ou alors ne le font qu'en surface ; ils ignorent le contenu autant que possible. Ceci leur permet de ne pas être confrontés au risque que ces derniers respectent la démarche scientifique. Ils n'ont donc même pas besoin de connaître la méthodologie exacte des travaux qu'ils rabaissent. Ensuite, ils affirment simplement que les scientifiques qu'ils veulent discréditer ne sont pas ou ne sont plus reconnus par la communauté scientifique ; ils concentrent l'attention sur l'habit, le contenant. Ce faux scandale peut prendre

plusieurs formes différentes et réussit à faire croire à bon nombre de personnes que les travaux visés n'ont rien de scientifique. Le plus dommage est que, peu importe la faiblesse des arguments des faux sceptiques, la simple utilisation du mot « pseudoscience » suffise souvent à convaincre beaucoup de gens, même parmi les scientifiques.

Méthodes des faux sceptiques

Les méthodes exactes par lesquelles les faux sceptiques détournent l'attention des lecteurs du contenu vers le contenant, de l'évaluation de la scientificité vers des arguments fallacieux, peuvent prendre diverses formes et sont multiples. Ce qui est certain, c'est que la manipulation y est toujours bien présente, inconsciemment ou pas.

La méthode la plus courante, pratiquement omniprésente, est évidemment de tordre les arguments de ceux qu'on veut discréditer. Ceci se fait souvent en prenant des citations hors contexte pour leur faire dire le contraire de ce qui est transmis lorsqu'on se penche sérieusement sur la référence originale, ou encore leur faire dire toute autre chose, sans rapport avec ce qui se trouve dans le texte référencé. Cette technique – déjà utilisée d'ailleurs sur une version antérieure du livre que vous êtes en train de lire, nous attribuant une opinion complètement fausse, fabriquée de toutes pièces, sur un sujet ne faisant même pas partie du livre⁴⁶ – comprend donc souvent des propos qui sont mensongers, mais qui ont l'air juste puisqu'une citation est utilisée. Le résultat est que les lecteurs obtiennent une impression faussée du message principal de l'ouvrage en question. Ceci est

souvent accompagné par une manipulation visant à convaincre les lecteurs que se renseigner sur le sujet ne vaut même pas la peine, qu'il faudrait balayer de la main toute considération sérieuse du sujet¹⁰⁹ ; naturellement, si les gens allaient s'informer, ils verraient que les propos soigneusement sélectionnés par les faux sceptiques sont mensongers, ou du moins qu'ils représentent faussement l'ensemble du travail ainsi discrédité.

Avoir recours à la moquerie est une autre technique typique des faux sceptiques et qui n'a rien à voir avec l'évaluation de la scientificité. Il s'agit plutôt d'une manipulation émotionnelle visant les lecteurs. En effet, un lecteur lambda n'étant pas encore familier avec les notions abordées par un sujet qui est ridiculisé par un faux sceptique, ou pas encore familier avec l'évaluation de la scientificité en générale, peut d'emblée avoir peur (souvent inconsciemment) d'être associé avec les moqueries utilisées avant même d'avoir pu se forger une opinion informée sur le sujet. Dans un tel cas, il est à craindre que sa quête d'information se limitera au texte du faux sceptique plutôt que de continuer en comparant les idées et les arguments des deux camps. C'est ça le but de la moquerie, qui n'a en réalité rien à faire dans un texte scientifique.

Relié au principe du recours à la moquerie est l'utilisation du point d'exclamation (!). Ceci peut à première vue paraître anodin ou même ridicule, mais il s'agit en réalité d'un outil efficace et pernicieux dans les mains d'un faux sceptique. Cette technique consiste généralement à décrire une partie de l'idée derrière les travaux que l'on veut discréditer ou, encore mieux, à utiliser une citation de ces travaux et à finir par un point d'exclamation. Par exemple : « *Les auteurs disent que X est relié avec Y !* » Cette technique sera parfois même accompagnée d'une phrase visant à renforcer le point d'exclamation : « *Les auteurs disent que X est*

relié à Y. Imaginez-vous ! » ou encore : « Les auteurs disent que X est relié à Y. La belle affaire ! »

Nous ne sommes évidemment pas contre l'utilisation du point d'exclamation. Cependant, dans le contexte du faux scepticisme, l'utilisation du point d'exclamation constitue souvent, encore une fois, une manipulation des lecteurs, visant particulièrement ceux qui ne sont pas familiers avec le sujet. En effet, en utilisant cette technique, les faux sceptiques se libèrent du besoin d'apporter une justification scientifique à leurs propos (ou quelconque justification d'ailleurs), car le lecteur comprend par lui-même qu'il est supposé croire que les notions citées sont invraisemblables. Ceci se produit tout seul, naturellement, sans jamais avoir évalué si la citation de base était vraiment incroyable ou invalide.

L'absence de support scientifique ou d'argumentation logique tout en utilisant un jargon scientifique complexe est d'ailleurs une caractéristique essentielle pour reconnaître le discours des faux sceptiques. Cette absence de support scientifique est particulièrement flagrante lorsqu'ils utilisent la moquerie ou le point d'exclamation. Ils ne peuvent justement pas apporter d'arguments scientifiques sur le sujet qu'ils discréditent, car pour ce faire, il faudrait qu'ils se penchent sur la méthodologie et qu'ils la décrivent correctement, ce qu'ils veulent normalement éviter à tout prix.

Une autre technique qui sera utilisée par tout faux sceptique si cela est possible, c'est de dire qu'il n'y a pas ou peu de publications avec révision par les pairs sur le sujet. Nous sommes d'accord ici que ce fait peut être un bon signal d'alarme qui devrait éveiller le sain scepticisme. Cependant, si la réflexion

s'arrête là et que ce seul critère permet à quelqu'un de dire impunément qu'une théorie ou une discipline est une pseudoscience, nous sommes d'avis que c'est une grave erreur de jugement scientifique, représentant à l'extrême l'idée que l'habit serait plus important que le moine. Plusieurs éléments justifient notre opinion sur le sujet, entre autre le fait que la publication scientifique avec révision par les pairs peut elle-même parfois contredire les principes fondamentaux de la démarche scientifique. Nous explorerons cette question en long et en large dans les prochaines sections du livre.

Plusieurs autres techniques sont utilisées par les faux sceptiques, comme par exemple l'association de mots entre les idées à discréditer et des termes à connotation négative ou à caractère « anti-scientifique » tels que : « haine », « blasphème », « violence », « créateur », « gourou », etc. La plupart de ces mots n'ont que rarement une place pertinente dans un texte scientifique et sont souvent utilisés gratuitement par les faux sceptiques, sans justification appropriée. Ceci ne signifie pas que l'utilisation de ces mots, ou l'utilisation des autres techniques présentées ici, constitue une preuve en soi de faux scepticisme, mais ce sont de bons signaux d'alarme pour les lecteurs avertis, indiquant qu'une attention particulière pour la détection de ces manipulations est requise.

Ironiquement, un faux sceptique – qui par notre définition tente de faire passer une théorie pour pseudoscience alors que c'est faux ou que ce n'est pas encore démontré – et un « pseudoscientifique » – qui tente de faire passer pour scientifique une théorie qui relève plutôt de la pseudoscience – ont beaucoup en commun ; ils utilisent tous deux un jargon scientifique complexe sans véritable support scientifique et ils

essaient tous deux de faire oublier aux autres le contenu en les poussant à se concentrer sur le contenant.

Nous tenons à préciser que le concept du faux scepticisme en est un que nous avons développé nous-mêmes à partir de nos propres observations, de nos expériences et de celles de nos collègues et de notre revue de littérature sur la science. Grâce à cette notion, nous croyons pouvoir aider à éclaircir les confusions entourant l'évaluation de la scientificité puisque nous croyons que ce que nous appelons faux scepticisme contribue en grande partie à ces confusions. Il faut tout de même noter que les exemples et les descriptions que nous avons utilisés jusqu'ici pour parler de ce phénomène constituent des représentations clairement définies, où nous ne laissons pas beaucoup de place au doute ou à la nuance quant aux motivations et aux méthodes des faux sceptiques. Bien que ce genre de situations très claires puissent effectivement exister, dans la plupart des cas il est plus probable que les indices et les motivations du faux scepticisme ne soient pas si évidentes et faciles à trancher. C'est pourquoi nous recommandons à toute personne vraiment intéressée à un sujet qui est entouré d'une controverse de toujours rechercher et lire l'information disponible sur les *deux* côtés de la controverse et de les comparer par rapport à la démarche scientifique. Que les textes paraissent pseudoscientifiques ou empreints de faux scepticisme, les lire et les comparer de façon objective en se basant sur la démarche scientifique est l'unique façon de se forger une opinion informée. Naturellement, ceci exige beaucoup de temps et d'efforts, ce qui requiert absolument un grand intérêt pour le sujet en question.

Pour les lecteurs néophytes, Lilienfeld et coauteurs¹¹⁷ proposent une liste d'éléments qui devraient éveiller le

scepticisme de façon saine. Cependant, si le lecteur n'est pas à l'aise d'évaluer lui-même la scientificité d'un travail, qu'il soit crédule ou sceptique, il ne pourra qu'émettre une opinion embrouillée sur la contribution d'un article pour la science. En effet, même en utilisant une liste de points à vérifier, rien ne prévaut sur une bonne réflexion quant à l'application de la démarche scientifique utilisée dans le travail qu'on lit. C'est ce que nous détaillons dans la prochaine section.

Application de la démarche scientifique

La démarche scientifique a été popularisée en réponse au dogmatisme du clergé. Ce qui fait sa force est qu'elle repose à la fois sur la logique^{Note 3} inductive¹³ et hypothéticodéductive^{155,156}. De plus, la démarche scientifique est devenue populaire en relation avec une méthodologie de recherche particulière : le design expérimental classique, où toutes les variables sauf la variable dépendante (Boîte 1) sont contrôlées et mesurables. Ce type de méthodologie fonctionne donc très bien dans des conditions de laboratoire et pour certains domaines (Boîte 1). Cependant, il existe plusieurs phénomènes qui échappent aux cinq sens et aux outils qui les prolongent, ou encore qui sont contrôlés par tellement de facteurs différents qu'il est impossible

³ La logique inductive se base sur l'expérimentation et l'observation de cas particuliers ou d'échantillons pour tirer des conclusions plus générales en utilisant des probabilités. Par exemple, une urne contient 100 billes, rouges ou bleues. Nous tirons aléatoirement 10 billes, 3 sont rouges, 7 sont bleues. Il y a donc une certaine probabilité que 30% des billes dans l'urne soient rouges. Cette probabilité augmente et devient encore plus fiable si nous tirons 50 billes et trouvons que 15 de ces billes sont rouges. Quant à la logique hypothéticodéductive, elle est déjà présentée à la page 24.

de les circonscrire précisément en laboratoire. Les scientifiques peuvent tout de même appliquer les étapes logiques de la démarche scientifique et des expériences de laboratoire pour étudier ces phénomènes. Ils le font en utilisant des méthodologies de recherche adaptées à leur objet d'étude et à leurs hypothèses et pour ce faire, un grand éventail de méthodologies de recherche est à leur disposition (Fig. 2).

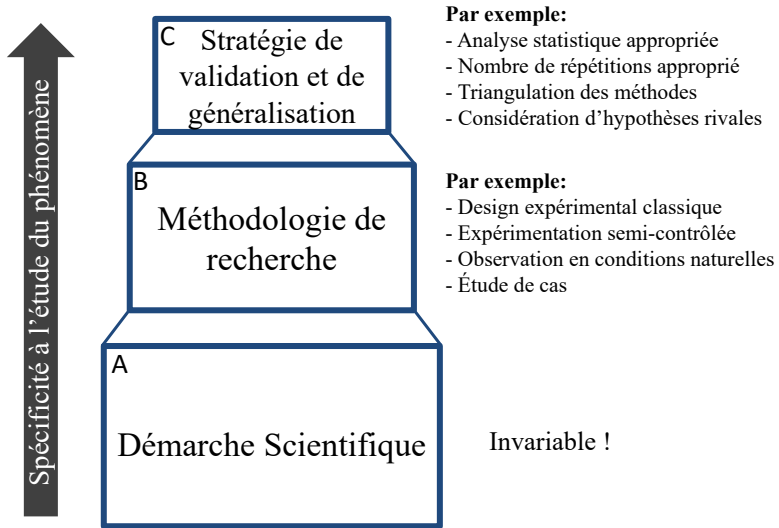


Figure 2. Hiérarchie de spécificité des concepts par rapport à l'étude scientifique d'un phénomène donné. La démarche scientifique est générale ; elle s'applique à toutes les recherches scientifiques (A). Au sein de la démarche scientifique et dépendamment du phénomène à l'étude, diverses méthodologies de recherche s'offrent aux scientifiques (B). Puis, pour chaque méthodologie de recherche, diverses stratégies de validation et de généralisation des résultats peuvent être appropriées (C). Notez qu'une stratégie donnée peut être appropriée pour plus d'un type de méthodologie.

Boîte 1. Le design expérimental classique

Variable dépendante, indépendante et covariable

Généralement, dans le design expérimental classique (comme dans la plupart des expériences scientifiques d'ailleurs), l'objectif est de tester un lien supposé entre une variable d'intérêt, qu'on appelle variable dépendante, et une variable d'effet, qu'on appelle variable indépendante, car cette dernière ne dépend pas de la variable d'intérêt. L'idée est qu'un changement de valeur pour la variable indépendante devrait affecter la valeur de la variable dépendante de façon prévisible si l'hypothèse de départ tient la route.

Ainsi, pour tester l'hypothèse, une expérience sera généralement constituée de plusieurs traitements où la variable indépendante aura une valeur fixe et connue à l'intérieur d'un même traitement, mais différente entre chaque traitement. On mesure ensuite la valeur de la variable dépendante pour chaque traitement, ce qui permet de déterminer si elle aussi change entre les traitements, si ce changement est corrélé avec les changements de la variable indépendante et si cette corrélation est en accord avec l'hypothèse de départ. Chaque traitement contiendra plusieurs répliques, donc des répétitions indépendantes du même traitement.

Comme les phénomènes comprenant uniquement deux variables (la variable dépendante et la variable indépendante) sont rares, surtout dans les sciences du vivant, il faut en général prendre en compte d'autres variables qui pourraient aussi influencer notre variable d'intérêt, c'est-à-dire notre variable dépendante.

On peut séparer en trois grandes catégories les façons de traiter ces variables confondantes pouvant affecter les résultats de l'expérience au-delà du lien qu'on veut tester :

1) La première situation en est une où l'on peut les « contrôler », c'est-à-dire les fixer sur une valeur déterminée et invariable pour tous les réplicas et tous les traitements.

2) Dans le cas où on ne pourrait pas les fixer, il faudra alors s'efforcer de les mesurer et de les intégrer dans les analyses statistiques en tant que covariables pour voir quelle part de la variation de la variable dépendante est expliquée par ces covariables ou par l'interaction entre ces covariables et la variable indépendante. Tout cela dans le but d'isoler la part de variation expliquée uniquement par la variable indépendante, c'est-à-dire celle dont on veut tester les effets.

3) Parfois, des variables influençant vraisemblablement la variable dépendante ne peuvent être contrôlées ni mesurées. Dans ce cas, il faut les inclure dans les limitations possibles de l'étude et en tenir compte dans l'interprétation des résultats, surtout lorsqu'on parle de généralisation. Il va sans dire que ce sont ces cas-ci qui apportent le plus d'incertitudes aux interprétations.

Expérience de laboratoire : exemple en écologie des plantes

Connaître, contrôler et mesurer toutes les variables confondantes d'une expérience est faisable en laboratoire, surtout dans des domaines tels que la physique ou la chimie. Lorsqu'on arrive dans le domaine du vivant cependant, identifier ces variables confondantes peut devenir difficile et souvent impossible à intégrer dans une

expérience de laboratoire, car le vivant est typiquement influencé par une myriade d'interactions entre beaucoup de facteurs différents. Une expérience en laboratoire peut alors avoir du mal à représenter la situation naturelle dans la vraie vie, car elle nécessite d'exclure plusieurs facteurs qui ne peuvent être contrôlés, mesurés, ou même présents en laboratoire.

Prenons un exemple tiré de l'écologie des plantes pour illustrer ces notions méthodologiques. Les plantes font généralement de la photosynthèse durant la journée, c'est-à-dire qu'elles utilisent l'énergie du soleil, l'eau absorbée par les racines et le dioxyde de carbone atmosphérique (CO_2) pour produire des molécules de sucre (glucides) et rejeter de l'oxygène. Avec l'augmentation de la concentration en CO_2 dans l'atmosphère, nous pourrions nous demander comment ceci influencera la photosynthèse des plantes. Nous pourrions par ailleurs faire l'hypothèse que plus il y aura de CO_2 (variable indépendante), plus les plantes pourront faire de photosynthèse (variable dépendante). Imaginons pour commencer une expérience simple où seules ces deux variables sont considérées. Nous pourrions choisir une espèce de plante et en placer des groupes dans des chambres climatiques en laboratoire avec différentes concentrations de CO_2 dans chaque chambre. Ensuite, pour chaque plante de chaque chambre (chaque traitement), nous pouvons mesurer la quantité de CO_2 entrant et sortant des feuilles pour déterminer le taux de photosynthèse.

Bien sûr, comme il s'agit de systèmes vivants, plusieurs facteurs peuvent influencer les résultats. Par exemple, le taux d'humidité dans l'air, la quantité d'eau disponible dans le sol, la quantité de lumière. Toutes ces variables peuvent

être contrôlées, c'est-à-dire que nous fixons leur valeur au même niveau dans toutes les chambres. Évidemment, ce sont aussi des variables qui peuvent interagir avec la concentration de CO_2 pour influencer le taux de photosynthèse. Il pourrait alors être intéressant d'en inclure quelques-unes comme covariables dans les analyses. Pour ce faire, il faut aussi faire varier leur valeur. Supposons que dans l'expérience de départ nous avons 3 concentrations différentes de CO_2 , avec 3 réplicas chacune (ce qui est très peu), pour un total de 9 chambres climatiques. Maintenant, on veut aussi tester l'impact de l'humidité dans l'air, aussi avec 3 niveaux différents. Il faut alors tripler l'expérience de départ pour avoir 9 chambres climatiques par niveau d'humidité dans l'air, pour un total de 27... déjà difficile à réaliser pour des laboratoires de taille moyenne. Si nous voulions ajouter d'autres covariables pour mieux comprendre ce qui se passe en réalité dans la nature, cela deviendrait vite impossible pour la plupart des laboratoires. Et la liste de variables à considérer n'est ici pas exhaustive...

À cela s'ajoute toutes les variables impossibles à contrôler ou à mesurer, comme toutes les interactions entre les différentes espèces de plantes dans une communauté naturelle ou encore les interactions entre les plantes et les organismes du sol, et ainsi de suite. Tout cela est impossible à reproduire de façon satisfaisante en laboratoire et doit être exclu. Même si des expériences comme celle décrite ici peuvent être très utiles et nécessaires, elles ne sont pas suffisantes pour l'étude de plusieurs phénomènes. Les scientifiques doivent donc utiliser d'autres méthodes que le design expérimental classique pour suivre la démarche scientifique et tenir en compte l'incertitude venant avec des variables non-mesurées. Nous vous référons au texte principal pour plus de détails sur la question.

Une démarche scientifique, plusieurs méthodologies de recherche

Le design expérimental classique, où toutes les variables sont contrôlées, est la méthodologie qui permet la plus grande certitude quant aux liens de cause à effet^{156,162} (Fig. 3). Cependant, lorsqu'on étudie des systèmes qui dépassent une certaine échelle spatiale ou temporelle, il devient difficile, voire impossible de contrôler toutes les variables. Ceci s'accroît encore lorsqu'on travaille avec le vivant, où certaines variables semblent même impossibles à mesurer directement à l'aide des cinq sens et des outils qui les prolongent. Dans un tel contexte, utiliser le design expérimental classique peut encore donner des résultats quantitatifs solides, mais ceci nécessite l'exclusion de variables non-contrôlables ou le contrôle de variables dont la variation serait nécessaire pour bien comprendre le phénomène (Boîte 1). Plus le nombre de variables contrôlées ou exclues sera grand, moins l'expérience ainsi simplifiée sera représentative de la réalité plus complexe (Fig. 3).

Il est toutefois possible de conduire une expérience semi-contrôlée, où certaines variables sont manipulées et mesurées, mais où certaines autres restent non-contrôlées ou même inconnues. Comme pour le design expérimental classique, c'est l'analyse statistique des probabilités d'observer ces résultats qui permettra ici de vérifier la validité de l'hypothèse (Fig. 2). Voir Crawley⁴¹ et Quinn et Keough¹⁵⁶ pour des exemples d'analyses statistiques appropriées pour différents types de données récoltées.

Un autre type de méthodologie de recherche consiste à observer et documenter (quantitativement ou qualitativement) un phénomène qui se produit en conditions naturelles. À partir de ces documentations et suivant la démarche scientifique, des hypothèses et des prédictions sont formulées par rapport au phénomène observé. Ici, c'est le degré de correspondance entre

les prédictions et les futures observations qui permettra de vérifier la validité des hypothèses¹⁵⁶. Toutefois, cette méthodologie apporte un certain degré d'incertitude par rapport aux causes du phénomène observé puisque plusieurs variables non-contrôlées et non-mesurées pourraient aussi être des causes, que les observations soient en accord ou non avec l'hypothèse de départ¹⁵⁶.

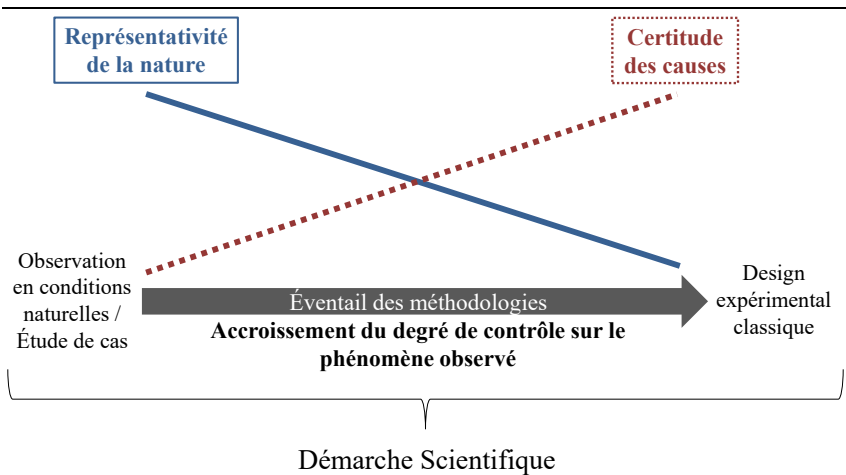


Figure 3. Éventail de méthodologies pour étudier les phénomènes comportant des variables inconnues ou incontrôlables. Plus les variables du phénomène sont contrôlées (ou exclues si le contrôle est impossible) au sein d'une expérimentation, plus la certitude des liens de cause à effet (ligne rouge pointillée) est élevée, mais plus la représentativité du phénomène naturel (ligne bleue continue) est incertaine et potentiellement faible. L'inverse se produit à mesure que les conditions d'observation deviennent plus naturelles, moins manipulées. Tout ceci sous l'égide de la démarche scientifique.

Entre le design expérimental classique et l'observation en conditions naturelles, il y a donc tout un éventail de degrés de contrôle appliqué au phénomène à l'étude¹⁶² (Fig. 3).

Dépendamment du phénomène en question, chaque méthodologie le long de cet éventail est potentiellement appropriée puisque chacune utilise – et est régie par – les étapes de la démarche scientifique (Fig. 1). Plus on se rapproche de l'extrémité « design expérimental » de cet éventail méthodologique, plus on est certain des causes de l'observation, mais plus l'observation risque de s'éloigner de la réalité naturelle du phénomène. Le contraire se produit lorsqu'on s'approche de l'extrémité « incontrôlée » de l'éventail.

Étant donné le compromis spécifique à chaque type de méthodologie – c'est-à-dire le compromis entre certitude des causes et représentation des conditions naturelles (Fig. 3) – chacune a aussi des stratégies spécifiques de validation et de généralisation des résultats^{41,156}. Ces stratégies sont nécessaires, car les domaines de recherche où toutes les variables ne sont pas contrôlables ou connues sont confrontés à trois difficultés majeures, chacune apportant des doutes quant aux résultats d'une étude.

Premièrement, il y a le problème de la reproductibilité invariable du phénomène. Plus le phénomène à l'étude est difficile à circonscrire et à reproduire avec certitude, plus les concepts théoriques soutenant la recherche devront être solides et plus les hypothèses de recherche devront être claires et justifiées^{12,79}. Ces hypothèses devront aussi être testées un plus grand nombre de fois et dans une plus grande diversité de conditions expérimentales afin de pouvoir généraliser les résultats.

Deuxièmement, il y a le problème de l'échantillonnage. Comme le phénomène reste incontrôlable, les résultats varient d'étude en étude. Cette variation invite les scientifiques à répéter

l'expérience sur plusieurs échantillons différents. Ceci pose une fois de plus la question de la généralisation, mais ici c'est la représentativité de l'échantillon qui doit être évaluée. En d'autres mots, en utilisant des analyses statistiques, les scientifiques doivent estimer la part de variation du phénomène attribuable aux caractéristiques des sujets de l'échantillon que le scientifique n'a pas pu contrôler. Encore une fois, un nombre important de répliques, d'expérimentations indépendantes ou d'études de cas servira à renforcer ces analyses et les inférences qui en sont tirées¹².

Troisièmement, il y a le problème de la subjectivité du scientifique lui-même¹⁸. De fait, les scientifiques doivent faire un grand effort d'abstraction pour émettre des hypothèses en lien avec les théories existantes, et pour étudier et interpréter les informations publiées tout en tenant compte de ce qui échappe ou pourrait échapper aux contrôles expérimentaux. Dans tout ceci, de nombreuses positions personnelles et culturelles, conscientes et inconscientes, influencent leurs conclusions. La subjectivité du scientifique ne s'arrête cependant pas là. Dans le domaine des sciences humaines par exemple, les études sur la validité de la recherche suggèrent que les participants ne devraient même pas être en contact avec un expérimentateur qui est au courant de la véritable nature de l'étude. En effet, des subtilités dans la communication non consciente entre l'expérimentateur et les participants sont suffisantes pour influencer les réponses des participants¹⁶⁶. La méthodologie doit alors contenir les stratégies de validation appropriées permettant de gérer au mieux cette difficulté¹⁶⁶.

L'étude de cas

Dans l'éventail des méthodologies de recherche scientifique, l'étude de cas représente l'extrémité où les variables sont le moins contrôlées ; elles ne le sont pas du tout¹² (Fig. 3). Avec l'étude de cas, il y a une description approfondie et très représentative de la réalité du cas, mais beaucoup d'incertitudes potentielles quant aux variables en action. L'objectif n'est alors pas nécessairement la représentativité générale, mais bien plus l'enrichissement théorique suivant l'interprétation de l'étude de cas : « ... [on] ne vise donc pas une généralisation statistique, mais une généralisation analytique¹². »

En amont de la généralisation analytique, la répétition est souvent une stratégie de validation importante pour les études de cas, permettant l'analyse de différents cas et la comparaison des résultats pour renforcer leur portée théorique¹³³. Ceci permet autant de constater et d'étudier les processus et phénomènes qui sont communs et transférables d'un cas à l'autre que de comprendre ceux qui les séparent. Plus le nombre de cas comparables augmentera, plus l'interprétation des résultats pourra enrichir la théorie et les recherches subséquentes¹² et plus on se rapprochera d'une généralisation analytique.

Plusieurs autres stratégies de validation et de généralisation des études de cas existent, touchant autant à la récolte et l'analyse des données qu'à l'interprétation des résultats (voir Ayerbe et Missonier¹² pour une discussion détaillée sur le sujet). Mais peu importe les stratégies utilisées, il ne faut pas oublier que ce n'est pas le cas en tant que tel qui est étudié ou extrapolé, mais bien ses processus et mécanismes sous-jacents¹².

D'ailleurs, ce point reste valide pour les autres méthodologies de recherche le long de l'éventail¹⁶².

Toutes ces méthodologies – étant encadrées par la démarche scientifique – impliquent un processus itératif entre i) la construction d'hypothèses par rapport à des processus ou des mécanismes, eux-mêmes ancrés dans la théorie ou dans de nouvelles observations et ii) la vérification des prédictions en conditions plus ou moins naturelles^{12,162}. Notez que les méthodologies de recherche et les stratégies de validation et de généralisation présentées ici ne sont pas mutuellement exclusives. Dans des domaines comme l'écologie ou la psychologie, elles peuvent même être complémentaires et nécessaires au développement d'une théorie solide dans laquelle s'ancreront les futures hypothèses de recherche^{12,162}. Notez aussi que le choix de la méthodologie de recherche et des stratégies associées, bien que déterminant, ne concerne qu'une seule étape de la démarche scientifique (Fig. 1D).

Notez finalement que la présentation des différentes méthodologies et stratégies n'est ici ni exhaustive, ni détaillée ; ce n'est pas l'objectif de cet ouvrage. Plutôt, il s'agit de bien faire comprendre au lecteur la différence entre méthodologie de recherche et démarche scientifique et de donner une idée de comment différentes méthodologies peuvent être utilisées dans le contexte de la démarche scientifique. En effet, cette différenciation est essentielle pour être capable de faire une évaluation de la scientificité qui est compétente, car la démarche scientifique est souvent confondue avec le design expérimental classique et les analyses statistiques qui en résultent. Toutefois, ni ce type d'expérience, ni ce type d'analyses statistiques ne sont nécessaires pour suivre la démarche scientifique.

Pourtant, il arrive parfois que des travaux soient critiqués parce qu'ils ne présentent pas la méthodologie de choix de la personne qui critique. Ceci peut être justifié, mais justement, il faut une justification expliquant pourquoi telle méthode est inappropriée dans ce cas précis et pourquoi telle autre méthode aurait été plus appropriée (voir Ioannidis^{93,97}) pour des exemples de comment ceci est possible à faire de façon respectueuse et professionnelle). Malgré la clarté et la logique de ce principe, il arrive encore souvent que des travaux soient dénigrés ou refusés sans autre justification qu'ils ne suivent pas la méthodologie préférée des critiques. Ceci est encore une fois analogue à se concentrer sur le contenant (une méthode particulière) plus que sur le contenu (l'adéquation des méthodes considérées). D'après notre expérience, ceci se produit le plus souvent par ignorance, puisque chacun est généralement expert d'un domaine relativement étroit et puisque le monde des inférences et des statistiques est vaste, complexe et étonnamment mal compris. Il n'en reste pas moins qu'il faille faire attention et s'assurer qu'une critique soit scientifiquement justifiée avant de la considérée comme valide.

Un exemple : les pensées

Les pensées sont un exemple de phénomène qui rend compte de la créativité dont doivent faire preuve les scientifiques qui étudient ce qui échappe aux cinq sens. Plusieurs théories existent sur la nature et le fonctionnement des pensées, car il est impossible de connaître tous les facteurs qui les influencent, ni de les mesurer tous à l'aide des cinq sens et des outils qui les prolongent. Les scientifiques doivent donc élaborer avec précision

des concepts en lien avec les pensées, construire des outils de mesure pour ces concepts et évaluer l'efficacité de ces outils⁹. Les résultats de ces outils de mesure indirecte sont donc soumis à des épreuves statistiques, dont les normes sont établies par des experts rompus aux statistiques⁸.

Par conséquent, différents scientifiques étudiant les pensées étudient des concepts différents avec des outils différents. L'étude des pensées est donc bien affectée par les trois difficultés mentionnées plus haut, c'est-à-dire la reproductibilité invariable du phénomène, la représentativité des échantillons et la subjectivité des scientifiques. Il est effectivement impossible de déterminer sur la base des cinq sens que le phénomène appelé « pensée » qui s'est produit dans une étude est identique à celui qui s'est produit dans une autre étude. D'où l'importance d'une description détaillée et approfondie de la théorie à l'origine de l'hypothèse et du phénomène observé.

Pour un phénomène aussi incontrôlable et influençable que la pensée, les stratégies de validation et de généralisation permettant de pallier à ces trois difficultés (reproductibilité, échantillonnage, subjectivité) sont particulièrement importantes. Comme discuté ci-haut, ces stratégies seront alors spécifiques à la méthodologie choisie (Fig. 2). C'est ainsi que les scientifiques ingénieux réussissent à appliquer la démarche scientifique à des phénomènes tels que la pensée. Les résultats permettent de très nombreuses applications dans des domaines pratiques variés : enquête, publicité, santé mentale, communication, éducation, sociologie et ainsi de suite. Tout cela, sans qu'aucune publication scientifique révisée par les pairs n'ait encore pu prouver l'existence ni même l'universalité des pensées. Au contraire, ce

sont les applications pratiques en lien avec les pensées qui en démontrent l'existence.

Du moment que l'expérience subjective du scientifique d'avoir des pensées suffise à lui en prouver l'existence, il pourra s'efforcer d'appliquer la démarche scientifique à l'étude des pensées. Si, par contre, ce n'est pas suffisant, il risque de juger qu'appliquer la démarche scientifique aux pensées est impossible ou inutile. Ceci est important à considérer. Un tel scientifique risque de déclarer que la recherche sur les pensées est une pseudoscience sans même s'informer réellement sur les méthodes employées. Il tente alors de faire oublier que la scientificité est évaluée par la rigueur et l'ingéniosité avec laquelle la démarche scientifique est appliquée. Dans le discours des faux sceptiques règne donc une confusion entre d'un côté l'évaluation de l'application de la démarche scientifique et de l'autre, de faux arguments ayant presque tous à voir avec le contenant et la forme, bref avec l'habit du moine.

Un exemple récurrent qui illustre à merveille cette obsession du contenant au détriment du contenu est le faux argument selon lequel la publication dans une revue scientifique avec révision par les pairs serait nécessaire pour qu'un travail soit considéré comme scientifique. Une telle publication serait alors non seulement suffisante pour démontrer la scientificité d'une étude, mais en serait aussi le critère royal. Cette association est cependant bien loin d'être automatique, et bien comprendre ce point semble nécessaire pour permettre aux lecteurs d'évaluer la scientificité d'une étude. C'est pourquoi les différences entre ce qui appartient à la démarche scientifique et ce qui appartient à la communication scientifique ou aux différents milieux de

recherche seront l'objet principal de notre attention pour le reste de ce livre.

L'INFLUENCE DU SYSTÈME DE PUBLICATION SUR LA COMMUNICATION SCIENTIFIQUE

La communication scientifique est un des mécanismes d'autocontrôle dans le domaine de la recherche. Elle n'est cependant pas une étape à part entière de la démarche scientifique. Elle est un outil permettant i) la rencontre entre scientifiques travaillant dans le même domaine, conduisant idéalement à leur collaboration, ii) de faire avancer l'état des connaissances plus rapidement en évitant aux scientifiques de faire les mêmes expériences sans le savoir, iii) d'effectuer des revues de littérature, facilitant ainsi grandement la deuxième étape de la démarche scientifique (Fig. 1B) et iv) de filtrer en partie les travaux scientifiques présentant des failles évidentes.

Un des moyens privilégiés de communication est la publication dans des revues scientifiques avec révision par les pairs. Malgré les nombreux bénéfices qu'apporte ce type de communication, le système de publication en œuvre aujourd'hui apporte aussi avec lui plusieurs écueils plus ou moins bien reconnus et qui peuvent contrevenir avec les idées mises de l'avant par la démarche scientifique¹¹. Cette contrevenance potentielle est d'ailleurs l'une des raisons pour lesquelles nous présentons ici – contrairement à d'autres auteurs^{45,136,166} – la communication scientifique comme une étape extérieure à la démarche scientifique.

En effet, la scientificité d'un travail est évaluée par rapport à son contenu, c'est-à-dire par la rigueur et l'ingéniosité de

l'application de la démarche scientifique (Fig. 1), peu importe si ce travail est communiqué ou non. Le système de publication scientifique représente plutôt le contenant ; ce n'est que la forme généralement choisie par les personnes au sein de la communauté scientifique pour communiquer entre elles.

Naturellement, il y a un certain contrôle de ce qui est publier dans une revue scientifique avec révision par les pairs et donc ici le contenu et le contenant peuvent être plus étroitement reliés qu'ailleurs. Cependant, nous verrons plus tard que ce contrôle n'est pas très fiable et peut même aller à l'encontre de la démarche scientifique, d'où l'importance de continuer à évaluer le contenu lui-même plutôt que de lui inférer une valeur en se basant sur son contenant. En réalité, ceux-là mêmes qui incluent la communication dans la démarche scientifique lui accordent souvent une place particulière, telle une greffe à notre figure 1, son importance semblant alors plutôt découler de l'incapacité des scientifiques à suivre correctement la démarche scientifique (voir par exemple Robert¹⁶⁶, chap. 1).

Étant donné la confusion entre communication et démarche scientifique, et étant donné l'importance qu'ont pris les revues scientifiques dans le monde scientifique, le système de publication actuel doit donc lui-même être évalué d'un point de vue scientifique afin de pouvoir évaluer la scientificité d'une publication, d'un travail ou même d'une discipline.

Les rôles du système de publication des revues scientifiques

Le système de publication des revues scientifiques fait actuellement partie intégrante de la recherche fondamentale (c'est-à-dire la recherche sans but commercial immédiat, ou non-contextualisée). Ainsi, en plus de servir à avancer l'état des connaissances par la communication, ce système est aussi devenu l'outil principal pour évaluer la renommée et la performance des scientifiques et de leurs instituts d'affiliation. Cette double fonction du système de publication des revues scientifiques voit donc émerger un conflit d'intérêts entre d'un côté ce que nous pourrions appeler des conditions optimales de recherche et de partage des connaissances scientifiques et de l'autre côté les intérêts des différents partis jouant un rôle dans la publication scientifique (auteurs, instituts de recherche, éditeurs, réviseurs, organismes subventionnaires, etc.).

Dans les prochains paragraphes, nous explorerons comment ces intérêts partiellement contradictoires s'agencent et pourquoi « scientificité » n'est pas égal à « publication avec révision par les pairs ». Nous avons identifié trois phénomènes qui contribuent à cette dissociation entre la scientificité et la publication scientifique : i) la privatisation de la publication scientifique ; ii) les pressions de la part des organismes subventionnaires ; et iii) les obligations pour faire carrière en recherche.

Privatisation de la publication scientifique

Premièrement, il faut savoir qu'à la suite d'une privatisation intense de la publication scientifique, une grande part des revues scientifiques sont maintenant publiées par des maisons d'édition privées et concurrentes. En 2010, cinq géants se partageaient déjà à eux seuls plus de 40% du marché des publications en science de la nature, technique et science médicale⁴⁰.

Des entreprises à but lucratif dirigent donc en grande partie les modalités de la publication scientifique. Ceci apporte des difficultés allant à l'encontre même des principes de la communication scientifique^{11,34} (Fig. 4). Par exemple, le prix des revues a augmenté jusqu'ici trois fois plus rapidement que le coût moyen de la vie, entraînant une diminution de leur accessibilité – le but premier de la communication scientifique ! – puisque les universités ou les individus n'ont pas les moyens de s'abonner à toutes les revues⁴⁰. Cette situation a mené au développement d'un marché pour des revues en accès libre, donc gratuits pour les lecteurs. Cependant, la grande majorité des revues prestigieuses restent en mode payant.

La cession des droits d'auteur aux éditeurs de ces revues constitue une barrière supplémentaire à la communication optimale des connaissances, pouvant empêcher les scientifiques de communiquer leurs propres travaux en toute liberté avec collègues, étudiants, ou le public⁴⁰. De plus, pour gagner de l'espace et ainsi augmenter leurs gains financiers, les revues imposent généralement des normes très strictes empêchant souvent les auteurs de présenter leurs travaux aussi bien qu'ils le désireraient ou même qu'il le faudrait, ce qui se traduit par une

baisse de la qualité des articles publiés et donc de la communication^{112,166,189}. Depuis le début des années 2000, un fort mouvement en faveur du partage libre des données scientifiques a émergé afin d’améliorer la situation. Cependant, son succès est encore limité et diffère fortement d’un domaine à l’autre⁴⁰.

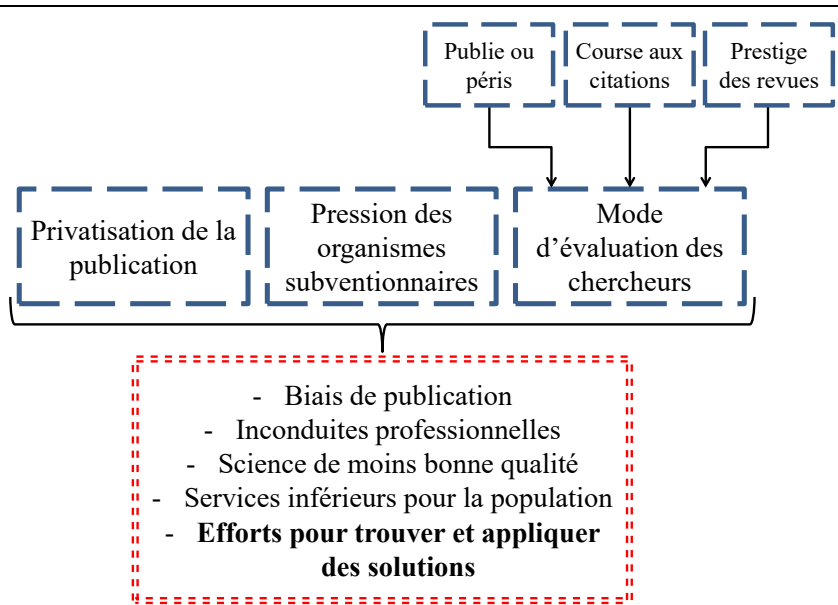


Figure 4. Résumé schématique des facteurs allant, au moins partiellement, à l’encontre d’une communication scientifique qui serait au service du bon fonctionnement de la démarche scientifique et de l’avancement des sciences (lignes simples discontinues, bleues), et les impacts d’une telle constellation (ligne double pointillée, rouge).

Pressions des organismes subventionnaires

Deuxièmement, il faut savoir que les intérêts des organismes subventionnaires exercent eux aussi une très forte pression sur la recherche et donc sur la science qui sera publiée au final⁴⁰. Par exemple, pour obtenir du financement, les scientifiques auront tendance à mettre dans leur proposition de recherche plus d'emphasis sur les intérêts des organismes subventionnaires⁹⁴. En écologie par exemple, parler de changements climatiques et de dioxyde de carbone dans les propositions de recherche semble augmenter les chances d'être financé. Nous avons nous-mêmes pu observer des propositions de recherche où les changements climatiques y étaient intégrés malgré le fait que la recherche qui devait être conduite n'avait en réalité pratiquement rien à voir avec ce sujet. Ceci se combine à un climat de compétition extrême où l'originalité, l'audace et la coopération peuvent être négligées et supprimées⁴. Les intérêts de recherche des scientifiques entrent alors potentiellement en contradiction avec leurs obligations de carrière, car ils doivent parfois choisir entre d'un côté ce qui selon leur avis est le plus pertinent et intéressant pour l'avancement de leur domaine de recherche et de l'autre côté ce qui a le plus de chances d'être financé.

