

Olivier Sibony

*Les « études » ne sont pas
ce que vous croyez...*

Supplément à
La diversité n'est pas ce que vous croyez !

Flammarion, 2025

Introduction

La diversité n'est pas ce que vous croyez ! s'ouvre sur un constat : dans leur ensemble, les travaux de recherche ne soutiennent pas l'idée que la diversité dans les entreprises produise des effets économiques bénéfiques. Il n'est pas exact d'affirmer, de manière générale, que « la diversité est bonne pour le business ». Rappelons que ce n'est évidemment pas une raison pour y renoncer : une obligation éthique n'a pas besoin de justification économique. Paradoxalement, c'est même le contraire : en utilisant un argument utilitariste, à la fois fragile et inutile, on donne des armes aux adversaires des politiques de diversité, qui ont beau jeu de dénoncer des promesses non tenues.

Il n'en demeure pas moins que ce constat est, pour beaucoup de lecteurs, une surprise. Car les journaux et les réseaux sociaux qui parlent quasi quotidiennement du « *business case* » pour les politiques de diversité citent non seulement des témoignages et des exemples frappants, mais aussi quantité d'études et de rapports qui prétendent démontrer, chiffres à l'appui, des bénéfices tangibles.

Ces études ne sont pas des *fake news*. Ce sont généralement des travaux menés avec soin par des personnes compétentes et de bonne foi. Ils fournissent un exemple frappant d'un phénomène plus général, non seulement en management mais dans toutes les sciences sociales (et au-delà) : toutes les « études » dont les conclusions nous semblent séduisantes ne sont pas également dignes de foi.

Ce phénomène, qu'on désigne souvent sous le nom de « crise de la répliquabilité », a des causes multiples et complexes. Ce court supplément s'efforce d'en expliquer les grandes lignes, pour aider des non-spécialistes à répondre à une question essentielle, et pas seulement sur les sujets de diversité : *à quelles études peut-on se fier ?*

1

Comment démontrer que la diversité rapporte

Si les travaux de recherche sont peu concluants sur le *business case* en faveur de la diversité, si la proportion de femmes dans les conseils d'administration ou les équipes de direction n'a que peu d'effets, comment se fait-il que tant d'études, de rapports et de livres affirment le contraire ? Pour le comprendre, faisons une expérience de pensée : vous allez, le temps d'un chapitre, vous mettre dans la peau de l'auteur d'une de ces études.

Vous travaillez dans un cabinet de conseil, dans un institut d'études, ou encore dans une banque d'affaires. La mixité dans les instances de direction des entreprises est un sujet qui vous tient à cœur. Vous en parlez souvent avec vos clients. Hélas, ils ne sont pas convaincus. « La mixité, la diversité, c'est bien, vous disent-ils ; mais c'est un objectif de plus, et nous avons déjà un *business* à gérer ! »

L'objection vous semble absurde. La diversité est une chance pour les entreprises ! Ce n'est pas une charge supplémentaire, mais au contraire un facteur de succès. Vous en êtes convaincu, parce que vous l'avez vu dans vos propres équipes. Mais vous vous rendez bien compte que votre témoignage personnel ne suffira pas à convaincre des dirigeants qui ont d'autres expériences et qui travaillent dans des secteurs différents. Pour répondre en professionnel à ces professionnels, ce qu'il vous faut, ce sont des faits et des chiffres.

« Faisons une étude ! »

Vous en parlez à vos collègues, et bientôt la décision est prise : vous allez réaliser une étude pour mesurer le lien entre diversité et performance. C'est un investissement important, mais vos collègues vous soutiennent. Même ceux qui ne partagent pas votre enthousiasme pour la cause se disent qu'une belle plaquette sur ce

sujet d'actualité ne peut pas faire de mal à votre réputation. À condition, bien sûr, que ce soit un travail sérieux. Pas question de vous contenter de quelques exemples et d'un calcul de coin de table : vous allez étudier une masse de données, sur une longue période. Pour ce faire, vous affectez aussitôt à ce beau projet une de vos meilleures équipes.