Lorsque les organismes subventionnaires sont publics, les propositions de recherche sont normalement évaluées par des experts indépendants. Dans ces cas, ce sont les conflits d'intérêts entre scientifiques et l'ouverture d'esprit des évaluateurs qui limiteront plus ou moins la liberté de recherche des scientifiques en demande de financement (voir la section « Suppression de l'innovation »). Un autre type de pression provenant d'organismes subventionnaires publics tels que les universités et les instituts de

recherche publics est le fait que ces organismes allouent leur financement en se basant sur des indices de performance tels que le nombre de publications, ce qui mène à diverses inconduites professionnelles, tant au niveau individuel qu'au niveau institutionnel⁵³. Nous parlerons de cet aspect plus en détails dans la prochaine section. Il peut aussi arriver qu'un gouvernement semble décider de couper le financement en ciblant des équipes de recherche dont les résultats vont à l'encontre de ses intérêts (voir par exemple l'éditorial publié dans *Nature* en 2012 sur le cas des Régions des Lacs Expérimentaux (RLE) en Ontario, Canada⁵²).

Le pouvoir des organismes subventionnaires peut causer des difficultés plus substantielles et évidentes lorsque ceux-ci sont privés^{83,205}. Il est par exemple possible pour des scientifiques universitaires de signer des contrats dans lesquels les intérêts de l'industrie sont protégés, mais pas ceux du public. Ceci peut aller jusqu'à des clauses de confidentialité empêchant la divulgation de certains résultats indésirables pour l'industrie¹⁸⁵, tout cela en accord avec les normes universitaires^{197,205}. Des cas où certaines compagnies semblent tout simplement « commander » des résultats sont aussi répertoriés³¹. Cette pression des organismes subventionnaires pour manipuler la recherche et la publication des résultats est un problème extrêmement sérieux où l'éthique et le bien-être de la société contredisent l'aspect purement légal⁹⁴ (Fig. 4) et qui se manifeste de façon plus ou moins évidente dépendamment des cas.

Dans le domaine de la recherche médicale par exemple, l'industrie pharmaceutique influence et contrôle aujourd'hui pratiquement toutes les étapes, allant de la définition des questions à poser jusqu'aux décisions de « où », de « comment » et de « quoi » sera publié¹⁸⁸. Des conflits d'intérêts évidents sont

donc à l'œuvre, les conclusions de ces recherches étant d'ailleurs souvent défaillantes comme indiqué par l'apparition sur le marché de nombreux médicaments qui doivent ensuite être retirés¹⁸⁸.

Un exemple concret concerne le domaine des antidépresseurs, où une revue de littérature portant sur 12 différents antidépresseurs a trouvé que les articles publiés sur le sujet démontraient presque toujours l'efficacité du médicament, c'est-à-dire que les résultats négatifs ne sont généralement pas publiés⁹³. Cette même revue de littérature démontre en plus que l'efficacité reportée dans la littérature scientifique est en plus exagérée dans tous les cas étudiés (entre 11 et 69% d'exagération de l'effet) et qu'en analysant tous les données, c'est-à-dire aussi ceux qui n'ont pas été publiés, seuls de petits effets modestes peuvent être démontrés. De plus, il semble que les essais cliniques sur les antidépresseurs (et d'autres médicaments?) ne sont pas développés pour détecter et évaluer les effets secondaires négatifs possibles – ceux-ci étant alors détectés après commercialisation – mais bien pour détecter l'effet souhaité par les producteurs du médicament qui financent ces essais, afin de mettre au plus vite ces produits sur le marché. Ces essais ne représentent donc pas vraiment une activité scientifique ni une preuve d'un bénéfice clinique pour les patients, mais plutôt une activité visant à démontrer certains critères légaux standards à des fins commerciales⁹³. Pourtant, le fait que ces essais cliniques soient publiés avec révision par les pairs, combiné avec le jargon scientifique nécessaire, fait croire à la scientificité des résultats et des conclusions. Au contraire, nous serions personnellement même tentés de dire que ce type de recherche démontre dangereusement beaucoup de caractéristiques associées aux pseudosciences. Malheureusement, juste parce que c'est publié, tout ceci passe largement inaperçu jusqu'à ce qu'il y ait de vrais

dommages. Cet exemple sur les antidépresseurs⁹³ nous montre bien comment, dans certains cas, la publication avec révision par les pairs et la scientificité peuvent être deux choses très différentes.

Ce genre de situation est largement provoqué par la façon dont le financement de la recherche est fait et comment il motive les scientifiques. Ces derniers sont obligés de jouer selon les règles pour recevoir du financement et, ultimement, conserver leur emploi ou être promus. Évidemment, les conflits d'intérêt entre les organismes subventionnaires et la recherche ne sont pas toujours contre la volonté des scientifiques lorsque ceux-ci s'attendent aussi à un gain financier par rapport au produit qu'ils décrivent dans leur publication²⁰⁶. Tout de même, il semble que les scientifiques soient généralement honnêtes et qu'ils souffrent plus qu'ils ne gagnent de cette situation.

De fait, le mode de financement actuel dans le monde des essais cliniques encourage de petites études rapidement publiables au détriment de questions plus intéressantes ou plus utiles pour le public, mais qui ne se conforment que difficilement aux standards des organismes subventionnaires⁹⁷. Ce focus étroit et peu pertinent d'un point de vue clinique résulte souvent en une démarche scientifique bâclée. Ceci ne se produit pas nécessairement (ou pas seulement) au niveau de l'expérimentation, mais avant, lors du positionnement et de la justification des questions et des hypothèses (Fig. 1). Cet état d'affaire représente une atmosphère de recherche bien désagréable pour les scientifiques impliqués et qui aimeraient faire un travail de qualité. Ça montre aussi encore une fois clairement la dissociation entre le monde de la publication et la démarche scientifique.

Ces conflits d'intérêts entre l'industrie et une communication scientifique optimale ont aussi récemment provoqué une énorme hausse dans la production de méta-analyses^{Note 4} en biomédecine, où de la publicité pure et simple pour des produits de l'industrie vient embourber la publication scientifique avec des méta-analyses redondantes et de piètre qualité⁹⁶. Parallèlement, des compagnies se spécialisent dans la production de méta-analyses et sont engagées par l'industrie pour aider les décisions d'investissement dans la recherche et le développement⁹⁶. Ces méta-analyses sont le plus souvent pertinentes et de bonne qualité, mais elles ne sont généralement pas publiées, par manque de motivation. En effet, ces professionnels ont plus gagné en investissant leur temps dans un autre contrat de méta-analyse qu'en préparant un manuscrit pour le processus de publication. Parfois aussi l'industrie ne souhaite tout simplement pas que les résultats soient publiés. Ainsi, nous voyons ici encore comment la publication et la scientificité ne sont pas synonymes ; des méta-analyses de piètre qualité n'ayant pas grand-chose à voir avec la rigueur scientifique sont souvent publiées, alors que des méta-analyses de bonne qualité ayant beaucoup de valeur pour la science ne le sont pas.

Il devient alors clair, comme nous l'avons mentionné dans la première section de cet ouvrage, que la publication n'a

⁴ Une méta-analyse est une méthode d'analyse systématique où plusieurs études indépendantes portant sur une question commune sont combinées afin de pouvoir tirer des conclusions plus globale sur la question. Par exemple, une méta-analyse sur l'effet d'un médicament combinera les données et les résultats de toutes (du moins plusieurs) les études ayant testé cet effet, ce qui permettra d'avoir une meilleure idée de si ce médicament a un véritable effet ou non, au-delà des limites d'échantillonnage de chaque étude individuelle. Les méta-analyses font partie intégrante de la recherche et sont essentielles pour avancer dans la bonne direction au sein d'un domaine de recherche donné.

absolument rien à faire dans un schéma représentant la démarche scientifique.

Notez que dans cette section, nous parlons beaucoup de la recherche biomédicale et des essais cliniques. La raison en est fort simple : étant donné l'importance de cette recherche pour la santé et la sécurité du public et étant donné les dangers lorsque cette recherche est conduite de façon inappropriée, c'est dans ce domaine que l'information sur les conflits d'intérêts est la mieux documentée. Ce domaine offre donc les données les plus fiables pour démontrer les points dont nous avons discutés, mais sachez que les conséquences de la pression des organismes subventionnaires sont vraisemblablement similaires partout où des conflits d'intérêts sont présents. Sachez aussi que bien qu'un contrôle accru ait amélioré la situation dans certains cas de recherche biomédicale⁹⁹, il reste encore énormément de travail à faire pour régler la situation¹⁸⁶.

Obligations carriéristes des scientifiques

Il faut finalement savoir que pour faire carrière en recherche fondamentale, les scientifiques misent avant tout sur leur capacité à obtenir des subventions de recherche, ce qui est principalement lié – mis à part leur thème de recherche – au nombre de leurs publications, au prestige des revues qui les ont publiées et au nombre de citations obtenues par leurs travaux.

La communauté scientifique en général sait que beaucoup d'efforts personnels de la part d'un évaluateur doivent être investis pour mener à bien l'évaluation d'un scientifique puisque,

même lorsque combinés, ces critères de sélection ne permettent pas de faire le travail adéquatement¹¹² (Fig. 4). Cependant, les scientifiques ou bureaucrates qui servent d'évaluateurs ne suivent aucune formation à la révision. De plus, ils sont débordés. Ainsi, bien que les institutions de recherche assurent généralement le contraire, les trois critères de sélection mentionnés ci-haut semblent avoir encore énormément de poids sur les décisions d'emploi, de financement, de promotion et de récompense^{53,155}, ou du moins c'est ce que croient 70% des scientifiques professionnels¹. C'est ainsi que même si ces critères de sélection ne jouaient pas de rôle majeur dans les décisions institutionnelles – contredisant l'expérience de beaucoup de scientifiques – ce qui compte est ce que les scientifiques pensent qui se passe, car ils modifieront leur comportement, leurs relations et l'ambiance de recherche en conséquence de ce qu'ils croient¹, affectant ainsi la qualité de la recherche qui se fait.

Le focus – individuel ou institutionnel – sur ces trois critères de sélection peut en effet amener avec lui plusieurs répercussions néfastes sur la science qui est publiée^{53,112}. La raison principale pour ça est que, comme ce sont par ces critères que les scientifiques et les instituts sont évalués, optimiser ces indices devient alors l'objectif principal, et non plus faire de la science de bonne qualité. Ces deux objectifs ne sont pas nécessairement contradictoires, mais ils ne sont pas non plus nécessairement équivalents, comme nous le verrons en détails plus loin.

De plus, selon la loi de Goodhart – nommée ainsi d'après l'économiste Britannique l'ayant décrite pour la première fois – aussitôt qu'une mesure numérique est choisie comme un indicateur de performance, elle s'ouvre à la corruption et à l'exploitation et cesse donc de fonctionner en tant

qu'indicateur^{22,182}. C'est d'ailleurs la raison pour laquelle une majorité de scientifiques se disent insatisfaits par le système d'évaluation en place, car ils craignent que leurs collègues manipulent ces indices pour « gagner la compétition »¹.

Le fait que les instituts disent que leurs décisions ne sont pas ou peu affectées par les indices de performance, mais que la majorité des scientifiques vivent des situations où ces indices ont clairement beaucoup d'importance, crée aussi une ambiance de flou qui rend les scientifiques insécurs et insatisfaits par rapport aux évaluations de performance¹. Tout ceci ne signifie pas que les scientifiques soient contre l'utilisation d'indices numériques pour évaluer leur performance. En effet, lorsqu'on leur demande¹, la plupart d'entre eux *veulent* être évalués par de tels indices, malgré la méfiance qu'ils ont vis-à-vis de la facilité avec laquelle ces critères peuvent être manipulés (loi de Goodhart). Nous voyons deux raisons principales pour cette prise de position de la part des scientifiques.

Premièrement, il semble qu'ils perçoivent ces indices de performance quantitatifs comme moins subjectifs que la consigne générale « d'évaluer la qualité de la recherche », qui elle peut varier énormément d'un évaluateur à l'autre.

Deuxièmement, il est probable qu'une fois qu'un scientifique ait investi beaucoup d'énergie pour améliorer certains indices de performance, même si ceux-ci ne sont pas adéquats, il n'ait pas envie que le système d'évaluation soit modifié, car il pourrait alors perdre l'énergie investie au profit d'autres critères de performance dans lesquels il n'aurait pas mis autant d'efforts.

Quoiqu'il en soit, l'utilisation de ces indicateurs de performance (nombre de publication, citations obtenues, prestige des journaux) a des répercussions pouvant réduire la qualité de la science produite aujourd'hui et il est important de bien comprendre ces répercussions et la forme qu'elles peuvent prendre pour être en mesure d'évaluer plus clairement les scientifiques et la scientificité en général. Dans les prochaines sections, nous explorons donc en détails le « côté sombre » de chacun de ces indices et les conséquences pour l'évaluation de la science.

Publie ou périclisse

Examinons tout d'abord le besoin des scientifiques de se concentrer sur la quantité, entre autre en publiant le plus grand nombre d'articles possible. Ce phénomène est appelé la culture du « publie ou périclisse ». Il semble que beaucoup de scientifiques souffrent de cette culture dans leur carrière, ressentant une grande pression pour publier en quantité afin de recevoir du financement¹⁹⁹. Beaucoup perçoivent en effet que l'évaluation de leur performance se base plus sur des mesures de quantité que sur des mesures de qualité. Ceci provoque une atmosphère générale de frustration et de ressentiment vis-à-vis de la culture de publication et d'évaluation actuelle, puisque la plupart des scientifiques préféreraient se concentrer sur la qualité de leur travail^{10,199}.

De plus, pour les professeurs ou tout autre personnel supervisant des étudiants, la pression pour la quantité n'est pas limitée aux publications, mais se transfère aussi aux nombres de doctorants qu'ils supervisent avec succès¹⁹⁹. Ceci a aussi des

répercussions néfastes sur la science, réduisant les standards de qualité des doctorants – qui font d’ailleurs souvent le travail que des étudiants à la Maîtrise pourraient faire – et provoquant une surproduction de scientifiques par rapport au nombre d’emplois dans le domaine⁵³.

En théorie, l’importance attribuée au nombre d’articles publiés vise à augmenter la productivité d’une recherche de qualité⁵³. Cependant, l’atmosphère actuelle de « publiés ou périss » risque plutôt de diminuer la qualité moyenne des articles publiés, puisque les scientifiques seront tentés de découper leurs travaux afin de publier l’unité publiable la plus petite – aussi appelé *salami-slicing* – produisant une grande quantité d’articles de moins en moins profitables pour la science^{40,53,73,112,189}. Cette tendance semble aussi encouragée par les éditeurs, qui imposent des limites de longueur strictes pour sauver de l’espace, même si en réalité la redondance entre plusieurs articles courts qui auraient pu être écrits en un seul gaspille les ressources des revues, des réviseurs, des auteurs et des lecteurs^{186,189}.

Cependant, les effets exacts de cette pression pour publier en quantité ne sont pas si faciles à circonscrire. Fanelli et Larivière⁶⁵ ont par exemple démontré qu’en moyenne, depuis l’année 1900, le nombre de publications par auteur individuel n’a que peu ou pas augmenté. Ceci signifie que l’idée très répandue du *salami slicing* causé par une pression pour publier trop importante serait, sinon fausse, du moins bien exagérée.

Ces mêmes auteurs présentent aussi des exceptions à leurs résultats principaux, entre autre le fait que pour plusieurs pays non-anglophones, le nombre d’articles par auteur aurait grandement augmenté, parfois même triplé. La raison pour une

telle différence reste incertaine. Par exemple, publier en anglais devient de plus en plus nécessaire pour les scientifiques voulant être reconnus, du moins au niveau international. Étant donné que l'outil de recherche d'articles scientifiques « *Web of Science* », utilisé par Fanelli et Larivière, répertorie moins bien les articles non-anglophones, ceci pourrait donner l'impression qu'il y a beaucoup plus d'articles publiés dans ces pays, alors que seule l'habitude de publication aurait changée. Une autre explication pourrait être que la pression pour publier en quantité (du moins pour les étudiants) est typiquement moins grande dans les pays non-anglophones, ce qui cependant change rapidement dans un contexte de recherche de plus en plus international.

Quoiqu'il en soit, notez bien que ces résultats ne signifient pas que la littérature scientifique ne soit pas inondée d'articles médiocres ; seules les relations de cause à effet ont été remises en question. Par exemple, le fait qu'il y ait une pression pour superviser le plus de doctorants possible mène inévitablement à une surproduction de scientifiques et à une baisse moyenne du standard de ceux-ci^{53,199}. Comme chaque scientifique vit bel et bien sous la pression de publier un certain nombre d'articles (même si ce nombre ne variait pas avec le temps), cette constellation crée automatiquement une « inondation » d'articles dont la qualité moyenne diminue.

Il y a donc actuellement bel et bien une augmentation exponentielle de la quantité d'articles – qu'ils soient médiocres ou de qualité – et de revues scientifiques qui les publient. Ceci rend les revues de littérature fastidieuses et souvent impossibles à faire complètement et correctement^{18,73}. De surcroît, les scientifiques agissant en tant que réviseurs sont complètement surchargés par ce trop grand nombre d'articles soumis aux revues¹⁰. Cette

situation accroît aussi les problèmes financiers liés à l'accès des publications scientifiques, puisque sécuriser un accès à toutes les revues et tous les articles disponibles devient extrêmement coûteux^{40,166}.

Une autre conséquence moins discutée de la culture du « publie ou périclisse » est qu'elle rend les scientifiques vulnérables aux revues dites prédatrices, celles-ci ne faisant aucun contrôle de qualité et étant surtout motivées par les paiements reçus de la part des auteurs²³. En effet, comme chaque jeune scientifique doit publier pour continuer sa carrière, si la recherche produite n'est pas acceptable pour les journaux bien cotés et reconnus (et ce, pour quelque raison que ce soit), certains scientifiques peuvent devenir désespérés et préférer publier des résultats douteux ou inutiles dans des revues tout aussi douteuses plutôt que de ne pas publier.

Pour tenter de réduire le nombre d'articles publiés, Statzner et Resh¹⁸⁹ discutent d'un retour vers la publication de longs manuscrits. Cette solution n'est cependant pas particulièrement populaire, car les scientifiques ne veulent généralement pas lire de longs articles, préférant souvent ne lire que le titre et regarder les images^{Note 5}. C'est d'ailleurs aussi la

⁵ Un jour que je (Jessy) présentais mes travaux (avant publication) à un scientifique américain très renommé dans mon domaine, il me conseilla de réduire de beaucoup la présentation de mes données en une image facile à comprendre et transmettant le message que je voulais donner, car selon lui les scientifiques ne lisent plus les articles : ils lisent le titre, le résumé si on a de la chance et regardent les images. Je lui ai alors expliqué que c'était impossible de présenter tous les résultats et les nuances essentielles de l'étude en une image simple, puisque ma recherche était une analyse synthétisant une dizaine de variables avec plusieurs niveaux hiérarchiques et une dizaine de sites d'étude et que, justement, c'était les nuances qui étaient importantes. Sa réponse : « Veux-tu que les gens citent ton article ou non ? » Il est difficile de dire à quel point cette attitude est

raison pour laquelle les indices de performance sont si influents et nombreux : tout le monde veut éviter de lire des articles⁵⁴. Une solution encore plus drastique provient de l'écologiste aquatique Gernot Bretschko qui souhaitait que les écologistes limitent leur carrière à cinq longs articles, chacun présentant l'entièreté de la recherche le concernant. Il va sans dire qu'une telle résolution nécessiterait un ajustement radical de la part des scientifiques, des instituts de recherche, des revues scientifiques ainsi que des organismes subventionnaires. De plus, elle ne serait certainement pas adéquate pour certains domaines où la communication rapide des résultats est essentielle.

Une idée plus réaliste et certainement digne d'attention est de limiter les scientifiques à un seul article par année¹³⁷. Ceux-ci auraient donc intérêt à produire un article qui compte vraiment, c'est-à-dire de grande qualité, et ils auraient le temps nécessaire pour ce faire¹³⁷. Notez cependant que bien que ce type de solution relativement simple soit attrayant, il ne pourrait pas à lui seul solutionner les problèmes présentés ici et pourrait même exacerber les problèmes associés aux autres mesures de performance si d'autres changements ne l'accompagnent pas (voir Fanelli et Larivière⁶⁵).

D'autres facteurs influencent la quantité et la qualité des articles dans la littérature scientifique. Par exemple, il est attendu des étudiants et résidents en médecine qu'ils fassent de la recherche, même si plus tard les médecins n'en font pas nécessairement, ou même pas généralement. Cependant, ces

généralisée (mes superviseurs n'étaient absolument pas d'accord), mais d'après notre expérience, elle est certainement répandue.

derniers ne sont pas jugés sur la qualité de leur recherche ni même sur son utilité, mais bien sur leur capacité à produire rapidement quelque chose de publiable, ce qui n'est souvent pas compatible avec de la bonne recherche clinique⁹⁷. Les universités exigent ces publications, sans focuser sur l'utilité clinique pour le public⁹⁷.

Ainsi, bien que le nombre d'articles par personne ne soit pas influencé par l'obligation pour les futurs médecins à publier quelques articles avant de se concentrer sur la pratique de la médecine, la façon dont le système fonctionne résulte en de nombreux articles scientifiques relativement inutiles...et coûteux. Ioannidis⁹⁷ plaide pour un changement de procédure, où le focus sur une grande quantité de publications rapide serait remplacé par un focus sur l'entraînement des futurs médecins à reconnaître et comprendre la recherche clinique de qualité, point essentiel pour une pratique médicale de qualité. Nous reconnaissons dans cette recommandation notre analogie de l'habit ne fait pas le moine, où il est beaucoup plus important de reconnaître les vrais moines que de se fier aux soutanes.

Il apparaît donc que, même si en moyenne les scientifiques ne publient pas plus qu'avant, la pression pour publier est bel et bien présente, se manifeste de différentes manières et est fortement ressentie par les scientifiques, surtout chez les jeunes. Aussi, bien que plusieurs hypothèses existent sur le sujet, il reste difficile de voir clairement comment cette pression de publication affecte vraiment la science qui se fait⁶⁵. Le climat de recherche créé par cette pression est néanmoins difficile à supporter pour beaucoup de personnes¹⁹⁹ et, en tant que communauté scientifique, nous devrions investir plus d'efforts pour étudier la question.

La course aux citations

Le nombre de citations obtenues par les travaux d'un scientifique est un autre point important de l'évaluation de pratiquement tous les aspects de sa carrière¹⁷⁴. Le nombre de citations obtenues pour chaque travail est souvent perçu comme une mesure de performance et de qualité plus intéressante que le nombre d'articles publiés, car le nombre de citations donne un indice de comment le travail en question a influencé la communauté scientifique. De fait, avoir publié 100 articles que personnes n'a lus ne démontre pas un grand impact scientifique alors qu'avoir publié 10 articles cités chacun 100 fois, oui. Le nombre d'articles publiés et le nombre de citations obtenues sont donc souvent combinés en un seul indice, notamment avec l'indice h ^{Note 6}.

L'objectif d'évaluer les scientifiques par rapport aux nombres de citations qu'ils obtiennent est de récompenser la recherche de qualité qui influence les autres scientifiques⁵³. Cependant, comme toutes les autres mesures quantitatives servant d'indice de performance, le nombre de citations est aussi sujet à la manipulation et à la dénaturation²². C'est en partie pourquoi on retrouve aussi plusieurs biais de citation qui n'ont plus rien à voir avec la qualité de la recherche qui se fait.

⁶ L'indice h est calculé en combinant la productivité d'un scientifique (nombre d'articles publiés) et son impact (nombre de citations). Il est égal au plus grand nombre possible d'articles (x) ayant chacun un nombre de citations égal à x . Ainsi, 1000 articles cités 1 fois chacun donne une valeur de h égale à 1, tout comme 1 seul article cité 1000 fois. Cinq articles cités chacun 5 fois donnera une valeur de h égale à 5. Plus de citations ne changeront pas cette valeur, car il n'y a que 5 articles. Pour augmenter à 6, il faudra au moins un autre article et attendre qu'au moins 6 articles aient au moins 6 citations.

Shadish et coautrices¹⁷⁴ ont évalué les articles cités par des auteurs scientifiques en psychologie. Ils ont trouvé que les articles cités étaient en moyenne de plus haute qualité et/ou représentaient un exemple classique dans leur domaine, quoique seulement 2% des auteurs aient listé « qualité » comme une raison principale pour citer un article. Shadish et coautrices ont aussi démontré qu'à qualité égale, les articles plus créatifs sont moins cités, défavorisant les scientifiques qui font preuve de créativité et d'innovation.

Plusieurs autres raisons n'ayant rien à voir avec la qualité de la recherche peuvent faire qu'un article est plus ou moins cité. Si un scientifique travaille dans un champ de recherche étroit ou s'il ne publie pas en anglais, ses travaux seront moins cités¹⁰⁷. Un article présentant des résultats significatifs, particulièrement s'ils étaient prédits par les auteurs, aura plus de chances d'être cité^{49,103,185}. En moyenne, les résultats significatifs donnent 32% plus de chance à un article d'être cité⁶¹, quoique ceci varie beaucoup entre les différents domaines de recherche, allant de 0% en physique ou en chimie⁶¹ jusqu'à 600% dans le domaine des essais cliniques¹⁵⁸. Les articles qui présentent des idées provocantes, qui « font des vagues » ou qui sont au cœur d'un débat scientifique seront beaucoup plus cités, ce qui explique pourquoi les revues les plus prestigieuses convoitent ce genre d'articles et influencent ainsi la recherche¹⁷¹.

Certains éditeurs de ce genre de revues « suggèrent » même qu'une liste d'articles récemment publiés dans leur revue soit rajoutée aux références d'un article soumis pour acceptation afin d'augmenter leur facteur d'impact¹⁸⁹. Certains réviseurs font de même, demandant à ce que certains de leurs propres travaux soient cités dans un article qu'ils révisent, tout ceci gonflant de

façon plus ou moins douteuse les listes de références des articles scientifiques⁵³.

Il existe d'autres exemples de biais de citation dont la raison d'être s'éloigne encore plus de la qualité de la recherche. Les auteurs scientifiques auront par exemple tendance à citer des articles où le premier auteur a la même nationalité qu'eux¹⁰⁴. Un autre exemple, plutôt farfelu, est que plus l'initiale du nom du premier auteur d'un article arrive tôt dans l'alphabet, plus cet article semble avoir de chances d'être cité²⁰². La cause de cette tendance est incertaine, mais il se pourrait qu'un auteur qui cherche des références simplement pour supporter son point de vue dans sa liste de références – classées en ordre alphabétique – en trouve suffisamment bien avant d'arriver à la lettre Z.

Une autre indication que le nombre de citations est un indice de performance douteux est le fait que plusieurs articles qui sont rétractés – peu importe la raison – continuent d'être cités comme si les auteurs qui les citent n'étaient pas au courant de la rétraction, malgré le fait que ce soit clairement indiqué^{134,204}. Ceci nous montre que certains auteurs ne portent pas toujours beaucoup d'attention à ce qu'ils citent. C'est d'ailleurs un problème général avec le fonctionnement de l'écriture d'articles scientifiques ; la seule chose qui compte, c'est d'avoir un article publié à citer pour supporter une affirmation. Que l'article en question soit de qualité ou non est souvent d'importance secondaire et n'est que rarement vérifié, même par les auteurs qui le citent, surtout pour les affirmations qui ne sont pas dérangeantes. Ce biais de citation peut même maintenir en vogue certaines affirmations ou théories même si celles-ci sont clairement fausses ou non-fondées¹⁹⁴. Ceci se produit malgré la bonne volonté des scientifiques, tout simplement parce que trop

souvent dans le système de publication actuel, le focus sur la présentation, sur le contenant donc, prend une place démesurée, parfois plus grande que celle attribuée à la qualité du contenu.

Quoiqu'il en soit, il apparaît que le type de résultats (positifs et en accord avec l'hypothèse de départ) et le facteur d'impact de la revue dans laquelle un article est publié influencent plus les probabilités de citation que la rigueur et la qualité scientifique d'une étude⁴⁹. En d'autres mots, les affirmations faites par les auteurs et la revue dans laquelle celles-ci sont publiées sont souvent plus importants que les données récoltées et les analyses exécutées⁴⁹. Ceci nous ramène encore une fois à l'idée que dans la publication scientifique, le contenu peu être négligé au profit du contenant, alors qu'en science ce devrait être le contraire. Il est donc évident que le nombre de citations ne donne que peu d'information quant à la qualité d'un article et que seule une lecture détaillée de la part de l'évaluateur permette un jugement informé¹⁷⁴.

Le prestige des revues

Le prestige des revues dans laquelle les articles scientifiques sont publiés est un autre point important dans l'évaluation de la performance des instituts de recherche et des scientifiques, malgré qu'il y ait un consensus général sur le fait qu'il s'agisse d'un mauvais indice de performance¹³¹. C'est aussi le point pour lequel notre analogie de « l'habit ne fait pas le moine » est le plus important à bien comprendre pour ne pas se faire aveugler par un critère non-scientifique lors de notre évaluation de la scientificité.

En effet, cette culture de prestige des revues représente vraiment bien le problème survenant lorsqu'on évalue la scientificité par rapport à la publication scientifique, c'est-à-dire lorsqu'on se concentre sur le contenant plutôt que sur le contenu. Tout ceci à cause d'un lien supposé entre la qualité d'un article et le prestige d'une revue, lien qui n'est malheureusement que trop rarement vérifié et qui n'a absolument aucune base scientifique. En fait, plusieurs types d'évidences indiquent même que la fiabilité des résultats dans les revues prestigieuses n'est pas plus grande que la moyenne²⁵.

Pourtant, afin d'améliorer leur renommée, les instituts exigent souvent de leurs scientifiques qu'ils publient dans les revues les plus prestigieuses^{90,94,130,189}, c'est-à-dire celles dont le facteur d'impact^{Note 7} est le plus élevé tel que rapporté annuellement par le *Journal Citation Reports*²⁰⁷. Au Royaume-Uni par exemple, il semble même y avoir une règle non écrite dictant que pour être considéré pour un emploi attrayant, il faille avoir au moins une publication dans une revue avec un facteur d'impact supérieur à 5⁵⁴.

Pour les scientifiques, publier dans ces revues constitue donc souvent un impératif pour améliorer leur propre renommée et ainsi obtenir du financement et des promotions^{69,131,199}. Ainsi, dans leurs décisions d'où publier leurs travaux, ils perçoivent souvent que le facteur d'impact d'une revue a plus d'importance même que le champ d'étude visé par cette dernière¹⁹⁹. Pour

⁷ Le facteur d'impact d'une revue est calculé en divisant le nombre de citations obtenues par les articles publiés dans la revue, par le nombre d'articles publiés. Ce calcul s'effectue au sein d'une période de temps limitée, généralement les deux dernières années, ce qui permet au facteur d'impact d'évoluer avec le temps et de rester actuel.

arriver à publier dans de telles revues, plusieurs finissent donc par commettre toutes sortes d'inconduites¹¹², plus ou moins inconsciemment, ce que nous explorerons davantage dans les prochaines sections. Ce qui est certain, c'est que tout cela force des intérêts contradictoires sur les scientifiques, qui veulent d'un côté pouvoir utiliser et diffuser efficacement et librement leurs travaux, et de l'autre côté veulent bonifier leur renommée en publiant dans les revues les plus prestigieuses⁴⁰. Ces revues sont cependant souvent les plus chères à l'achat et celles qui imposent les plus fortes restrictions¹¹², rendant l'utilisation et la diffusion des travaux moins facile. Comme la publication dans ces revues est toutefois nécessaire pour faire carrière, c'est cet aspect qui prime en importance.

C'est ainsi qu'on entend souvent des choses comme : « C'est un article publié dans *Science*!^{Note 8} » pour faire entendre que c'est un article particulièrement important ou innovateur. Pourtant, l'utilisation du facteur d'impact des revues pour juger un article scientifique semble absurde puisque seulement 10 à 20% des articles publiés par une revue sont responsables de 80 à 90% de la valeur de son facteur d'impact¹³¹. En général donc, seul un relativement faible pourcentage d'articles publiés dans une revue prestigieuse ont un fort impact de part l'énorme nombre de citations qu'ils obtiennent. Les autres articles ne sont que peu ou pas cités. Le prestige de la revue dans laquelle un article est publié ne donne donc absolument aucune garantie sur la qualité de chaque article individuellement.

⁸ *Science* est une des revues scientifiques les plus connues, avec un des facteurs d'impact les plus élevés. D'autres exemples incluent : *Nature*, *The Lancet*, *Cell*, *Journal of the American Medical Association* (JAMA), ou *New England Journal of Medicine*.

Brembs et coauteurs²⁴ font une revue de littérature sur les effets d'un système basé sur le facteur d'impact des revues, normalement calculé comme suit pour chaque revue :

$$\text{Éq. 1} \rightarrow \frac{\text{nombre de citations des 2 dernières années}}{\text{nombre d'articles publiés}}$$

Ils exposent plusieurs tendances inquiétantes. Par exemple, le facteur d'impact d'une revue semble être un meilleur prédicteur du manque de fiabilité scientifique que de l'importance scientifique. De plus, ce système coûte cher et viole les standards scientifiques les plus élémentaires ; le calcul du facteur d'impact semble biaisé de plusieurs façons, notamment mathématiquement, mais certains biais ne peuvent même pas être correctement évalués puisque la méthode exacte n'est pas divulguée²⁰³. Un biais connu et consternant est que le facteur d'impact puisse être négocié entre les revues et l'entité faisant les calculs²⁴. Ceci se produit par exemple en décidant d'exclure des articles considérés comme « non aptes » à être cités (donc le dénominateur de l'éq. 1 diminue), mais en considérant quand même les citations issues de ces articles (le numérateur de l'éq. 1 augmente). Ceci permet donc de hausser le facteur d'impact sans avoir aucun lien avec l'impact réel d'une revue^{24,203}.

Malgré cela, certains professionnels, chercheurs ou praticiens, se concentrent sur les articles publiés dans les revues les plus prestigieuses, pensant que ce sont les plus importants. Cet état d'affaire est par exemple démontré dans le monde médical où certains médecins cliniques préfèrent lire uniquement les articles publiés dans des revues avec un haut facteur d'impact⁹⁷. Cependant, ces revues ne couvrent qu'une faible

minorité des articles qui pourraient leur être utile pour prendre des décisions et suivre des pratiques cliniques informées⁹⁷.

De plus, et bien que les études publiées dans ces revues prestigieuses soient généralement faites à plus grande échelle, lorsqu'il s'agit d'études à petite échelle, les résultats positifs – pour l'utilisation d'un médicament par exemple – sont plus grandement exagérés dans les revues prestigieuses par rapport à quand ce type d'études sont publiées dans des revues moins « importantes »¹⁸¹. Les revues prestigieuses peuvent donc avoir de plus grands impacts négatifs sur la santé des gens, d'autant plus qu'elles jouissent d'une visibilité accrue⁹⁷. Un focus démesuré sur le prestige des revues – c'est-à-dire sur le contenant présentant la science – représente donc bel et bien un comportement malsain pour la science comme pour la santé publique. D'autant plus qu'une recherche centrée sur la pratique et les patients est aussi rare et peu encouragée dans les grandes revues que dans les moins connues⁹⁷.

Une atmosphère de compétition insoutenable

Les obligations carriéristes pour obtenir du financement résultent en ce que la renommée et la compétition prennent de plus en plus d'espace dans l'esprit des scientifiques^{10,14,143,199}. Par exemple, plusieurs scientifiques sont convaincus que sans la renommée octroyée par la publication dans des revues prestigieuses, ils pourraient « découvrir le sens de la vie » sans que personne ne leur porte la moindre attention¹⁰.

Nous sommes là loin de l'image qu'on se fait de la science impartiale, mais plutôt dans un monde où la publication et la réputation (le contenant) prennent le dessus sur l'information (le contenu). Les scientifiques sont donc soumis à une compétition intense pour publier dans des revues prestigieuses^{10,112} et obtenir des promotions attrayantes ou du financement de qualité^{10,199}. De plus, les processus régissant cette compétition pour publier et être financé sont perçus comme subjectifs et facilement manipulables^{1,199}. Si l'on combine tout ça avec des chances de succès très faibles puisqu'il y a beaucoup plus de demandes que de ressources, on obtient un énorme gaspillage d'efforts et d'énergie, mais aussi un gaspillage de personnes hautement qualifiées^{10,80}. Ce trop-plein de scientifiques formés et la compétition intense qui en résulte créent une série de motivations perverses dans le système académique, rendant les postes universitaires de moins en moins attrayants et poussent de plus en plus de jeunes scientifiques (comme les moins jeunes) vers l'industrie⁵³. Le résultat est que la science et la connaissance deviennent de plus en plus difficile à conserver en tant que biens publics, tout cela au profit de gains privés⁵³.

Pourtant, il faut noter que cette atmosphère de compétition n'est pas perçue uniquement de façon négative au sein de la communauté scientifique. Au contraire, lorsque questionnés, plusieurs la considèrent même bénéfique et nécessaire pour l'avancement et le progrès de la science^{10,199}. Cependant, pour que ceci soit vrai, il faut que les critères de sélection soient constants, justes et objectifs, ce qui n'est évidemment pas toujours le cas. Nous croyons aussi que cette attitude est basée sur une fausse croyance selon laquelle la compétition est la meilleur motivatrice. De plus, cette attitude tend d'après nous à sous-estimer les effets néfastes de la

compétition, autant pour la science que pour la qualité de vie des scientifiques. En effet, un tel climat de compétition apporte avec lui plusieurs problèmes et désavantages.

Il y a premièrement une grande peur de se faire voler son travail et ses efforts par les compétiteurs¹⁰. Ceci se traduit par une aversion pour le partage complet des données et des méthodes ainsi que par toutes sortes de clauses de confidentialité. Nous verrons plus loin comment un manque de transparence et de partage peut nuire à la science puisque ces aspects sont nécessaires pour la réplication et la vérification de ce qui se fait. La compétition est en plus négativement corrélée avec l'esprit communautaire : une compétition agressive est notoirement connue comme nécessitant un focus particulièrement fort sur ses propres intérêts personnels¹⁰. Plusieurs scientifiques sentent donc qu'une ambiance de jalousie et de rancœur de la part des collègues règne autour de chaque succès¹⁹⁹. Cet impact de la compétition est bien souvent négligé, puisque ces effets ne sont pas encore bien étudiés, ni compris. Pourtant, ce phénomène semble conduire à la destruction des relations entre scientifiques travaillant dans le même domaine, détruisant ainsi des collaborations qui auraient pu être bénéfiques¹⁰.

D'ailleurs, lorsque questionnés sur l'augmentation observée du nombre d'auteurs par article, des scientifiques dans le domaine de la recherche biomédicale n'avaient rien de positif à dire sur le sujet, rien par rapport aux avantages du travail d'équipe¹⁹⁹. Les seuls commentaires concernaient plutôt la frustration et les difficultés supplémentaires venant avec plus de collaborateurs, incluant évidemment les conflits pour déterminer l'ordre des auteurs (facteur important pour la carrière de chacun). Notez que même s'il ne s'agit pas de la majorité des interactions

en science, il semble que chaque scientifique puisse raconter une anecdote où la compétition ait provoqué de tels conflits, plus ou moins graves pour les scientifiques ou pour la science qui se fait¹⁰.

De surcroît, lorsqu'Anderson et coautrices¹⁰ ont demandé à des scientifiques de décrire leur environnement de travail, beaucoup d'entre eux ont utilisé pour ce faire un vocabulaire guerrier, tel que « *The Art of Warfare* est un manuel très utile », « c'est un combat », ou « Il faut leurs [les compétiteurs] donner des coups bas ». Plusieurs d'entre eux décrivent aussi des collègues qui ont abandonné la recherche non pas parce qu'ils n'étaient pas compétitivement viables, mais parce qu'ils n'avaient pas envie de jouer ce jeu de guerre ; ils auraient simplement voulu faire de la science. Une corrélation a d'ailleurs été démontrée entre les signes d'épuisement professionnel et la pression pour publier ressentie par des professeurs du domaine médical¹⁹⁸. Un autre danger du focus démesuré sur une compétition quantitative malsaine est donc de créer une sélection naturelle dénaturée au sein des scientifiques, où les plus altruistes tendent à se retirer ou à être exclus, et où ceux qui sont le plus confortables avec ces motivations perverses excellent⁵³.

Il semble d'ailleurs que les jeunes scientifiques paraissent plus naïfs et idéalistes par rapport à comment l'activité scientifique fonctionne ou devrait fonctionner, les scientifiques plus âgés acceptant plus facilement les failles du système et étant plus réalistes, mais aussi plus cyniques¹⁹⁹. Évidemment, investir autant d'énergie dans un système compétitif encourageant toutes sortes d'inconduites et de coup bas assurera certainement cette transition de l'idéalisme vers le cynisme. Parallèlement, ceux qui veulent conserver leurs hauts idéaux doivent abandonner la partie, produisant une sélection pour ceux qui sont prêt à modifier

leur attitude ou leur comportement pour supporter le combat^{53,199}.

L'obligation de suivre les modes ou les vagues d'engouement en recherche pour obtenir du financement semble aussi être une conséquence de la compétition intense régissant l'activité scientifique¹⁰. Ceci peut se faire au détriment de questions intéressantes qu'il serait aussi important d'étudier. Par exemple, les scientifiques interviewés dans l'étude d'Anderson et coautrices¹⁰ parlent d'une situation où des scientifiques étudiant le cancer du sein se tournent vers la recherche sur le cancer de la prostate lorsque la mode le dicte, alors qu'ils sont en réalité moins bien équipés pour ce faire et que la recherche sur le cancer du sein reste importante. Cependant, ils savent (ou du moins suspectent) qu'ils n'obtiendraient pas de financement s'ils continuaient avec le cancer du sein. Ce genre de modes et leurs conséquences sont le résultat, entre autre, d'un climat de compétition trop à l'avant-plan dans la recherche.

Inconduites professionnelles

Cette très forte compétition pour publier dans les revues prestigieuses provoque nécessairement des effets indésirables sur le comportement professionnel et éthique des scientifiques^{80,199}. Les inconduites professionnelles, tant au niveau individuel qu'institutionnel, sont nombreuses et peu étudiées, car ceci coûterait très cher⁵³. Il existe pourtant plusieurs types d'inconduites et de conflits qui réduisent la qualité et la fiabilité de la publication scientifique^{10,182} (Fig. 4). Selon un sondage en Chine par exemple, un tiers des scientifiques répondants ont déjà

fraudé en plagiant ou en fabriquant des données¹⁵⁵. Il apparaît aussi que ceux qui perçoivent le plus de pression pour publier soient aussi ceux qui rapportent le plus avoir commis des inconduites¹²³ (les sondages étant anonymes).

Il existe aussi une tendance épidémique à manipuler ou à choisir uniquement ce qui convient le mieux dans la présentation des hypothèses, des données et des résultats, transformant en quelque sorte les résultats non-significatifs en résultats significatifs ou gonflant l'importance de ces derniers, tout ceci se faisant par des pratiques plus que douteuses (torture des données) et inacceptables en science^{6,18,58,59,94,120,143}. Les scientifiques sont en effet encouragés à choisir, consciemment ou non, les analyses qui leur donneront les résultats voulus, que ce soient des résultats significatifs leur permettant de publier, ou des résultats non-significatifs pour invalider une hypothèse indésirée⁷.

De plus, il est bien connu que les articles scientifiques présentent le déroulement de l'étude comme elle aurait dû se passer et non pas comme elle s'est réellement passée, omettant toute erreur ou cul-de-sac pouvant apporter un trop grand doute sur l'étude⁸⁰. Tout ceci à cause de la pression pour publier, d'un biais de publication contre les résultats non-significatifs et des instructions trop rigides de la part des revues^{7,70,199}, phénomènes complètement aberrants et hors de contrôle des scientifiques¹⁹⁹.

D'après d'autres sondages auprès des scientifiques, Fanelli⁵⁷ conclut que de façon générale et conservative, 2% des scientifiques ont déjà plagié ou fabriqué des données au moins une fois (14% admettent avoir observé ces comportements chez des collègues) et jusqu'à un tiers ont déjà commis des inconduites moins graves, mais qui endommagent tout de même l'activité

scientifique (jusqu'à 72% admettent avoir observé ces inconduites chez des collègues). Ces résultats sont corroborés par un sondage auprès des scientifiques en écologie et en évolution⁷⁰. Toujours selon des sondages, Martinson et coauteurs¹²⁵ ont montré qu'au moins 33% des répondants ont déjà commis des fautes professionnelles réprimandables par la loi.

Tout de même, les inconduites vraiment graves telles que la fraude restent un phénomène rare qui se produit presque toujours dans le domaine de la biomédecine⁸⁰. Toutefois, lorsqu'il s'agit de préserver l'intégrité de la science, il ne faut pas seulement établir des mesures pour solutionner le problème des fraudes, mais aussi pour toutes ces inconduites « moins graves » qui apportent elles aussi des faussetés dans la littérature scientifique¹²⁵. De plus, certains⁵⁴ considèrent avec raison qu'il n'y a pas grande différence entre la fraude et, par exemple, l'usage continu de mauvaises statistiques une fois le problème connu...ce qui est aujourd'hui le cas, au moins au niveau institutionnel.

En effet, bien que la science soit conceptuellement auto-correctrice, dépistant et rejetant éventuellement ces faussetés, un trop grand nombre de faux résultats dans la littérature viendrait mettre fin à ce caractère auto-correctif⁸⁰. D'autant plus que les motivations perverses régissant pratiquement tout le système poussent beaucoup de scientifiques à prétendre que les inconduites décrites ici n'existent pas ou n'ont pas d'effets importants, réduisant davantage le potentiel auto-correctif de la science⁵³. Malheureusement selon la tendance actuelle, la compétition, les fraudes et les autres inconduites deviendront probablement plus importantes dans le futur de la recherche⁸⁰.

Toutes ces inconduites peuvent être en effet retracées jusqu'au climat de compétition omniprésent, lui-même accentué par les universités, les organismes de subventions et les éditeurs de revues qui aiguillent tous les scientifiques dans des directions différentes et plus ou moins déconnectées de la science¹⁴¹. En conséquence, plus un sujet d'étude est populaire, plus les résultats publiés sont susceptibles d'être faux puisque le besoin d'être le premier devient plus important que celui de faire les vérifications nécessaires et d'avoir raison^{10,18,92}. Les instituts de recherche semblent contribuer à cette tendance, plusieurs scientifiques croyant que l'objectif n'y est plus l'intégrité scientifique, mais bien la grosseur des projets et leur potentiel pour mettre les instituts « sur la carte »¹⁰. Les revues sont aussi responsables du comportement des scientifiques, exigeant généralement des articles novateurs et à résultats significatifs⁵⁴. Certaines revues telles que PLoS ONE tentent de soutenir de meilleures pratiques de recherche, par exemple en ne requérant ni de nouveauté ni de résultats significatifs, mais ce genre de revues sont rares, particulièrement parmi les plus prestigieuses.

Les instituts de recherche scientifique seraient encore les mieux placés pour régler la situation, mais ils choisissent plutôt d'en profiter¹⁸. Les politiques institutionnelles promeuvent en effet la compétition non seulement en accentuant l'importance de la quantité d'articles publiés dans des revues prestigieuses, mais aussi en formant beaucoup plus de scientifiques qu'il n'y a d'emplois attractifs^{4,6,10,53,182}. Cet excédent de scientifiques performants et bien éduqués, en compétition féroce pour des ressources limitées (publications des revues, postes, promotions, financement)^{10,80}, assurent une main-d'œuvre bon marché et une productivité maximale pour les laboratoires de recherche, qui ressentent eux aussi la pression de publier et

profitent de la situation en abusant de l'idéalisme encore présent chez les jeunes gradués¹⁰. Telle une structure pyramidale, ces jeunes scientifiques travaillent donc au profit d'un petit nombre de scientifiques situés en haut de la hiérarchie¹⁰. Ce genre de structure où on récompense de façon disproportionnée ceux qui ne sont que marginalement plus performants¹⁰ encourage et propage les inconduites et la publication d'une science de moins bonne qualité⁵³.

Comme le faisait déjà remarquer Altman⁶ en 1994, tous ces problèmes sont bien connus et l'inaction générale de la communauté scientifique à leur égard est troublante. À l'heure actuelle, peu d'indices laissent croire que la situation ait grandement changée depuis 1994. Il est d'ailleurs intéressant d'observer que plusieurs scientifiques ressentent non seulement la pression de présenter des résultats significatifs de façon brève et sexy, mais qu'ils perçoivent eux mêmes un rapport complet et détaillé comme étant moins attractif puisqu'un tel rapport serait presque assurément plus désordonné et compliqué⁷⁰ et que personne ne désire passer beaucoup de temps à lire des articles⁵⁴. Cet état d'esprit, plus ou moins bien justifié dépendamment des pratiques⁷⁰, facilite certainement la propagation de ces inconduites moins graves que la fraude, puisqu'elles paraissent acceptables ou même désirables.

Ainsi, malgré ses nombreux effets indésirables, il semble que ce climat de compétition en science soit impossible à arrêter, qu'il puisse être vu comme positif ou même nécessaire¹⁰, et que plusieurs de ses conséquences non-scientifiques soient perçues comme étant bénéfiques⁷⁰. Pourtant, il reste important que nous nous demandions si ces perceptions sont le résultats d'une compréhension profonde du système, ou bien celui de l'habitude

qui tend à faire paraître normales ou indispensables des pratiques qui sont en réalité néfastes. Anderson et coautrices¹⁰, suite à l'analyse de leurs interviews auprès des scientifiques par rapport aux effets pervers de la compétition en recherche, apportent un questionnement similaire :

« Nous suggérons cependant que les prémisses fondamentales d'un système duquel le régime de récompenses, qu'il soit intentionnel ou pas, mène à des choix naturels et rationnels de la part des individus au sein dudit système, choix qui ensuite compromettent l'intégrité même du système, devraient être soumises à un examen minutieux. » ^{Note}

9

Ces autrices soulignent donc l'importance de scruter avec soin les raisons pour lesquelles nous conservons certaines pratiques ainsi que les conséquences qui s'en suivent. Elles nous offrent aussi le transcrit de leurs interviews, une discussion que chacun qui est intéressé à mieux comprendre ce que les scientifiques perçoivent et vivent dans leur vie professionnelle devrait lire.

⁹ Traduction libre de la part des auteurs. Citation originale : *“We suggest, however, that the fundamental premises of a system whose incentives, intentional or not, lead to natural and rational choices by individuals in the system that then compromise the integrity of the system must be subjected to intense scrutiny.”*

L'invention de faits à travers les citations

La forte compétition entre scientifiques, le trop grand nombre de scientifiques formés et la quantité d'articles qui s'en suit sont trois facteurs qui jouent un rôle important dans la fabrication (intentionnelle ou pas) de faits à travers la citation. Ceci est un phénomène assez peu discuté ou étudié, mais qui peut avoir des répercussions vraiment indésirables en science. En effet, comme la quantité d'articles à lire dans un domaine donné augmente sans arrêt et de façon excessive, toute revue de littérature devient extrêmement chronovore.

Les scientifiques, désirant publier le plus d'articles possibles, le plus rapidement possible, n'ont pas le temps aujourd'hui de faire une revue de littérature complète et de lire dans leur entièreté tous les articles qui seraient importants pour leurs travaux. De plus, la nature payante de beaucoup de littérature scientifique limite l'accessibilité de certains ouvrages. Le résultat est que beaucoup d'articles qui sont cités dans les travaux scientifiques n'ont pas été lus au complet par les auteurs, le résumé de l'article étant parfois la seule section lue avec attention. Combinée avec le besoin ou le désir d'établir ses idées et ses hypothèses à tout prix au sein de cette atmosphère de compétition, la façon dont les articles sont cités dans la littérature peut mener à plusieurs types de confusions plus ou moins consciemment appliquées.

Par exemple, à travers la citation d'articles, certaines affirmations, vraies ou fausses, peuvent finir par être acceptées et considérées comme des faits seulement parce qu'elles sont reprises et citées à répétition dans la littérature, sans être supportées par des données et, après un temps, sans même

référer à l'article original contenant les données^{81,194}. C'est ce qu'on appelle l'amplification, phénomène provoquant l'expansion d'une croyance à travers les citations seulement, sans données scientifiques⁸¹. L'affirmation, vraie ou fausse, se répand donc dans la littérature comme un fait alors qu'il faudrait évaluer la qualité de l'article original pour savoir ce qui en est, ce qui n'arrive malheureusement pas toujours. C'est ainsi que la structure des citations dans la littérature peut grandement influencer les croyances scientifiques, même sans avoir de réel support scientifique⁸¹.

La transmutation par citation est un phénomène similaire à la simple amplification, sauf qu'ici l'affirmation de base n'est qu'une hypothèse encore non-testée, reprise et citée en tant que fait par d'autres auteurs, subissant donc une amplification, mais n'ayant aucune donnée empirique supportant ladite hypothèse⁸¹. Cette dernière en vient pourtant peu à peu à être acceptée comme un fait.

Des problèmes de citations influençant les croyances scientifiques sans base scientifique peuvent prendre plusieurs autres formes (voir Greenberg⁸¹ pour une revue sur le sujet). Il y a par exemple la « citation en cul-de-sac » ou la diversion par citation, où les auteurs citent un ouvrage pour supporter une affirmation, mais où le contenu de l'ouvrage cité ne traite pas du sujet mentionné, ou encore il en traite, mais offre une toute autre interprétation. Ceci peut se faire par accident ou incompréhension, mais aussi dans le but d'inventer un fait ou de détourner l'attention des lecteurs de travaux importants. C'est aussi une technique appréciée par les faux sceptiques qui veulent discréditer un ouvrage ou un ensemble de travaux ou de personnes. Cette technique à d'ailleurs déjà été utilisée sur une

version antérieure du livre que vous lisez en ce moment même⁴⁶, quoiqu'il reste difficile de dire si, dans ce cas précis, l'incident s'est produit par malveillance ou par incompetence.

Un autre phénomène est l'invention de fait au travers du titre ou du résumé d'un article. Ceci se produit lorsque les auteurs d'un article incluent dans le titre ou le résumé, consciemment ou non, une affirmation qui n'est pas démontrée par le corps de l'article⁸¹. Étant donné qu'il arrive souvent que les scientifiques ne lisent que le titre et le résumé d'un article qu'ils citent, ces affirmations sont reprises et citées comme des faits alors qu'elles n'ont jamais été démontrées. Les raisons pour lesquelles un article peut être cité sans vérification adéquate sont multiples et incluent la perception de manquer de temps, ou encore tout simplement parce que l'affirmation citée cadre bien avec les préconceptions de l'auteur, alors il n'y pense pas à deux fois.

Greenberg⁸¹ a développé un modèle d'analyses statistiques (analyse de pistes) pour détecter ce genre de distortion des faits par la citation dans la littérature scientifique. Bien que son efficacité reste encore à déterminer, cet outil est certainement très intéressant à considérer dans l'objectif de séparer le bon grain de l'ivraie au sein de la masse des « faits » scientifiques.

Nous voyons donc de plus en plus clairement que la compétition, la pression pour publier et l'importance du facteur d'impact des revues ont plusieurs répercussions sur le développement de la carrière des scientifiques, sur la science et sur le public. Nous voyons aussi beaucoup de négativité, de frustration, de ressentiment et de cynisme par rapport à la culture de publication scientifique¹⁹⁹. Nous n'avons cependant toujours

pas discuté en profondeur d'un type de conséquences qui revêt une importance toute particulière puisque ses effets – nombreux, néfastes et insidieux – peuvent longtemps passer inaperçus, mais peuvent aussi se répercuter directement et sévèrement sur la science qui est reconnue et valorisée. L'existence de ce type de conséquences provient encore une fois de la façon dont la science est communiquée et dont les scientifiques sont évalués, sans rapport avec la scientificité. Il s'agit des biais de publication (Fig. 4).

Biais contre les résultats non-significatifs

Il existe de nombreux types de biais de publication, mais le plus souvent discuté et peut-être le plus fréquent est la tendance à ne pas publier les résultats non-significatifs, c'est-à-dire qui ne corroborent pas l'hypothèse du scientifique^{60,103,149}. Par exemple, si l'hypothèse est que l'utilisation d'un fertilisant augmentera la productivité du potager, la probabilité que les résultats soient publiés sera beaucoup plus grande s'ils démontrent une différence de productivité entre le potager fertilisé et celui non-fertilisé, que si les deux produisent de façon semblable. Dans le second cas, les résultats pourront être considérés moins intéressants, difficiles à expliquer, ou difficile à publier. C'est ainsi que se développe un biais de publication en faveur des résultats positifs.

Ce critère n'est cependant pas en lien avec la rigueur scientifique, mais bien avec l'attrait des résultats. Pourtant, ce biais de publication est observable dans presque tous les domaines de recherche^{58,60,94,120,132,149,172,185,191,196}. En conséquence, beaucoup de

données scientifiques (jusqu'à 60% !) tombent dans les oubliettes^{50,83}, tout simplement parce qu'ils sont jugés inintéressants et donc non publiables.

Il est important de préciser ici que le qualificatif « inintéressant » ne se réfère pas à l'utilité pour l'avancement de la science ni même pour le public, mais bien au sex-appeal du résultat, souvent considéré comme plus important que la qualité de la recherche lorsque vient le temps de publier¹⁹⁹. En d'autres mots, un résultat non-significatif ne se vend pas aussi bien et ne sera pas cité aussi souvent qu'un résultat significatif^{49,158}, malgré qu'il puisse être tout aussi important pour la science. Il est par exemple important de savoir si l'utilisation d'un médicament apporte un effet ou non, et ne publier que les études qui ont démontré un effet crée un biais important pour la pratique médicale. Ce biais de publication est donc un exemple classique où le contenant devient plus important que le contenu, l'habit plus important que le moine, et ceci se passe au détriment de la science.

Un exemple qui illustre très bien ce point nous est fourni par Ioannidis⁹³ avec une revue des données relatifs aux effets de différents anti-dépresseurs. La moitié des essais cliniques considérés démontraient des effets positifs de ces médicaments, tous sauf un étant publiés. L'autre moitié démontraient des effets non-significatifs. Deux tiers des essais à résultats non-significatifs n'ont pas été publiés, alors que l'autre tiers a été publié, mais de façon à ce que les résultats semblent à première vue significatifs. De plus, ce genre d'essai clinique est généralement planifié pour détecter le plus petit effet possible, donc cliniquement insignifiant, c'est-à-dire que pour la vie des patients, le médicament ne change pratiquement rien même si les résultats

sont statistiquement significatifs. Ainsi donc, il semble que l'approbation d'un anti-dépresseur spécifique se fasse par l'atteinte d'un critère de signification statistique pré-déterminé au sein d'un groupe restreint, sans vraiment tenir compte des bienfaits réels et pratiques pour les patients en général⁹³ – ceci ne voulant pas dire qu'aucun anti-dépresseur n'ait de bienfaits réels. Ces essais cliniques ne représentent donc pas tellement une activité scientifique, mais bien une entreprise planifiée pour démontrer un effet à des fins légales⁹³. En d'autres mots, nous avons là un texte décrivant une activité qui semble scientifique, utilisant un jargon scientifique et faisant croire à de la science de qualité... on se rapproche dangereusement de la définition qu'on donne souvent à la pseudoscience. Malgré tout cela, les anti-dépresseurs (et pas seulement ceux qui fonctionnent) sont parmi les médicaments les plus prescrits du monde, leurs effets secondaires néfastes à longs termes étant beaucoup trop peu étudiés⁹³. Tout ceci à cause du sceau de confiance démesurée qu'apporte la publication scientifique, parce qu'on confond « science » avec « publication », parce qu'on préfère regarder l'habit plutôt que la personne qui veut se faire passer pour un moine.

Pour remédier à la situation du biais de publication contre les résultats non-significatifs, plusieurs revues dédiées exclusivement à la publication de tels résultats ont vu le jour⁷⁸. Ces revues ont cependant toutes des facteurs d'impact très bas⁸³ ne se retrouvant même pas dans le plus récent rapport de *Clarivate Analytics – Web of Science*²⁰⁷, une des entités se chargeant de calculer et de classer les facteurs d'impact des différentes revues. Elles ne sont donc pas particulièrement attrayantes pour les scientifiques et leur succès reste à déterminer. Plusieurs d'entre elles – *New Negatives in Plant Science* ou *Journal of Negative*

Results in Biomedicine, par exemple – ont d'ailleurs déjà fermé leurs portes.

Certains prétendent que la cause principale du biais de publication contre les résultats non-significatifs est que les scientifiques décident de ne tout simplement pas prendre le temps d'écrire sur ces résultats^{47,132,145,185}. Que les scientifiques n'allouent pas de temps pour écrire des articles sur ces résultats est certainement vrai, mais il ne s'agit assurément pas de la cause principale du phénomène. Étant donné leurs ressources limitées en temps et en argent, il est d'ailleurs tout à fait compréhensible que ce ne soit pas leur priorité s'ils s'attendent à ce que l'article soit rejeté ou difficile à publier^{59,185,196,199}, surtout s'ils considèrent les résultats significatifs nécessaires pour obtenir du financement¹⁹⁹. De plus, il ne faut pas oublier qu'en science, les auteurs, les réviseurs et les éditeurs font tous partie du même groupe de personnes et donc ce que les auteurs croient représente nécessairement, au moins en partie, la réalité.

Les éditeurs semblent effectivement moins susceptibles de publier des articles à résultats non-significatifs, perçus comme moins intéressants^{83,132,149,173,196}, allant même jusqu'à les décourager par leurs politiques¹⁸⁵. Pour eux, la survie financière du journal et la compétition des citations²⁰⁷ sont souvent les principaux soucis¹⁰³. En fait, il apparaît plus juste de considérer que ce biais de publication est le résultat de l'interaction entre auteurs, réviseurs et éditeurs, tous à court de temps et souvent aussi d'expertise^{112,189}. De façon générale et consciente, les exigences en termes de rigueur scientifique sont donc plus élevées pour les articles présentant des résultats non-significatifs¹⁷³, bien que la nécessité de publier ces résultats ne soit pas nécessairement moindre¹⁴⁰.

Biais contre les résultats non-significatifs : la valeur-p et son influence

Dans cette section sur le biais de publication contre les résultats non-significatifs, nous avons défini les résultats significatifs comme étant des résultats corroborant l'hypothèse initiale des scientifiques. Mais comment fait-on pour déterminer si l'hypothèse est corroborée ou non, si un résultat est significatif ou non? Cette question paraît triviale pour beaucoup de personnes et, par conséquent, elle est généralement traitée de façon trop simpliste. En effet, le constat de « résultats significatifs » est plus complexe que ce qui est généralement assumé par le public ou même par les scientifiques. Malgré la nature très technique du sujet, nous trouvons donc très important de l'explorer en détails pour bien comprendre à quel point le biais de publication contre les résultats non-significatifs peut être problématique et ainsi démontrer encore une fois à quel point la publication scientifique peut être affectée par d'autres facteurs que la scientificité des travaux publiés.

La question de la significativité des résultats concerne la plupart du temps la valeur-p, utilisée en statistique pour exprimer la probabilité d'observer les résultats si l'hypothèse nulle est vraie (Boîte 2). On appelle cette procédure un « test de significativité de l'hypothèse nulle » et cette méthode d'analyse statistique est de loin la plus commune à travers toutes les sciences^{87,194}. De plus, il faut savoir que la plupart des scientifiques sont exclusivement éduqués avec cette méthode et tendent parallèlement à l'interpréter de façon erronée¹⁹⁴. Brièvement, la valeur-p est obtenue à l'aide de tests statistiques et est comparée à un seuil de significativité arbitrairement choisi, communément fixé à 0,05. Si la valeur-p est égale ou inférieure au seuil de 0,05, on dit

généralement qu'un effet a été détecté et les résultats sont considérés significatifs. Si la valeur-p est supérieure au seuil de 0,05, on les considère non-significatifs. Voyons un exemple pour mieux comprendre ce que tout ceci signifie.

Supposons que nous étudions l'effet d'un médicament contre les maux de tête et que nous administrons le médicament à un groupe de personnes, alors qu'un autre groupe ne reçoit qu'un placebo (une pillule de sucre). Supposons aussi que dans le groupe ayant reçu le médicament, 60% des patients aient ressenti une amélioration de leur condition contre 40% dans le groupe du placebo. Il faut alors analyser statistiquement ces résultats pour savoir s'ils sont significatifs ou non, c'est-à-dire si le médicament à bel et bien eu un effet positif plus fort que celui du placebo. Nous calculerons donc la valeur-p en utilisant un test statistique approprié. La valeur-p dépendra non seulement de la différence entre les deux groupes, mais aussi du nombre de personnes dans chaque groupe et de la variabilité intra-groupe^{Note 10}.

Supposons que nous obtenons une valeur-p de 0,05. On considère généralement que ceci signifie que même si le médicament n'apporte aucun bienfait, nous devrions observer ces mêmes résultats ou des différences encore plus grande entre les groupes dans 5% des études si nous reproduisons l'étude originale un grand nombre de fois. Comme la valeur-p est égale ou inférieure au seuil de significativité fixé par convention, les résultats sont considérés significatifs, c'est-à-dire que nous

¹⁰ Notez bien que plusieurs autres facteurs et considérations entrent en ligne de compte lors d'une analyse statistique et de la détermination de la valeur-p (Boîte 2). En expliquer chaque détail n'est pas l'objectif ici, mais bien offrir une idée générale de l'utilisation de la valeur-p.

pouvons considérer qu'un effet réel a été détecté. Effectivement, comme nous avons des ressources limitées et que nous ne pouvons pas reproduire une étude infiniment, ce seuil de 5% est généralement considéré comme un compromis satisfaisant entre les faux-positifs (détection d'un effet qui n'existe pas) et les faux-négatifs (ne pas détecter un effet qui en réalité existe). Notez quand même que pour accepter un médicament sur le marché cependant, une seule étude ne suffit pas et plusieurs stades de recherche avec plusieurs études doivent être conduits, même si au final il n'y ait pas toujours autant de réplifications qu'on pourrait le souhaiter.

L'utilisation de la valeur-p semble à première vue logique et bien réfléchie. Toutefois, plusieurs problèmes sont associés avec un focus démesuré sur une valeur-p et un seuil de significativité fixé par convention (Boîte 2). Bien comprendre ces écueils n'est pas une tâche facile, même pour les scientifiques^{16,86,194}, mais cette compréhension nous paraît essentielle pour séparer convenablement ce qui appartient aux biais de publication de ce qui appartient à la démarche scientifique et à une activité scientifique en santé.

Un premier problème vient du fait qu'il est maintenant pratique courante en science de considérer des résultats significatifs – c'est-à-dire avec une valeur-p inférieure à un seuil de significativité arbitrairement choisi – plus précieux que des résultats non-significatifs⁸⁷. Ceci mène au biais de publication dont nous parlons dans cette section et nous verrons pourquoi il est insensé que la valeur-p ait une influence si forte sur ce qui se fait publier (voir aussi Boîte 2). Un exemple de cela est que la valeur-p est souvent érronément interprétée comme étant proportionnelle à la force de la relation entre deux variables.

Boîte 2. Le Test de Significativité de l'Hypothèse Nulle (TSHN) et la valeur-p

Les TSHN ont été popularisés par Fisher⁶⁸ et Neyman et Pearson¹³⁸. Ils ont été développés à des fins d'indication, pour faciliter des décisions et, surtout, avec la compréhension que leur utilité nécessite un grand nombre de réplifications¹⁹⁴. **Aucun verdict concret n'est donc permis par un seul ou peu de TSHN.** Fisher notait d'ailleurs qu'on ne devrait jamais dire qu'un fait scientifique soit empiriquement prouvé, à moins qu'une expérience appropriée n'échoue que rarement à arriver au niveau de signification¹⁰⁸. Pour ça par contre, nous avons besoin de réplifications! Nous explorerons ici quelques-unes de ces questions, mais pour les lecteurs qui veulent creuser plus loin, nous conseillons de lire Balluerka et coautrices¹⁶ (une excellente revue sur le sujet), Szucs et Ioannidis¹⁹⁴ (une critique impressionnante des TSHN) et Hagen⁸⁶ (une brillante réponse aux critiques les plus communes).

Hypothèse nulle (H_0) et hypothèse alternative (H_1)

H_0 stipule que la différence entre les données récoltées au sein de deux (ou plusieurs) groupes est due à de la variabilité aléatoire entre les échantillons et que les groupes proviennent en fait d'une même population. Elle est normalement comparée à H_1 , stipulant qu'il y a une vraie différence entre les groupes à l'étude. Un exemple de H_1 serait : « Si nous ajoutons un purin d'ortie (riche en azote) à nos choux, il y aura une production de la biomasse des feuilles plus élevée. » H_0 quant à elle stipule qu'il n'y aura pas d'effet de notre traitement au purin comparé à des choux qui n'en reçoivent pas. La comparaison s'effectue avec un TSHN où nous calculons la probabilité d'observer

la différence mesurée ou une différence encore plus grande entre les groupes de choux si H_0 est vraie. C'est cette probabilité qu'on appelle la valeur-p.

Valeur-p, seuil de significativité et paramètres négligés

L'objectif d'un TSHN est de vérifier s'il est plausible de rejeter H_0 , donc de rejeter l'idée que les différents groupes proviennent d'une même population. En d'autres mots, nous comparons la valeur-p obtenue par notre test avec un seuil de significativité, qui est lui-même dans la très grande majorité des cas arbitrairement fixé à 0,05. Si, pour revenir sur notre exemple, nos choux traités au purin ont produit en moyenne 15% plus de biomasse, une valeur-p de 0,05 signifierait qu'il y ait 5% de probabilité d'obtenir une telle différence – ou une différence encore plus grande – entre les deux groupes même si le purin n'a aucun effet. L'hypothèse nulle serait donc tout de même rejetée et le purin serait considéré comme effectif.

Le problème avec un seuil de significativité qui est fixé à 0,05 sans autres considérations est que ce seuil devrait en réalité être déterminé par notre compréhension du système à l'étude¹⁹⁴. La détermination informée d'un seuil de significativité se fait avec plusieurs autres paramètres importants. Entre autres, la probabilité de commettre une erreur de type I (α) – c'est-à-dire d'avoir un faux-positif, par exemple démontrer une différence entre nos deux groupes de choux sans qu'il y en est une – est généralement fixée à 0,05 et c'est la base pour y fixer aussi notre seuil de significativité.

Cependant, ceci ne représente pas la vraie probabilité d'obtenir des faux-positifs à longs termes, qui elle peut être beaucoup plus élevée¹⁹⁴ et dépend i) de la taille de l'effet

qu'on cherche (dans notre exemple, 15% plus de biomasse), ii) de la puissance statistique de notre test qui elle-même dépend de la taille de l'effet recherché et de la probabilité de commettre une erreur de type II (β), c'est-à-dire produire un faux-négatif, iii) de la taille de l'échantillon qu'on utilisera dans nos analyses, ce qui devrait dépendre de la puissance statistique et du seuil de significativité choisi et iv) des probabilités précédant l'expérience que H_0 et H_1 soient vraies, respectivement.

Comme vous pouvez le voir, décréter si un résultat est « significatif » ou non, avec toute la portée qui vient aujourd'hui avec une telle affirmation, n'est pas une simple affaire de valeur-p fixe. La réalité générale en recherche est premièrement que la taille de l'effet qu'on cherche à démontrer est rarement connue et difficile à estimer avant l'expérience. Il peut être par exemple difficile d'estimer précisément le surplus de croissance de nos choux par l'application de purin d'ortie. Même la taille d'intérêt minimale est rarement évaluée : par exemple, à partir de quel pourcentage de surplus de croissance serait-il intéressant d'appliquer du purin? Dans ce cas-ci, il est bien possible de faire le calcul, par exemple en calculant les coûts de produire le purin et de l'appliquer et en les comparant avec la marge de profit supplémentaire minimale que l'on souhaite d'un surplus de production. Toutefois, ce genre de calcul n'est que rarement fait pour améliorer la qualité des analyses statistiques : on veut souvent seulement démontrer un effet statistiquement significatif. Cela sans compter le fait que la plupart des scientifiques ne comprennent pas exactement le fonctionnement des analyses statistiques utilisées¹⁹⁴.

Deuxièmement, la puissance statistique est elle aussi rarement connue ou évaluée (quoique ceci dépende des domaines) puisqu'on ne peut souvent pas évaluer exactement la probabilité β d'obtenir des faux-négatifs sans connaître la taille d'effet qu'on recherche ou dont on aurait besoin pour que les différences observées soient d'intérêt pour nos recherches. Ceci apporte toutes sortes de difficultés pour la crédibilité des analyses produites en science puisque dans les TSHN, la puissance statistique est largement négligée pour se concentrer uniquement sur la valeur-p et sur α ¹⁹⁴. La taille de l'échantillonnage nécessaire pour tester les hypothèses est donc souvent inconnue, ce qui pousse des scientifiques à tout simplement récolter des données jusqu'à atteindre une valeur-p de 0,05, peu importe si cet objectif les informe ou non sur le système à l'étude, pensant qu'un résultat significatif vient annuler l'importance de la puissance statistique^{115,194}.

Finalement, avoir une bonne idée des probabilités que H_0 et H_1 soient vraies est souvent très difficile à obtenir et nécessiterait la publication de tous les données ayant traité la question antérieurement, que les résultats aient été significatifs ou non¹⁹⁴. Cependant, comme nous l'avons vu dans cette section, le biais de publication contre les résultats non-significatifs ainsi que la manipulation *post-hoc* de la présentation des résultats rendent cette évaluation difficile, voire impossible. Cette évaluation est d'ailleurs négligée pour les TSHN, contrairement aux analyses Bayésiennes qui la nécessitent absolument.

La non-spécification de l'hypothèse alternative (H_1)

Non seulement les probabilités pré-expérience de H_0 et de H_1 ne sont généralement pas considérées dans les TSHN,

mais en réalité les TSHN ne concernent que H_0 . Ceci signifie que tout ce que permet un TSHN est de rejeter H_0 , sans rien dire de précis par rapport à une quelconque H_1 ¹⁹⁴. Ceci est particulièrement problématique si H_1 reste très vague, avec aucune précision de la taille d'effet attendue, ce qui représente la majorité des cas. C'est problématique puisque le TSHN ne permet pas de différencier entre plusieurs H_1 plausibles, mais, généralement, les scientifiques considèrent à tort que la réjection de H_0 donne du support à leur H_1 préférée sans avoir besoin de la comparer avec d'autres H_1 , ni même besoin de vérifier si leur H_1 explique mieux les données récoltées que H_0 ¹⁹⁴. Pourtant, les TSHN n'offrent absolument aucune base scientifique pour ce faire¹⁹⁴.

Ceci découle premièrement de la mésinterprétation très commune de ce que représente un TSHN, mais aussi de comment les résultats donnés par les programmes statistiques sont représentés : on nous montre l'effet de la variable d'intérêt et une valeur-p associée, mais la valeur-p ne dit rien par rapport à cet effet ou par rapport à une H_1 particulière, elle ne concerne que H_0 . Tout ceci n'est cependant pas représenté de façon claire, tout est plus ou moins « caché » dans la « boîte noire » de la fonction utilisée dans le programme. Ainsi, sans une compréhension profonde du test qu'on utilise, il devient très facile d'interpréter de façon erronée les résultats présentés.

Utiliser la valeur-p

Il est important de préciser que les TSHN ne sont pas un problème en soi et peuvent être très utiles ; le problème est la façon plus ou moins mal informée avec laquelle ces tests sont majoritairement utilisés.

Pour améliorer la situation, certains ont proposé de tout simplement réduire le seuil de significativité à 0,005 pour justifier la publication de nouvelles découvertes²⁰. Cependant, il a été démontré que cette simple stratégie ne peut pas solutionner (ou pourrait même aggraver) les problèmes liés à l'incompréhension générale des TSHN et aux biais de publication^{108,201}. Utiliser des méthodes alternatives et complémentaires aux TSHN telles que les statistiques Bayésiennes et les intervalles de compatibilité devrait nous aider à obtenir des résultats plus fiables²⁰¹, tout en gardant à l'esprit qu'aucune mesure individuelle ne devrait représenter le critère principal avec lequel les résultats scientifiques sont jugés²⁰¹. En effet, d'autres méthodes comme les statistiques Bayésiennes ne sont pas immunes aux mêmes biais de publication que les TSHN¹⁸⁰, ces biais découlant simplement de la nécessité de présenter des effets forts et « significatifs » pour publier⁸⁷.

Une autre façon de valoriser les TSHN est de produire des courbes de distribution des valeur-p calculées par toutes les études faites sur une question précise^{87,194}. Ce type de méta-analyse semble être l'unique façon de vérifier de manière fiable, avec les valeurs-p, si l'effet qu'on cherche est vraiment significatif ou non¹⁹⁴, ou pour reprendre la définition de Fisher, si la démonstration de l'effet n'échoue que rarement à donner des résultats significatifs. Ces représentations graphiques permettent aussi de vérifier la présence de *p-hacking*. Par exemple, un nombre disproportionné d'études présentant une valeur-p qui est tout juste significative indique que la récolte, ou du moins la présentation des données a été manipulée dans ce but précis^{87,115}.

Dans ces conditions, certains scientifiques plaident carrément pour l'abandon des TSHN comme fer-de-lance de l'activité scientifique¹⁹⁴, ou même pour l'abandon de la terminologie « statistiquement significatif »^{7,108}. Dans tous les cas, une utilisation continue et utile des TSHN semble nécessiter d'accomplir un grand travail de répliques ainsi que la publication de tous les résultats, significatifs ou non¹⁹⁴ – deux critères contre lesquels les biais de publication sont les plus évidents.

Aussi, au lieu d'avoir un seuil de significativité fixe et généralement accepté, il serait beaucoup plus utile que les scientifiques justifient à l'aide de la théorie et de la logique chaque décision prise lors des analyses, incluant leur seuil de significativité¹⁰⁸. En effet, ce dernier devrait varier dépendamment des paramètres dont nous avons parlé ici – qui devraient idéalement aussi être spécifiés avec justification – mais aussi selon les disciplines ou le type de recherche^{108,194} (exploratoire ou descriptive). En fait, dans le meilleur des mondes, il faudrait évaluer les coûts associés à obtenir un faux-positif, un faux-négatif, un vrai-positif et un vrai-négatif et ainsi déterminer le seuil de significativité et la puissance statistique avec lesquels on veut opérer¹⁹⁴. Tout ceci devrait d'ailleurs se produire avant la collecte des données et être pré-enregistré pour éviter la manipulation *post-hoc* des données et des résultats^{108,194}. Bref, sans ces précautions, une valeur-p ne nous dit pratiquement rien sur la probabilité que nos résultats soient vrais ou faux¹⁹⁴.

Cependant, même si une valeur-p est très faible (par exemple 0,0000001) indiquant des résultats « très significatifs », cela ne signifie tout de même pas que l'effet détecté soit très fort, ni même précis⁸⁷, car la valeur-p dépend d'autres paramètres tels

que la taille de l'échantillon. Il est donc possible que ces résultats hautement significatifs, souvent perçus comme particulièrement intéressants, soient en pratique quasi insignifiants¹⁹⁴ (Boîte 2).

Pourtant, obtenir une valeur-p « significative » est maintenant devenu un objectif principal dans la plupart des domaines scientifiques, sans quoi publier devient difficile ou même improbable lorsqu'on s'adresse à des revues prestigieuses. Selon Szucs et Ioannidis¹⁹⁴, la facilité avec laquelle les tests de significativité de l'hypothèse nulle peuvent donner des résultats significatifs (et donc publiables) – peu importe la valeur pratique de ces résultats – est d'ailleurs l'une des raisons pour laquelle ces tests dominent la recherche biomédicale.

Par conséquent, on observe dans la littérature scientifique une hausse de résultats présentés comme étant tout juste significatifs avec une valeur-p de 0,05 exactement^{115,172}, souvent arrondie vers le bas plutôt que vers le haut – c'est-à-dire que 0,051 ou même 0,056 devient 0,05 dans la publication¹¹⁵. Ceci dénote une obligation de présenter des résultats significatifs pour publier, mais aussi le fait que les scientifiques peuvent aujourd'hui facilement suivre leurs données et les analyser à mesure qu'ils les ramassent, s'arrêtant lorsque les résultats sont finalement significatifs¹¹⁵. Ce comportement va à l'encontre de la démarche scientifique, mais semble être sélectionné comme gagnant dans le monde de la publication scientifique. Comme le fait remarquer Poldrack¹⁵³, si la majorité des études publiées montrent des résultats significatifs, soit nous faisons de la science insignifiante aux résultats prévisibles, soit nous utilisons des méthodes biaisées pour démontrer des effets même quand il n'y en a pas.

Un autre problème majeur avec une utilisation mal comprise de la valeur-p est qu'il devient très facile de mettre le poids d'inférence et de réflexion sur le test de significativité plutôt que sur les scientifiques^{108,194}. La procédure est en effet trompeusement simple : on se limite à vérifier si les résultats sont significatifs, après quoi une observation peut devenir un « fait scientifique » alors que la réalité peut être toute autre¹⁹⁴.

De toute évidence, l'objectif n'est pas non plus de publier absolument tout ce qui se fait, car dans ce cas nous serions ensouvelis sous une montagne d'articles encore plus grande et insurmontable que celle que nous avons déjà aujourd'hui, avec une marée de résultats insignifiants dans tous les sens du terme rendant difficile de trier ce qui est bon de ce qui est mauvais¹³⁷. C'est d'ailleurs en partie ce qui pousse certains auteurs¹⁷⁸ à dire que le biais contre les résultats non-significatifs est non seulement tolérable, mais même souhaitable. Nous sommes en effet d'accord, comme d'autres^{129,137}, que ce ne sont pas tous les résultats non-significatifs qui devraient être publiés, comme par exemple ceux provenant d'études avec un devis de recherche médiocre, aux hypothèses nulles et alternatives mal circonscrites, ou tout simplement issues de mauvaises idées (quoique ce dernier critère soit tellement subjectif qu'il ne devrait probablement pas être utilisé pour prendre des décisions).

En effet, comme le pointe brillamment Nelson et coauteurs¹³⁷, les études jugées de qualité et importantes finissent probablement toujours par être publiées, donc tout simplement rendre la publication plus facile augmenterait nécessairement de façon disproportionnée la publication de mauvaises études et les faux-positifs. Nous sommes tout à fait d'accord avec ce point, quoiqu'il soit intéressant de mentionner que si les scientifiques

étaient jugés par la qualité de leurs travaux plutôt que par des indices de performance, cette argumentation perdrait soudainement beaucoup de poids.

De plus, ce que les auteurs acceptant le biais contre les résultats non-significatifs semblent souvent négliger, c'est que nous devrions avoir les mêmes standards de rigueur scientifique pour les résultats significatifs que pour les résultats non-significatifs. Ce n'est pourtant pas ce que nous observons en réalité, car il semble falloir avoir une bien meilleure étude pour publier des résultats non-significatifs. En conséquence, nous pouvons aujourd'hui trouver dans la littérature scientifique des « preuves » statistiquement significatives pour pratiquement toutes les affirmations, ce genre de résultats étant beaucoup trop facile à publier^{172,179}. Il est pourtant évident que ce ne sont pas toutes les affirmations possibles qui sont vraies.

De plus, il s'avère que qualifier une étude de « robuste » peut devenir surprenamment difficile lorsqu'on prend en compte l'ignorance généralisée par rapport aux pré-requis des tests communément utilisés (Boîte 2). Par exemple, lorsqu'une étude inclut une grande taille d'échantillon, les résultats sont généralement considérés plus robustes et d'une plus grande importance théorique. La vérité est cependant plus complexe. À cause de la manière dont les tests de significativité de l'hypothèse nulle sont calculés, il semble qu'ils ne soient pas très bien adaptés aux grandes banques de données, car ils garantissent alors la détection de la moindre différence entre les groupes, même pour des différences qui ne sont pas nécessairement en relation avec les variables à l'étude¹⁹⁴.

Une autre idée commune qui peut être trompeuse est qu'obtenir des résultats significatifs malgré des erreurs de mesures importantes serait particulièrement impressionnant, suivant la logique que l'effet détecté serait encore plus fort si les conditions de mesures auraient été optimales. Cependant, il semble que cette idée soit généralement fausse pour de petites tailles d'échantillonnage¹¹⁹, une occurrence courante en recherche.

De plus, selon Simmons et coauteurs¹⁷⁹, les degrés de liberté décisionnelle des scientifiques seraient responsables de beaucoup de résultats faussement significatifs, ceci ne requérant aucune intention malhonnête pour se produire. En effet, il est rare que les scientifiques prennent – ou même puisse prendre – toutes les décisions relatives à une expérience *avant* que celle-ci ait eu lieu ou ait commencé. Ainsi, au cours d'une étude, il se produit un grand nombre de décisions par rapport à la récolte des données et à leurs analyses. C'est ce que ces auteurs appellent degrés de liberté des scientifiques, car une certaine ambiguïté couvre plusieurs de ces décisions, celles-ci restant matière à débat. Ceci, combiné avec l'importance d'obtenir des résultats significatifs pour sa carrière et à l'habileté que nous avons tous de justifier des conclusions en accord avec nos désirs^{143,179}, mène à des décisions favorisant la significativité statistique. D'après leur modélisation, Simmons et coauteurs¹⁷⁹ trouvent que si seulement 4 degrés de liberté très communs^{Note 11} sont ainsi utilisés, il en résulterait que

¹¹ Ces degrés de liberté des scientifiques incluent des pratiques très courantes et acceptées en recherche, telles que tester plusieurs covariables, plusieurs modèles différents, ou plusieurs catégorisations de la variable dépendante sans présenter l'effet de tous ces différents tests. Aussi incluse est la décision de la taille d'échantillonnage, qui peut varier à mesure que l'expérience avance, sans présenter cette variation et la justification pour ce faire.

le taux de faux-positifs (communément perçu comme se trouvant autour de 5%) pourrait facilement monter à 60%.