Vos collaborateurs sont aussi enthousiastes que vous : c'est pour cela qu'ils ont accepté cette mission inhabituelle ! Vous leur expliquez donc leur tâche : mesurer, dans un échantillon d'entreprises, la mixité du management d'une part, la performance économique d'autre part, afin de démontrer la corrélation entre ces deux données. Par exemple, suggérez-vous, prenons les entreprises du CAC 40, mesurons le nombre de femmes dans leurs comités exécutifs, et voyons s'il est corrélé avec leur rentabilité.

« Mais quelle rentabilité ? » demande l'analyste qui pense déjà à sa feuille Excel. Le ratio résultat sur chiffre d'affaires ? Ou la rentabilité des capitaux ? Bonne question, bien sûr, qui, comme tous les managers du monde le savent, ne peut appeler qu'une seule réponse : « Essaie les deux ! » Impressionnés par cette démonstration de votre sagesse et de votre ouverture d'esprit, vos collaborateurs se mettent au travail.

Deux jours plus tard, les voilà de retour dans votre bureau. Perplexes, ils vous montrent le résultat de leur analyse. Il n'est guère convaincant. Sur les graphiques qu'ils vous présentent, le taux de mixité des entreprises figure en abscisse et la rentabilité en ordonnée. Les quarante entreprises de l'échantillon sont représentées par quarante points. Et ces quarante points forment un nuage d'où ne se dégage aucune logique évidente – en tous cas, pas la belle droite que vous aviez espérée. Qu'on représente en ordonnée la marge opérationnelle ou le rendement des capitaux n'y change rien.

Heureusement, votre équipe a des idées. Peut-être, suggèrent-ils, faudrait-il élargir l'échantillon ? Quarante entreprises, ce n'est pas beaucoup. Vous êtes d'accord, bien sûr. Et en y réfléchissant, une autre idée vous vient : ces mesures de rentabilité sont beaucoup trop rudimentaires ! Il y a dans votre échantillon des entreprises de secteurs très différents. Leur rentabilité varie pour bien d'autres raisons que la présence ou non de femmes dans le top management.

Vous proposez donc de remplacer la rentabilité brute par une mesure de l'écart entre la rentabilité de l'entreprise et la moyenne de son secteur. « Et vous allez voir, ça va marcher », expliquez-vous.

Voilà qui met du baume au cœur de votre équipe. Après une collecte méticuleuse et une analyse rigoureuse des nouvelles données, ils reviennent pourtant dans votre bureau en faisant grise mine : le nuage de points, qui comprend maintenant une centaine d'entreprises, refuse obstinément de s'aligner.

« Persévérons ! »

Diable ! Ce problème va finalement vous demander plus de réflexion que vous ne le pensiez. Mais au fond, vous dites-vous, c'est assez logique. La rentabilité varie non seulement d'un secteur à l'autre, mais aussi d'année en année. En utilisant seulement l'année en cours, comme vous l'avez fait jusqu'à maintenant, vous avez introduit dans vos données une variabilité qui masque l'effet de la diversité sur la performance. Il faut trouver un moyen de neutraliser cette variabilité – par exemple, en prenant la moyenne de la rentabilité sur plusieurs années. Votre équipe a même une meilleure idée : pourquoi ne pas utiliser la performance boursière, qui met toutes les entreprises sur la même ligne de départ ? Cela exclut les entreprises non cotées, mais au moins, vous serez certains de la qualité des données.

Ce brainstorming vous fait d'ailleurs penser à d'autres possibilités. Peut-être que la mixité n'influe pas sur l'efficacité de manière linéaire. On vous a souvent parlé d'une sorte d'effet de seuil, une « masse critique » de femmes dans une équipe qui change sa performance ^{1,2}. Ce ne serait donc pas le pourcentage de femmes dans le comité exécutif qu'il faudrait mesurer, mais leur nombre absolu. Peut-être aussi – ce serait logique – l'effet de la mixité sur la performance n'est-il pas immédiat, et faut-il introduire dans votre analyse un décalage dans le temps : la mixité de l'année n produirait des effets sur le profit des années $n + 1$ à $n + 4$, par exemple.