Vous vous demanderez peut-être pourquoi de telles décisions ne sont pas mieux réglementées si elles sont si importantes. La raison est que les scientifiques ne croient généralement pas que ces décisions aient vraiment un effet sur la véracité des résultats en général. Cependant, la modélisation de Simmons et coauteurs¹⁷⁹ montre de façon conservatrice que cette croyance n'est pas juste et qu'elle représente probablement encore une fois notre prédisposition à tirer des conclusions qui nous sont bénéfiques plutôt qu'objectives^{143,179}.

Ainsi donc, il faudrait évaluer avec soin n'importe quel résultat, significatif ou non, et la façon dont il a été obtenu pour en déterminer son importance pour la recherche et son droit à la publication. Ce n'est pas ce qui se passe actuellement. Ce double-standard de rigueur scientifique entre résultats significatifs et non-significatifs est donc la source du biais dont nous parlons, et ce biais est en effet problématique pour l'activité scientifique (voir la section « Inconduites et biais de publication : les effets sur la science »). Une incompréhension générale du fonctionnement des tests de significativité de l'hypothèse nulle (Boîte 2) est probablement responsable de cette attitude plutôt frivole méprenant ce biais de publication pour quelque chose d'acceptable ou même de souhaitable¹⁹⁴. Ainsi, étant donné les piliers vacillants supportant l'empire standardisé de la valeur-p en science, il est vraiment aberrant qu'elle soit elle-même souvent la base pour décider si un résultat est significatif ou non, étant donc aussi la source du biais de publication le plus répandu.

Notez tout de même que ni nous, ni les auteurs que nous citons ici n'argumentons pour l'abandon de la valeur-p et des tests de significativité de l'hypothèse nulle qui, malgré tout, peuvent donner de bons résultats⁸⁷. Le lecteur attentif aura pu voir que les critiques les plus importantes concernant cette pratique ne touchent pas tellement la logique qui la soutient ni même son utilité. Plutôt, il s'agit de la façon dont elle est utilisée et du fait que l'information qui en découle soit souvent mal comprise et donc aussi mal interprétée, ce que même les défenseurs les plus ardents de cette pratique reconnaissent^{16,86}.

L'objectif de cette section est donc plutôt de remettre le poids de la réflexion et de la justification sur le dos des données et des scientifiques, et non sur celui d'une procédure standardisée pour tous avec un processus de décision catégorique de type « oui/non »^{7,108,194}. Il s'agit de faire remarquer que le système de publication actuel utilise la valeur-p pour créer des biais qui affectent toute la science de manière non-scientifique, au détriment des scientifiques. Ces biais sont une conséquence de se concentrer sur l'habit (le contenant) sans apporter grande attention au supposé moine (au contenu).

Étroitement reliée à cette attitude et tout aussi erronée est l'idée que la publication dans les revues scientifiques avec révision par les pairs fasse partie de la démarche scientifique et qu'elle soit donc nécessaire pour conférer le sceau de la scientificité. La publication dans les revues et la démarche scientifique sont deux concepts indépendants l'un de l'autre, c'est-à-dire que chacun peut exister sans égards pour l'autre – ce qui se produit au moins partiellement – même si nous souhaiterions tous que les deux travaillent main dans la main.

Il est aussi important de clarifier ici que plusieurs écueils étayés dans cette section-ci ne sont pas spécifiques aux tests de significativité de l'hypothèse nulle^{180,201}, mais découlent bien du biais de publication en lien avec la significativité et d'un focus quasi complet sur une seule procédure, quelle qu'elle soit. Il n'y a donc pas d'alternatives qui pourraient prendre la place des tests de significativité de l'hypothèse nulle de façon satisfaisante⁸⁶. Des solutions impliqueraient plutôt l'utilisation concomitante des tests de significativité avec d'autres méthodes statistiques – par exemple des analyses Bayésiennes et des intervalles de compatibilité – donnant chacune différents types d'information complémentaire et permettant donc une interprétation plus complète et moins biaisée des résultats obtenus^{16,194,201}.

D'autres pistes de solution incluent : 1) exiger la publication détaillée de toutes les données, décisions et analyses reliées à l'étude afin de permettre la vérification des analyses et des choix pris par les auteurs^{137,194} ; 2) encourager les réplifications^{16,137,194}, quoique ceci soit bien difficile (voir la prochaine section sur les biais de publication contre les réplifications) ; et 3) promouvoir des données et des résultats de qualité peu importe leur statut de significativité. Ce dernier point impliquerait d'améliorer de façon substantielle l'éducation statistique des scientifiques pour leur permettre de planifier des expériences et de récolter et d'analyser des données de façon réfléchie et justifiée... ou du moins d'améliorer leur éducation au niveau où ils sont en mesure de réaliser qu'ils ont besoin d'aide, le cas échéant^{16,194}. Il existe des programmes d'éducation complémentaire à cet effet, mais leur succès semble encore bien loin de ce dont nous avons besoin¹²³.

Une autre facette des solutions à apporter a à voir avec l'exemple que donne les scientifiques plus âgés, les superviseurs, et dont la carrière n'est plus en danger¹⁵³. Ceux-ci doivent soutenir les doctorants et les post-doctorants qui désirent faire de la meilleure science, même si ces derniers n'obtiennent pas de résultats significatifs (apparemment encore souvent un pré-requis pour graduer)¹⁵³. Les superviseurs doivent aussi montrer l'exemple en démontrant qu'il est plus important de publier une science qui est juste que de publier dans une revue prestigieuse¹⁵³ – notez que les deux options ne sont pas mutuellement exclusives.

Pour finir, nous tenons aussi à souligner que nous ne mettons pas la faute de cette situation sur les scientifiques, qui essaient majoritairement de simplement publier leurs travaux au mieux de leur connaissance. Beaucoup de scientifiques croient d'ailleurs que les tests de significativité et la valeur-p sont les seuls outils à leur disposition puisque les alternatives ne sont généralement pas enseignées, ou pas suffisamment pour bien les comprendre¹⁹⁴. Si nous voulions vraiment nommer un responsable – quoiqu'il n'y ait certainement pas de responsable unique – il s'agirait plutôt de systèmes de publication et de récompenses institutionnelles qui ne tiennent pas en priorité le bien des scientifiques, de la science ou du public^{18,182}.

Biais contre la reproduction

Commençons cette section par quelques définitions. Lors de l'évaluation de la reproductibilité des études scientifiques, le statut de non-reproductibilité des résultats est attribué de trois manières principales : 1) lorsque les auteurs ne fournissent pas

suffisamment d'information pour même tenter de reproduire les résultats, ce qui arrive très souvent⁹⁸, mais ne signifie rien quant à la véracité des résultats¹¹⁴. On parlera alors de non-reproductibilité du protocole. 2) Lorsqu'en utilisant la même méthodologie, les résultats ne sont pas reproduits significativement, ou 3) lorsqu'en utilisant exactement les mêmes données de base fournies par les auteurs, on n'arrive pas à reproduire les résultats publiés. Ainsi, la reproductibilité englobe à la fois la possibilité de reproduire les résultats (point 1), ainsi que l'occurrence avec laquelle les résultats sont reproduits avec succès lorsque c'est tenté (points 2 et 3).

Dans ce qui suit, nous utiliserons le terme « reproduction » pour parler du principe général et le terme « réplique » pour parler de la reproduction des résultats (selon Goodman et coauteurs⁷⁹), c'est-à-dire lorsqu'il y a vraiment une tentative de reproduire (à partir de maintenant, répliquer) les résultats. Notez que les répliques sont notoirement difficiles à réaliser et surtout à interpréter étant donné toutes les sources de variation d'une étude à l'autre^{114,151}.

Le biais de publication contre les répliques est beaucoup moins subtil que celui contre les résultats non-significatifs ; il est explicitement ou implicitement encouragé par les éditeurs de plusieurs revues – particulièrement pour les revues prestigieuses – dans leurs instructions aux auteurs et aux réviseurs^{124,132,148,149,185}. Dans les instructions des revues de psychologie par exemple, 1% refusent explicitement les répliques, 33% les découragent implicitement, 63% ne les mentionnent pas et seulement 3% indiquent spécifiquement qu'ils les acceptent¹²⁴.

Pourtant, la reproductibilité fait partie intégrante de la démarche scientifique¹⁴⁸ (Fig. 1D) et est nécessaire pour vérifier ce qui a déjà été fait. En effet, comme nous avons pu le voir dans la section sur la valeur-p, les tests statistiques ne sont (pratiquement) jamais fiables à 100%^{92,126,142,194}. La reproduction des protocoles et les répliques sont donc absolument essentielles pour déterminer la probabilité d'une hypothèse d'être vraie ou non^{126,194} ainsi que pour produire des méta-analyses, ces dernières étant essentielles pour évaluer l'état des connaissances sur un sujet et diriger les futures recherches.

Les répliques sont particulièrement nécessaires pour les phénomènes plus « flexibles »^{126,149}, c'est-à-dire ceux qui échappent aux cinq sens et aux outils qui les prolongent ou qui sont contrôlés par des facteurs inconnus et/ou incontrôlables, car les résultats des tests statistiques y sont alors encore plus incertains. La reproduction prend une importance toute particulière pour des recherches plus descriptives n'utilisant pas de statistiques, telles que certaines études de cas¹³³, car seule la comparaison d'observations répétées peut y éclairer les rapports de cause à effet. Pour ces cas toutefois, nous parlerons de reproduction des inférences qualitatives de l'étude puisque les résultats ne peuvent pas être répliqués exactement^{12,79}.

Malheureusement, étant donné le dédain montré par les éditeurs des revues prestigieuses envers la réplique, il devient difficilement justifiable pour les scientifiques d'y réserver des ressources¹³². Par ailleurs, beaucoup de scientifiques ont du mal à publier les recherches visant à répliquer des résultats¹³⁰ ou ne s'y essaient même pas¹⁴. Cette situation explique probablement en partie pourquoi les faux-positifs, une fois arrivés dans la littérature scientifique, y restent de façon si persistante¹⁷⁹. Certains

prétendent donc que nous nous trouvions présentement dans une crise générale de reproductibilité en science^{14,140,143}.

Démonstrations empiriques

Ce constat de crise ne se base pas seulement sur le biais des revues scientifiques, mais aussi sur quelques évaluations empiriques de la reproductibilité en science, quoique ces évaluations soient encore beaucoup trop rares⁹⁸ pour tirer des conclusions fiables. Certains exemples d'évaluation de la reproductibilité sont alarmants, d'autres moins. En regroupant par exemple plusieurs études ayant testé la reproductibilité des travaux dans diverses branches de la biomédecine, entre 51 et 90% des résultats publiés semblent irreproductibles ou non-répliquables¹⁹. Dans le domaine des études précliniques, entre 51% et 89% des résultats ont échoué leur test de reproductibilité ou de répliquabilité⁷¹.

En sciences sociales, Camerer et coauteurs³³ ont pu répliquer 62% des 21 études qu'ils ont évaluées, toutes publiées dans *Nature* ou *Science* entre 2010 et 2015. Selon eux et leur revue de littérature sur le sujet, il semblerait raisonnable de considérer que le taux de non-répliquabilité des études en sciences sociales se situe entre 35% et 75%.

Dans le domaine de la psychologie, un effort de réplification n'a pu répliquer que 39 études sur un total de 100, avec une taille des effets répliqués en moyenne deux fois moins grande que les effets originaux¹⁴⁸. Cependant, ce très faible taux de réplification a été remis en question avec l'argument que tout dépend de la

définition que l'on donne à une réplication réussie. Par exemple, ce taux de réplication de 39% est basé sur des tests de signifiacnce et la valeur- p ¹⁴⁸ ; nous avons bien vu dans la section précédente comment cet indice peu être délicat à interpréter.

En comparaison, Patil et coauteurs¹⁵¹ se sont penchés sur les tailles d'effets calculés dans ces mêmes répliations. Ils ont démontré que 77% des ces tailles d'effets tombaient à l'intérieur de l'intervalle de prédiction des tailles d'effets des études originales, avec une probabilité de 95%. Il semble donc que le vrai taux de réplication pour ces 100 études se situe entre 39%¹⁴⁸ et 77%^{114,151}. Ceci est remarquablement semblable à l'intervalle donné par Camerer et coauteurs³³ de 35-75%. Il s'agit d'un intervalle large nous indiquant que la question devrait être étudiée davantage afin que nous puissions y voir plus clair.

Reproductibilité protocolaire

Maintenant, d'après toutes ces informations, nous sommes en droit de se demander pourquoi observe-t-on ces problèmes de reproductibilité des études scientifiques et de réplication des résultats? Commençons avec la non-reproductibilité des études, c'est-à-dire que les études publiées ne contiennent généralement pas suffisamment de détails pour reproduire exactement ce que les auteurs ont fait. Ce problème est particulier puisque normalement, fournir ces détails protocolaires est un pré-requis pour publier. C'est un problème assez bien reconnu et qui est probablement responsable des plus hauts taux de non-reproductibilité des études publiées.

Pour remédier à la situation, de plus en plus de revues ont répondu à l'appel de plusieurs auteurs^{19,79,95,194} militant sur l'importance des descriptions plus complètes et plus détaillées des protocoles et méthodes de recherche ainsi que pour un meilleur accès aux données utilisées pour toute étude publiée. De plus en plus de revues exigent donc que les auteurs fournissent des documents supplémentaires en ligne par rapport 1) à une description détaillée de leurs méthodes lorsque celle-ci ne peut pas être intégrée dans l'article principal pour cause de longueur, 2) aux données brutes utilisées dans les analyses et 3) aux codes exacts utilisés dans les programmes d'analyses.

C'est le cas par exemple pour la revue *Science*, depuis 2011. Stodden et coautrices¹⁹² ont testé les effets de ces mesures prises par *Science* en sélectionnant aléatoirement 204 études publiées depuis 2011. Elles ont pu voir une amélioration certaine par rapport au passé, et pour les articles où toutes les données ont été fournies, elles ont répliqué les résultats en utilisant les mêmes données et les mêmes codes d'analyse dans 100% des cas. Cependant, elles n'ont pas pu avoir accès aux données de plus de la moitié des études sous observation, les auteurs de ces études n'étant souvent pas même au courant qu'ils étaient supposés de partager leurs données.

Ainsi donc, même si beaucoup d'efforts sont déjà faits pour instaurer des mesures visant à améliorer la situation, il semble qu'implémenter ces mesures reste un défi de taille et qu'on ne vérifie pas si les auteurs s'y conforment vraiment. La raison principale pour laquelle les auteurs peuvent ne pas vouloir partager toutes leurs données est probablement la peur de se faire « voler » les fruits de leur travail. Par exemple, si quelqu'un utilise leurs données pour publier avant eux sur le même thème,

ou s'ils ont développé de nouvelles analyses leur donnant un avantage de recherche. Il s'agit donc d'une peur de perdre un avantage compétitif, par exemple pour l'obtention de financement de recherche. Étant donné la collaboration de plus en plus serrée entre les scientifiques et l'industrie (ce que nous explorerons en détail plus loin), il se peut aussi que des clauses de confidentialité et des avantages compétitifs sur le marché viennent troubler davantage cette situation à l'avenir.

Réplication des résultats

Penchons-nous maintenant sur le taux de non-réplication des résultats, lequel se situe vraisemblablement entre 35% et 75%^{33,114}. Quelles peuvent être les causes de ce phénomène? Déjà, ce type de problèmes est favorisé par un système de publication et de récomense privilégiant les affirmations audacieuses au détriment de la rigueur scientifique¹⁹. Aussi, comme nous l'avons vu précédemment (Boîte 2), le fait que les procédures expérimentales ainsi que les méthodes statistiques et leur interprétation soient largement incomprises et mal appliquées contribue certainement à cet état d'affaire avec, entre autre, la publication d'une quantité importante de faux-positifs^{6,15,71,79,88,92,94,119,143,179,194}. D'ailleurs, le fait que les programmes statistiques soient de plus en plus faciles à utiliser, sauvant beaucoup de temps et d'énergie, est dans cette situation une épée à deux tranchants ; les scientifiques n'ont souvent qu'à cliquer sur un bouton pour analyser leurs données sans besoin de vraiment comprendre si ce qu'ils font est justifié ou justifiable¹⁴³.

Plusieurs autres pratiques courantes supportent probablement fortement la non-répliquabilité⁷⁹. Celles-ci incluent la présentation d'une hypothèse qui a été formulée une fois les résultats obtenus, ou encore l'application de différents tests statistiques jusqu'à l'atteinte d'un résultat significatif pour ensuite ne présenter que le test qui a « fonctionné ». Il y a aussi le fait que beaucoup de jeunes scientifiques se sentent pressés par leurs supérieurs pour publier des résultats dont ils doutent¹³⁰. Un autre problème est que les résultats attendus ne sont généralement jamais remis en question, alors que ceux qui contredisent l'hypothèse favorite sont souvent simplement ignorés ou critiqués de façon démesurée^{18,143}.

Notez bien qu'encore une fois, les scientifiques honnêtes ne sont pas à l'abri de telles pratiques puisque plusieurs mécanismes inconscients nous rendent susceptibles – nous pourrions même dire qu'ils nous prédisposent – à nous tromper nous-mêmes et à tirer de mauvaises conclusions de façon instinctive^{143,179}. D'ailleurs, les scientifiques semblent avoir des opinions contradictoires quant à la présence des résultats non-répliquables dans la littérature (ils pensent qu'il y en a beaucoup) et à la confiance qu'ils y accordent (ils démontrent une grande confiance aux articles publiés)¹⁴.

Mise en garde et pistes pour l'avenir

Nous tenons à revenir ici sur le fait que la mesure sur laquelle on se base pour évaluer les répliquations a un énorme impact sur la considération qu'une réplication ait été réussie ou non. La significativité statistique par exemple n'est pas un critère

satisfaisant pour évaluer le succès de réplication d'une étude donnée. En effet, un résultat non-significatif peut avoir plusieurs causes qui n'ont pas nécessairement à voir avec l'absence d'effets^{7,194}, particulièrement si la représentation graphique des résultats montre une différence entre différents groupes⁷. Ceci est important à considérer lorsque les scientifiques parlent d'études contradictoires en ne se basant que sur la significativité des résultats, alors que les effets vont dans la même direction⁷. Il ne s'agirait alors pas d'un échec de réplication, mais bien d'une réussite, ou du moins d'une comparaison incertaine.

Parallèlement, il est important de faire plus de recherche empirique sur le sujet de la reproductibilité et de la réplication des études, car sans ces données, nous ne pourrions jamais vraiment comprendre ce qui se passe. L'estimation selon laquelle plus de 80% des résultats scientifiques publiés en recherche médicale seraient faux^{88,92} est par exemple beaucoup plus élevée que les seules données empiriques que nous ayons sur le sujet, qui elles se situent entre 14 et 24% et varient d'une branche de recherche à l'autre¹¹⁴. Ainsi, certains auteurs^{7,100,114} ne croient pas que nous nous trouvions dans une crise de réplication ; seulement que nous définissions mal ce qui constitue une réplication réussie, situation causée principalement par un focus démesuré et injustifié sur les valeurs-p et la significativité des résultats. Notez bien, encore une fois, qu'un focus démesuré et injustifié sur n'importe quelle autre méthode nous mènerait certainement aux mêmes problèmes^{Note}

¹².

¹² Par exemple, si nous utilisions les intervalles de confiance au lieu des tests de signification, il faudrait aussi faire attention à ne pas écarter des résultats juste parce que la valeur 0 fait partie de l'intervalle. En effet, admettons que nous étudions les effets secondaires d'un médicament. Nous ne pouvons pas conclure qu'il n'y a pas d'effets

Il faudrait donc surtout réfléchir à ce que les résultats nous montrent sans utiliser des catégorisations standardisées de type « oui/non », évaluer les évidences cumulées de différentes façons et évaluer la susceptibilité des études aux différents biais possibles^{7,79,194}. Patil et coauteurs¹⁵¹ ont d'ailleurs développé une procédure utilisant plusieurs de ces recommandations pour évaluer la réplication des études. Par exemple, ils incluent dans leurs analyses la robustesse de l'étude d'origine et celle de la réplication, et ils obtiennent avec leur méthode de bien meilleurs taux de réplication. Par contre, ces mêmes auteurs nous mettent en garde que répliquer des études peu robustes (par exemple avec une taille d'échantillon trop petite) avec leur méthode ne signifie pratiquement rien, puisque plus le design d'une étude laisse à désirer, plus l'intervalle des prédictions acceptables pour une réplication sera grand et plus il sera difficile de démontrer que l'étude n'a pas été répliquée.

Ainsi, les affirmations les plus alarmistes par rapport à la crise de réplication ne sont peut-être pas justifiées, mais pour causes : il y a trop d'études qui sont si pauvrement réalisées qu'on ne peut pas dire si elles peuvent être répliquées ou pas. À quel point ceci est mieux qu'une crise de réplication, nous ne sommes pas certains, mais il s'agit certainement d'une nuance importante à apporter. Quoi qu'il en soit, la méthode développée par Patil et coauteurs¹⁵¹ reste un nouvel outil extrêmement intéressant pour le futur et nous aide à souligner l'importance d'études bien

secondaires seulement parce que 0 fait partie de l'intervalle si la valeur (hypothétique) 30 en fait aussi partie⁷.

planifiées et bien conduites si nous voulons obtenir des résultats informatifs.

Il semble donc que le constat d'un état de crise en science soit discutable et doit être nuancé. Il en résulte que certains auteurs n'aiment pas que des termes comme « crise » soit utilisés pour parler de la situation^{100,114} et ce, même s'ils pensent eux-mêmes qu'une grande quantité d'études souffrent de failles importantes¹¹⁴.

Jamieson¹⁰⁰ par exemple préférerait que nous représentions les scientifiques comme des héros sur le chemin d'une quête sacrée et les fraudeurs comme des vilains, plutôt que de parler d'une crise systémique en science. Sa rhétorique – utilisée par plusieurs autres¹⁷⁶ – inclut le fait qu'il y ait de la science extraordinaire qui continue à se faire tous les jours malgré les problèmes et que dresser un portrait négatif de la science avec des rapports de crise ne peut que nuire à l'activité scientifique. De plus, pour elle, des phénomènes tels que les taux de rétractions qui augmentent et les soucis par rapport aux taux de réplification ne sont pas les indices d'une crise, mais bien les premières étapes d'un système d'auto-correction qui fonctionne bien. Cette autrice apporte d'excellents points, par exemple qu'il faille être juste dans notre analyse et ne pas seulement se concentrer sur les problèmes, mais aussi reconnaître et applaudir tous les efforts qui sont faits pour régler ces problèmes. Cependant, contrairement à elle, nous trouvons particulièrement important de mettre l'accent sur certains problèmes, peu importe l'image que ça donne, et ce, pour plusieurs raisons.

Premièrement, bien qu'il y ait des efforts pour régler la situation et que ce doit être reconnu, cela n'empêche pas que la

majorité des problèmes exposés ici sont connus depuis bien longtemps et que des efforts généraux et au niveau institutionnel se produisent avec un délais marqué⁶, de plusieurs décennies au moins. Ceci met un bémol sur l'idée d'un système auto-correctif fonctionnant bien, et fait plutôt penser à un système où les vrais changements ne se produisent qu'à cause d'une dénonciation de plus en plus fréquente et virulente des nombreuses failles, rendant insoutenable de continuer comme avant.

Deuxièmement, la présence d'efforts pour régler une situation n'implique pas qu'il n'y ait pas de crise, justement, à régler. Ainsi, vouloir se concentrer seulement sur les points positifs pour donner l'image d'une entreprise scientifique remplie de héros avec quelques vilains^{100,176}, n'est pas juste non plus. Ceci donne la fausse impression que le fonctionnement de la publication scientifique et que les motivations qu'on donne aux scientifiques ne sont pas truffés de problèmes profonds, causes des biais dont nous parlons ici.

Troisièmement, le fait que de la science de qualité continue à être produite à chaque jour ne signifie pas que les multiples ressources qui sont gaspillées par de nombreuses études non-réplicables et de piètre qualité ne constituent pas une crise ou un problème profond¹⁹⁴. Il est effectivement beaucoup plus facile de constater les progrès qui existent que d'estimer ceux qui n'arrivent pas ou qui sont retardés à cause de problèmes importants. Nous croyons donc qu'il soit tout aussi dommageable de faire croire au public que la science a tout bien en main lorsque ce n'est pas le cas. En effet, le système de publication scientifique est trop non-scientifique et l'ignorance quant aux analyses statistiques est trop répandue pour que ce soit le cas. D'ailleurs, penser que ce dernier point ne justifie pas de donner l'alerte

dénote justement de cette incompréhension des analyses statistiques, contribuant par le fait même à la non-reproductibilité des études scientifiques¹⁹⁴. C'est tout le système qui apparaît moins certain et inébranlable qu'il le laisse paraître aux yeux du public. Ceci est important à comprendre pour ne pas prendre ou soutenir des décisions trop hâtives.

Quatrièmement, on oublie souvent que pour les jeunes scientifiques, il y a encore de grandes disparités entre les standards de rigueur exigés pour une science de qualité et les obligations pour faire carrière¹⁵³. Ainsi, bien que la grande majorité des jeunes scientifiques souhaitent certainement produire la meilleure science possible, ils sont aussi inquiets quant aux conséquences qu'une telle approche apportera pour leur carrière^{27,153,199}. En d'autres mots, ils craignent que mettre plus d'efforts pour produire de la science de meilleure qualité les désavantagera compétitivement puisqu'ils seront susceptibles d'avoir moins de résultats significatifs, moins de publications dans les revues prestigieuses et moins de citations de leurs travaux.

Pour être juste donc, il faut absolument faire la distinction entre la science et la publication scientifique en pointant les biais importants créés par cette dernière, biais qui mettent en évidence les conflits entre le système actuel de publication et la démarche scientifique. Il faut voir les failles de la publication scientifique et du système de récompenses scientifiques, reconnaître que ces failles ont souvent une place centrale en science et dans la carrière des scientifiques, et finalement bien sûr reconnaître les efforts pour trouver et appliquer des solutions, ainsi que les progrès qui en découlent.

D'ailleurs, trouver de bonnes solutions nécessite d'être au clair avec les causes des problèmes liés à la reproductibilité et aux réplifications. Nous en avons déjà parlé en abondance dans cette section, mais en résumé nous pourrions dire que les causes principales sont i) les directives des revues scientifiques (biais contre les réplifications), ii) des devis de recherche de piètre qualité¹¹⁴, iii) l'incompréhension des méthodes statistiques utilisées par les scientifiques^{108,114,119,194} et iv) le système de motivation et de récompense pour les scientifiques^{114,182}. Des solutions possibles incluent donc l'obligation de comparer des hypothèses rivales, coopérer avec des scientifiques rivaux et, probablement le plus important, le développement de meilleures compétences statistiques chez les scientifiques^{95,114,143,153,194}.

Favoriser les travaux de réplification en modifiant les attentes institutionnelles pour les scientifiques et les pratiques dans les revues scientifiques serait aussi une direction importante à suivre. Néanmoins, nous devons faire attention ici, c'est-à-dire faire preuve de raison et de flexibilité. En effet, comme les ressources sont limitées, il est souvent judicieux de produire des études qui rajoutent des conditions à des fins de généralisation plutôt que de simplement démontrer que les résultats d'une étude donnée sont répliquables ou non¹⁷⁶. Ceci s'applique surtout lorsque les coûts de se tromper ne sont pas importants.

Quant aux jeunes scientifiques, ils ont besoin que les plus vieux décrient haut et fort l'injustice de la situation dans laquelle ils se trouvent¹⁵³. Ils pourraient aussi créer des groupes dédiés à la reproductibilité en science et rajouter dans leurs CV une section démontrant leur implication pour une science reproductible^{27,153}. Bien sûr, ces actions n'auront qu'un effet à long terme, et

seulement si elles sont soutenues par les superviseurs et les instituts de recherche.

Évidemment, le but n'est pas de complètement bloquer l'avancement de la science avec une infinité de répliques des mêmes études¹¹⁴ ou en exigeant des devis de recherche tellement pointus qu'ils en deviendraient trop coûteux pour faire de la recherche exploratoire¹⁷⁶. Le message est plutôt d'être conscient de ce qu'une étude publiée représente (ou ne représente pas), d'être conscient de comment le monde de la recherche fonctionne, tout ceci afin de pouvoir tirer des conclusions plus éclairées, plus nuancées et surtout plus justes. Il s'agit d'évaluer ce que nous lisons, non pas par la revue de publication, mais bien par la méthode utilisée, par la solidité des analyses, par la logique décrite et par les applications permises. Encore une fois, il s'agit de ne pas juger le moine par son habit, de ne pas juger le contenu par son contenant.

Inconduites et biais de publication : les effets sur la science

Les méta-analyses

Les conséquences des biais de publication et des inconduites scientifiques sont communes, néfastes et multiples, affectant l'intégrité et la crédibilité même de la science^{10,18,95,103,194}. Un très bon exemple concerne les méta-analyses (voir page 63 pour une définition), qui sont compromises par tout biais de publication puisque le bilan sera lui-même biaisé^{60,129,149,185}. Ceci est un point crucial à intégrer, puisque ce sont les méta-analyses

de qualité qui nous permettent de tirer les conclusions les plus fiables par rapport à un questionnement scientifique, pour ensuite avancer dans la bonne direction. Surtout que le financement de la recherche n'encourage pas encore les méta-analyses prospectives, qui pourraient donner les meilleurs résultats, et le niveau de collaboration actuel ne les permet de toute façon pas encore⁹⁶. Ainsi, les méta-analyses se font de façon rétrospective et souffrent donc particulièrement des biais de publication.

Par exemple, nous avons déjà discuté du manque de réplifications en science, ce qui fait que très peu d'études peuvent en fait être comparées adéquatement au sein d'une méta-analyse. Cet effet du manque de réplification est accentué par un manque encore plus grand de réplifications formelles, c'est-à-dire que même si beaucoup d'études investiguent des questions qui ont déjà été étudiées, le besoin de présenter l'étude comme une nouveauté extraordinaire rend difficile la tâche de savoir quelles études peuvent aller ensemble lors d'une méta-analyse⁹⁶. Par conséquent, une majorité d'études et de données ne sont jamais incluses dans les méta-analyses et perdent donc une grande partie de leur potentiel pour la science^{96,98}.

De plus, non seulement les études comparées dans les méta-analyses sont soumises aux divers écueils discutés dans le présent ouvrage, mais les méta-analyses elles-mêmes sont susceptibles aux biais de publication et à l'impératif de publication. Le biais de publication en faveur des résultats significatifs par exemple s'applique aussi aux méta-analyses, une évaluation de 461 méta-analyses dans le domaine des essais cliniques ayant trouvé que 16 à 37% représentaient des faux-positifs, ces derniers montrant en plus les plus importantes inflations de l'effet étudié¹⁵².

Tout de même, les méta-analyses restent essentielles et portent beaucoup de poids dans la balance de l'information scientifique. Elles peuvent donc être publiées dans des revues prestigieuses et sont en général beaucoup citées⁹⁶. Étant donné les obligations carriéristes des scientifiques dont nous avons déjà parlé, publier une méta-analyse devient alors très attrayant. Les méta-analyses sont aussi attrayantes pour donner du crédit à une opinion particulière par la sélection biaisée d'articles qui supportent cette opinion, ou encore pour l'industrie en tant qu'instrument de marketing – seules 31% des méta-analyses présentant un produit ont quoique que ce soit de négatif à dire dans leur résumé⁹⁶.

Tout ceci mène à une augmentation disproportionnelle des méta-analyses dans la publication scientifique par rapport à d'autres types de publications, du moins dans le domaine de la recherche biomédicale⁹⁶. Cependant, lorsqu'on évalue ces méta-analyses, il semble que la présence de failles importantes et une piètre qualité soient des caractéristiques communément retrouvées^{96,152}. Il s'en suit que dans ce domaine, il y a une quantité importante de méta-analyses (44%) qui sont redondantes, inutiles, ou non-informatives puisqu'elles ne présentent souvent aucune nouvelle information comparées aux méta-analyses antérieures⁹⁶. De leur côté, les méta-analyses qui utilisent des méthodes désuètes et trompeuses, ou qui sont tout simplement complètement défaillantes, représentent 33% des méta-analyses en recherche biomédicale⁹⁶.

Un autre 20% sont probablement de bonne qualité, mais ne sont tout simplement jamais publiées puisqu'elles sont produites par des compagnies privées spécialisées sur la production de méta-analyses pour l'industrie⁹⁶. Ces compagnies

travaillent donc à contrat et n'ont aucune motivation pour publier, à cause de deux raisons principales. Premièrement, l'industrie ne souhaite pas voir publiés certains de ces résultats qui pourraient aller à l'encontre de leurs profits. Deuxièmement, le processus de publication avec révision par les pairs est complexe et prend beaucoup trop de temps et d'énergie, soi disant pour assurer la qualité des publications (nous discuterons plus loin de cet aspect et d'à quel point son efficacité est médiocre). N'ayant pas besoin d'être évaluées par leurs publications, ces compagnies préfèrent donc employer leur temps à faire d'autres contrats.

Ainsi, il semble que dans le domaine de la recherche médicale, seul 3% des méta-analyses soient à la fois bien faites et cliniquement utiles⁹⁶. Cet estimé est alarmant, car les méta-analyses constituent les publications scientifiques inspirant le plus de confiance et ayant le plus de poids pour des décisions de toutes sortes. D'autant plus que plusieurs scientifiques – auteurs, réviseurs ou éditeurs – n'ont pas la compétence pour différencier une méta-analyses de mauvaise qualité d'une méta-analyse de haute qualité⁹⁶. Il est donc grand temps de changer nos pratiques et de faire une transition vers la production de méta-analyses prospective, pré-enregistrées et indépendantes de l'industrie⁹⁶. Notez aussi que plusieurs techniques existent afin de corriger les biais de publication lors d'une méta-analyse, comme par exemple le graphique en entonnoir^{102,116,149}. Cependant, l'interprétation de telles techniques est souvent trop simpliste et trompeuse^{102,185} et les méthodes plus complexes et potentiellement plus fiables ne sont encore que très peu utilisées¹⁸⁵.

Le gaspillage des ressources

Les inconcutes, grandes ou petites, et les biais de publication ont aussi comme conséquence une énorme quantité de ressources gaspillées, par exemple lorsque des scientifiques testent les mêmes hypothèses sans le savoir parce que les résultats non-significatifs ne sont pas publiés^{126,129,185}, ou lorsqu'ils basent leurs travaux sur des publications dont les résultats représentent des faux-positifs^{18,71,179,194}. Ces faux-positifs, ayant une prévalence beaucoup plus grande qu'on s'imagine généralement^{88,92,94,179}, peuvent par exemple mener à des efforts de recherche mal dirigés ou à des décisions politiques défaillantes¹⁷⁹.

Dans le domaine de la recherche biomédicale et des essais cliniques, il est estimé qu'on gaspille 85% de l'argent qui y est investi³⁷. Ceci serait causé non seulement par une quantité importante de résultats non-réplicables, mais aussi par l'importance que prennent les « modes de recherche » et les biais de publication dans la planification et la publication des études, au lieu de baser la recherche sur les besoins des patients et sur les questions les plus pertinentes^{18,37,71,94}. Il en résulte que la recherche médicale est non seulement ralentie et truffée de faux-positifs, mais aussi qu'une grande partie des vrais positifs sont inutiles puisque les études sont planifiées pour être publiées et non pas pour être pragmatiques, informatives, ou utiles aux patients⁹⁷. Ces estimations sont troublantes puisque pour le public, les conséquences les plus graves d'une recherche sous-optimale sont évidemment reliées au domaine médical puisque ceci touche directement à leur santé^{185,188}.

La sélection de la mauvaise science

Les inconduites scientifiques et les biais de publication mènent aussi à la sélection et à la propagation des pratiques statistiques gonflant les faux-positifs^{126,182} (Fig. 5) et même à la « canonisation » de faits qui sont en réalité faux si les résultats non-significatifs sont suffisamment défavorisés¹⁴⁰. En effet, le biais de publication contre les résultats non-significatifs poussent les scientifiques à choisir des procédures de recherche et des analyses statistiques favorisant les résultats significatifs, que ces méthodes soient appropriées ou non¹⁷⁹.

Ceci est un phénomène que Smaldino et McElreath¹⁸² appellent la sélection naturelle de la mauvaise science. Selon leur modèle, cette sélection ne requiert aucun stratagème ou malhonnêteté de la part des scientifiques, seulement des idées préconçues et des habitudes qui favorisent la publication et en alimentent les biais. Par exemple, une incompréhension de certaines méthodes statistiques qui augmente les chances d'obtenir des résultats significatifs peut être sélectionnée et propagée à travers la publication, pour ainsi devenir générale^{15,142} (Fig. 5). Publier de la mauvaise science devient alors facile⁶. Ainsi, il semble que même si les fraudes conscientes étaient mieux contrôlées et mises hors jeu, les négligences continueront à être sélectionnées puisque des faux-positifs ont plus de succès que des résultats non-significatifs⁸².

Tout cela est déplorable, d'autant plus que dans le domaine de la santé par exemple, des résultats peuvent être statistiquement non-significatifs, mais cliniquement significatifs si l'ineffectivité d'un traitement déjà en vigueur est impliquée. À l'inverse, un traitement peut être statistiquement significatif, mais

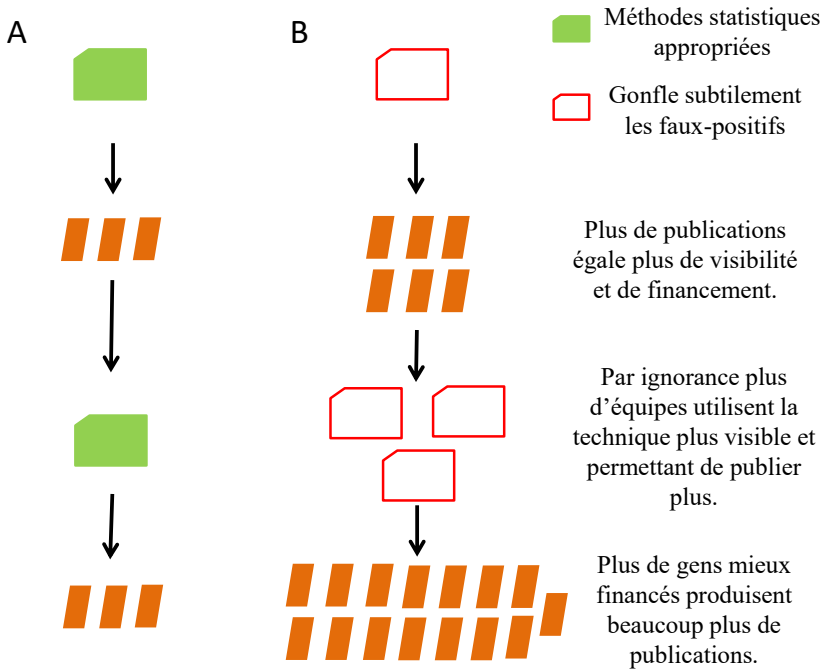


Figure 5. Schéma de la sélection de la mauvaise science Imaginons deux équipes de recherche travaillant sur un même phénomène. Imaginons aussi que ces deux équipes ont une productivité similaire. Une première (A : vert plein), utilise des méthodes statistiques appropriées à leurs données. La deuxième (B : rouge vide) utilise par ignorance ou négligence des méthodes gonflant les faux-positifs. Si produire ces faux-positifs donne deux fois plus de publications (losanges oranges), il s'en suit naturellement une propagation disproportionnellement élevée des mauvaises méthodes et des faux-positifs, autant par un financement accru des équipes semblant plus productives que parce que d'autres équipes leur emboîteront le pas (toujours par ignorance). Ceci crée une « mode méthodologique », aux dépens des méthodes appropriées (A). C'est la sélection de la mauvaise science, elle-même permise par les biais de publication.

cliniquement non significatif. Par exemple, une faible amélioration de la condition du patient détectée au niveau statistique peut ne pas valoir les efforts ou ne pas compenser les effets secondaires du traitement, choses qui ne sont souvent pas considérées dans les essais cliniques⁹³.

Certains considèrent que le biais de publication par rapport aux résultats non-significatifs est inhérent au processus de recherche scientifique, et qu'il suffirait donc d'être conscient du problème pour éviter les conséquences les plus graves¹⁷⁸. D'autres probablement plus réalistes^{120,140,194} croient que cette position est naïve et ne reflète pas l'ensemble de la réalité des scientifiques. Il semble effectivement que dépendamment du type d'étude et des conditions de recherche, une publication des résultats non-significatifs non biaisée soit nécessaire pour que l'autocorrection scientifique fonctionne^{60,63,126,140}.

Fanelli⁶³ développe d'ailleurs un cadre théorique d'équations pour mesurer quantitativement l'importance de publier tous les résultats non-significatifs. Bien que ses équations nécessiteraient de circonscrire précisément des paramètres qui sont, selon nous, aussi difficiles à circonscrire que la question elle-même, elles nous offrent un contexte facilitant la réflexion sur le sujet. Selon ce cadre théorique, il semble que l'importance de publier tous les résultats non-significatifs soit particulièrement grande dans des domaines où la méthodologie est claire et bien codifiée et où les hypothèses testées sont clairement définies. Ceci inclut des domaines comme la physique (où les résultats non-significatifs sont en effet particulièrement valorisés) et les essais cliniques (où le biais de publication contre ces résultats est particulièrement présent, nuisible et fortement décrié). Dans d'autres domaines où la méthodologie reste très flexible car

sujette à interprétation et où les hypothèses sont difficiles à circonscrire précisément, comme par exemple certaines branches des sciences humaines ou de l'écologie, il semble que l'importance des résultats non-significatifs diminuent par rapport à celle des résultats significatifs⁶³.

Notez que ceci ne signifie tout de même pas que les résultats non-significatifs dans ces domaines soient inutiles ou non-informatifs. Nous le répétons : pour que la publication des résultats non-significatifs fonctionne sans embourber la littérature scientifique d'articles insignifiants, il nous faut avoir les mêmes standards de qualité pour les résultats significatifs que pour les résultats non-significatifs. Ceci est important à considérer, car l'embourbement de la littérature est un argument commun contre l'acceptation plus facile des résultats non-significatifs pour publication^{63,137,178}. Cependant, cet argument n'est souvent pas considéré pareillement pour les résultats significatifs provenant d'études médiocres, embourbant ainsi eux aussi la littérature. Nous pourrions donc résumer l'objectif à atteindre comme suit : il faut faire place aux résultats non-significatifs de bonne qualité en réduisant la présence de résultats significatifs de piètre qualité.

De plus, les auteurs qui « défendent » le biais de publication contre les résultats non-significatifs en disant qu'il ne faut pas embourber la littérature de résultats non-informatifs ne considèrent généralement pas un point très important – ou du moins ils n'en parlent pas. Il s'agit du point dont nous avons parlé au début de cette section, c'est-à-dire la sélection de la mauvaise science¹⁸². En effet, ce biais de publication ne fait pas que cacher de l'information qui, dépendamment des domaines, serait plus ou moins importante pour l'avancement de nos connaissances.

Comme nous l'avons expliqué, il modifie aussi le comportement et les habitudes des scientifiques (Fig. 5), qui sont malgré eux aux prises avec leurs obligations carriéristes^{88,179,182}. Cette sélection forcée de la mauvaise science embourbe certainement la littérature avec des articles aux résultats significatifs, mais qui sont toutefois non-informatifs ou même nuisibles à l'avancement de la recherche⁸⁸.

Les rétractions

Étroitement lié à la sélection de la mauvaise science est le phénomène des rétractions, où une revue peut rétracter un article lorsqu'une erreur importante a été trouvée après publication, invalidant les conclusions de l'article. La découverte d'une telle erreur peut se faire par les auteurs mêmes ou encore par des lecteurs. La majorité des rétractions sont le résultats d'inconduites avec une intention de duper le lectorat, quoiqu'environ un tiers soient le résultats d'erreurs tout à fait honnêtes^{134,204}. Il est donc toujours extrêmement important de vérifier les causes d'une rétraction avant de tirer des conclusions sur l'intégrité des auteurs. Toujours en recherche biomédicale, entre les années 2000 et 2015, les causes principales de rétraction étaient une révision par les pairs compromise – ce dont nous discuterons abondamment plus loin dans le livre – le plagiat et la non fiabilité des données¹³⁴.

Le nombre absolu de rétractions augmente actuellement de façon exponentielle, les revues prestigieuses étant en tête de ligne de ce phénomène²⁴, quoiqu'elles soient rapidement rattrapées par d'autres²⁰⁴. En réalité, cette augmentation n'est

cependant pas reliée à une inconduite quelconque⁶², comme nous en faisons l'hypothèse dans la version précédente de cette ouvrage. Il s'agit plutôt de plus d'efforts et de volonté pour identifier une grande diversité d'erreurs nécessitant une rétraction^{134,190,204}, mais surtout d'un plus en plus grand nombre de revues acceptant de pratiquer la rétraction d'articles⁶². En effet, nous pouvons observer que le taux de rétraction reste quasi constant pour les revues permettant les rétractions depuis longtemps⁶². Ainsi, cette augmentation du taux de rétraction peut être vu avec raison comme un bon signe, comme une augmentation de l'intégrité scientifique et non pas comme une augmentation des inconduites^{62,100}. Parallèlement, il ne faut pas oublier que la présence des rétractions reste tout de même un indice que des inconduites sévères se produisent, et qu'elles étaient déjà là, actives, bien avant que l'intégrité scientifique soit augmentée.

D'ailleurs, lorsqu'on regarde les données avec une plus grande résolution, certains motifs commencent à émerger. Par exemple, il semble que le taux de rétraction soit plus élevé dans les pays offrant des récompenses monétaires pour le succès des scientifiques, ou dans les pays ayant des politiques de punition plus strictes, peut-être parce qu'ils sont alors meilleurs pour découvrir les fraudes⁶⁵. Aussi, plus la réputation dépend de l'opinion des collègues ou plus il y a de collaboration pour écrire un article, moins il y a de rétractions ou d'exagérations des résultats^{65,66}. Finalement, il semble que les jeunes scientifiques soient responsables de plus de rétractions, possiblement à cause d'une plus grande pression pour publier⁶⁵. Il semble donc que les rétractions (honnêtes ou pas) dans les revues scientifiques soient déterminées principalement par des facteurs socio-culturels et par l'environnement de recherche⁶⁵. Il ne semble avoir aucun lien

entre d'un côté le nombre de rétractions malhonnêtes ou l'exagération des résultats, et de l'autre côté la performance des scientifiques qui pratiquent ces fautes professionnelles – bien au contraire^{65,66}.

Cette démonstration mène Fanelli et coauteurs^{65,66} à conclure que la pression pour publier ne semble pas avoir d'effet sur l'intégrité scientifique, même s'il faut dire qu'ils n'ont aucune mesure d'intégrité, seulement des indices d'intégrité ne représentant absolument pas l'ensemble de la chose. Tout de même, comme ils sont parmi les seuls à avoir empiriquement vérifié ce lien entre « pression pour publier » et « intégrité scientifique », il est recommandable de prendre leur conclusion au sérieux, quoique nous trouvons qu'elles ne cadrent pas bien avec toutes les données provenant des sondages que nous avons présentés ici quant à la pression ressentie par les scientifiques pour publier ou être promus, ni avec les inconduites professionnelles qui se produisent et semblent courantes. Nous avons une hypothèse alternative qui pourrait correspondre à la fois avec les données de Fanelli et coauteurs, et avec les données de ces sondages.

La raison pour laquelle la tendance à comettre des fautes professionnelles pouvant mener à des rétractions est moins forte ou égale pour les scientifiques les plus productifs pourrait être que ce sont les scientifiques les moins compétents ou performants qui sont tentés de tricher, pas ceux qui sont déjà particulièrement bons et prospères. En effet, on a tendance à associer les fraudes avec du succès, probablement à cause de quelques cas très célèbres et publicisés où c'était effectivement comme ça. Par contre, il est plus probable que les scientifiques les moins performants, ayant le plus de mal à monter les échelons de

la carrière académique, aient le plus à gagner et le moins à perdre de comettre une fraude. N'étant pas particulièrement doués, même avec la fraude, ils n'arrivent tout de même que rarement parmi les meilleurs.

Cette hypothèse, de laquelle nous n'avons aucune preuve empirique, pourrait expliquer ce manque de relation entre les conduites graves comme la fraude et la performance des scientifiques. Plus de recherche sur le sujet sera nécessaire pour en dire plus. Il faut aussi être conscient que ceci vaut justement pour le genre de comportement menant à des rétractions. Toutes les autres failles et conduites moins grandes – ou même généralement acceptées – que nous avons étayées dans d'autres sections ne sont pas portées principalement par les auteurs et les revues moins réputés ; il s'agit de phénomènes généraux qui touchent tous les niveaux, exactement comme la recherche d'importance et de qualité, qui se retrouve elle aussi à tous les niveaux, pas seulement dans les revues prestigieuses.

L'aspect social

Comme nous l'avons déjà vu dans la section « Une atmosphère de compétition insoutenable », la compétition, la pression pour publier et les conduites qui en découlent peuvent avoir un grand impact sur l'aspect relationnel de la vie des scientifiques. On décompte notamment des abus de pouvoirs de la part de superviseurs envers leurs étudiants ou des querelles parfois malhonnêtes entre deux compétiteurs, situations tragiques puisqu'elles détruisent des relations professionnelles et des collaborations qui auraient pu bénéficier à la science¹⁰.

Par ailleurs, l'existence de groupes dédiés uniquement à soutenir et à aider les jeunes scientifiques qui se font intimidés et poussés vers la production de science de piètre qualité²⁷ dénote clairement la présence non-marginale de tels comportements, mais aussi, heureusement, les efforts pour régler la situation. Ces initiatives sont d'autant plus importantes que plusieurs jeunes scientifiques ne ressentent justement pas de lien de confiance ou de confidentialité suffisamment fort pour parler d'inconduites avec leurs collègues plus âgés¹⁹⁹.

Les conséquences des inconduites professionnelles sur la vie relationnelle et personnelle des scientifiques ne représentent pas un sujet souvent discuté en profondeur. Pourtant, elles sont la source de grandes injustices, parfois dévastatrices, et il semble que chaque scientifique (incluant nous-mêmes) ait été témoin d'au moins une telle situation, plus ou moins grave. Pourtant, les scientifiques n'aiment pas faire de vagues avec cela, par exemple en faisant des plaintes formelles. Il nous semble que la peur des représailles ou même d'une perte de bonne réputation soit à l'origine de cette inactivité troublante contre ces injustices. Les agresseurs seraient alors mieux protégés que les agressés. Le désir de ne pas « perdre du temps avec ça » pourrait aussi expliquer une partie de cette inactivité. Ces hypothèses viennent de notre propre expérience, de celle de nos collègues et de nos lectures. Il serait donc souhaitable pour le futur que ce sujet, que nous considérons très sérieux, soit évalué davantage et pas seulement dans la littérature scientifique, mais aussi ouvertement dans les universités.

Bilan et perspectives

Pour terminer cette section sur l'influence de la publication scientifique en recherche, il est crucial de noter que le réel problème n'est pas l'absence de publication. Il s'agit bien plus de la publication sélective et subjective en faveur de certains résultats ou de certaines idées préconçues¹⁷², ce qui peut induire la communauté scientifique en erreur. Dans un tel contexte, le besoin des scientifiques de publier dans les revues les plus prestigieuses, qui elles-mêmes établissent clairement des biais de publication, est un fléau dont la communauté scientifique gagnerait à se débarrasser^{94,103,132,171}.

La résolution de ces problèmes doit cependant débiter au niveau institutionnel par la transition de l'évaluation actuelle des scientifiques vers une évaluation basée sur le contenu de leurs écrits (et non le contenant) et sur la qualité de leur recherche et de leurs relations professionnelles^{18,43,103,131}. Ceux-ci pourront ensuite modifier leur comportement en conséquence. Ceci est un point crucial, que nous ne saurions souligner avec trop d'emphasis.

Tout à fait relié au fléau des revues prestigieuses est le problème de la perception que l'on a d'un article qui est publié. Nous parlons ici de la fausse impression souvent répandue selon laquelle un article publié est nécessairement de bonne qualité, alors qu'un article non-publié ne l'est pas. Comme nous l'avons vu, qualité – elle-même difficile à évaluer – et publication ne représentent pas une équivalence en soi (voir aussi la section « Révision par les pairs »). Déjà les biais de publications sont des facteurs s'assurant de cette non-équivalence. Reprenons par exemple le biais contre les résultats non-significatifs. En réalité, ce

qui est important n'est pas la significativité des résultats, ni non plus de tout publier ce qui se fait ; l'important est de ne publier que les études de qualité, peu importe la significativité. Ceci nous permettrait de publier moins d'articles et de faire plus confiance aux résultats dans la littérature scientifique. Ceci permettrait aussi au caractère autocorrectif de la science de mieux faire son œuvre.

Vous remarquerez que depuis le début de cette section, nous utilisons beaucoup le terme « qualité », sans le définir précisément. La définition d'une étude de qualité est en effet très difficile à circonscrire, particulièrement pour la recherche exploratoire. Nous postulons donc que le meilleur moyen de régler cette situation serait de changer radicalement non seulement la façon dont les scientifiques sont évalués, mais aussi celle dont ils sont éduqués. Déjà, il faudrait une bien meilleure éducation par rapport aux statistiques afin que chacun comprenne ce qu'il fait¹⁹⁴. Ensuite, évaluer les scientifiques par rapport à la qualité de leurs travaux, mais aussi par rapport à la qualité de leurs relations avec leurs collègues les responsabiliserait à s'assurer eux-mêmes de cette qualité.

Pour les instituts, il s'agira d'abandonner les indices de performance pour un retour vers une évaluation concrète de performance, et récompenser les comportements bénéfiques pour la science plutôt que ceux qui sont bénéfiques pour l'individu. Ces deux aspects seront certainement cruciaux pour le futur d'une recherche scientifique basée sur l'intégrité de la démarche scientifique plutôt que sur le système actuel de publication.

Nous tenons à souligner que le fait que « évaluer les scientifiques selon la qualité de leurs travaux et le contenu de

leurs écrits » soit une recommandation sérieuse et répandue montre bien à quel point le système actuel dans lequel la recherche évolue s'est égaré, et à quel point l'évaluation des scientifiques est généralement pauvre. Nous ne trouvons pas normal qu'il faille recommander que le contenu et la qualité des travaux soient des éléments décisifs dans l'évaluation des scientifiques ; ça devrait aller de soi, et c'est d'ailleurs certainement le cas dans l'esprit des gens en général.

Cependant, le système de publication actuel et l'importance malade des facteurs d'impact viennent déformer et embrouiller la situation en créant des approximations non fiables de la qualité des travaux. Ces approximations ne sont pas fiables puisqu'elles deviennent des buts en soi, au détriment de l'objectif originel qui était de faire de la recherche de qualité et non biaisée. La conséquence est cette déconnexion substantielle entre d'un côté les besoins de la société et d'une activité scientifique saine, et de l'autre côté le système de motivations et de récompenses pour les scientifiques^{88,97,131,141}.

Des solutions concrètes

Nous sommes restés jusqu'ici relativement vagues par rapport aux solutions proposées, nous contentant de recommandations générales – mais tout à fait essentielles – par rapport aux changements nécessaires dans l'éducation, l'évaluation et comment on récompense les scientifiques. Ces questions importantes pour la recherche sont toutefois connues depuis longtemps, et de nombreux auteurs ont déjà proposé

plusieurs solutions concrètes aux différents problèmes concrets. C'est ce que nous explorerons dans cette section-ci.

Par rapport au problème des revues prestigieuses et de leur « monopole » sur l'importance scientifique, une solution serait de créer une alliance entre scientifiques dans le but de boycotter ces grandes revues et de changer les habitudes de publication^{112,171}. Brembs et coauteurs²⁴, argumentant que n'importe quel système serait préférable à l'actuel, suggèrent même l'abandon complet des revues scientifiques en faveur d'un système de publication basé sur des librairies savantes virtuelles. L'objectif serait ici que seule la qualité des travaux publiés pourrait être utilisée pour évaluer les scientifiques, éliminant par le fait même plus rapidement ceux qui produisent de la recherche médiocre. Dans ce contexte, mais aussi généralement, l'idée d'une limite d'un article par année et par scientifique¹³⁷ aiderait certainement tout le monde à naviguer toute transition importante, autant pour ne pas submerger les évaluateurs d'articles que pour donner le temps aux auteurs de produire de meilleurs articles. Évidemment, pour des domaines où la recherche avance très rapidement, il se pourrait que ceci doive être complémenté avec des banques d'archivages libres, dont nous discuterons plus loin.

Nous avons aussi déjà beaucoup parlé du système de récompenses pour les scientifiques, et comment celui-ci n'est actuellement pas aligné avec les valeurs et l'intégrité tenues par la démarche scientifique et les scientifiques en général. Il serait effectivement grand temps de commencer à récompenser les comportements que l'on souhaite – c'est-à-dire qui sont connus pour améliorer la qualité de la recherche¹³¹ – si on veut que ces comportements deviennent monnaie courante. Il faut inclure dans

l'évaluation des scientifiques, pour des promotions ou du financement, des critères d'intégrité, de collaboration et de transparence, critères que chaque scientifique pourrait intégrer dans une section à cet effet dans ses demandes de financement et dans son CV. Concrètement, ceci inclut des comportements favorisant la transparence et la reproductibilité de la recherche, comme par exemple le fait de pré-enregistrer ses expériences, rendre public tous les résultats peu importe leur significativité, ou partager publiquement toutes ses données et tous ses codes d'analyses^{131,141}. Lorsque justifié, la publication de toutes les données pourrait même être accompagnée d'un certain copyright afin de garantir aux auteurs de récolter les fruits de leurs efforts¹⁷⁷.

En parallèle avec le partage complet des données, Simmons et coauteurs¹⁷⁹ ajoutent qu'il serait important d'encourager aussi la divulgation complète des méthodes et de chaque décision prise durant une étude, dans le but de réduire les prises de décision subjectives et biaisées de la part des auteurs, honnêtes ou pas. Ils fournissent d'ailleurs une liste de six points à suivre pour les auteurs, ne coûtant que peu de temps supplémentaire et qui, assumant l'honnêteté générale des scientifiques, devraient réduire de beaucoup les faux-positifs dans la littérature publiée. Cette liste inclut de présenter toutes les variables récoltées et analysées, tous les tests essayés et utilisés, et surtout une justification pour toute décision par rapport au design expérimental, décisions qui devraient idéalement être prises avant le début de l'expérience. Ils résument l'effet escompté et l'importance de leurs propositions ainsi :

« Ces solutions ne libéreront pas les chercheurs de la pression pour publier, mais elles limiteront ce que les auteurs seront capables de justifier comme

acceptable envers les autres et envers eux-mêmes. Nous devrions accueillir à bras ouverts ces critères de divulgation, comme si la crédibilité de notre profession en dépendait. Car c'est le cas. » ^{Note 13}

Évidemment, tout ceci ne pourrait fonctionner qu'avec une vigilance professionnelle de la part des réviseurs¹⁷⁹. Il faudrait aussi s'assurer que les scientifiques qui veulent participer à cette amélioration de la publication scientifique ne soient pas pénalisés pour ce faire, ce qui pourrait se produire si leur description des méthodes est perçue comme trop longue, complexe, ou désordonnée. Au contraire, il faudrait que leur participation soit valorisée lorsqu'ils se font évaluer. Encore une fois, pour que tout cela fonctionne, il faudrait des réviseurs qui font preuve de beaucoup de réflexion et de raison, ce qui n'est pas gagné d'avance. D'ailleurs, dans un système où la révision par les pairs est si importante, produire des révisions compétentes et respectueuses devrait aussi être encouragé et faire partie de l'évaluation des scientifiques¹³¹. Voir la section « Révision par les pairs » pour plus de détails sur tout ce qui a trait à ce sujet.

Par rapport à ne pas pénaliser les scientifiques pour faire la bonne chose, la même ligne de pensée s'applique également aux rétractions d'erreurs honnêtes⁶⁷. En effet, bien souvent les rétractions sont automatiquement reliées à de l'inconduite, chose qui n'est pas du tout certaine. Il est donc important de changer cette perception en indiquant toujours de façon claire et précise

¹³ Traduction libre de la part des auteurs. Citation originale: "These solutions will not rid researchers of publication pressures, but they will limit what authors are able to justify as acceptable to others and to themselves. We should embrace these disclosure requirements as if the credibility of our profession depended on them. Because it does."

les causes de chaque rétraction, facilitant la tâche pour les revues et les auteurs de corriger ces erreurs honnêtes⁶⁷. Sans ce changement, certains sont alors nécessairement dissuadés de rapporter leurs erreurs découvertes une fois un article publié, par peur des effets négatifs sur leur réputation²⁰⁴, ce qui nuit à l'avancement de la science.

Quant aux incompréhensions statistiques, qui sont de toutes évidences très répandues¹⁹⁴, il serait non seulement souhaitable de fournir une meilleure éducation statistique aux scientifiques, mais aussi d'engager les services de statisticiens professionnels, autant dans les revues pour la révision^{6,206} qu'à l'intérieur de chaque laboratoire pour aider directement les scientifiques dans leurs analyses¹⁷⁷. Il n'y a cependant pas encore assez de motivations pour ce faire et, par conséquent selon le principe d'offre et de demande, pas assez de statisticiens.

La réallocation des fonds de recherche est aussi une solution concrète qui devrait être considérée très sérieusement et sous plusieurs angles. Déjà, un financement accru de la recherche visant à mieux comprendre comment les scientifiques et les revues sont évaluées, ainsi que l'impact de ces évaluations serait un bon point de départ pour améliorer tout le processus¹³¹. Il y a aussi l'importance de financer des idées originales, nouvelles et risquées dans le sens où elles peuvent prendre beaucoup de temps avant de se transformer en publications¹³¹. Si seule la publication et ses dérivés servent de mesure de performance pour les scientifiques, toutes sortes d'idées ne sont alors jamais poursuivies et étudiées, car aucun financement n'y est alloué.

Un autre aspect crucial de l'allocation du financement concerne les déconnexions mentionnées entre la recherche

produite et le bien-être de l'activité scientifique et de la société. Dans ce contexte, très illustrateurs sont les domaines où la recherche débouche directement sur des produits lucratifs, comme dans la recherche biomédicale par rapport au développement de médicaments, d'ailleurs souvent fortement financée par l'industrie⁹⁷. Dans ces cas, les questions qui seraient les plus intéressantes et pragmatiques pour les cliniciens et donc pour le public sont souvent en conflit avec l'objectif de maximiser les profits. Par exemple, il y a des branches de recherche où l'octroi de brevet et le potentiel de commercialisation sont absents, mais qui néanmoins seraient très utiles pour améliorer les soins aux patients⁹⁷. Réallouer en grande partie les fonds publics de recherche biomédicale – qui eux-mêmes pourraient être renfloués à partir ne serait-ce que d'un minime pourcentage des profits de l'industrie pharmaceutique – à de tels sujets de recherche constituerait certainement une solution améliorant la recherche en générale et les soins apportés à la société⁹⁷.

D'ailleurs, de façon générale, des critères d'utilité de la recherche pour la société pourraient être établis dans le but d'évaluer les scientifiques et d'encourager ce que nous souhaitons vraiment¹³¹. Par exemple, dans les décisions de promotion et d'allocation de financement, l'utilité de la recherche proposée ou déjà faite par un scientifique pourrait peser plus lourd dans l'évaluation. Ce serait une belle façon d'encourager une science utile à tous, quoiqu'il faille faire bien attention dans la détermination de ces critères d'utilité et ne pas tomber encore une fois dans un indice de performance hermétique et inchangeant, ou encore opresser la recherche exploratoire dont nous parlions précédemment, où l'utilité n'est pas toujours si facile à circonscrire. Il faut alors ici faire preuve de beaucoup de

précautions et de sagesse, les critères d'utilité et leur sévérité devant changer d'un domaine de recherche à l'autre.

Quoi qu'il en soit, les problèmes présentés dans cette section représentent un enjeu complexe à plusieurs facettes (auteurs, éditeurs, instituts, organismes subventionnaires) et les solutions devront donc impliquer toutes ces facettes de façon coordonnée et complémentaire si nous voulons avoir une chance de succès^{18,141,164}. De plus, bien que la plupart des propositions de solutions ne discutent jamais de l'évaluation de l'adoption de ces solutions¹³¹, un plan d'évaluation est absolument nécessaire pour tout changement du système si nous voulons en suivre les effets et savoir plus rapidement comment ajuster le tout au fil du temps. Dans tous les cas, il est évident que tout changement important représentera un grand défi et nécessitera d'abord des changements au niveau institutionnel pour que les scientifiques aient ensuite la liberté de faire les changements nécessaires à leur niveau^{18,131}.

LA RÉVISION PAR LES PAIRS

Aujourd'hui, la révision par les pairs est souvent perçue comme nécessaire et comme une garantie de la qualité ou de la pertinence d'un article scientifique. En tant que tel, elle représente souvent le fer-de-lance de l'argumentation des faux sceptiques lorsque possible^{44-46,109,110}, puisqu'ils semblent considérer que seule la révision par les pairs garantisse la scientificité et, de par une lacune de logique scientifique, que son absence garantisse la non-scientificité.

Toutefois, la révision par les pairs n'a été introduite de façon systématique et plus ou moins générale qu'à partir des années 1950³⁰. De fait, avant cette systématisation, les éditeurs (ou comités éditoriaux) prenaient souvent eux-mêmes rapidement la décision de publier et rejetaient un article seulement s'il y avait présence évidente de contenu inapproprié³⁰. Le recours à un avis extérieur se produisait, mais ce n'était pas la norme, et ces avis extérieurs provenaient de véritables experts spécifiquement choisis pour leur expertise supérieure à celle de l'éditeur par rapport à l'article considéré³⁰. Ainsi la révision par les pairs est apparue peu à peu, au cas par cas, sans ordre ni obligation, différemment pour chaque revue et avec beaucoup de résistance. La revue *Nature* par exemple n'a introduit la révision par les pairs de façon formelle dans son processus de publication qu'en 1973¹⁹⁵.

La révision par les pairs s'est imposée en réponse à deux phénomènes³⁰ : 1) le degré de spécialisation méthodologique ne cessant d'augmenter, les éditeurs devaient de plus en plus demander d'opinions extérieures, et surtout 2) le nombre de

scientifiques et d'articles soumis a connu une telle explosion que le besoin de sélectionner et de rejeter beaucoup d'articles qui auraient auparavant été publiés devenait trop important. Le concept de la révision par les pairs était tout de même à ses débuts une activité consciente entre scientifiques, soutenue par les mêmes valeurs d'intégrité régissant la recherche elle-même. Avec l'augmentation de sa popularité et avec sa généralisation grandissante, elle s'est cependant peu à peu transformée en un outil commercial, permettant aux entreprises privées de publication scientifique de justifier l'existence de leurs revues et de vendre ce nouveau service aux scientifiques mêmes qui fournissent gratuitement la main d'œuvre pour ce faire¹⁹⁵.

Aujourd'hui donc, il semble que la révision par les pairs soit appréciée principalement parce qu'elle différencie de façon exclusive les articles scientifiques d'autres types de publications, donnant une sorte de renom tout à fait particulier aux scientifiques¹⁹⁵. De plus, la recherche par rapport à l'impact de la révision par les pairs sur l'activité scientifique n'est pas évidente à faire et est coûteuse, ce qui fait qu'il y en a très peu et donc que nous ne savons pas grand-chose sur son utilité réelle^{101,195}.

De plus, il existe la perception souvent répandue que la révision par les pairs est un phénomène invariable qui a toujours fait partie de la science. Elle fait donc maintenant partie intégrante de l'« identité scientifique », affectionnée par la plupart des scientifiques malgré – ou peut-être à cause de – l'ignorance générale à son sujet. Dans ce sens, la révision par les pairs s'apparente plutôt à une norme sociale qu'à un processus scientifique¹⁹⁵. C'est probablement pourquoi elle reste largement incontestée en tant que source principale de confiance pour

l'information scientifique, malgré les incertitudes qui l'entourent¹⁹⁵.

Notez que le besoin de rejeter beaucoup d'articles, le point qui fût déterminant pour l'établissement de la révision par les pairs, est en lien avec l'économie de temps, d'argent et d'investissement personnel, mais n'a rien à voir avec la rigueur scientifique. En d'autres mots, ça n'a rien à voir avec le contenu. Notez aussi que la pression issue de la spécialisation méthodologique et du nombre d'articles impossibles à publier est en hausse, donnant de plus en plus de pouvoir aux réviseurs quant aux décisions de publication¹⁸⁹. Dans les prochaines sections, nous explorerons donc plus en détails le fonctionnement de la révision par les pairs, les différents types de révisions qui existent déjà, leur efficacité et leurs effets sur la science, ainsi que des alternatives possibles. Pour les lecteurs intéressés, nous suggérons aussi la lecture de « *A multi-disciplinary perspective on emergent and future innovations in peer review* »¹⁹⁵, un manuscrit dédié uniquement à la révision par les pairs et qui en donnent une description plus détaillée, donnant aussi plus d'exemples par rapport à l'histoire de son développement ainsi que par rapport aux différentes questions qui l'entourent. Ce manuscrit a de plus lui-même passé au travers d'une révision ouverte où tous les commentaires des réviseurs sont disponibles, un bel exemple d'une alternative intéressante.

Fonctionnement et appréciation

La révision par les pairs est maintenant omniprésente, un pré-requis pour tout article scientifique qui veut être reconnu. Le

principe est relativement simple et est sensiblement le même partout, avec quelques rares exceptions dont nous parlerons plus loin. En bref, un éditeur scientifique qui reçoit un article en soumission pour être publié choisira et demandera à deux (ou trois) autres scientifiques ou étudiants dans le même domaine de recherche de lire, réviser et commenter l'article en question par rapport à son importance, sa qualité et, ultimement, son mérite à être publié. La sélection des réviseurs reste mystérieuse dans bien des cas³⁴, mais les auteurs sont souvent tenus de suggérer des réviseurs potentiels pour alléger la tâche de l'éditeur qui sinon doit en trouver lui-même. La décision finale reste à la discrétion de l'éditeur.

La procédure la plus commune est que les réviseurs restent anonymes, mais connaissent l'identité des auteurs¹³⁵. Après examen de l'article, les réviseurs donnent leur opinion sur la nécessité de le publier, le modifier ou le rejeter. En se basant sur ces opinions, l'éditeur prend la décision finale et la communique aux auteurs, soit rejetant l'article ou demandant de faire les modifications suggérées par les réviseurs. Un article est très rarement accepté sans modifications ; nous n'avons jamais entendu parler d'un tel cas.

La révision par les pairs sert donc de contrôle de qualité et détermine essentiellement quels travaux seront publiés, lus et cités et quels scientifiques seront financés et promus^{139,186}. Les scientifiques se disent généralement satisfaits par la performance de la révision par les pairs^{135,139,186}, quoique certaines sources rapportent que moins de 50% des scientifiques en recherche biomédicale voient le système de révision par les pairs comme étant juste, transparent ou scientifique⁸⁹.

Si certains perçoivent la révision par les pairs comme le pilier inébranlable sans lequel la science s'écroulerait, d'autres la considèrent – dans sa forme actuelle du moins – comme une nuisance à l'avancement de la science (voir Campanario^{34,35} et Souder¹⁸⁶ pour des revues de littérature sur les deux camps). La seule chose sur laquelle tous seront peut-être d'accord est que la révision par les pairs est imparfaite. Il paraît en effet maintenant évident que ce système est truffé de problèmes affectant la recherche scientifique^{11,187,206}. La procédure même est remarquablement non-scientifique^{11,161} : un échantillon faible et non-aléatoire (les réviseurs), plusieurs biais probables et des intervalles de confiance énormes... voilà un design expérimental qui n'augure rien de bon.

Ces problèmes ne sont pas d'hier, le premier congrès pour en parler ayant eu lieu en 1989¹⁶¹. Malgré tout, il semble y avoir une perception générale selon laquelle un résultat ne peut être sérieux sans révision par les pairs, cette dernière lui conférant non seulement son droit d'existence, mais aussi une preuve de sa fiabilité^{34,135}. Cette attitude dangereusement naïve porte certains scientifiques à prendre pour acquis qu'un article révisé par les pairs ne contient pas d'erreur majeure au niveau de l'application de la démarche scientifique ou de la présentation et de l'interprétation des résultats. Et malgré de nombreux travaux publiés au sujet de la révision par les pairs, il est surprenant de constater que très peu d'études ont effectivement testé la performance de ce système^{101,195}. Son efficacité n'a d'ailleurs jamais été formellement démontrée, ni réfutée¹⁵⁰. Ce manque de données scientifiques est d'autant plus surprenant que la révision par les pairs occupe aujourd'hui une place fondamentale dans la recherche scientifique et invoque un respect que nous pourrions presque qualifier d'indécent.

En 2002, Jefferson et coauteurs¹⁰¹ ont fait une revue de littérature sur le sujet et ont trouvé que les failles des quelques études tentant de mesurer directement la fiabilité de la révision par les pairs rendent toute généralisation impossible. Une étude de qualité sur la question nécessiterait un investissement substantiel en argent et en ressources^{101,186,195} et leur conclusion finale est que :

« Jusqu'à ce que de telles études soient entreprises, la révision par les pairs devrait être considérée comme un processus non vérifié menant à des résultats incertains. » ^{Note 14}

Depuis 2002, nos lacunes de connaissance quant à l'efficacité de la révision par les pairs ne semble pas avoir bien changé¹⁹⁵. Néanmoins, plusieurs études se penchant indirectement sur le fonctionnement et l'évaluation de la révision par les pairs existent, et c'est de ces études dont il sera question dans les prochains paragraphes.

Contrôle de qualité

Rejet d'articles médiocres

Le rôle le plus important de la révision par les pairs devrait être le contrôle de qualité des articles soumis afin de rejeter ceux

¹⁴ Traduction libre de la part des auteurs. Citation originale: “*Until such research is undertaken, peer-review should be regarded as an untested process with uncertain outcomes.*”

qui présentent des failles importantes, et d'améliorer puis de publier ceux qui présentent les résultats les plus pertinents. Explorons d'abord la performance des réviseurs par rapport au rejet d'articles médiocres, incomplets ou erronés.

Un grand défi pour ce faire est d'identifier les situations où les méthodes statistiques sont inadéquates, par exemple lorsque le pouvoir statistique d'une étude est gonflé^{103,142}. Malheureusement, nous avons déjà vu dans les sections précédentes comment des incompréhensions des méthodes statistiques couramment utilisées sont très répandues au sein de la communauté scientifique¹⁹⁴. Les réviseurs, qui ne sont d'autres que les scientifiques constituant cette communauté, ne sont pas entraînés, encadrés, ni rémunérés^{11,38,186} et ils sont par conséquent très souvent trop ignorants ou négligents dans leur révision des méthodes statistiques^{35,112,144,206}. Ceci ouvre les portes de la publication à beaucoup de faux-positifs qui auraient dû être rejetés^{34,92} (Fig. 5). Une solution serait que les revues aient recours à des consultants en statistique, ce qui augmentent clairement la qualité des articles publiés lorsque c'est fait²⁶. Cependant, cela reste rare^{6,206} et rien ne laisse penser que ça changera bientôt.

Bien entendu, la révision est souvent inutile contre les fraudes et l'intention de tromper puisque les réviseurs ne peuvent évidemment pas détecter le mensonge. La révision devrait par contre pouvoir s'assurer que les articles publiés sont reproductibles. Comme nous l'avons vu dans la section « Biais contre la reproduction » cependant, ceci est loin d'être le cas. Il est en effet facile pour les auteurs de ne pas partager la méthode complète de leurs études afin que d'autres ne puissent pas répéter leurs expériences, ce qui se fait souvent dans un souci de garder son avantage compétitif¹⁰. La révision par les pairs ne

semble donc pas non plus en mesure de contrôler l'exactitude et l'intégrité des protocoles de recherche. Notez que nous trouvons tout à fait normal que la révision par les pairs ne puisse pas faire cela, car ceci nécessiterait des connaissances que la plupart n'ont pas ainsi qu'un investissement de temps impossible à fournir pour les scientifiques. Ce qui est anormal, c'est le respect frôlant l'adoration que la révision par les pairs inspire trop souvent.

Une étude très parlante par rapport à la performance des réviseurs dans le rejet d'articles médiocres a été conduite par Godlee et coautrices⁷⁷. Elles ont rajouté à un article qui avait déjà été accepté pour publication huit erreurs importantes par rapport à la méthodologie ou aux concepts. Elles ont ensuite fait réviser cet article transformé par 221 réviseurs. En moyenne, 2 erreurs ont été détectées, avec 16% des réviseurs n'ayant détecté aucun problème. 33% ont recommandé la publication avec modifications mineures contre 30% qui ont recommandé le rejet. Seulement 10% ont détecté plus de 3 erreurs. Nylenna et coauteurs¹⁴⁴ ont mené une étude similaire et obtenu des résultats très similaires.

Même si plus d'études semblables seraient nécessaires pour tirer des conclusions solides, les résultats déjà disponibles indiquent que des études clairement défailtantes peuvent facilement être acceptées pour publication. Rennie¹⁶¹ résume la situation de la révision par les pairs ainsi :

« Il semble n'y avoir aucune étude trop fragmentée, aucune hypothèse trop triviale, aucune référence de littérature trop biaisée ou trop intéressée, aucun design de recherche trop pervers, aucune méthodologie de recherche trop

bousillée, aucune présentation des résultats trop inexacte, trop obscure et trop contradictoire, aucune analyse trop intéressée, aucun argument trop circulaire, aucune conclusion trop insignifiante ou trop injustifiée et aucune grammaire ou syntaxe trop offensante pour qu'un article finisse par être publié. » ^{Note 15}

Acceptation d'articles importants

Si la révision par les pairs semble avoir du mal à rejeter les travaux médiocres, incomplets ou erronés, assure-t-elle au moins la publication d'articles de qualité et d'importance? Malheureusement, l'incompétence fréquente des réviseurs par rapport aux statistiques est une lame à deux mauvais tranchants. Par exemple, l'existence de « modes » pour l'analyse des données¹⁶⁶ fait que beaucoup de scientifiques préfèrent une certaine méthode et l'exigent lors d'une révision, qu'elle soit appropriée ou non^{189,206} (voir Fig. 5). Cela met souvent les auteurs dans la situation désagréable d'avoir deux recommandations méthodologiques bien différentes et une demande générale de l'éditeur de satisfaire aux commentaires des réviseurs^{112,189}. Au mieux, ceci entraîne une redondance de la présentation des résultats dans l'article, au pire un ajout de méthodes

¹⁵ Traduction libre de la part des auteurs. Citation originale : “*There seems to be no study too fragmented, no hypothesis too trivial, no literature citation too biased or too egotistical, no design too warped, no methodology too bungled, no presentation of results too inaccurate, too obscure, and too contradictory, no analysis too self-serving, no argument too circular, no conclusions too trifling or too unjustified, and no grammar and syntax too offensive for a paper to end up in print.*”

inappropriées et le retrait de résultats plus représentatifs de la réalité.

L'incompétence – parfois aussi la négligence – de la part des réviseurs est d'autant plus dommageable que les éditeurs s'en remettent de plus en plus à leurs recommandations^{112,163}. De fait, si un des réviseurs recommande de rejeter l'article (pour quelque raison), le manuscrit a très peu de chance d'être accepté^{167,189}. Pourtant, les réviseurs ne sont pas d'accord sur leur décision ou sur la priorité de publication plus souvent que ce qu'on pourrait s'attendre si leur opinion était choisie au hasard^{34,74,167}. Si la révision doit vraiment se faire, le meilleur compromis entre sa pertinence et son coût semble être d'utiliser cinq réviseurs de façon simultanée¹⁸⁴. En fait, la raison pour laquelle un si petit nombre de réviseurs est utilisé est qu'en utiliser plus serait une charge trop pénible pour la communauté scientifique. Il va sans dire que ceci n'a rien à voir avec la scientificité du processus.

De plus, pour que la révision puisse porter ses fruits, peu importe le nombre de réviseurs, il faudrait que les éditeurs prennent le temps d'examiner soigneusement le contenu de l'article pour pouvoir bien comprendre les commentaires des réviseurs. Nous revenons encore une fois à l'importance du contenu, à l'importance de la personne portant l'habit du moine. Ce n'est pas parce qu'un réviseur (ou plusieurs) indique qu'il ne trouve pas l'article pertinent que l'article ne l'est pas. Malheureusement, les éditeurs ne semblent pas avoir souvent le temps (ou l'envie?) de faire ce travail et c'est pourquoi l'avis des réviseurs a tellement de poids lors d'une soumission d'article.

Suppression de l'innovation

Le problème considéré comme le plus dommageable pour la science par rapport au rejet non-fondé après une révision par les pairs est le biais contre l'innovation¹⁶⁹. Les réviseurs sont effectivement plus susceptibles de recommander la publication d'articles qui sont favorables à leurs idées préconçues, à leur carrière ou à leur égo, et de recommander le rejet d'articles qui ne le sont pas^{11,18,35,149,166,185}. Il y a beaucoup d'exemples où des articles innovateurs, finalement considérés comme des points tournants de la recherche dans leur domaine respectif, ont premièrement été rebutés par la révision par les pairs³⁴. Voir Horrobin⁹¹ et Spier¹⁸⁷ pour des listes d'exemples.

Si le biais contre l'innovation n'est peut-être pas la norme, il n'est certainement pas exceptionnel^{11,91,169}. Le physicien et philosophe Thomas Kuhn, dans son livre *The Structure of Scientific Revolutions*¹⁰⁶, explore et développe les concepts de paradigme et de révolution scientifique, lesquels peuvent aider à comprendre cette situation. Kuhn met en lumière des cycles de recherche dite « normale », où les questions et les méthodes de recherche sont limitées par le paradigme en vigueur, ne servant plus à « découvrir », mais bien à avancer au maximum notre compréhension dudit paradigme. Ces cycles sont entrecoupés par des changements de paradigme suivant une accumulation d'anomalies et de découvertes inattendues. Ici, le paradigme scientifique se différencie du dogme scientifique en ce qu'il est sous-entendu qu'il sera supplanté par un paradigme tenant compte des nouvelles découvertes.

Cependant, le fait que les paradigmes scientifiques soient implicitement sujets à révolution n'empêche pas certains

individus de pratiquer le dogmatisme avec ces paradigmes. Ainsi, pour un scientifique donné, le paradigme dans lequel il a évolué peut facilement devenir une sorte de doctrine scientifique influençant grandement son opinion en tant que réviseur. Ceci peut décourager l'innovation des scientifiques voulant à tout prix publier leurs travaux dans les revues prestigieuses⁴³. De plus, l'idée selon laquelle tous les scientifiques seraient des « pairs » ne semble pas pouvoir s'appliquer dans le contexte de l'innovation⁹¹ ; au contraire, la personne se croyant experte est tentée de croire qu'elle peut juger de ce qui sera découvert ou non.

Ceci étant dit, les articles importants finissent probablement toujours par être publiés, même si le processus peut prendre des années et de nombreuses resoumissions¹⁸⁷. En effet, beaucoup d'articles – et pas seulement des perles – sont premièrement refusés une ou plusieurs fois puis acceptés par d'autres réviseurs dans une autre revue^{11,187}. Loin d'être rassurant, ceci exprime clairement que la révision par les pairs ne permet pas d'éviter la publication des articles qui font défaut à l'application de la démarche scientifique^{11,35}. Il semble donc que, bien que l'innovation ne soit pas bienvenue ni supportée par la révision par les pairs, les véritables obstacles à la publication soient la persévérance des scientifiques et leur capacité à supporter la frustration^{166,187}.

Il en est tout autrement pour ce qui est de l'allocation du financement, qui est aussi régie par la révision par les pairs. Effectivement, si l'innovation est bloquée au niveau de la demande de financement, alors la recherche ne se fait pas et il y a une réelle suppression de la créativité et de l'originalité^{4,91,187}. Ce qui a été dit sur la publication est alors encore plus important dans le contexte de l'attribution des subventions de recherche, où

les scientifiques les plus innovateurs savent qu'ils auraient peu de chance d'obtenir du financement s'ils disaient clairement ce qu'ils prévoient faire⁹¹. Ainsi, une énorme perte de temps se produit toujours lors de la préparation d'une demande de financement, car les scientifiques tentent désespérément d'intégrer les mots, les sujets ou les méthodes qui sont en vogue à ce moment-là^{4,187}. Plusieurs décisions importantes pour la recherche sont donc dirigées non pas par les évidences présentées par les scientifiques ou par leur créativité (le contenu), mais par les conventions et les modes⁹⁵ (le contenant). Un scientifique en demande de financement peut alors se sentir limité dans sa liberté de recherche si ses travaux l'emmènent au-delà des paradigmes en vigueur. Ainsi, si l'innovation résulte en effet des subventions, il est plus probable que ce soit malgré la révision par les pairs qu'à cause d'elle^{168,187}.

Au final, il semble que le bénéfice principal de la révision par les pairs soit d'améliorer la qualité de la présentation des articles qui sont finalement publiés^{51,101,135}, quoique certains questionnent même ce bénéfice par rapport aux coûts de la révision et des multiples soumissions souvent nécessaires pour publier²⁴.

Quoi qu'il en soit, notez que tout le monde est biaisé d'une manière ou d'une autre, plus ou moins inconsciemment, et qu'il est très difficile pour n'importe qui de questionner ses propres idées préconçues. Le jugement des scientifiques est donc bien loin d'être fiable à 100% et il est facilement manipulable^{11,32,186}. Ceci a des répercussions autant pour le rejet non-fondé de certaines idées que par rapport à une sorte de contrat d'erreur¹³ entre d'un côté les auteurs, qui présentent l'information non pas pour qu'elle soit la plus facilement évaluable, mais la plus crédible

possible, et de l'autre côté les réviseurs qui préfèrent croire que l'information – si elle est cohérente avec leurs idées préconçues – est vraie plutôt que d'en douter et d'investiguer.

Biais de révision

Nous avons déjà vu que plusieurs biais de publication existent et influencent ce qui est publié et donc aussi la recherche scientifique qui se fait. Comme les réviseurs sont les « gardiens » de la publication scientifique, cela ne vous étonnera pas d'apprendre qu'une partie importante de ces biais se joue au moment de la révision. En effet, les réviseurs portent moins d'attention à la méthodologie lorsque les résultats sont significatifs³⁴ et sont plus susceptibles de recommander le rejet des résultats non-significatifs^{34,103,191} (voir cependant Song et coauteurs¹⁸⁵). Pis encore, les réviseurs semblent défavoriser les articles présentant des répliques de résultats^{34,149}, quoiqu'il serait difficile de déterminer si ceci n'est pas tout simplement le résultat des instructions données par les éditeurs¹⁴⁹, comme nous l'avons vu précédemment. Ces tendances biaisées de la part des réviseurs, qui vont dans le sens des tendances éditoriales, vont à l'encontre de la démarche scientifique. Combinées avec le manque d'entraînement et de temps des réviseurs, elles peuvent amoindrir plus encore la contribution scientifique de la révision par les pairs.

Plusieurs autres types de biais spécifiques à la révision existent et sont susceptibles d'affecter la publication d'un article selon d'autres critères que sa valeur scientifique^{135,163} (Tableau 1). Ces biais sont une illustration claire de ce qui se passe lorsqu'on

se concentre sur le contenant de l'information plutôt que sur le contenu. Par exemple, il y a un biais clair en faveur des auteurs qui vivent aux États-Unis d'Amérique ou en Europe¹⁸⁶ et dont la langue maternelle est l'anglais^{189,206}. Notez que le biais par rapport au langage concerne l'attitude des réviseurs envers l'article, même pour les articles écrits avec un bon anglais. Un biais par rapport à la nationalité s'exprime par une tendance des réviseurs à favoriser les auteurs de leur pays³⁵. Un autre biais de révision est en lien avec l'âge des réviseurs : les jeunes scientifiques, souvent assidus, produisent normalement de meilleures révisions que leurs collègues plus âgés⁷⁷.

Il y a aussi le biais par rapport à la réputation d'un auteur et à son autorité intellectuelle¹³⁰ : les idées innovatrices d'un scientifique bien connu qui travaille pour un institut prestigieux ont plus de chances d'être acceptées que celles d'un scientifique inconnu^{10,11}. Pourtant, comme nous le répétons depuis le début, l'habit ne fait pas le moine et nous serions en droit de s'attendre à ce que l'évaluation d'idées scientifiques se fasse dans la reconnaissance de ce simple fait. Vous aurez cependant déjà bien compris qu'il y a plusieurs facteurs influençant la révision et qui n'ont rien à voir avec la contribution scientifique d'un travail, c'est-à-dire qui n'ont rien à voir avec son contenu. Notez que puisque la recherche sur le sujet n'en est encore qu'à ses débuts⁷⁴, il est réaliste de penser que d'autres biais de révision non répertoriés existent et influencent la recherche.

Tableau 1. Liste de critères recensés comme ayant un impact sur le succès d'une étude lors de la révision par les pairs, mais n'ayant rien à voir avec la rigueur scientifique. Notez que cette liste n'est probablement pas exhaustive.

Les chances d'une réception <u>positive</u> augmentent lorsque...	Les chances d'une réception <u>négative</u> augmentent lorsque...
• l'étude reste dans les limites du paradigme en vigueur. • les résultats sont significatifs. • l'auteur est Américain ou Européen. • la langue maternelle de l'auteur est l'anglais. • la nationalité de l'auteur est la même que celle du réviseur. • l'auteur et/ou son institut d'affiliation a une réputation d'importance. • le réviseur a peur des représailles ou espère des faveurs.	• l'étude déroge des idées préconçues du réviseur. • les résultats sont non-significatifs. • le réviseur est un compétiteur. • la langue maternelle de l'auteur n'est pas l'anglais. • l'étude présente une reproduction de résultats. • l'auteur est inconnu.

Inconduites professionnelles et biais de révision

Les conflits d'intérêts

Un type particulier de biais de révision concerne les conflits d'intérêts, lesquels abondent et sont souvent dus à la compétition, donc à la concurrence financière et/ou idéologique entre scientifiques^{4,10,11,94,112,166,186,187,199,206}. Ces conflits, dans un contexte de révision par les pairs, peuvent mener à diverses inconduites plus ou moins sérieuses de la part des réviseurs. Ceci est accentué par leur énorme pouvoir, puisque dans la plupart des cas ils restent anonymes et ne sont jamais tenus responsables de leurs commentaires ou de leurs demandes^{11,112,186}.

Imaginons par exemple un scientifique qui supervise un étudiant au doctorat travaillant dans son laboratoire. On lui soumet un article en révision. Cet article a le même contenu que la thèse de son étudiant. Il se retrouve en conflit d'intérêts parce qu'en approuvant la publication de l'article qui lui a été soumis, il compromet la publication de l'article sur lequel il travaille avec son étudiant. Il est aussi en conflit d'intérêts si les résultats présentent une vision qui va à l'encontre de celle de ses travaux²⁰⁶. Il est aussi en conflit d'intérêts s'il concurrence l'auteur de l'article pour les mêmes sources de subventions^{112,150,206}. Il est également en conflit d'intérêts s'il pense se faire reconnaître ; il peut alors avoir peur des représailles²⁰⁶ ou espérer des faveurs^{35,112}.

Bref, il est impossible pour un éditeur de prévoir tous les biais de révision en lien avec les conflits d'intérêts. Ces biais sont si nombreux que chaque scientifique peut probablement raconter une ou plusieurs anecdotes où un certain conflit d'intérêts a entravé la publication de ses travaux¹¹ ; nous pourrions

certainement le faire. Malencontreusement, il semble que sous les conditions actuelles du système de publication, un pourcentage relativement faible de réviseurs agissant par rapport à leurs conflits d'intérêts serait suffisant pour diminuer de façon importante la qualité de la science qui est publiée¹⁵⁰.

Un constat aberrant par rapport aux conflits d'intérêts en lien avec la révision par les pairs est que même si les commentaires ou demandes des réviseurs sont souvent subjectifs, vagues et accompagnés de justifications minimales, ils ont généralement plus de poids que l'opinion des auteurs¹¹. Ces demandes peuvent donc facilement être inappropriées, impossibles à satisfaire ou même irrationnelles, sans aucune conséquence pour ceux qui les font^{11,112,189}. Une telle constellation est assurément une terre fertile pour agir sur ses conflits d'intérêts. D'ailleurs, dans une discussion de groupe entre scientifiques menée par Anderson et coautrices¹⁰, les participants avaient l'impression générale que la révision par les pairs représentait une opportunité unique et même protégée pour les compétiteurs de profiter d'eux.

Une peur très présente chez les participants était que, dans un processus de demande de financement, les idées exposées dans leur proposition de recherche soient volées par leur compétiteurs, qui en même temps s'assureraient que la demande de financement soit refusée. Il arrive en effet que des réviseurs dénigrent ou retardent la publication ou le financement d'un projet sous révision afin de publier soi-même sur le même sujet^{10,35,150,163}. Dans ce sens, les scientifiques trouvent particulièrement frustrant ce différentiel de pouvoir entre les arguments des auteurs et celui des réviseurs¹⁰.

Dans l'extrême, ces conflits d'intérêts peuvent mener à une véritable campagne de salissage professionnel de la part d'un scientifique influent pour empêcher une idée d'être publiée ou financée. Horrobin⁹¹ en récite un exemple horrifiant et décrit à quel point la communauté scientifique resta inerte devant un tel phénomène, ne faisant rien pour arrêter ces comportements pathologiques de scientifiques influents. Notez que ce type de comportement est inhabituel et que peu de réviseurs semblent consciemment malhonnêtes¹¹². Cependant, ce n'est pas un phénomène rarissime non plus, et le fait que ce problème soit relégué au statut d'exception négligeable a sans aucun doute créé de grandes injustices et beaucoup de souffrance⁹¹.

Il faut aussi mentionner des actes relevant d'une éthique exceptionnelle et qui ne sont pas rarissimes non plus. Par exemple, il y a des cas où des scientifiques retardent leurs propres travaux pour qu'un article sur le même sujet, soumis pour révision à leur chef de laboratoire, soit publié en premier même si le chef essaie de retarder la révision tant qu'il peut¹⁰. Ce genre de comportement est exemplaire et montre bien à quel point les scientifiques, en fait, préféreraient se concentrer sur la recherche de qualité plutôt que de jouer le jeu de la compétition acharnée, un jeu soutenu par les systèmes de publication et d'évaluation actuels.

Les conflits d'intérêts ne proviennent cependant pas seulement de la part des réviseurs. En effet, l'idéologie et les tendances des réviseurs sont parfois bien connues des éditeurs. Ces derniers sont alors à même de s'en servir selon leurs propres intérêts. Par exemple, ils peuvent sélectionner des réviseurs plus ou moins sévères dépendamment de leur propre perception de l'article. En choisissant des réviseurs particulièrement sévères, ils

agissent comme ces réviseurs malhonnêtes que nous venons tout juste de décrire. D'un autre côté, en choisissant des réviseurs particulièrement cléments, ils peuvent aussi créer un réseau d'éditeurs, de réviseurs et d'auteurs profitant tous du favoritisme³⁵. Évidemment, le favoritisme n'a pas nécessairement besoin de l'aide des éditeurs pour se produire, amis et collègues pouvant facilement en profiter puisque les auteurs doivent suggérer des réviseurs en soumettant leur article³⁵. Tout ceci représente aussi des conflits d'intérêts biaisant le processus de révision par les pairs.

Un tout autre type d'inconduite reliée aux conflits d'intérêts se produit lorsque des réviseurs demandent aux auteurs de rajouter des références superflues¹⁶³. Ceci arrive lorsque les réviseurs veulent augmenter le nombre de citations de leurs travaux et profitent de leur position de pouvoir sur les auteurs pour ce faire. Comme les auteurs veulent faire plaisir aux réviseurs pour que ceux-ci donnent un avis favorable, ils ajoutent souvent toutes références demandées. Ce genre de situation se produit aussi lors de la demande de financement. Les auteurs ressentent alors le besoin d'intégrer en référence – même si ce n'est pas pertinent – les travaux de scientifiques influents dans le domaine, des grands noms, de ceux qui sont susceptibles d'être choisis comme réviseurs, sous peine de ne pas obtenir le financement s'ils ne s'adonnent pas à cette mascarade¹⁰.

Fabrication de réviseurs et rudesse

En plus des conflits d'intérêts, plusieurs autres types d'inconduites entourent la révision par les pairs. Un très bon

exemple a trait à la manipulation des personnes qui seront choisies comme réviseurs. En effet, nous avons déjà mentionné que lors de la soumission d'un article, une liste de réviseurs potentiels est demandée aux auteurs, qui ne suggèrent alors évidemment pas des personnes qui seront particulièrement critiques. Ces suggestions varient dans leur contenu ainsi que dans leur honnêteté. Il peut s'agir i) de personnes que les auteurs considèrent particulièrement compétentes pour juger de leurs travaux, ii) de personnes avec lesquelles les auteurs ont déjà eu de bonnes expériences, iii) d'amis ou de collègues qui pourraient potentiellement pratiquer le favoritisme, ou encore iv) de vraies personnes, mais accompagnées d'une fausse adresse e-mail, cette dernière menant à un compte créé par l'auteur afin de se donner soi-même une révision positive.

D'ailleurs, en 2015, la source principale de rétraction d'article au sein de la maison d'édition *Biomed Central* était la manipulation du processus de révision par les pairs¹³⁴. Il existe même des agences d'aide pour auteurs scientifiques, à buts lucratifs, qui se spécialisent dans la manipulation du processus de révision par les pairs^{39,134}. Officiellement, ces agences offrent des services tels qu'améliorer l'écriture en anglais ou s'occuper du processus fastidieux de soumission d'article. Certaines d'entre elles sont cependant spécialisées – à l'insu des auteurs ou non – dans la fabrication de l'information de contact de vrais scientifiques afin de les suggérer comme réviseurs, garantissant des révisions positives pour les auteurs^{39,134}. Il semble donc que la révision par les pairs soit bien facilement manipulable.

À cela s'ajoutent plusieurs inconduites telles que des commentaires agressifs ou rudes^{186,189}, la proportion de tels commentaires étant plus élevée envers les auteurs dont la langue

maternelle n'est pas l'anglais^{189,206}. Un sondage auprès de 220 scientifiques en recherche biomédicale a d'ailleurs révélé que 18% avaient déjà reçu des attaques personnelles de la part de réviseurs¹⁶³. Une conséquence de ce manque de professionnalisme est que plusieurs scientifiques ayant subi des révisions persécutrices pensent honnêtement qu'ils doivent aller dans ce sens lorsque c'est à leur tour de réviser, que c'est « comme ça que ça fonctionne »³⁵. Leurs révisions sont alors beaucoup plus dures et sévères qu'elles devraient l'être.

Bilan

La révision par les pairs est difficile à évaluer du point de vue scientifique à cause du manque d'information quant à sa contribution pour la science et à cause de la difficulté surprenante avec laquelle le processus se définit^{183,195}. Ce processus semble en fait naviguer sur une ligne plus ou moins large entre la suppression de l'innovation et la suppression de la réplication des résultats (Tableau 1). Bien qu'elle permette certainement souvent de filtrer des articles médiocres ainsi que de pousser la publication d'articles importants, elle semble tout aussi susceptible de permettre la publication d'articles défaillants et d'empêcher (au moins pour un temps) la publication d'articles pertinents. À son meilleur, elle procure une amélioration de la présentation de la recherche, mais d'une façon qui apparaît plutôt inefficace en vu des coûts impliqués. Smith¹⁸³ explore le sujet en profondeur et affirme que :

« [...] nous n'avons que peu de raisons de croire en l'efficacité de la révision par les pairs, mais nous

avons des preuves considérables de ses tares. En plus d'être médiocre dans la détection de failles évidentes [...] la révision par les pairs est lente, coûteuse, prodigue du temps des chercheurs, hautement subjective, apparente à une loterie, encline aux biais et facilement abusée. » ^{Note 16}

Ces affirmations contrastent fortement avec la confiance irrationnelle en la révision par les pairs, confiance observée autant chez les scientifiques que dans le public. Ce système est abondamment vanté en tant que champion du caractère autocritique de la science, alors qu'il n'est lui-même pas soumis à la démarche scientifique. Campanario³⁴ dénote une raison apparente de ce paradoxe de façon tout à fait explicite :

« Une fois un article accepté, nous désirons tous croire que son contenu est valide. » ^{Note 17}

Ce désir de croire en l'efficacité et en l'objectivité de la révision par les pairs se rapproche en fait plutôt d'une croyance aveugle^{11,161}. Cet état des faits est causé en partie par une ignorance générale sur le sujet. Par exemple, le facteur d'impact des revues est généralement perçu comme étant proportionnel à la qualité de la révision par les pairs qui s'y fait¹³⁹. À la lumière des

¹⁶ Traduction libre de la part des auteurs. Citation originale : “[...] we have little evidence on the effectiveness of peer review, but we have considerable evidence on its defects. In addition to being poor at detecting gross defects [...] it is slow, expensive, profligate of academic time, highly subjective, something of a lottery, prone to bias, and easily abused.”

¹⁷ Traduction libre de la part des auteurs. Citation originale : “We all want to believe that once a paper is accepted, its content is valid.”

recherches faites dans ce domaine, il est évident que ceci est une mésinterprétation de la réalité et que dans l'intérêt de la science, il faudrait fortement remettre en question non seulement l'utilisation du facteur d'impact, mais aussi sa signification¹²².

Pourquoi cette apparente ignorance du système règne-t-elle parmi les scientifiques? D'abord, le besoin et le défi de développer sa propre carrière en recherche rendent l'idée de questionner le système peu attrayante. Suivre le courant devient plus pratique¹²². D'ailleurs, les problèmes d'intégrité scientifique dans le processus de révision sont connus depuis longtemps, mais les changements significatifs ne se produisent qu'extrêmement lentement et avec de grandes difficultés^{186,195}, démontrant une certaine réticence de la part de la communauté scientifique pour agir contre ces problèmes. Cette réticence est en partie justifiable, puisque des changements profonds du système doivent être évalués et appliqués avec beaucoup de soin si nous ne voulons pas rempirer la situation. Cependant, selon notre expérience, cette réticence est aussi en partie due au fait que les scientifiques ne désirent pas investir temps et énergie pour s'informer des problèmes et pour participer à trouver et implémenter des solutions.

De plus, ceux qui écrivent en faveur du système de révision par les pairs ne discutent souvent pas des mêmes points que ceux qui soutiennent le besoin d'un changement plus radical¹⁶⁸. La plupart des scientifiques sont effectivement réticents à inclure certains problèmes dans leur évaluation de la révision par les pairs et rares sont ceux qui en présentent correctement plusieurs. Séparer les différents problèmes sans tenir compte de leur interaction est un autre facteur déterminant par rapport à la vision qu'on a de la révision par les pairs. Ceci a souvent pour effet de

minimiser la gravité de la situation et d'apporter de la confusion au sein des débats.

Par exemple, il y a un débat sans consensus à ce jour quant aux effets découlant du fait que les réviseurs ne sont pas plus souvent d'accord que ce qui serait attendu par la chance¹⁵⁰. Certains prétendent que c'est une bonne chose puisque les éditeurs ont alors accès à plusieurs points de vue différents^{105,113}. D'autres considèrent ce fait comme un indice de la non-fiabilité de la révision^{74,167}. Pourtant, ce débat n'a que très peu d'intérêt lorsque l'ensemble de la situation est considéré. Dans les faits, les éditeurs – surtout ceux des revues prestigieuses – se contentent le plus souvent de simplement rejeter un article si les réviseurs ne sont pas unanimes sur son importance^{167,189}, sans évaluer les différents points de vue présentés par les réviseurs. Il est donc inutile dans ce contexte de se demander si la présence de plusieurs points de vue est en soi une bonne chose ou non.

Les défenseurs de la révision par les pairs discutent souvent eux-mêmes de certains problèmes du système, avec références à l'appui. Toutefois, un phénomène bien curieux est que lorsque vient le temps de se prononcer sur les bienfaits ou la nécessité dudit système, ces défenseurs s'en remettent alors à des croyances non scientifiques, puisqu'aucune preuve de sa performance n'existe¹⁸³. Ils vont par exemple affirmer que malgré ses biais, la révision par les pairs est essentielle pour le bon fonctionnement de la science, ceci n'étant supporté que par le désir d'y croire ou des arguments triviaux (voir par exemple Kassirer et Champion¹⁰⁵ et Lee et collègues¹¹³). Un exemple très parlant est la revue de littérature sur la révision par les pairs effectuée par Lee et coauteurs¹¹³. Dans leur conclusion, ils ne mentionnent que les biais qui étaient peu supportés par la

littérature et omettent complètement ceux dont l'existence était claire selon leur propre revue, c'est-à-dire les biais de publication et les biais en lien avec l'influence de la subjectivité des réviseurs. Ils négligent donc de se prononcer sur les biais les plus dommageables pour la science malgré que leur propre recherche en démontrait l'existence.

Ainsi, les défenseurs du système de révision actuel semblent déterminés à ignorer i) toute vision d'ensemble, ii) l'historique de la révision par les pairs et iii) le fait que plusieurs organismes subventionnaires n'ont jamais utilisé la révision par les pairs, mais produisent d'excellents résultats¹⁶⁸. Dans ces conditions, il semble étrange que la communauté scientifique mette autant d'effort et d'énergie dans un système dont les bases sont si douteuses¹¹. Smith¹⁸³ conclut de façon éloquente sur le sujet :

« Ainsi la révision par les pairs est un processus défectueux, emplie de tares facilement identifiables et avec peu de preuves de son efficacité. Néanmoins, il est probable qu'elle restera centrale pour la science et pour les revues, car il n'y a pas d'alternative évidente et parce que les scientifiques et les éditeurs continuent à croire en elle. Comme il est étrange que la science soit ancrée dans une croyance. » ^{Note 18}

¹⁸ Traduction libre de la part des auteurs. Citation originale : *“So peer review is a flawed process, full of easily identified defects with little evidence that it works. Nevertheless, it is likely to remain central to science and journals because there is no obvious alternative, and scientists and editors have a continuing belief in peer review. How odd that science should be rooted in belief.”*

Certains prétendent que les critiques de la révision par les pairs émergent d'un idéal irréalisable¹¹⁶. En réalité, c'est tout le contraire. Ces critiques sont une réponse adéquate et nécessaire à la confiance exagérée en la révision par les pairs qui découle elle-même d'une vision idéale et irréaliste du système. Une prise de position réaliste serait justement de reconnaître que la révision par les pairs n'offre aucune garantie et que chaque personne a toujours la responsabilité d'évaluer elle-même la scientificité d'un travail qu'elle lit, peu importe s'il a été révisé par les pairs ou non.

Atkinson¹¹ conclut que le problème principal de la révision par les pairs se compose i) du fait que les réviseurs ne sont ni encadrés ni tenus responsables de leurs révisions, ii) de la négligence des faiblesses connues du système et iii) de l'admission sans répercussions de jugements non-fondés et inacceptables. Il en résulte qu'un tiers des décisions de publication sont estimées être prises pour d'autres raisons que la rigueur ou l'importance scientifique³⁵. Dans un tel contexte, il est bien malheureux que la simple mention du terme « révision par les pairs » invoque une telle aura d'autorité. C'est d'ailleurs ce phénomène inexplicable qui est responsable du focus aberrant sur la contenant plutôt que sur le contenu lors de l'évaluation d'une source d'information. La révision par les pairs est devenue un contenant sans lequel aucun contenu ne peut être pris au sérieux.

Quoi qu'il en soit, la grande majorité des scientifiques, qu'ils soient pour ou contre le système actuel de révision par les pairs, s'accordent pour dire qu'il faille le faire évoluer et non l'éliminer^{72,116,135,183}. Un consensus sur la procédure à suivre pour ce faire se laisse cependant encore désirer¹³⁹.

Solutions et alternatives au système habituel

Formation à la révision et différentes motivations

Au cours du temps, plusieurs solutions ont été proposées et partiellement appliquées pour plusieurs des problèmes de la révision par les pairs. Une proposition qui semble faire l'unanimité est de former les scientifiques pour faire des révisions^{101,135}, et certaines revues telles que *BioTechniques* ou *British Medical Journal (BMJ)* offrent maintenant ce service^{51,183}. Cependant et pour l'instant du moins, la grande majorité des scientifiques reste sans formation^{135,186} et les rares études évaluant l'effet de la formation sur la qualité des révisions sont pour le moins mitigées^{26,101,183,195}. Une méta-analyse sur le sujet²⁶ démontre que la formation à la révision n'améliore que très peu la qualité du processus et n'aide pas à diminuer le temps pour faire la révision. Le seul impact notable était que le taux de réjection ait augmenté, ce qui pourrait être un développement dans la bonne direction comme dans la mauvaise.

Par rapport au manque de civilité de certains réviseurs, Souder¹⁸⁶ propose de tout simplement retirer les réviseurs impolis ou agressifs du système de révision. Cependant, ceci ne représente pas une grande motivation, certains pouvant même être tentés de se faire sortir de la liste des réviseurs, car faire cela exige souvent un grand investissement de temps et d'énergie. D'ailleurs, certains des scientifiques les plus productifs font aussi le moins de révision⁹⁰.

Onitilo et coauteurs¹⁴⁶ propose un questionnaire destiné à guider plus précisément la révision. Ce questionnaire améliore le consensus entre réviseurs et oblige la justification des

commentaires, mais il reste à démontrer si ces questions permettent une sélection adéquate des recherches. Les indices d'augmentation de la qualité de la recherche révisée en utilisant un questionnaire précis sont aussi mitigés et plus de recherche doit être faite sur la question²⁶.

Simmons et coauteurs¹⁷⁹ ont quant à eux développé quatre recommandations à suivre pour les réviseurs afin d'améliorer la qualité des révisions et de diminuer la quantité de faux-positifs dans la littérature. Brièvement, il s'agit de s'assurer que les auteurs présentent correctement et complètement leur protocole de recherche, qu'ils y justifient de façon adéquate chaque décision prise, et qu'ils démontrent que leurs résultats ne dépendent pas de décisions arbitraires quant à la récolte ou à l'analyse de données. Les réviseurs devraient aussi encourager la présentation complète du protocole de recherche en étant plus tolérant si cette présentation paraît plus désordonnée que ce à quoi ils sont habitués. Finalement, dans les cas où les justifications des auteurs ne sont pas suffisantes, les réviseurs devraient exiger une réplication exacte de l'étude. Nous trouvons ces recommandations tout à fait légitimes et raisonnables. Cependant, elles nous apparaissent difficiles à mettre en place, étant improbable que beaucoup de réviseurs puissent faire tout cela avec compétence. De plus elles augmentent dangereusement le pouvoir des réviseurs déjà trop grand. Par exemple, comment décider si les justifications des auteurs sont satisfaisantes ou non? Est-ce que les réviseurs sont vraiment mieux placés pour ce faire que les auteurs? Notre expérience nous dit que non, assurément pas dans tous les cas.

Un autre exemple de solution est le système ingénieux proposé par Aktipis et Thompson-Schill³, lequel est basé sur la

motivation positive – autre que l’argent – et la coopération plutôt que la compétition. Ici, les éditeurs établiraient une liste de critères simples pour évaluer la qualité et la rapidité des révisions et chaque réviseur recevrait un score. Les articles de chaque scientifique sont ensuite révisés par des réviseurs de même qualité que celle de l’auteur principal. Voilà un système de motivation qui récompenserait un bon travail de qualité et pénaliserait les plus négligents et insolents.

Modifications du système lui-même

D’autres solutions consistent à modifier directement le processus de révision le plus courant, c’est-à-dire celui où les réviseurs restent anonymes, mais connaissent l’identité des auteurs. Une solution évidente et souvent privilégiée par la majorité des scientifiques est que les auteurs aussi restent anonymes^{135,186}. Cependant, les données montrent que les effets de cette solution sont variables et complexes, mais qu’en moyenne elle n’affecte pas la performance des réviseurs^{26,35,77}, peut-être parce que l’identité des auteurs est souvent facile à deviner^{35,206} ou que le plagiat est alors plus difficile à détecter¹³⁵. Quoi qu’il en soit, cette solution reste matière à débat, avec quand même plusieurs scientifiques étant catégoriquement contre l’anonymité en général par peur qu’on abuse de la situation¹⁹⁵.

Une autre alternative est donc de demander aux réviseurs de signer leur révision pour plus de transparence et qu’alors tous soient connus de tous. Lorsque questionnés, la plupart des scientifiques disent qu’ils refuseraient de réviser ouvertement, par peur de représailles¹¹⁶. Par contre, lorsque confrontés à

l'obligation de signer leurs révisions, comme c'est le cas pour la revue *BMJ*, il ne semble pas y avoir plus de refus de réviser, les scientifiques semblent simplement s'habituer à la nouvelle situation¹⁶¹. Encore plus intéressant est l'idée de non seulement faire signer les révisions par les réviseurs, mais aussi de les publier avec l'article lorsque celui-ci est accepté. Le fait que les révisions soient rendues public et signées augmenterait certainement le professionnalisme des révisions et rendrait tout le processus beaucoup plus transparent¹⁹⁵.

La peur de représailles pourrait aussi diminuer puisque tout serait transparent, rendant plus difficile de « se venger » sans que ce comportement devienne lui aussi transparent et donc répréhensible. Parallèlement, nous pourrions nous attendre à ce que les scientifiques déclinent moins souvent de servir de réviseurs si leur nom apparaîtrait, permettant de mieux valoriser ce travail normalement anonyme¹⁹⁵.

Il y a aussi la révision complètement ouverte, où tout le monde peut directement réviser et commenter l'article avec l'option d'un processus normal de révision par la suite⁸⁴. Les tenants de la révision ouverte prétendent que ceci pourrait réduire le nombre de soumissions puisque les scientifiques seraient embarrassés de soumettre un article de piètre qualité qui serait visible à tous⁸⁴. Toutefois, dans un tel système d'ouverture, les jeunes scientifiques pourraient facilement être intimidés et l'influence des scientifiques connus serait encore plus grande^{72,112,116,186}. Néanmoins, certaines revues telles qu'*Atmospheric Chemistry & Physics* utilisent ce système avec succès depuis plus de 10 ans⁷². Dans ce cas, tous les documents (première version de l'article, tous les commentaires de révision) restent toujours disponibles au public. Une méta-analyse du côté

de la recherche biomédicale montre que la révision ouverte augmente en effet la qualité des révisions, tout en diminuant le taux de réjection²⁶, ce qui encore une fois peut être un bon comme un mauvais signe. D'autres cependant trouvent que la qualité de la révision n'est pas grandement affectée par ce mode ouvert^{77,186}.

La révision ouverte peut prendre plusieurs formes, telles qu'une révision post-publication, c'est-à-dire que chaque article est publié après vérification minimale de la part du comité éditorial. Des réviseurs ainsi que la communauté scientifique peuvent ensuite discuter de sa qualité et les auteurs peuvent modifier l'article au besoin, tout ce qui se fait et s'écrit restant toujours disponible avec la version finale de l'article. Ce système est utilisé par exemple par la revue *F1000 Research*.

Il est intéressant de noter que la peur des représailles n'est pas le seul défi d'une révision ouverte ; le désintérêt de la communauté scientifique pour commenter la qualité des articles est jusqu'ici aussi important⁷². Il semble en fait que les scientifiques auraient besoin de récompenses pour s'acquitter de cette tâche, un point qui fait d'ailleurs partie intégrante des solutions et alternatives proposées par Tennant et coauteurs¹⁹⁵ puisque sans obligations, ces auteurs assument que trop peu de scientifiques ne s'adonneraient à cette pratique. En résumé, tout le monde semble penser que réviser les articles est extrêmement important, mais personne n'a envie de le faire. Ceci étant dit, la révision ouverte est une opportunité intéressante de transformer le processus hermétique actuel de révision par les pairs en un processus social, communautaire, mené par l'ensemble de la communauté scientifique¹⁹⁵. Ce type d'efforts collectifs ont déjà fait leurs preuves en tant qu'outils d'auto-correction de la

recherche scientifique là où la révision par les pairs habituelle avait échoué (voir Meskus et coauteurs¹²⁸ pour un exemple célèbre).

En réalité, il est vraiment difficile de dire quel système de révision par les pairs est le plus efficace, ou lequel procure la plus grande qualité de recherche. Il n'existe tout simplement pas encore assez de données empiriques sur le sujet pour tirer des conclusions robustes²⁶. La revue *BMJ* par exemple a expérimenté plusieurs systèmes alternatifs de révision par les pairs et n'a pu observer que très peu de différences de qualité¹⁸³. Le processus y est toutefois de plus en plus ouvert et donc vérifiable, ce qui est déjà une grande amélioration.

Le fait que peu de différences de qualité soient observées entre différentes méthodes de révision^{26,183,186} pourrait indiquer que les faiblesses de chaque méthode annulent ses forces par rapport aux autres méthodes. De plus, le manque de recherche et de données empiriques sur le sujet rend l'acceptation à grande échelle d'un changement général du système vraiment difficile à appliquer puisqu'une bonne part de la communauté scientifique aurait besoin de voir des différences concrètes d'efficacité avant de vouloir changer quoique ce soit¹⁹⁵, ce qui est tout à fait compréhensible. Il est donc impératif de produire de telles études^{161,195}.

Il semble de toute façon bien difficile d'améliorer la qualité moyenne des révisions si les réviseurs conservent autant de pouvoir sans acquérir plus de compétences. La revue PLoS ONE fut la première à réduire significativement le pouvoir des réviseurs en ne faisant réviser que les méthodes, ne portant donc aucun jugement sur l'importance scientifique d'un article, laissant ce rôle

à la communauté de lecteurs^{72,135,189}. Qui plus est, la rigueur scientifique seule y est encouragée, peu importe si les résultats sont significatifs ou non. Cette revue connaît un grand succès, mais seules des revues plus spécialisées et moins connues lui ont emboîté le pas¹³⁵, peut-être parce que les changements sociaux sont si lents par nature et que, comme nous l'avons mentionné, encore trop peu d'études sur les différences d'efficacité entre les différents systèmes existent^{84,161,195}. Le lecteur intéressé pourra se référer à Fresco-Santalla et Hernández-Pérez⁷² pour une exploration plus approfondie et exhaustive de ces différents systèmes.

Les banques d'archivage

Une alternative plus radicale est la banque d'archivage ouverte à tous (voir Couture et coauteurs⁴⁰ pour une synthèse du développement des banques d'archivage). Celles-ci sont des serveurs en ligne où les scientifiques peuvent déposer, modifier ou retirer leurs travaux ainsi que lire les travaux de leurs collègues, tout ça gratuitement et sans l'entremise de compagnies privées de publication. De plus, tout le monde peut commenter sur ces prépublications (*E-print*), comme dans le principe de la révision ouverte.

La banque d'archivage la plus connue et certainement celle qui a le plus grand succès est arXiv, utilisée principalement par des physiciens, des mathématiciens et des informaticiens^{121,200}. Cette plateforme, présentée en détail par Luce¹²¹, a été créée par les scientifiques pour répondre aux besoins des scientifiques et permet une communication

scientifique gratuite, extrêmement rapide et ouverte. Dans ce sens, sauf pour la rapidité, elle se rapproche beaucoup plus de ce à quoi la « révision par les pairs » devait ressembler à ses débuts, avant qu'elle ne devienne un symbole d'identité scientifique utilisé par les revues pour justifier leur importance. Au sein de cette banque d'archivage, il n'y a pas de révision par les pairs telle qu'on la connaît aujourd'hui, mais un tri très modéré pour filtrer les articles évidemment défaillants. Ainsi, les lecteurs doivent utiliser leur jugement et leurs commentaires servent de révision. Chaque prépublication est accompagnée d'une date, prévenant le vol d'idées. De plus, ce système est suffisamment flexible pour soit coexister avec le processus traditionnel de publication, ou pour le faire évoluer vers quelque chose de mieux adapté aux besoins des scientifiques. Cette banque d'archivage est beaucoup utilisée et ses standards sont élevés, 90% de ses articles finissant dans des revues normales ou des actes de conférence²⁰⁰.

Certains prévoient que l'archivage remplacera le système actuel de publication⁸⁴, surtout dans les domaines où la communication rapide des résultats est essentielle, et la publication déjà une formalité^{84,200}. Effectivement, l'amélioration du système, souhaitée par tous, consiste peut-être en un processus de révision minimale et rapide, laissant la communauté scientifique décider de l'importance et de la qualité de chaque étude¹⁸³. Les scientifiques en astronomie et en physique – les principaux utilisateurs d'arXiv – valorisent d'ailleurs moins la révision par les pairs que les scientifiques d'autres domaines¹³⁵. L'archivage prendra probablement de l'importance dans ces autres domaines scientifiques, mais la rapidité avec laquelle ceci se produira dépend certainement des domaines^{139,200} et de notre capacité à modifier la façon dont les scientifiques sont évalués

pour leur carrière. Dans tous les cas, comme l'affirme Couture et coauteurs⁴⁰ :

« Rien ne permet de croire que le progrès des connaissances dans les domaines où les prépublications jouent un rôle important, voire de premier plan, souffre du fait que les chercheurs prennent connaissance des travaux de leurs collègues avant que ceux-ci n'aient été évalués. »

Déjà, les jeunes scientifiques sont particulièrement intéressés par l'utilisation de plateformes digitales sans révision par les pairs, même si une transition complète des pratiques et des croyances reste futuriste¹³⁹. Ce qui toutefois pourra aider à faciliter cette transition sera de prendre exemple sur des plateformes qui existent déjà, qui sont connues et attrayantes. Tennant et coauteurs¹⁹⁵ fournissent une liste de telles plateformes, chacune avec ses différentes caractéristiques, et proposent une plateforme hybride regroupant les fonctions utiles à la publication scientifique. Quoi qu'il en soit, le système de révision par les pairs changera certainement vers plus d'ouverture et d'indépendance par rapport au système de publication des revues⁷². Ce nouveau système devra être motivants pour les scientifiques, assurer au moins le même contrôle de qualité que l'actuel (ce qui ne devrait pas poser de problèmes), et devra avoir suffisamment bonne réputation pour être largement accepté¹⁹⁵.

En conclusion, le système actuel de révision par les pairs offre une illusion de contrôle de la qualité, rassurant les moins bien informés et offrant une image de perfection irréaliste de la publication scientifique aux yeux des crédules. Pis encore, cette autorité artificielle alimente une immonde plaidoirie chez certains

faux sceptiques⁴⁵, ce qui contribue au dogmatisme scientifique. Cependant, il est rassurant de voir le large éventail de solutions en implantation, qui démontre que les biais de révision et de publication sont une préoccupation pour de très nombreux scientifiques. Les solutions les plus drastiques par rapport au système actuel témoignent quant à elle de la nécessité d'un changement majeur.

DIFFÉRENTS MILIEUX DE RECHERCHE ET LEUR FAÇON D'ABORDER LA COMMUNICATION SCIENTIFIQUE

Étant donné les écueils du système de publication scientifique présentés dans les sections précédentes, il faut se demander s'il y a d'autres moyens de valoriser la recherche. Certains faux sceptiques⁴⁴⁻⁴⁶ tenteront de faire croire que non, mais la réponse dépend beaucoup du milieu de recherche dans lequel un scientifique évolue.

Traditionnellement, on sépare en grandes catégories les différents milieux de recherche comme suit : recherche fondamentale (*basic research*) versus recherche appliquée (*applied research*) ; recherche financée par les contribuables versus recherche financée par le privé. La recherche fondamentale est généralement associée au milieu public alors que la recherche appliquée est associée au privé, même si en réalité tous ces milieux interagissent de façon de plus en plus étroite. Chacun de ces quatre milieux comporte des priorités et des devoirs spécifiques, influençant la façon dont la recherche est valorisée et communiquée. Il s'en suit que l'application scrupuleuse de la démarche scientifique – commune à toute recherche – est indépendante des impératifs de publication qui eux appartiennent aux différents milieux de recherche^{136,165}. Ceci représente une nuance extrêmement importante à comprendre. Il devient donc nécessaire de bien nommer ces différences entre milieux de recherche pour bien distinguer ce qui appartient à la démarche scientifique de ce qui appartient aux particularités d'un

milieu donné. Dans les prochains paragraphes, chacun de ces milieux sera donc détaillé dans ce sens.

Recherche fondamentale versus recherche appliquée

En recherche fondamentale, les objectifs principaux sont l'avancement des connaissances et la satisfaction de la curiosité naturelle et de l'intérêt personnel des scientifiques qui la pratiquent^{160,165,208}. Elle répondra donc à des questions qui débutent typiquement par « Pourquoi » et « Comment » : « Pourquoi avons-nous des pensées? » ; « Comment l'univers a-t-il commencé? » On ne peut généralement pas prédire précisément les résultats ou les applications potentielles qui découleront de la recherche fondamentale puisqu'elle explore des terrains inconnus^{160,165}. Cependant, elle est considérée comme un capital intellectuel et comme le moteur de l'innovation et de la création de nouveaux savoirs. Elle est donc la base sur laquelle s'appuient toutes les nouvelles applications et le bien-être économique d'une nation^{165,208}.

Le but premier des scientifiques qui travaillent en recherche fondamentale est – en plus de satisfaire leur curiosité – d'être compris et acceptés par leurs collègues¹³⁶. Ils visent donc à publier leurs résultats dans les revues scientifiques, exposant au besoin les faiblesses et les problèmes méthodologiques de leur étude¹³⁶. Ici, la capacité des scientifiques à donner libre cours à leur curiosité et à publier tous leurs résultats est considérée comme essentielle au bon fonctionnement de ce type de

recherche, même si le système de publication en vigueur entrave cette liberté.

Dans le milieu de la recherche appliquée, la priorité est l'application des connaissances pour résoudre des problèmes, combler des besoins, développer des marchés et/ou développer de meilleurs services pour la société^{160,165,208}. La recherche appliquée répondra donc plutôt à des questions débutant par « Comment fait-on pour... » et les résultats se traduisent par des produits, des biens et des services pour lesquels il y a un marché¹⁶⁵. Notez aussi que si la liberté et l'indépendance des scientifiques sont au premier plan en recherche fondamentale, l'éthique devrait normalement être une priorité de la recherche appliquée¹⁶⁵, car la responsabilité du scientifique ne s'arrête pas à la publication¹³⁶.

On considère souvent que la recherche appliquée ne peut survivre sans la recherche fondamentale^{147,165}. Cependant, il ne faut pas penser que la relation fondamentale-appliquée soit unidirectionnelle. Plusieurs exemples existent où des développements appliqués ont ensuite permis à la recherche fondamentale d'avancer^{159,160,165}. De plus, en recherche biomédicale du moins, l'importance de la recherche fondamentale comparée à la recherche appliquée a été remise en question par Chalmers et coauteurs³⁶, qui soulignent que le fait d'ignorer les besoins des patients résulte en un gaspillage énorme des ressources chaque année.

L'interrelation entre la recherche appliquée et fondamentale s'accroît de plus en plus depuis les années 1970, où la pression pour l'applicabilité de la recherche fondamentale a commencé à augmenter^{21,147,165}. Les deux sont donc de plus en

plus pratiquées en partenariat pour développer de nouvelles connaissances avec des applications plus directes, réduisant le laps de temps entre la recherche fondamentale et la recherche appliquée^{17,21,165}. Précisons que la recherche appliquée permet des applications, mais qu'elle est souvent plus spécifique et donc moins généralisable, alors que la recherche fondamentale tente souvent d'extraire des principes généraux à partir de l'analyse des phénomènes observables, ce qui met en évidence la complémentarité entre les deux²⁰⁸.

Ce partenariat plus étroit est forcément important pour l'industrie en général, mais aussi dans les domaines de l'aérospatial et de la défense nationale¹⁶⁵, et jusque dans les universités^{21,147}. Aujourd'hui, la grande majorité des scientifiques universitaires combinent donc recherche fondamentale et appliquée dans leurs activités et sont en moyenne plus engagés dans le monde appliqué. Évidemment, ceci peut varier beaucoup d'une discipline à l'autre et d'un pays à l'autre, mais la distinction entre les instituts de recherche fondamentale et appliquée tend généralement à devenir de plus en plus floue²¹. Après tout, ces deux milieux de recherche sont essentiels et complémentaires pour le développement économique et l'avancement de la science^{42,147,160}. Qui plus est, ils utilisent la même démarche scientifique^{136,147}.

Valorisation de la recherche : une question de financement

La valorisation et la communication de la recherche peuvent prendre différentes formes et ceci semble surtout

dépendre du milieu de recherche, c'est-à-dire du type de recherche – fondamentale ou appliquée – et de la source du financement – public ou privé (Fig. 6). Nous explorons ces différentes relations et leurs conséquences dans les prochaines sections.

Recherche fondamentale financée par le public

Traditionnellement, la recherche fondamentale – principalement représentée en Amérique par les universités – est en majeure partie financée par les contribuables à travers des subventions gouvernementales. Le système de publication scientifique avec révision par les pairs, discuté précédemment, y est la méthode privilégiée pour valoriser et pour communiquer la recherche⁵⁵, la seule méthode acceptable selon certains auteurs¹⁶⁵. Ces mêmes auteurs¹⁶⁵ font cependant ouvertement la promotion de l'égo, de l'orgueil et de certaines des failles les plus néfastes de ce système, laissant un arrière-goût douteux à leurs conclusions sur le sujet. Leur opinion est néanmoins courante et partagée, la publication avec révision par les pairs étant souvent pour les scientifiques académiques la seule avenue pour faire avancer leur carrière. Ce système de publication influence donc grandement la pratique de la science dans la recherche fondamentale publique.

Ce milieu de recherche jouit d'une réputation de liberté et d'ouverture complète puisque les organismes subventionnaires ne devraient pas influencer la direction de la recherche²⁰⁸ et que chaque résultat devrait être rapidement rendu public à travers la

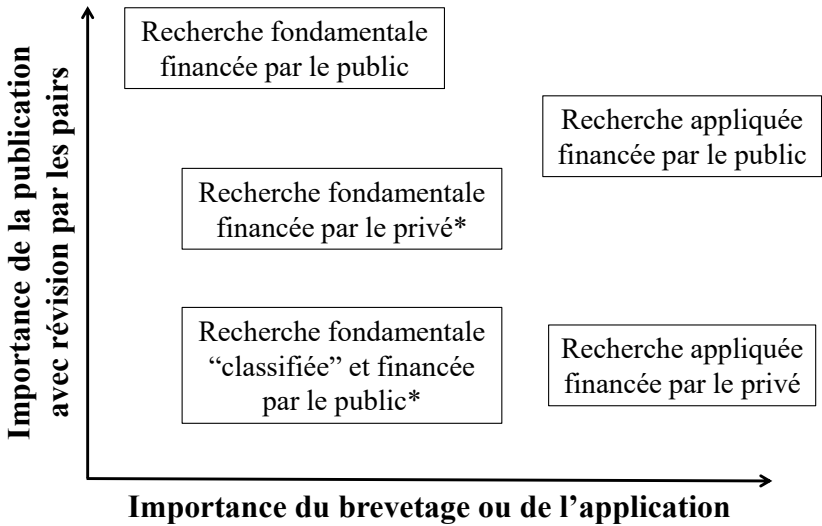


Figure 6. Comparaison entre l'importance de la publication dans des revues avec révision par les pairs et l'importance du brevetage ou des applications dans la communication et la valorisation de la recherche selon différents milieux de recherche. Cette comparaison est une approximation qualitative que nous avons faite à partir de notre revue de littérature sur le sujet. Aussi, cette approximation est une généralisation et ne serait pas représentative de tous les domaines de recherche considérés individuellement.

*L'importance de la publication serait beaucoup plus élevée si l'on prenait le point de vue des scientifiques seulement.

publication⁵⁵. L'accès public aux résultats de la recherche est d'ailleurs une obligation morale et sociale lorsque celle-ci est financée par les contribuables¹⁷⁵. Cependant, le lecteur attentif aura peut-être déjà pu comprendre que le contribuable moyen devra dépenser d'importantes sommes d'argent pour avoir accès aux articles de plusieurs revues scientifiques. De plus, ces articles sont souvent incompréhensibles pour le contribuable moyen et même pour des scientifiques hors-domaine. Il serait donc plus juste de dire que le milieu de la recherche fondamentale publique

fait preuve d'ouverture au sein de cercles plus ou moins fermés d'universitaires et d'instituts de recherche.

Néanmoins, des alternatives comme l'accès libre et le serveur arXiv commencent à prendre de la place et représentent un système de diffusion plus cohérent avec l'obligation morale de rendre au contribuable ce pour quoi il paye^{40,55}. La vulgarisation scientifique étant une voie importante de diffusion, on investit aussi beaucoup d'efforts pour vulgariser certains travaux scientifiques en parallèle du système de publication qui lui s'adresse principalement aux universitaires d'un même domaine.

Il y a des exceptions quant aux moyens de valorisation et de communication dans le milieu de la recherche fondamentale financée par le public. Dans des domaines comme le nucléaire, le militaire et plus récemment la virologie, la publication de la recherche fondamentale est typiquement restreinte pour des fins de sécurité^{42,166}, surtout aux États-Unis d'Amérique^{175,193}. Ces contrôles de la recherche fondamentale impliquent que le gouvernement puisse déclarer certains résultats comme étant « classifiés ». L'accès à ces résultats est alors limité à un groupe restreint. Il peut aussi y avoir des clauses dans les octrois de subvention concernant les prépublications, où le gouvernement se réserve le droit de réviser les articles pour déterminer s'ils peuvent être publiés ou non. La publication y est soit retardée, soit facultative et laissée à la discrétion des scientifiques. Le choix des scientifiques de publier ou non sera alors influencé seulement par des préoccupations morales ou carriéristes.

En général, les scientifiques sont contre de telles mesures puisque i) elles peuvent retarder l'avancement de la science, de l'innovation et du développement économique du pays, ii) elles

sont de moins en moins efficaces dans le contexte actuel de mondialisation et donc iii) elles gaspillent de précieuses ressources. D'ailleurs, certaines universités refusent tout simplement de faire de la recherche classifiée. Vous pouvez vous référer à Shea¹⁷⁵ et Stowsky¹⁹³ pour plus de détails sur la question.

Recherche fondamentale financée par le privé

Depuis maintenant 30-40 ans, des coupures dans le financement public de la recherche fondamentale ont forcé les scientifiques universitaires à trouver d'autres sources de subvention, créant des relations toujours plus étroites entre universités et industries privées^{17,21,55,147,157,197}. Évidemment, l'implication de l'industrie dans la recherche fondamentale universitaire affecte presque assurément la recherche qui s'y fait puisque les industries sont surtout intéressées par les applications potentielles, la commercialisation des résultats et donc aussi leur confidentialité^{147,157}. De fait, des clauses de non-divulcation sont monnaie courante lorsque des entreprises privées sont impliquées, et ceci est conforme aux règlements universitaires¹⁹⁷.

L'objectif principal de l'industrie n'est effectivement pas l'avancement de la science pour le public, ni la publication, mais bien de développer de nouvelles applications qu'elle peut ensuite commercialiser. Ceci fait que la recherche à contrats est en réalité mieux adaptée – et se fait d'ailleurs plus souvent – en recherche appliquée qu'en recherche fondamentale¹⁶⁵. Pour le lecteur intéressé, Ioannidis⁹⁵ présente à ce sujet une liste des divers protagonistes en recherche scientifique, révélant leurs intérêts

parfois contradictoires par rapport à la publication, au financement, ou aux opportunités d'affaires.

Par rapport aux relations entre collègues, l'implication de l'industrie et cette ligne de plus en plus floue entre la recherche à buts lucratifs ou non lucratifs rendent de plus en plus difficile pour les scientifiques de favoriser la collaboration plutôt que la compétition¹⁰. Ainsi, des collaborations qui auraient pu être bénéfiques pour la recherche sont arrêtées au profit de la compétition. Si cela apporte de plus grands profits monétaires reste une question ouverte.

Les industries investissent elles aussi dans leurs propres laboratoires de recherche privés et en tirent des bénéfices importants^{42,147}. Cependant, les politiques de publication peuvent y être bien différentes⁸⁵. Par exemple, il y a souvent un délai avant d'obtenir l'autorisation de publier. La décision de soumettre des articles pour publication peut être encore une fois laissée à la discrétion des scientifiques, ou encore cela peut leur être explicitement proscrit dès l'embauche⁸⁵. Ceci pose évidemment un problème pour les scientifiques, puisque la publication reste la voie principale d'évaluation de leur performance lorsqu'ils appliquent pour un nouvel emploi.

Des compagnies privées spécialisées dans la production de méta-analyses pour l'industrie existent également. Ces méta-analyses sont de très haute qualité, professionnelles, et très utiles pour l'industrie, mais ne sont souvent pas publiées dans des revues avec révision par les pairs⁹⁶. La raison est que l'industrie n'a souvent pas d'intérêt que les résultats soient publics, et les scientifiques travaillant dans ces compagnies n'ont aucune motivation pour publier, car ils perdraient alors du temps précieux

dans un processus fastidieux et trop complexe, temps qu'ils pourraient plutôt employer à faire un autre contrat⁹⁶. Dans ces compagnies privées donc, la publication n'est aucunement valorisée pour juger de la qualité scientifique du travail qui s'y fait.

Il existe par contre aussi des instituts privés de recherche fondamentale qui sont extrêmement ouverts dans la publication des résultats, mais qui ne se formalisent pas avec la révision par les pairs, préférant utiliser la technique d'archivage, plus rapide et mieux adaptée aux besoins des scientifiques¹²¹. Ceci se produit surtout dans des domaines où la recherche évolue rapidement et où la concurrence est féroce. Un tel cas est présenté de façon très parlante par Gura⁸⁴ :

«« Une fois le processus de révision par les pairs terminé, mes travaux sont déjà obsolètes » affirme Giulio Ruffini, physicien théoricien et directeur de *Starlab Barcelona*, un centre de recherche privé accordant peu d'importance à la publication dans des revues avec révision par les pairs. « L'archivage est rapide, facile et utilise un système d'horodatage sur votre travail afin que vous puissiez plus tard en revendiquer la propriété. » » ^{Note 19}

¹⁹ Traduction libre de la part des auteurs. Citation originale : ““By the time I go through peer review, my work is old news” says theoretical physicist Giulio Ruffini, director of *Starlab Barcelona*, a private research centre that places little importance on publishing in peer-reviewed journals. “Archiving is fast and easy and puts a time stamp on what you have done so that you can later claim ownership.” ”

Recherche appliquée financée par le public

Depuis les années 1970, il s'est produit une augmentation de la pression pour l'applicabilité des résultats de la recherche fondamentale, mettant un accent de plus en plus fort sur la recherche appliquée^{21,55,165}. Les universités sont donc de plus en plus entrepreneuriales, tentant de retirer des gains financiers maximum de leurs activités de recherche et provoquant un mélange sans précédent entre recherche fondamentale et appliquée^{17,55,76}.

Ce transfert d'importance vers la recherche appliquée pour valoriser la recherche publique au-delà de la publication est non seulement encouragé par la triple hélice gouvernement-universités-industries^{21,76,157}, mais aussi par l'opinion publique qui veut voir des justifications sociales à l'existence de coûteux laboratoires de recherche^{55,76}.

Cependant, cette transition de plus en plus importante vers des universités publiques menées comme des entreprises à buts lucratifs crée un système de récompenses et de motivations pour les scientifiques de plus en plus problématique et malsain⁵³. Comme nous l'avons déjà expliqué dans la section « Une atmosphère de compétition insoutenable », ce genre d'environnement corrosif n'encourage pas nécessairement une science optimale. Ainsi, bien qu'en tant que contribuables nous soyons tout à fait en droit de s'attendre à ce que la recherche que nous finançons avec nos impôts soit ajustée à nos besoins, il faut faire attention à l'environnement de recherche que nous créons pour nos scientifiques, sans quoi il se peut que nous obtenions l'effet contraire. C'est le cas notamment avec la recherche sur le développement de certains médicaments comme les

antidépresseurs⁹³, dont nous avons aussi discuté dans la section « Pression des organismes subventionnaires ».

Pour la recherche appliquée en milieu universitaire, la publication reste importante, mais elle n'est plus la seule voie de valorisation de la recherche, ni même nécessairement la plus importante. Les résultats en tant que tels et leurs applications pratiques (*donc le contenu*) peuvent être considérés comme plus importants que la publication¹⁶⁵ (*c'est-à-dire le contenant*). Les universitaires qui sont surtout impliqués dans la recherche appliquée publient donc beaucoup moins que leurs collègues¹⁵⁷ et ont tendance à privilégier le brevetage, mieux adapté à leurs besoins pour ce type de recherche⁵⁵. Comme la publication, une demande de brevet coûte de l'argent et du temps. Elle est aussi accompagnée d'une description détaillée du processus ou du produit qu'on cherche à breveter, laquelle doit être rendue publique 18 mois après la demande de brevet. L'octroi d'un brevet garantit alors à son possesseur l'exclusivité commerciale pour son produit pour une période de temps limitée, normalement 20 ans, et n'empêche pas la publication des résultats une fois le brevet accepté.

Depuis 30 ans, le nombre de demandes de brevet par les universitaires monte en flèche. Ce contexte de commercialisation des résultats favorise la protection de la propriété intellectuelle au détriment de la publication^{40,55}. Il est donc courant que les scientifiques retardent la publication de leurs travaux en faveur de la procédure de brevetage, ou encore pour rester en tête de ligne d'une découverte scientifique importante^{42,185}. Cette tendance est accentuée par le partenariat avec l'industrie et les clauses de non-divulcation venant avec lui. Tout ceci s'agence mal avec la libre circulation de l'information qui est souvent perçue comme un

sceau de la recherche universitaire⁴⁰. Les directives départementales tentent de satisfaire ces intérêts contradictoires, mais les politiques à ce sujet diffèrent d'une université à l'autre, ce qui permet difficilement d'avoir une vue d'ensemble sur la question⁴⁰.

D'autres problèmes sont associés avec cette situation. Par exemple, des retards dans la publication des thèses doctorales ou encore le phénomène de « brevète ou pèris⁴⁰ » – l'équivalent de phénomène de « publie ou pèris » qui vient avec la pression de publier et dont nous avons déjà abondamment discuté. Les mêmes problématiques causent et accompagnent donc la pression pour breveter.

Un autre aspect des brevets qui peut être problématique est qu'ils bloquent, pour un temps, la recherche sur et l'utilisation de nouvelles découvertes et inventions pour tous les autres scientifiques. Bien que ce soit souvent précisément le but du brevetage, ceci pose un problème moral surtout lorsque la recherche est financée par des fonds publics. Une pratique alternative consiste donc à demander un brevet pour ensuite l'abandonner. Le scientifique s'assure ainsi que l'information reste dans le domaine public et que personne ne puisse breveter la même chose dans le futur et ainsi bloquer la recherche⁵⁵.

Comme pour la recherche fondamentale, des laboratoires de recherche appliquée sur la défense militaire ou sur le nucléaire sont financés publiquement. Cependant, en recherche appliquée, la divulgation des résultats est encore plus strictement entravée que dans le monde de la recherche fondamentale^{42,175,193}. Ce type de recherche appliquée dite « fermée » est fortement critiquée, car l'utilité de ces restrictions est remise en cause puisque

l'industrie surpasse souvent le gouvernement en matière de technologie (voir Stowsky¹⁹³ pour plus de détails). Il faut cependant noter que les besoins des gouvernements peuvent être bien différents de ceux de l'industrie et du marché public, ce qui pourrait peut-être encore justifier l'existence de ces laboratoires gouvernementaux.

Il existe aussi des instituts gouvernementaux de recherche appliquée qui restent ouverts quant à la divulgation des résultats. La recherche en agriculture est un exemple, les résultats étant prévus pour être directement applicables par le public⁴². Dans ces cas, la présentation des résultats prendra une forme beaucoup plus vulgarisée que les articles scientifiques. Ceci pourrait être une autre raison pour laquelle les scientifiques en recherche appliquée publient moins dans les revues. En effet, dans ce milieu de recherche, la forme des rapports scientifiques est bien différente de celle des articles scientifiques écrits en recherche fondamentale^{42,136}. Ceci s'explique par le fait que leurs publications respectives ne s'adressent pas aux mêmes personnes. En recherche fondamentale, le lecteur cible est un collègue du même domaine ; en recherche appliquée, c'est souvent un employeur ou un travailleur non rompu à la méthodologie de recherche¹³⁶. La priorité n'est donc pas de publier dans des revues scientifiques, mais de développer un nouveau marché ou d'ajuster pragmatiquement une façon de travailler.

Cette différence teinte aussi les demandes de subventions. En recherche appliquée, la justification pour la recherche concerne directement les applications potentielles alors qu'en recherche fondamentale, des applications potentielles sous-entendent de nombreuses recherches ultérieures¹³⁶.

Recherche appliquée financée par le privé

Pour aller chercher un financement supplémentaire, les universités favorisent de plus en plus la recherche appliquée financée par le privé¹⁷. Cette tendance est d'ailleurs renforcée par les politiques gouvernementales^{21,157}. La recherche appliquée en milieu universitaire a déjà été détaillée du point de vue des universitaires dans la section précédente. Cependant, le point de vue de l'industrie influence lui aussi la recherche et la publication. Les industries préfèrent donner plus de place aux brevets, à la propriété intellectuelle et à la concrétisation de l'innovation. Le partage des connaissances se fait alors avec la vente d'un produit, l'offre d'un nouveau service ou avec le système des brevets.

Plusieurs industries ont aussi leur propre laboratoire de recherche appliquée. Il est intéressant de constater que les revenus générés par la commercialisation des brevets industriels sont en général beaucoup plus importants que ceux des brevets universitaires⁴⁰. L'absence de publication n'est donc certainement pas un signe que les travaux effectués n'ont pas de valeur scientifique inhérente. Dans ce milieu, l'évaluation de la performance des scientifiques se base sur leurs brevets et sur la création de nouveaux produits. Ces scientifiques investissent donc une plus petite partie de leurs efforts à la publication dans des revues scientifiques⁸⁵.

Au sein des industries, lorsqu'il n'est pas possible ni désirable de breveter les résultats de la recherche appliquée, ceux-ci peuvent devenir des secrets de commerce. Ceci nécessite souvent toute une infrastructure de management à cet effet⁵⁶. Les secrets de commerce sont moins chers et obligent un concurrent à découvrir par lui-même les résultats pour en connaître

l'existence. Cependant, si un concurrent arrive aux mêmes résultats, il peut les breveter ou les publier, et récolter les bénéfices.

Bilan

L'importance accordée à la publication scientifique avec révision par les pairs change selon les milieux de recherche (Fig. 6). Dans le milieu de la recherche fondamentale universitaire, elle est nécessaire pour faire carrière. Dans d'autres milieux tels que la recherche appliquée industrielle, la publication avec révision par les pairs peut être facultative ou même interdite. Ce sont deux extrêmes entre lesquels existent plusieurs degrés d'importance (Fig. 6).

Cependant, peu importe le mode de diffusion des résultats (publications révisées par les pairs, rapports internes, services et produits offerts au public, etc.), ce qui détermine la scientificité d'un travail est l'application de la démarche scientifique. Le scientifique critique se doit donc de distinguer la démarche scientifique du mode de diffusion des résultats. En résumé, les intérêts carriéristes et commerciaux influencent énormément le choix des sujets de recherche, le type de résultats, leur accessibilité et leur voie de communication. Ces choix n'ont rien à voir avec la rigueur scientifique, ni avec le partage des connaissances. Ces choix ont trait au contenant de la recherche. La démarche scientifique a trait à son contenu. C'est le contenu qui nous permet donc de juger de la scientificité d'un travail quelconque.

Voici une anecdote illustrant à merveille la nécessité de maintenir des alternatives au système de publication avec révision par les pairs pour optimiser l'avancement de la science, peu importe le milieu de recherche. En 2014, le statisticien retraité Thomas Royen a eu une inspiration surprenante lui permettant de découvrir la preuve mathématique de la conjecture de l'inégalité de la corrélation Gaussienne, un problème mathématique connu depuis 40 ans, mais qui était resté jusque-là non solutionné. N'ayant aucune motivation carriériste et voulant éviter le processus fastidieux de la révision par les pairs, il a tout simplement publié sa solution dans le serveur arXiv¹⁷⁰. Le serveur arXiv est ouvert à tous, ce qui permet aux intéressés d'en profiter. Son ingéniosité a retenu l'attention, et elle est maintenant reconnue grâce à ceux qui ont pris le temps de la vérifier.

REMARQUES FINALES

« L'habit ne fait pas le moine. » Voilà un impératif de la démarche scientifique, que nous avons aussi choisi comme sous-titre de cet ouvrage. Nous pourrions aussi dire qu'il ne faut pas se fier aux apparences, ou, comme nous l'avons fait tout au long de notre texte, que le contenu est plus important que le contenant.

Dans cet ouvrage, nous avons voulu démontrer que le fait de publier dans une revue scientifique et le fait de suivre la démarche scientifique ne sont pas deux choses qui sont nécessairement équivalentes. Nous avons voulu démontrer que le fait de publier dans une revue scientifique ne représente pas une garantie de rigueur scientifique, mais bien plus une garantie que certaines normes et conventions sociales – par exemple un test de significativité, une révision par les pairs – ont été respectées.

À elle seule cette constatation devrait disqualifier la publication dans les revues scientifiques avec révision par les pairs comme mesure de ce qui est scientifique ou non. Ceci ne veut pas dire que ce qui est publié dans ces revues n'est pas scientifique, seulement que la publication n'est pas une mesure de scientificité adéquate. Les faux sceptiques tentent, lorsque possible, de faire oublier cette simple évidence et d'embrouiller les lecteurs quant à l'évaluation de la scientificité. Ces confusions ont très souvent rapport avec la publication scientifique, qu'ils essaient de présenter comme un processus scientifique.

C'est pourquoi nous avons présenté ici les systèmes de publication et de révision en vigueur aujourd'hui et que nous les avons situés par rapport à l'application de la démarche

scientifique. Nous les avons aussi contextualisés selon les principaux types de milieux de recherche.

Ceci vous aura permis, nous l'espérons, de constater les nombreuses incompatibilités entre d'un côté les systèmes de publication et de révision, et de l'autre côté la rigueur scientifique. Ces incompatibilités restent souvent floues, même pour les scientifiques, ce qui explique certainement en partie l'attitude d'inertie générale au sein de la communauté scientifique, laquelle néglige ces incompatibilités et laisse à penser qu'elles ne représentent pas de problèmes importants pour la science. Cette attitude décourage d'ailleurs tout effort pour changer les choses. De plus, elle favorise les inconduites professionnelles ; il n'y aurait, après tout, pas de conséquences graves...non?

Alternatives et perspectives par rapport aux systèmes actuels

Plusieurs auteurs ont proposé de nombreuses solutions pour différents problèmes et nous avons tenté d'en faire un portrait représentatif tout au long de cet ouvrage. Bien que notre objectif soit plutôt de conscientiser, une précision est importante à mentionner : toute solution n'a que peu de chance de succès sans l'accompagnement d'une réforme globale du système^{4,18,141}. Nous ne pouvons effectivement pas nous attendre à ce que les scientifiques modifient leur comportement par rapport à un seul point précis, car tous les aspects du système sont reliés ensemble. Il faut d'abord établir les priorités, et ensuite déterminer quels sont les changements à faire à travers le système en entier pour y arriver.

Ébauche d'un plan d'action

Les prochains paragraphes reprennent les idées qui émergent à la lumière du présent ouvrage pour établir une ébauche de plan d'action possible. Cette description n'est pas absolue et devrait surtout servir de piste de réflexion.

Nous avons déjà mentionné que changer la façon dont les scientifiques sont évalués est probablement le défi le plus important auquel la communauté scientifique fait face actuellement, tout en étant aussi le plus difficile à réaliser²¹⁰. En résumé, disons qu'il faudrait évaluer chaque scientifique non pas avec des indices indirects de performance comme le facteur d'impact, mais par une lecture soigneuse de ses travaux, ainsi qu'en discutant avec ses collègues et employeurs pour aussi pouvoir évaluer son intégrité^{43,112,131}.

Pour que ceci soit réalisable, il faudrait limiter le nombre d'articles par scientifique, comme par exemple avec un maximum d'un article par personne par année¹³⁷, tout en encourageant la présence de longs articles et en favorisant la coopération entre scientifiques. Ceci permettrait de produire des articles de haute qualité et donnerait même parfois la chance de produire les réplifications nécessaires avant la publication. Pour les domaines où la communication rapide des résultats et des méthodes est essentielle, la coopération entre scientifiques, en plus d'un système d'archivage pour les données, les résultats et les méthodes, pourraient au moins partiellement solutionner le problème.

Les universités devraient réduire le nombre de scientifiques qu'elles forment et rallonger officiellement la durée

du doctorat. En Amérique du Nord par exemple, les doctorants en sciences naturelles ont souvent besoin de 5 ans, mais leur financement n'est prévu que pour 3 ans. De plus, nous devrions mettre beaucoup plus d'efforts durant la formation des scientifiques pour améliorer leurs compétences quant au design, à l'analyse et à la présentation des expériences de recherche²¹⁰.

Tout aussi importantes sont la déprivatisation de la publication, la réduction du nombre de revues et l'abolition du facteur d'impact et du prestige des revues généralistes. Grâce à la numérisation de la publication, il pourrait n'y avoir qu'une seule revue par domaine, éliminant naturellement l'utilisation du facteur d'impact. Les publications pourraient aussi être toutes en accès libre, ce qui serait cohérent avec le fait que ce sont les contribuables qui paient pour une bonne partie de la recherche publiée et que ce sont les scientifiques qui produisent et valident cette recherche.

Quant à la révision des articles, puisque les scientifiques seraient évalués par la qualité de leurs œuvres, seule une révision minimale et rapide serait nécessaire, accentuée sur les méthodes. La révision serait complètement ouverte et nous aurions besoins de former beaucoup plus de statisticiens pour le contrôle des articles comme pour la formation des scientifiques.

Un autre grand thème est le préenregistrement^{Note 20} des études et les obligations par rapport à leur pouvoir statistique et

²⁰ Le préenregistrement d'une étude consiste à divulguer avant le début de l'étude quels en sont les objectifs, quelles sont les hypothèses nulles et alternatives, quelle est la taille d'échantillon prévue, quelles analyses sont prévues, et ainsi de suite. Le but est d'empêcher les scientifiques de modifier le tout en cours de route dans le but d'atteindre la significativité des résultats. Le but est aussi que les études ne disparaissent pas tout simplement si les résultats ne sont pas significatifs. En effet, même si de telles études ne

aux réplifications. Certains auteurs sont de grands advocates du préenregistrement et nous partageons certainement cette opinion. Il faut cependant apporter la nuance que ceci ne devrait être obligatoire et stricte que dans certains types de recherche où il est possible de fournir des hypothèses de recherche précises et où les coûts de se tromper sont importants¹⁷⁶ – les essais cliniques, par exemple.

Dans un monde où les ressources sont limitées, il ne serait toutefois pas souhaitable pour, par exemple, de la recherche exploratoire d'avoir des obligations trop strictes quant au préenregistrement ou au design expérimental. Effectivement, ça ne ferait aucun sens de ne pas accepter un résultat intéressant mais inattendu parce que la question à laquelle il répond n'était pas préenregistrée¹⁷⁶. Ça ne ferait pas de sens non plus d'exiger des designs expérimentaux pointus et coûteux pour faire de la recherche exploratoire, plutôt que de premièrement tâter le terrain de façon économique pour ensuite n'investir vraiment que dans les directions prometteuses¹⁷⁶. Après tout, les grandes découvertes sont souvent faites par accident et alimentée par la publication de recherche exploratoire qui, par définition, est incertaine.

Bref, la science doit rester libre, mais elle doit pouvoir justifier pourquoi elle fait ce qu'elle fait. En retour, les personnes qui évaluent ces justifications doivent avoir la flexibilité et surtout les compétences pour juger de la validité du contenu et, au besoin, faire des exceptions... tant que les principes de la

sont pas nécessairement publiées, du moins il reste un registre montrant qu'elles ont été conduites.

démarche scientifique sont respectés. Pour ne pas entraver l'avancement de la science, il nous faut établir un équilibre entre laisser libre cours à la créativité scientifique et instaurer des critères de travail trop rigoureux¹⁷⁶.

Toutes ces propositions et remarques ne représentent pas une liste exhaustive des changements à apporter simultanément ou des choses auxquelles réfléchir, ni une situation parfaite et sans problème. Elles ne sont qu'un point de départ pour de futures réflexions. Surtout, il ne faut pas appliquer des restrictions aveuglement et de façon standardisée pour tous, car les besoins sont assurément différents d'un domaine à l'autre, d'une branche de recherche à l'autre. Un manque de flexibilité dans nos solutions pratiques pourrait apporter plus de mal que de bien¹⁷⁶. Ce qui est certain, c'est que l'impact de toute modification devra être évalué scientifiquement et avec beaucoup d'attention⁹⁵.

De grands défis à surmonter

Pour de plus en plus de scientifiques – dont plusieurs ont été présentés avec leurs idées dans cet ouvrage – il devient clair que la science fait aujourd'hui face à de nombreux défis cruciaux si elle veut conserver une crédibilité et une fiabilité de façon continue. Plusieurs de ces défis concernent les inconduites professionnelles et les mauvaises pratiques scientifiques¹⁷⁶, lesquelles sont causées principalement par la façon dont les scientifiques sont évalués et par de profondes incompréhensions des statistiques communément utilisées.

Un autre grand défi concerne la quantité absurde de scientifiques formés, d'articles écrits et de revues qui les publient, menant à une quantité de données impossible à saisir pour un esprit humain¹⁷⁶. À cela s'ajoute des coûts de recherche toujours plus grands, mais pas nécessairement égalés par les possibilités de financement¹⁷⁶. Tout ceci est logique : dans un monde aux ressources limitées, une quantité toujours grandissante de scientifiques et de recherche crée des pénuries de financement. Ce sont d'ailleurs ces problèmes de financement qui impliquent toujours de plus en plus l'industrie en recherche, avec tous les effets dont nous avons déjà discuté dans une autre section.

Mais probablement le plus grand défi à surmonter est la réticence de la part de la communauté scientifique dans son ensemble à agir face aux problématiques étayées dans cet ouvrage, ou même à les reconnaître. Le fait de reconnaître un problème important constitue la base minimale pour vouloir apporter tout changement, et c'est pourquoi nous nommons cette réticence comme défi principal pour la science actuellement. Yarborough et coauteurs²¹⁰ ont identifié quatre croyances erronées et fortement répandues, lesquelles semblent pouvoir expliquer cette réticence étonnante. La première est l'idée selon laquelle questionner la fiabilité de la recherche scientifique serait la même chose que questionner l'intégrité des scientifiques eux-mêmes. En effet, bien que les conflits d'intérêts soient un aspect qui puisse toucher directement l'intégrité des scientifiques, ces derniers sont en général plutôt prisonniers malgré eux d'un système forçant et encourageant la majorité des problématiques présentées ici. C'est d'ailleurs de cette manière que nous avons tenté de présenter notre argumentation tout au long du texte, ne mettant pas la faute sur les individus, mais sur les pressions malsaines d'un système mésadapté à la recherche.

La deuxième croyance erronée, probablement fortement liée à la première, est que la majorité des problèmes proviennent de quelques individus malhonnêtes. Cependant, comme nous l'avons à maintes reprises indiqué, la plupart des plus grands problèmes – tels que l'incompréhension des statistiques ou les biais de publication – sont systémiques et affectent la communauté dans son ensemble, les scientifiques honnêtes n'en étant absolument pas à l'abri. Se concentrer uniquement sur la punition des conduites clairement malhonnêtes peut donc nous faire passer à côté de problèmes plus pressants²¹⁰.

La troisième croyance erronée est que la science s'auto-corrige toujours. Cette croyance est particulièrement répandue et souvent utilisée pour amoindrir l'importance de toute affirmation remettant en question la fiabilité de la science publiée. Cette croyance est d'ailleurs souvent mise en relation avec la révision par les pairs, alors que nous avons vu qu'il n'y a pas de liens concrets entre les deux. En effet, la publication d'erreurs se produisant plus fréquemment qu'on ne le pense – un fait étayé ailleurs dans cet ouvrage – combinée avec plusieurs conduites liées aux biais de publication, contrevient fortement à l'auto-correction de la science^{53,80,210}. Les erreurs publiées peuvent donc être citées et rester des décennies dans la littérature avant d'être découvertes, – si elles le sont un jour – ce qui arrivent plus par coïncidence que par un quelconque effort conscient²¹⁰. Il faudrait donc cesser de se fier aveuglément à cette mythique auto-correction si nous voulons conserver la crédibilité scientifique. Notez toutefois qu'il y a plusieurs cas où l'auto-correction de la recherche fonctionne (par exemple Meskus et coauteurs¹²⁸). C'est juste que ce n'est pas une valeur sûre, rien sur quoi nous devrions nous baser pour régler nos problèmes à notre place. Ce n'est qu'un outil, pas un caractère infaillible de la science.

La quatrième croyance erronée est que de simplement suivre des réglementations puisse régler les problèmes qui en ont créé la nécessité. La raison est que c'est d'une révision du système au complet dont nous avons besoin. En particulier, nous avons besoin d'un bien meilleur alignement entre la façon dont les scientifiques sont récompensés et de bonnes pratiques scientifiques²¹⁰ et c'est absolument incroyable que ce ne soit pas déjà le cas. Sans ce type de changements radicaux donc, seulement intégrer des réglementations pour, par exemple, diminuer la fabrication illégale de données ne peut apporter qu'une amélioration bien limitée. De plus, ces réglementations coûtent beaucoup de temps et d'argent pour en contrôler le respect et donnent un faux sentiment de sécurité par rapport aux problèmes, ce qui diminue la motivation de faire les changements qui seraient vraiment nécessaires²¹⁰.

À tout cela s'ajoutent les croyances erronées des scientifiques quant à l'impact de leurs propres décisions de recherche, c'est-à-dire qu'ils ne sont généralement pas conscients de l'ampleur avec laquelle chaque décision de récolte ou d'analyse de données affecte leurs résultats¹⁷⁹.

Toutes ces fausses croyances, qui n'épargnent pratiquement personne, aussi honnête que l'on puisse être, provoquent une distorsion de la vision de l'activité scientifique. Certains auteurs¹⁷⁶ discutant certains des problèmes détaillés dans cet ouvrage le font par exemple en les cadrant dans un contexte irréaliste (du moins de nos jours), où les pratiques et les scientifiques représenteraient toujours les plus hauts standards possibles, laissant de côté tout ce qui ne se laisse pas placer dans ce monde idéal. Certains^{100,176} plaident d'ailleurs pour qu'on fasse attention dans la manière dont on présente les problèmes en

recherche, l'objectif principal semblant alors être de ne pas ternir l'image de la science aux yeux du public. Nous trouvons cependant préférable de présenter une image réelle de la science afin que tous puissent être conscients des failles à éviter si possible et afin que tous puissent aussi participer de façon adéquate à la recherche et à l'implémentation de solutions. De fait, proposer des solutions invalides ou insignifiantes est une autre répercussion d'une vision trop idéaliste de la science ; on ne peut pas solutionner ce qu'on ne reconnaît pas.

Les changements nécessaires pour un processus scientifique plus fiable et plus ouvert sont donc très lents à se faire¹⁹⁵, une conséquence de ces fausses croyances difficiles à secouer :

« Si nous pouvions trouver une façon de se débarrasser de ces croyances erronées, nous pourrions devenir plus proactifs pour démontrer que nous méritons la confiance du public²¹⁰. »^{Note}

21

Yarborough et coauteurs²¹⁰ discutent aussi de plusieurs solutions très prometteuses^{Note 22} pour améliorer la situation, mais ils notent surtout le manque d'intérêt désarçonnant de la part de la communauté scientifique et le fait que le potentiel de telles

²¹ Traduction libre de la part des auteurs. Citation originale : "If we can find a way to shed these erroneous beliefs, we could become more proactive in showing how we deserve the public's trust."

²² Certaines de ces solutions ont aussi été présentées dans cet ouvrage-ci, mais pas toutes, puisque nos objectifs sont légèrement différents. Nous recommandons donc fortement aux lecteurs intéressés par ce sujet la lecture de Yarborough et coauteurs²¹⁰.

solutions n'est donc encore pratiquement pas exploité. Pourtant, comme il est certainement clair pour tous à ce point de la lecture, nous ne manquons pas de cris d'alarme par rapport aux problèmes actuels en recherche, ni de recommandations pour régler ces problèmes et ce, même de la part d'éminents scientifiques. Ces auteurs soulignent donc l'importance d'identifier et de bien comprendre ce qui rend l'abandon des conduites en recherche si peu attrayant pour les scientifiques, car sans cette compréhension, l'adoption et le succès de toute recommandation est bien peu probable. Pour eux, il s'agit des croyances erronées qu'ils ont identifiées. Nous sommes d'accord avec cela et renchérissons en pointant que d'autres facteurs présentés dans notre ouvrage y jouent certainement un grand rôle aussi. Ceci inclue en premier plan le climat de compétition malsaine régnant aujourd'hui sur la recherche et paralysant, semble-t-il, toute motivation collective pour œuvrer à changer le système.

Pour finir, Yarborough et coauteurs²¹⁰ suggèrent tout de même qu'une meilleure éducation des scientifiques en formation par rapport à la récolte et l'analyse des données, mais aussi par rapport aux croyances erronées – particulièrement en ce qui a trait à l'auto-correction scientifique – serait potentiellement la meilleure solution pour le moment. Ils concluent de façon générale, et nous secondons, que ce dont nous avons le plus besoin serait que la communauté scientifique dans son ensemble se concentre à mériter la confiance qui lui est si foncièrement accordée par le public. Cette prise de position, ce réarrangement des priorités, pourrait aider à adopter plus allégrement les réformes dont nous avons besoin, autant au niveau individuel qu'institutionnel. Pour nous, il s'agit de réaligner nos priorités et

l'éducation des scientifiques avec le contenu plutôt qu'avec le contenant, avec le moine plutôt qu'avec l'habit.

Vivre avec le paradoxe

Malgré toutes les entraves mentionnées, des travaux de qualité se font chaque jour, la science continue à faire des avancées extraordinaires, et l'estime des collègues a encore une place importante dans la carrière des scientifiques^{176,182}. Ceci apparaît comme un paradoxe, car s'il y a tellement de problèmes et que ces problèmes ont un si grand impact, pourquoi observe-t-on tout de même de si grands succès en science?

Cette impression de paradoxe est causée selon nous par trois facteurs principaux. Premièrement, le désir de produire de la science de qualité chez la majorité des scientifiques et les qualités intrinsèques de la démarche scientifique font que la recherche produit des travaux de qualité *malgré* l'environnement de compétition et de publication dans lequel elle est forcée. Nous croyons que sans tous les problèmes indiqués et sans le gaspillage omniprésent des ressources, la science avancerait beaucoup plus vite et beaucoup plus facilement. Les progrès qui se produisent sont en effet plus facile à voir que ceux qui ne se produisent pas, et ce déséquilibre alimente le paradoxe. Un exemple de cela a trait au partage de données. Nous avons déjà expliqué en quoi le partage libre de données est très important et nécessaire pour vérifier ce qui se fait. En même temps, le partage de données n'est pas toujours libre ou même il n'est parfois pas de mise dans certains contextes de recherche. Il semble pourtant évident que de la recherche de qualité puisse se faire dans ces

situations, donc avec un partage minimal. La science finit donc toujours par faire des progrès, sauf que dans une situation où le partage de données est faible ou absent, elle le fait moins efficacement, et la recherche de piètre qualité a le potentiel de perdurer et d'affecter la science plus longtemps.

Deuxièmement, il faut dire que ce ne sont pas tous les domaines qui sont sujets à de grandes controverses ou à des guerres idéologiques. Dans ces domaines moins polémiques donc, la recherche peut avancer avec beaucoup moins d'entraves. Parallèlement, bien que la compétition malsaine puisse y être aussi très forte et produise certainement de grandes injustices et beaucoup d'inconduites, elle ne peut évidemment pas empêcher toute recherche de qualité, surtout à longs termes. Par exemple, même si des idées sont volées, la recherche se fera. Ou même si des mensonges sont racontés, les compétiteurs les découvriront tôt ou tard¹²⁸.

Troisièmement, les désavantages et les failles des systèmes dans lesquels les scientifiques évoluent ne sont que très peu médiatisés, sauf dans des cas tellement gros qu'ils deviennent célèbres^{39,128}. Il y a en effet généralement une espèce de voile devant ces problèmes, presque des tabous, qui fait qu'on n'en parle généralement que très peu et qu'on préfère présenter les grands succès, les progrès, et les avantages. Ceci mène selon nous à une sous-estimation des difficultés, et à une sur-estimation des avancés scientifiques comparée à une situation où la science se ferait de façon plus optimale.

Tous ces facteurs sont particulièrement influents dans le domaine de la recherche appliquée, où i) les guerres idéologiques sont naturellement moins fréquentes (mais pas absentes)

puisqu'on n'y parle généralement pas de théories, mais bien d'applications, ii) les progrès sont particulièrement évidents et frappants pour le public, et iii) les statistiques y prennent un rôle beaucoup moins important, ou du moins beaucoup plus précis, ce qui fait que les incompréhensions statistiques sont moins dommageables.

Parallèlement (ou, encore une fois, paradoxalement), il y a quand même de plus en plus de discussions, de congrès et de malaises par rapport aux problèmes de la publication scientifique et du système d'évaluation des scientifiques. Tout cela est un très bon signe et dénote du désir et du besoin très forts et très présents d'améliorer les choses, du moins chez les scientifiques.

L'objectif principal de cet ouvrage n'est donc absolument pas de dire que le monde de la recherche soit brisé, invalide, ou quelconques absurdités de la sorte. Ce serait un mensonge, pur et simple. De la bonne science et de bons scientifiques, il y en a des tonnes. L'objectif est plutôt de défaire l'impression, toute aussi mensongère, selon laquelle le monde de la recherche et surtout celui de la publication scientifique est toujours critique, impartial et porteur de vérité. Car le problème, c'est justement cette différence entre cette vision idéale dont la science jouit¹⁷⁶ et son fonctionnement réel, qui lui n'est pas idéal²¹⁰. Cette impression erronée de la science est selon nous dommageable, car elle éteint le désir d'aller vérifier, elle permet de se complaire dans des certitudes qui sont non scientifiques et pourtant de faire croire – même à soi-même – qu'elles le sont. Ce comportement se rapproche pourtant plus du dogmatisme que de l'esprit scientifique.

Nous visons donc à permettre aux lecteurs, s'ils le désirent, de développer une opinion plus éclairée par rapport à l'information qui circule sur un domaine de recherche donné. La capacité à développer une opinion éclairée passe irrémédiablement par une prise de conscience des limites, de l'importance et de l'utilité de la révision par les pairs ainsi que par la compréhension des enjeux influençant la publication scientifique, car ces facteurs influencent grandement l'information qui circule sur un sujet donné.

Il faut sortir de la mentalité selon laquelle : « Si c'est publié avec révision par les pairs, c'est bon, sinon, ce n'est pas bon. », car plusieurs critères non scientifiques interviennent dans le processus de publication. Le fait qu'un article soit publié ou non ne signifie donc rien en soi. Il faut s'intéresser, lire avec attention et, s'il s'agit d'un sujet polémique, aller lire l'information originale provenant des deux camps. Il est donc clair, nous l'espérons, que développer une opinion scientifique informée nécessite beaucoup de temps et d'investissements personnels.

Un tel changement de la perception de la science publiée serait énormément bénéfique. En effet, même si rien d'autre ne changeait – c'est-à-dire qu'aucune des solutions proposées dans les sections précédentes n'étaient mises en place – que ce changement de perception, la situation pourrait tout de même être satisfaisante, car, comme nous venons de le mentionner, la science de qualité continuerait de se produire et cette nouvelle conscience de ce que représente la publication avec révision par les pairs aiderait grandement à filtrer les travaux ainsi que les scientifiques de mauvaise qualité. C'est aussi pourquoi nous mentionnions qu'un manque de flexibilité dans l'application de certaines solutions aurait autant de chance d'améliorer que de

remplir la situation. Rien ne prévaut une bonne compréhension du système et le désir de développer une opinion juste.

Le mythe des pseudosciences

L'habit ne fait pas le moine. Voilà une phrase bien simple, porteuse d'un message essentiel. Nous avons utilisé cette expression comme un fil conducteur tout au long du texte, dans le but de mieux faire comprendre ce qui est important à considérer lors de l'évaluation de la scientificité... et ce qui ne l'est pas. L'important, c'est de se concentrer sur le contenu, pas sur le contenant, car le contenant ne dit absolument rien sur la rigueur scientifique ou sur la validité de la méthode suivie.

Il s'en suit que nous avons pu démontrer qu'il est absurde de baser l'évaluation de la scientificité sur le mode de diffusion ou même d'inclure un mode de diffusion particulier dans la démarche scientifique. Le mode de diffusion ne représente que l'habit, le contenant dans lequel un contenu est placé. Effectivement, les résultats d'une étude scientifique représentent un savoir scientifique avant d'avoir été publiés, révisés ou reconnus par les pairs. Il serait tout aussi absurde de qualifier, par exemple, la recherche militaire de pseudoscience parce qu'elle ne produit pas ou peu de publications révisées par des pairs. La publication scientifique reste évidemment importante et nécessaire, mais surtout en tant qu'outil de communication dans la mesure où la diffusion est souhaitée. Par ailleurs, il existe d'autres modes de diffusion tout aussi valides pour communiquer la science.

L'évaluation de la scientificité nécessite donc des efforts considérables de la part de la personne qui cherche à se faire une opinion. Dans ce contexte, il est exact de dire que la seule et unique condition, nécessaire et suffisante, pour qualifier un travail de scientifique, est l'application de la démarche scientifique, laquelle peut être représentée par une grande variété de méthodologies de recherche et ne se retrouve que dans le contenu. Ainsi, qu'un travail ait été réalisé par un scientifique en milieu universitaire, aussi reconnu soit-il, ou par une personne n'ayant jamais mis les pieds à l'université, il doit simplement être évalué selon sa démarche de recherche avant d'être déclaré scientifique ou non. En d'autres mots, si un sujet n'intéresse pas assez une personne pour qu'elle s'en informe correctement, celle-ci doit avoir la décence de reconnaître que son opinion n'a rien de scientifique, peu importe la nature des travaux considérés. Ceci nous ramène aux concepts de pseudoscience et de faux scepticisme.

Le « mythe » des pseudosciences n'est pas une contribution scientifique, mais bien philosophique. On ne peut que très difficilement situer une discipline dans son entièreté sur le continuum philosophique entre science et pseudoscience. En effet, des travaux scientifiques et non scientifiques (peu importe les résultats et le statut de publication) ont vraisemblablement été produits pour toutes les disciplines existantes. Ce qu'on peut faire cependant, c'est d'évaluer la scientificité d'un travail, d'un article, d'une investigation, ou même d'un groupe de théories précises.

De la même façon, il serait tout aussi impertinent de parler de pseudo-scientifique ; un scientifique peut mener à bien une étude scientifique et, la même année, écrire un billet empli de préjugés, d'inexactitudes et d'opinions personnelles. Isaac

Newton faisait de l'alchimie et Johannes Kepler de l'astrologie⁵. Ce sont des scientifiques qui ont apporté une grande contribution à la science. On ne pourrait pas dire d'eux qu'ils sont des pseudo-scientifiques parce qu'ils ont étudié des domaines où les travaux n'ont pas apporté de résultats probants ou ne sont pas reconnus (nous ne connaissons pas ces domaines particulièrement bien, nous savons seulement que ce sont des domaines non reconnus par la communauté scientifique en général). Ceci démontre aussi qu'il est crucial d'évaluer soi-même la scientificité de chacun des travaux d'un scientifique pour départager sa contribution à la science de ses intérêts ou de ses idées.

Quant aux faux sceptiques, ils dépassent même le jugement de valeur dans leur démarche, car ils versent dans la supercherie. Lorsqu'on est face à quelqu'un qui se prononce négativement sur la scientificité d'un travail, ou d'un ensemble de travaux, sans expliciter son évaluation de la scientificité, on est face à un faux sceptique. Les stratégies sont en général la diffamation^{109,110}, la dissimulation de l'information⁴⁴, des faux arguments issus d'une incompréhension de l'application de la démarche scientifique^{44,45}, ou de tordre les arguments ou les travaux qu'ils veulent discréditer avec des mensonges⁴⁶. L'utilisation du mot « pseudoscience » est aussi souvent utilisé tel un argument massue par les faux sceptiques, bien qu'il n'apporte en soi aucune contribution à l'évaluation de la scientificité. Dans tout cela, la moquerie joue aussi souvent un rôle plus ou moins subtil. Ce qui compte avant tout pour les faux sceptiques, c'est de réussir à détourner l'attention des lecteurs du contenu vers le contenant.

En effet, le véritable sceptique refuse de croire avant de vérifier ; le faux sceptique peut avoir déjà vérifié, mais surtout il

veut empêcher que d'autres vérifient. Le faux sceptique tente donc de faire prendre position aux gens en omettant l'évaluation de l'application de la démarche scientifique. En d'autres mots, même si les motifs du faux sceptique peuvent être difficiles à circonscrire (incompréhension, orgueil mal placé, vengeance personnelle, intérêts commerciaux), le faux sceptique lui-même est facilement reconnaissable.

Il est toutefois intéressant de noter que la lecture des travaux des faux sceptiques en vaut quand même la chandelle. En effet, elle constitue une excellente pratique pour apprendre à reconnaître des arguments non scientifiques qui veulent se faire passer pour scientifiques, et à les différencier des arguments qui sont réellement scientifiques, c'est-à-dire qui se base sur l'évaluation de l'application de la démarche scientifique. De plus, comme un lecteur intéressé ne devrait jamais accepter un constat de pseudoscience sans vérification minutieuse, la même chose devrait s'appliquer à un constat de faux sceptique. Allez lire, vérifiez par vous-mêmes, basez-vous sur le contenu et pas le contenant, jugez par rapport à la démarche scientifique et non par rapport au mode de diffusion. Voilà les meilleures recommandations que nous puissions vous donner.

BIBLIOGRAPHIE

1. Abbott, A., Cyranoski, D., Jones, N., Maher, B., *et al.* (2010) Do Metrics Matter? *Nature*, 465, p. 860-862.
2. Afonso, A. S. et Gilbert, J. K. (2010) Pseudoscience: A meaningful context for assessing nature of science. *International Journal of Science Education*, 32:3, p. 329-348.
3. Aktipis, C. A. et Thompson-Schill, S. L. (2010) Positive assortment for peer review. *Adaptive Behavior*, 18:5, p. 448-450.
4. Alberts, B., Kirschner, M. W., Tilghman, S. et Varmus, H. (2014) Rescuing US biomedical research from its systemic flaws. *Proceedings of the National Academy of Sciences of the United States of America*, 111:16, p. 5773-5777.
5. Allchin, D. (2004) Pseudohistory and pseudoscience. *Science and Education*, 13:3, p. 179-195.
6. Altman, D. G. (1994) The scandal of poor medical research. *British Medical Journal*, 308:6924, p. 283-284.
7. Amrhein, V., Greenland, S., McShane, B., *et al.* (2019) Retire Statistical Significance. *Nature*, 567, p. 305-307.
8. Anastasi, A. (1968) *Psychological testing* (3rd ed.). New-York : Macmillan.
9. Anastasi, A. (1986) Envolving concepts of test validation. *Annual Review of Psychology*, 37:1, p. 1-16.
10. Anderson, M.S., Ronning, E.A., De Vries, R. et Martinson, B.C. (2007) The Perverse Effects of Competition on Scientists' Work and Relationships. *Science and Engineering Ethics*, 13, p. 437-461.
11. Atkinson, M. (2001) « Peer review » culture. *Science and Engineering Ethics*, 7, p. 193-204.
12. Ayerbe, C. et Missonier, A. (2007) Validité interne et validité externe de l'étude de cas: principes et mise en œuvre pour un renforcement mutuel. *Finance Contrôle Stratégie*, 10:2, p. 37-62.
13. Bacon, F. (1605) *The advancement of learning*. London, UK.
14. Baker, M. (2016a) Is there a reproducibility crisis? *Nature*, 533, p. 452-454.
15. Baker, M. (2016b) Statisticians issue warning on P values. *Nature*, 531, p. 151.
16. Balluerka, N., Gómez, J. et Hidalgo, D. (2005) The Controversy over Null Hypothesis Significance Testing Revisited. *Methodology*, 1:2, p. 55-70.
17. Banal-Estañol, A., Macho-Stadler, I. et Pérez-Castrillo, D. (2013) Research output

- from university-industry collaborative projects. *Economic Development Quarterly*, 27:1, p. 71-81.
18. Begley, C. G., Buchan, A. M. et Dirnagl, U. (2015) Robust research: Institutions must do their part for reproducibility. *Nature*, 525, p. 25-27.
 19. Begley, C. G. et Ioannidis, J. P. A. (2015) Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research*, 116:1, p. 116-126.
 20. Benjamin, D.J., Berger, J.O., Johannesson, M. *et al.* (2018) Redefine statistical significance. *Nat Hum Behav*, 2, p. 6-10.
 21. Bentley, P. J., Gulbrandsen, M. et Kyvik, S. (2015) The relationship between basic and applied research in universities. *Higher Education*, 70:4, p. 689-709.
 22. Biagioli, M. (2016) Watch out for cheats in citation games. *Nature*, 535, p. 201.
 23. Bohannon, J. (2013) Who's afraid of peer review? *Science*, 342, October, p. 60-65.
 24. Brembs, B., Button, K. et Munafò, M. (2013) Deep Impact: Unintended consequences of journal rank. *Frontiers in Human Neuroscience*, 7, June, p. Article 291: 1-11.
 25. Brembs, B. (2018) Prestigious Science Journals Struggle to Reach Even Average Reliability. *Front. Hum. Neurosci.* 12:37.
 26. Bruce, R., Chauvin, A., Trinquart, L., Ravaud, P. et Boutron, I. (2016) Impact of interventions to improve the quality of peer review of biomedical journals: a systematic review and meta-analysis. *BMC Med.*, 14:1, p. 85.
 27. <https://bulliedintobadscience.github.io/BulliedIntoBadScience.org/> (accessed December 2020)
 28. Bunge, M. (1984) What is pseudoscience? *The Skeptical Inquirer*, 9:1, p. 36-47.
 29. Burlamaque-Neto, A. C., Santos, G. R., Lisboa, L. M., Goldim, J. R., *et al.* (2012) Study of the comprehension of the scientific method by members of a university health research laboratory. *Brazilian Journal of Medical and Biological Research*, 45:2, p. 93-96.
 30. Burnham, J. C. (1990) The evolution of editorial peer review. *Journal of the American Medical Association*, 263:10, p. 1323-1329.
 31. Burtcher-Schaden, H., Clausing, P. et Robinson, C. (2017) *Glyphosat und Krebs: Gekaufte Wissenschaft*. Vienna, Austria : 72 p.
 32. Cain, D. M. et Detsky, A. S. (2008) Everyone's a little bit biased (even physicians). *Journal of the American Medical Association*, 299:24, p. 2893-2895.
 33. Camerer, C.F., Dreber, A., Holzmeister, F., Ho, T.H., *et al.* (2018) Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nat Hum Behav.*, 2:9, p. 637-644.

34. Campanario, J. M. (1998a) Peer review for journals as it stands today: Part 1. *Science Communication*, 19:3, p. 181-211.
35. Campanario, J. M. (1998b) Peer review for journals as it stands today: Part 2. *Science Communication*, 19:4, p. 277-306.
36. Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., *et al.* (2014) How to increase value and reduce waste when research priorities are set. *The Lancet*, 383:9912, p. 156-165.
37. Chalmers, I. et Glasziou, P. (2009) Avoidable waste in the production and reporting of research evidence. *Obstet Gynecol.*, 114:6, p. 1341-1345.
38. Chauvin, A., Ravaud, P., Baron, G., Barnes, C. et Boutron, I. (2015) The most important tasks for peer reviewers evaluating a randomized controlled trial are not congruent with the tasks most often requested by journal editors. *BMC medicine*, 13:1, p. 1-10.
39. COPE (Committee of publication Ethics) statement 2014 on inappropriate peer review processes:
<http://publicationethics.org/news/cope-statement-inappropriatemanipulation-peer-review-processes> (accessed December 2020)
40. Couture, M., Dubé, M. et Malissard, P. (2010) *Propriété intellectuelle et université: Entre la libre circulation des idées et la privatisation des savoirs*. Québec : Presses de l'Université du Québec.
41. Crawley, M. J. (2013) *The R Book* (2nd Editio.). Chichester, UK : John Wiley & Sons Ltd.
42. Dasgupta, P. et David, P. A. (1994) Toward a new economics of science. *Research Policy*, 23:5, p. 487-521.
43. de Rond, M. et Miller, A. N. (2005) Publish or perish: Bane or boon of academic life? *Journal of Management Inquiry*, 14:4, p. 321-329.
44. Denault, V. (2015) *L'incidence de la communication non verbale lors de procès: une menace à l'intégrité du système judiciaire?* Thèse de doctorat, Université du Québec à Montréal, 234 p.
45. Denault, V., Larivée, S., Plouffe, D. et Plusquellec, P. (2015) La synergologie, une lecture pseudoscientifique du langage corporel. *Revue de psychoéducation*, 43:2, p. 425-455.
46. Denault, V., Plusquellec, P., Jupe, L.M., St-Yves, M., *et al.* (2020) The Analysis of Nonverbal Communication: The Dangers of Pseudoscience in Security and Justice Contexts. *Anuario de Psicología Jurídica*, 30:1, p. 1-12
47. Dickersin, K., Chan, S., Chalmers, T. C., Sacks, H. S. et Smith, H. (1987) Publication bias and clinical trials. *Controlled Clinical Trials*, 8, p. 343-353.
48. Dunn, D., Halonen, J. et Smith, R. (2008) Teaching critical thinking in psychology:

A handbook of best practices. Oxford, UK : John Wiley & Sons Ltd.

49. Duyx, B., Urlings, M.J.E., Swaen, G.M.H., Bouter, L.M., *et al.* (2017) Scientific citations favor positive results: a systematic review and meta-analysis. *Journal of Clinical Epidemiology*, 88, p. 92-101.
50. Dwan, K., Gamble, C., Williamson, P. R. et Kirkham, J. J. (2013) Systematic review of the empirical evidence of study publication bias and outcome reporting bias - An updated review. *PLoS ONE*, 8:7, p. e66844.
51. Editor (2015) Benefits and burdens of peer-review. *BioTechniques*, 58:1, p. 5.
52. Editorial (2012) A different agenda. *Nature*, 487, p. 271-272.
53. Edwards, M.A. et Roy, S. (2017) Academic Research in the 21st Century: Maintaining Scientific Integrity in a Climate of Perverse Incentives and Hypercompetition. *Environmental Engineering Science*, 34:1, p. 51-61.
54. Eisner, D.A. (2018) Reproducibility of science: Fraud, impact factors and carelessness. *Journal of Molecular and Cellular Cardiology*, 114, p. 364-368.
55. European IPR Helpdesk (2015a) Fact sheet: Publishing v. patenting. Luxembourg : 12 p.
56. European IPR Helpdesk (2015b) Fact sheet: How to manage confidential business information. Luxembourg : 12 p.
57. Fanelli, D. (2009) How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4:5, p. e5738.
58. Fanelli, D. (2010a) Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*, 5:4, p. e10271.
59. Fanelli, D. (2010b) « Positive » results increase down the hierarchy of the sciences. *PLoS ONE*, 5:4, p. e10068.
60. Fanelli, D. (2012) Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90:3, p. 891-904.
61. Fanelli, D. (2013a) Positive results receive more citations, but only in some disciplines. *Scientometrics*, 94, p. 701-709.
62. Fanelli, D. (2013b) Why Growing Retractions Are (Mostly) a Good Sign. *PLoS Med*, 10:12, p. e1001563.
63. Fanelli, D. (2019). A theory and methodology to quantify knowledge. *R. Soc. open sci.*, 6: 181055.
64. Fanelli, D. et Larivière, V. (2015) Researchers' Individual Publication Rate Has Not Increased in a Century. *PLoS ONE*, 11:3, p. e0149504.
65. Fanelli, D., Costas, R. et Larivière, V. (2015) Misconduct Policies, Academic Culture and Career Stage, Not Gender or Pressures to Publish, Affect Scientific Integrity. *PLoS ONE*, 10:6, p. e0127556.

66. Fanelli, D., Costas, R. et Ioannidis, J.P.A. (2017) Meta-assessment of bias in science. *Proc Natl Acad Sci U S A.*, 114:14, p. 3714-3719.
67. Fanelli, D., Ioannidis, J.P.A. et Goodman, S. (2018) Improving the integrity of published science: An expanded taxonomy of retractions and corrections. *Eur J Clin Invest.*, 48:4.
68. Fisher, R. (1925). *Statistical Methods for Research Workers*, First Edition. Edinburgh: Oliver and Boyd.
69. Frank, E. (1994) Authors' criteria for selecting journals. *Journal of the American Medical Association*, 272:2, p. 163-164.
70. Fraser, H., Parker, T., Nakagawa, S., Barnett, A., Fidler, F. (2018) Questionable research practices in ecology and evolution. *PLoS ONE* 13:7, p. e0200303.
71. Freedman, L. P., Cockburn, I. M. et Simcoe, T. S. (2015) The economics of reproducibility in preclinical research. *PLoS Biology*, 13:6.
72. Fresco-Santalla, A. et Hernández-Pérez, T. (2014) Current and evolving models of peer review. *The Serials Librarian*, 67:2014, p. 373-398.
73. Gad-el-Hak, M. (2004) Publish or perish—An ailing enterprise? *Physics Today*, 57:3, p. 61-62.
74. Gallo, S. A., Sullivan, J. H. et Glisson, S. R. (2016) The influence of peer reviewer expertise on the evaluation of research funding applications. *PLoS ONE*, 11:10.
75. Garland, T. (2015) *The scientific method as an ongoing process*. College Lectures. Riverside : Department of Biology, University of California.
76. Gingras, Y. et Roy, L. (2006) *Les transformations des universités du XIIIe au XXIe siècle*. Québec : Presses de l'Université du Québec.
77. Godlee, F., Gale, C. R. et Martyn, C. N. (1998) Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: A randomized controlled trial. *Journal of the American Medical Association*, 280:3, p. 237-240.
78. Goodchild van Hilten, L. (2015) Why it's time to publish research « failures ». Elsevier.
79. Goodman, S. N., Fanelli, D. et Ioannidis, J. P. A. (2016) What does research reproducibility mean? *Science Translational Medicine*, 8:341, p. 341ps12.
80. Goodstein, D. (2002) Scientific misconduct. *Academe*, 88, 28–31.
81. Greenberg, S.A. (2011) Understanding belief using citation networks. *J Eval Clin Pract.* 17:2, p. 389-93.
82. Grimes, D.R., Bauch, C.T. et Ioannidis, J.P.A. (2018) Modelling science trustworthiness under publish or perish pressure. *R. Soc. Open Sci.*, 5: 171511.
83. Gumpenberger, C., Gorraiz, J., Wieland, M., Roche, I., *et al.* (2013) Exploring the bibliometric and semantic nature of negative results. *Scientometrics*, 95:1, p.

277-297.

84. Gura, T. (2002) Peer review, unmasked. *Nature*, 416, p. 258-260.
85. Gwynne, P. (2012) Assessing individuals in scientific teams. *Research Technology Management*, 55:5, p. 8-9.
86. Hagen, R.L. (1997) In Praise of the Null Hypothesis Statistical Test. *American Psychologist*, 52:1, p. 15-24.
87. Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. et Jennions, M.D. (2015) The Extent and Consequences of P-Hacking in Science. *PLoS Biol*, 13:3, p. e1002106.
88. Higginson, A.D. et Munafò, M.R. (2016) Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. *PLoS Biol*, 14:11, p. e2000995.
89. Ho, R. C.-M., Mak, K.-K., Tao, R., Lu, Y., Day, J. R. et Pan, F. (2013) Views on the peer review system of biomedical journals: an online survey of academics from high-ranking universities. *BMC Medical Research Methodology*, 13:1, p. 74.
90. Hochberg, M. E., Chase, J. M., Gotelli, N. J., Hastings, A. et Naeem, S. (2009) The tragedy of the reviewer commons. *Ecology Letters*, 12:1, p. 2-4.
91. Horrobin, D. F. (1990) The philosophical basis of peer review and the suppression of innovation. *Journal of the American Medical Association*, 263:10, p. 1438-1441.
92. Ioannidis, J. P. A. (2005) Why most published research findings are false. *PLoS Medicine*, 2:8, p. e124.
93. Ioannidis, J. P. A. (2008) Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philosophy, Ethics, and Humanities in Medicine*, 3:14.
94. Ioannidis, J. P. A. (2011) An epidemic of false claims. *Scientific American*, 304:June, p. 16.
95. Ioannidis, J. P. A. (2014) How to make more published research true. *PLoS Medicine*, 11:10, p. e1001747.
96. Ioannidis, J. P. A. (2016) The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank Quarterly*, 94: 3, p. 485-514.
97. Ioannidis, J. P. A. (2016) Why Most Clinical Research Is Not Useful. *PLoS Med* 13:6, p. e1002049.
98. Iqbal, S.A., Wallach, J.D., Khoury, M.J., Schully, S.D. et Ioannidis, J.P.A. (2016) Reproducible Research Practices and Transparency across the Biomedical Literature. *PLoS Biol*, 14:1, p. e1002333.
99. Ivanov, A., Kaczowska, B. A., Khan, S. A., Ho, J., et al. (2017) Review and analysis

- of publication trends over three decades in three high impact medicine journals. *Plos One*, 12:1.
100. Jamieson, K.H. (2018) Crisis or self-correction: Rethinking media narratives about the well-being of science. *Proc Natl Acad Sci U S A.*, 115:11, p. 2620-2627.
 101. Jefferson, T., Alderson, P., Wager, E. et Davidoff, F. (2002) Effects of editorial peer review: A systematic review. *Journal of the American Medical Association*, 287:21, p. 2784-2786.
 102. Jennions, M. D. et Møller, A. P. (2002) Publication bias in ecology and evolution: an empirical assessment using the 'trim and fill' method. *Biological Reviews of the Cambridge Philosophical Society*, 77:2, p. 211-222.
 103. Joobar, R., Schmitz, N., Annable, L. et Boksa, P. (2012) Publication bias: What are the challenges and can they be overcome? *Journal of Psychiatry and Neuroscience*, 37:3, p. 149-152.
 104. Joyce, J., Rabe-Hesketh, S. et Wessely, S. (1998) Reviewing the reviews: the example of chronic fatigue syndrome. *Journal of the American Medical Association*, 280:3, p. 264-266.
 105. Kassirer, J. P. et Campion, E. W. (1994) Peer review: Crude and understudied, but indispensable. *Journal of the American Medical Association*, 272:2, p. 96-97.
 106. Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*. (O. Neurath, Dir.) (2nd Ed., II). Chicago : The University of Chicago Press.
 107. Kumar, M. J. (2009) Evaluating scientists: Citations, impact factor, h-index, online page hits and what else? *IETE Technical Review*, 26, p. 165-168.
 108. Lakens, D., Adolfs, F.G., Albers, C.J. *et al.* (2018) Justify your alpha. *Nat Hum Behav*, 2, 168-171.
 109. Lardellier, P. (2008) *Arrêtez de décoder: pour en finir avec les gourous de la communication*. Édition de l'Hèbe.
 110. Lardellier, P. (2017) *Enquête sur le business de la communication non verbale: Une analyse critique des pseudosciences du « langage corporel »*. Caen : EMS Edition.
 111. Laudan, L. (1981) The pseudo-science of science? *Philosophy of the Social Sciences*, 11:2, p. 173.
 112. Lawrence, P. A. (2003) The politics of publication. *Nature*, 422:6929, p. 259-261.
 113. Lee, C. J., Sugimoto, C. R., Zhang, G. et Cronin, B. (2013) Bias in peer review. *Journal of the American Society for Information Science and Technology*, 64:1, p. 2-17.
 114. Leek, J.T. et Jager, L.R. (2017) Is Most Published Research Really False? *Annual Review of Statistics and Its Application*, 4, p. 2.1-2.14.

115. Leggett, N.C., Thomas, N.A., Loetscher, T. et Nicholls, M.E.R. (2013) The life of p: "Just significant" results are on the rise. *The Quarterly Journal of Experimental Psychology*, 66:12, p. 2303-2309.
116. Light, R. J. et Pillemer, D. B. (1984) *Summing up: The science of reviewing research*. Cambridge, MA : Harvard University Press.
117. Lilienfeld, S. O., Ammirati, R. et David, M. (2012) Distinguishing science from pseudoscience in school psychology: Science and scientific thinking as safeguards against human error. *Journal of School Psychology*, 50:1, p. 7-36.
118. Lilienfeld, S. O., Lynn, S. J. et Lohr, J. M. (dir.) (2014) *Science and pseudoscience in clinical psychology* (2nd editio.). New York, USA : Guilford Press.
119. Loken, E. et Gelman, A. (2017) Measurement error and the replication crisis. *Science*, 355:6325, p. 584-585.
120. Lortie, C. J. et Dyer, A. R. (1999) Over-interpretation: Avoiding the stigma of non-significant results. *Oikos*, 87:1, p. 183-184.
121. Luce, R. E. (2001) E-prints intersect the digital library: inside the Los Alamos arXiv. *Issues in Science and Technology Librarianship*, p. 12.1-12.10.
122. Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P. A., Al-Shahi Salman, R., *et al.* (2014) Biomedical research: Increasing value, reducing waste. *The Lancet*, 383:9912, p. 101-104.
123. Maggio, L., Dong, T., Driessen, E. et Artino, A. Jr. (2019) Factors associated with scientific misconduct and questionable research practices in health professions education. *Perspect Med Educ.*, 8:2, p. 74-82.
124. Martin, G.N. et Clarke, R.M. (2017) Are Psychology Journals Anti-replication? A Snapshot of Editorial Practices. *Front. Psychol.*, 8:523.
125. Martinson, B. C., Anderson, M. S. et de Vries, R. (2005) Scientists behaving badly. *Nature*, 435:7043, p. 737-738.
126. McElreath, R. et Smaldino, P. E. (2015) Replication, communication, and the population dynamics of scientific discovery. *PLoS ONE*, 10:8.
127. McGrew, J.H. et McFall, R.M. (1990) A Scientific Inquiry Into the Validity of Astrology. *Journal of Scientific Exploration*, 4:1, p. 75-83.
128. Meskus, M., Marelli, L. et D'Agostino, G. (2017) Research Misconduct in the Age of Open Science: The Case of STAP Stem Cells. *Science as Culture*, 27:1.
129. Mlinarić, A., Horvat, M. et Šupak Smolčić, V. (2017) Dealing with the positive publication bias: Why you should really publish your negative results. *Biochem Med*, 27:3, p. 030201.
130. Mobley, A., Linder, S. K., Braeuer, R., Ellis, L. M. et Zwelling, L. (2013) A survey on data reproducibility in cancer research provides insights into our limited ability

- to translate findings from the laboratory to the clinic. PLoS ONE, 8:5, p. e63221.
131. Moher, D., Naudet, F., Cristea, I.A., Miedema, F., *et al.* (2018) Assessing scientists for hiring, promotion, and tenure. PLoS Biol 16:3, p. e2004089.
 132. Møller, A. P. et Jennions, M. D. (2001) Tetsing and adjusting for publication bias. Trends in Ecology & Evolution, 16:10, p. 580-586.
 133. Moriceau, J. L. (2003) La répétition du singulier : pour une reprise du débat sur la généralisation à partir d'études de cas. Revue Sciences de Gestion, 36, p. 113-140.
 134. Moylan, E.C. et Kowalczyk, M.K. (2016) Why articles are retracted: a retrospective cross-sectional study of retraction notices at BioMed Central. BMJ Open, 6, p. e012047.
 135. Mulligan, A., Hall, L. et Raphael, E. (2013) Peer review in a changing world: An international study measuring the attitudes of researchers. Journal of the American Society for Information Science and Technology, 64:1, p. 132-161.
 136. Nafstad, H. E. (1982) Applied versus basic social research: A question of amplified complexity. Acta Sociologica, 25:3, p. 259-267.
 137. Nelson, L.D., Simmons, J.P. et Simonsohn, U. (2012) Let's Publish *Fewer* Papers. Psychological Inquiry, 23, p. 291-293.
 138. Neyman, J., et Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. Philos. Trans. R. Soc. Lond. Ser. A, 231, p. 289-337.
 139. Nicholas, D., Watkinson, A., Jamali, H. R., Herman, E., *et al.* (2015) Peer review: Still king in the digital age. Learned Publishing, 28:1, p. 15-21.
 140. Nissen, S. B., Magidson, T., Gross, K. et Bergstrom, C. T. (2016) Publication bias and the canonization of false facts. eLife, 5, p. e21451.
 141. Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., *et al.* (2015) Promoting an open research culture. Science, 348:6242, p. 1422-1425.
 142. Nuzzo, R. (2014) Statistical Errors. Nature, 506, p. 150-152.
 143. Nuzzo, R. (2015) Fooling ourselves. Nature, 526, p. 182-185.
 144. Nylenna, M. (1994) Multiple blinded reviews of the same two manuscripts. Journal of the American Medical Association, 272:2, p. 149-151.
 145. Olson, C. M., Rennie, D., Cook, D., Dickersin, K., *et al.* (2002) Publication bias in editorial decision making. Journal of the American Medical Association, 287:21, p. 2825-2828.
 146. Onitilo, A. A., Engel, J. M., Salzman-Scott, S. A., Stankowski, R. V. et Doi, S. A. R. (2014) A core-item reviewer evaluation (CoRE) system for manuscript peer review. Accountability in Research, 21:2, p. 109-121.
 147. Oosterlinck, A., Debackere, K. et Cielen, G. (2002) Balancing basic and applied

- research. *EMBO reports*, 3:1, p. 2-5.
148. Open Science Collaboration (2015) Estimating the reproducibility of psychological science. *Science*, 349:6251, p. aac4716.
 149. Palmer, A. R. (2000) Quasireplication and the contract of error: Lessons from sex ratios, heritabilities and fluctuating asymmetry. *Annual Review of Ecology and Systematics*, 31, p. 441-480.
 150. Paolucci, M. et Grimaldo, F. (2014) Mechanism change in a simulation of peer review: From junk support to elitism. *Scientometrics*, 99:3, p. 663-688.
 151. Patil, P., Peng, R.D. et Leek, J.T. 2016. What should we expect when we replicate? A statistical view of replicability in psychological science. *Perspect Psychol Sci.*, 11(4), p. 539-544.
 152. Pereira, T.V. et Ioannidis, J.P.A. (2011) significant meta-analyses of clinical trials have modest credibility and inflated effects. *J Clin Epidemiol.*, 64:10, p. 1060-9.
 153. Poldrack, R.A. (2019) The Costs of Reproducibility. *Neuron*, 101:1, p. 11-14.
 154. Popper, K. (1959) *The logic of scientific discovery*. London, UK : Hutchinson & Co.
 155. Qiu, J. (2010) Publish or perish in China. *Nature*, 463:January, p. 142- 143.
 156. Quinn, G. et Keough, M. (2002) *Experimental design and data analysis for biologists*. Cambridge, UK : Cambridge University Press.
 157. Ranga, L. M., Debackere, K. et von Tunzelmann, N. (2003) Entrepreneurial universities and the dynamics of academic knowledge production : A case study of basic vs. applied research in Belgium. *Scientometrics*, 58:2, p. 301-320.
 158. Ravnskov, U. (1992) Cholesterol lowering trials in coronary heart disease: Frequency of citation and outcome. *British Medical Journal*, 305:July, p. 15-19.
 159. Reinerman-Jones, L. E. (2010) The basic versus applied research dilemma. *Proceedings of the human factors and ergonomics society*, p. 215-217.
 160. Reinerman-Jones, L. et Lackey, S. (2011) Basic versus applied research: the final episode! Next season: transfer! *Theoretical Issues in Ergonomics Science*, 12:5, p. 395-396.
 161. Rennie, D. (2016) Make peer review scientific. *Nature*, 535, p. 31-33.
 162. Resetarits, W. J. et Bernardo, J. (dir.) (1998) *Experimental Ecology: Issues and perspectives*. New York, USA : Oxford University Press.
 163. Resnik, D. B., Gutierrez-Ford, C. et Peddada, S. (2008) Perceptions of ethical problems with scientific journal peer review: An exploratory study. *Science and Engineering Ethics*, 14:3, p. 305-310.
 164. Resnik, D.B. et Shamoo, A.E. (2017) Fostering Research Integrity. *Account Res.*, 24:6, p. 367-372.

165. Riazuddin, K., H. A., Hayee, I., Swati, S. N., Revol, J.-P., *et al.* (2007) Basic or applied: Dilemma of developing countries. (H. A. Khan, M. M. Qurashi, et I. Hayee, Dir.). Islamabad : M/s New United Printers.
166. Robert, M. (1988) *Fondements et étapes de la recherche scientifique en psychologie* (3e éd.). Saint-Hyacinthe : Edisem.
167. Rothwell, P. M. et Martyn, C. N. (2000) Reproducibility of peer review in clinical neuroscience: Is agreement between reviewers any greater than would be expected by chance alone? *Brain*, 123, p. 1964-1969.
168. Roy, R. (1985) Funding science: The real defects of peer review and an alternative to it. *Science, Technology & Human Values*, 10:3, p. 73-81.
169. Roy, R. et Ashburn, J. R. (2001) The perils of peer review. *Nature*, 414, p. 393-394.
170. Royen, T. (2014) A simple proof of the Gaussian correlation conjecture extended to multivariate gamma distributions. *arXiv*, p. 1408.1028v2.
171. Schekman, R. (2013) How journals like Nature, Cell and Science are damaging science. *The Guardian*.
172. Schoenfeld, J.D. et Ioannidis, J.P. (2013) Is everything we eat associated with cancer? A systematic cookbook review. *Am J Clin Nutr.*, 97:1, p. 127-34.
173. Senn, S. (2012) Misunderstanding publication bias: Editors are not blameless after all. *F1000Research*, 1.
174. Shadish, W. R., Tolliver, D., Gray, M. et Sen Gupta, S. K. (1995) Author judgements about works they cite: Three studies from psychology journals. *Social Studies of Science*, 25:3, p. 477-498.
175. Shea, D. A. (2003) CRS Report for Congress: Balancing scientific publication and national security concerns. 32 p.
176. Shiffrin, R.M., Börner, K. et Stigler, S.M. (2018) Scientific progress despite irreproducibility: A seeming paradox. *Proc Natl Acad Sci U S A.*, 115:11, p. 2632-2639.
177. Sijtsma, K. (2015) Playing with Data--Or How to Discourage Questionable Research Practices and Stimulate Researchers to Do Things Right. *Psychometrika*, 81:1, p. 1-15.
178. Silvertown, J. et McConway, K. J. (1997) Does « publication bias » lead to biased science? *Oikos*, 79:1, p. 167-168.
179. Simmons, J.P., Nelson, L.D. et Simonsohn, U. (2011) False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol Sci.*, 22:11, p. 1359-66.
180. Simonsohn, U. (2014) Posterior-Hacking: Selective Reporting Invalidates Bayesian Results Also. Available at SSRN: <https://ssrn.com/abstract=2374040>

181. Siontis, K.C.M., Evangelou, E., Ioannidis, J.P.A. (2011) Magnitude of effects in clinical trials published in high-impact general medical journals. *International Journal of Epidemiology*, 40, 1280–1291.
182. Smaldino, P. E. et McElreath, R. (2016) The natural selection of bad science. *Royal Society Open Science*, 3, p. 160384.
183. Smith, R. (2006) Peer review: a flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine*, 99:4, p. 178-182.
184. Snell, R. R. (2015) Menage a quoi? Optimal number of peer reviewers. *PLoS ONE*, 10:4.
185. Song, F., Parekh, S., Hooper, L., Loke, Y. K., *et al.* (2010) Dissemination and publication of research findings: An updated review of related biases (14). London, UK : 1-220 p.
186. Souder, L. (2011) The ethics of scholarly peer review: A review of the literature. *Learned Publishing*, 24:1, p. 55-74.
187. Spier, R. (2002) Peer review and innovation. *Science and Engineering Ethics*, 8:1, p. 99-108.
188. Stamatakis, E., Weiler, R. et Ioannidis, J. P. A. (2013) Undue industry influences that distort healthcare research, strategy, expenditure and practice: A review. *European Journal of Clinical Investigation*, 43:5, p. 469-475.
189. Statzner, B. et Resh, V. H. (2010) Negative changes in the scientific publication process in ecology: Potential causes and consequences. *Freshwater Biology*, 55:12, p. 2639-2653.
190. Steen, R. G., Casadevall, A. et Fang, F. C. (2013) Why has the number of scientific retractions increased? *PLoS ONE*, 8:7, p. e68397.
191. Sterling, T. D., Rosenbaum, W. L. et Weinkam, J. J. (1995) Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, 49:1, p. 108-112.
192. Stodden, V., Seiler, J. et Ma, Z. (2018) An empirical analysis of journal policy effectiveness for computational reproducibility. *Proc Natl Acad Sci U S A*, 115:11, p. 2584-2589.
193. Stowsky, J. (2004) Secrets to shield or share? New dilemmas for military R&D policy in the digital age. *Research Policy*, 33:2, p. 257-269.
194. Szucs, D. et Ioannidis, J.P.A. (2017) When Null Hypothesis Significance Testing Is Unsuitable for Research: A Reassessment. *Front. Hum. Neurosci.* 11:390.
195. Tennant, J.P., Dugan, J.M., Graziotin, D. *et al.* (2017) A multi-disciplinary perspective on emergent and future innovations in peer review [version 3; referees: 2 approved] *F1000Research*, 6:1151.

196. ter Riet, G., Korevaar, D. A., Leenaars, M., Sterk, P. J., *et al.* (2012) Publication bias in laboratory animal research: A survey on magnitude, drivers, consequences and potential solutions. *PLoS ONE*, 7:9, p. e43404.
197. Thompson, J., Baird, P. et Downie, J. (2001) Report of the committee of inquiry on the case involving Dr. Nancy Olivieri, the hospital for sick children, the University of Toronto, and Apotex Inc. Toronto : 527 p.
198. Tijdkink, J., Smulders, Y.M. et Vergouwen, A.C.M. (2014) Publication pressure and burn out among Dutch medical professors: a nationwide survey. *Clinical neuropsychopharmacology*, P.4.003.
199. Tijdkink, J.K., Schipper, K., Bouter, L.M., Maclaine Pont, P., *et al.* (2016) How do scientists perceive the current publication culture? A qualitative focus group interview study among Dutch biomedical researchers. *BMJ Open*, 6, p. e008681.
200. Till, J. E. (2001) Predecessors of preprint servers. *Learned Publishing*, 14:1, p. 7-13.
201. Trafimow, D., Amrhein, V., Areshenkoff, C.N., Barrera-Causil, C.J., *et al.* (2018) Manipulating the Alpha Level Cannot Cure Significance Testing. *Front. Psychol.*, 9:699.
202. Tregenza, T. (1997) Darwin a better name than Wallace? *Nature*, 385:6616, p. 480.
203. van Driel, M. L., Magin, P. J. et Del Mar, C. B. (2008) Journal impact factor and its importance for AFP. *Australian Family Physician*, 37:9, p. 770-773.
204. van Noorden, R. (2011) The trouble with retractions. *Nature*, 478, p. 26-28.
205. Vaucher, C., Bovet, E., Bengough, T., Pidoux, V., *et al.* (2016) Meeting physicians' needs: A bottom-up approach for improving the implementation of medical knowledge into practice. *Health Research Policy and Systems*, 14.
206. Wagner, A. K., Boninger, M. L., Levy, C., Chan, L., Gater, D. et Kirby, R. L. (2003) Peer review: Issues in physical medicine and rehabilitation. *American Journal of Physical Medicine and Rehabilitation*, 82:10, p. 790-802.
207. Web of Science Group (2019) *Journal Citation Reports – A Clarivate Analytics Company*.
208. Wisocki, A. B. (1983) Basic versus applied research: Intrinsic and extrinsic considerations. *Western Journal of Nursing Research*, 5:3, p. 217-224.
209. Wolfs, F. (1996) *Introduction to the scientific method. Physics Laboratory Experiments, Appendix E. Department of Physics and Astronomy, University of Rochester.*
210. Yarborough, M., Nadon, R. et Karlin, D.G. (2019) Four erroneous beliefs thwarting more trustworthy research. *Elife*, 8, p. e45261.

REMERCIEMENTS

Nous remercions tous les collègues, professeurs, chargés de cours et formateurs qui nous ont transmis la rigueur scientifique ainsi que l'analyse épistémologique : les deux valeurs principales qui créent le besoin pour lequel ce livre existe.

Notamment, Jessy remercie Bill Shipley pour l'avoir accompagné dans son apprentissage des statistiques et lui en avoir montré l'intérêt et la valeur. Jonathan remercie Marcos Balbinotti, expert en psychométrie, pour l'avoir éveillé à la valeur relative et intégrale des méthodes de collecte et d'analyse des données en sciences humaines. Jonathan remercie aussi René Marineau, psychologue qui a acquis une expérience théorique et pratique dans les différentes approches reconnues en psychologie, pour lui avoir brossé un portrait historique et inclusif de ces différentes approches et pour l'avoir sensibilisé à l'importance de tenir compte du contexte socioculturel où les grandes contributions à la science sont nées pour mieux en apprécier la valeur.

Jessy remercie son épouse, Hannah Loranger, pour son support lors de la réalisation de ce projet et pour toutes les conversations qui ont stimulé ses propres réflexions et qui ont grandement amélioré la présentation de ce livre.

Nous remercions Philippe Turchet, fondateur de la synergologie et docteur en science du langage, pour avoir accepté de lire le manuscrit et pour ses commentaires sur l'application de la démarche scientifique.

Nous remercions aussi Christine Gagnon, formatrice en synergologie spécialisée dans le domaine de la sécurité publique, pour son appui dans la réalisation de ce livre, et pour avoir mis Jonathan en contact avec le GRIQS, un groupe de recherche qui a stimulé ses réflexions.

Nous remercions finalement tous les amoureux de l'avancement des connaissances, qui ont écrit sans concession heuristique, et dont nous sommes les héritiers aujourd'hui.