Le ballet des hypothèses et des résultats se poursuit quelques semaines. Le travail est plus ardu que vous ne le pensiez, mais vos efforts n'auront pas été vains. Car à force de réflexion, vous avez fini

© Olivier Sibony, 2025

par trouver la bonne mesure de mixité, le bon échantillon, et le bon indicateur de rentabilité. Sur le graphique, les points ne sont pas parfaitement alignés (ce serait trop beau), mais ils dessinent une sorte de comète tout à fait convaincante. Vous avez d'ailleurs demandé à votre équipe de recalibrer les axes du graphique pour accentuer cet effet visuel. Bien entendu, comme vous avez quelques souvenirs de vos cours de statistiques, vous avez demandé qu'on calcule le coefficient de corrélation, et il est excellent. Il vous semble donc parfaitement légitime de revendiquer la découverte d'un lien statistique réel, et de le mettre en évidence visuellement. Avec tout le mal que vous vous êtes donné, ce serait dommage de gâcher le résultat par une erreur de communication.

Vous présentez ce travail à vos collègues. Comme vous, ils sont ravis du résultat et anticipent déjà ses retombées dans la presse. Mais le plus expérimenté d'entre eux a encore une suggestion : il vous demande de résumer tout cela par un chiffre choc qu'on pourra reprendre dans le communiqué de presse. Au fond, vous demanderont les journalistes, la mixité, ça rapporte combien ?

Vous êtes un peu réticent à réduire tout votre travail à une simple formule. Mais vous savez que votre collègue a raison. La patience des médias et des réseaux sociaux a ses limites. Vous calculez donc l'écart de résultat entre les 10 % d'entreprises qui ont le plus de diversité et la moyenne. Bien que positif, il est faible. Après quelques tâtonnements, vous trouvez une comparaison plus parlante : l'écart de rentabilité entre les 20 % d'entreprises les plus diverses et les 20 % les moins diverses, s'avère, lui, considérable. Faire ressortir le contraste entre les plus avancés et les traînants correspond parfaitement à votre message. Le titre de votre communiqué de presse est tout trouvé : « La mixité, c'est 10 % de profits en plus ! »

Tout faux, en toute bonne foi

Beau travail ? Hélas, il est éminemment suspect.

Naturellement, les protagonistes de cette histoire (fictive mais réaliste) seraient très choqués par ce jugement. Dans leur comportement n'entre aucune malhonnêteté ni aucune négligence. Ils ne sont mus que par les meilleures intentions et ne ménagent pas

© Olivier Sibony, 2025

leur peine. On voit bien, certes, qu'ils ont un parti pris, et que leurs analyses pourraient être influencées par le biais de confirmation. Mais enfin, ce sont des professionnels des chiffres ! Ils n'ont pas inventé les données ni travesti les faits. Ils avaient des hypothèses, bien sûr ; mais ces hypothèses, ils les ont vérifiées. Chacun sait ce qu'est une corrélation et comment la mesurer. Et si nous savons tous que « corrélation n'est pas causalité », cela appelle des précautions oratoires, mais ne retire rien à l'évidence du résultat. Donc, où est le problème ?

Le problème est subtil, en effet. Il l'est même assez pour avoir échappé pendant des années non seulement aux analystes « amateurs » que met en scène cette histoire, mais aussi aux chercheurs professionnels.

Ce que nous avons vu tout au long du processus d'analyse qu'a suivi notre équipe (presque) imaginaire, c'est qu'elle a fait une série de choix. Choix des variables qui rendent mesurables les concepts, en réalité vagues, de « diversité » et de « performance ». Choix des échantillons d'entreprises sur lesquels on les mesure. Choix des sociétés à inclure ou à exclure, par exemple quand certains chiffres sont manquants ou suspects. Choix des trimestres ou des années qu'on analyse, du nombre d'années qu'on prend en compte, des décalages dans le temps qu'on suppose entre les causes et les effets. Choix des méthodes d'analyse – ici une simple recherche de corrélation, mais on aurait pu en imaginer beaucoup d'autres. Choix des comparaisons à mettre en évidence ou à dissimuler dans la présentation des résultats. Aucun de ces choix n'est absurde : on trouve toujours une bonne raison pour les justifier. À chaque fois, même, on a l'impression de *comprendre* de plus en plus finement les mécanismes qu'on étudie. Apparemment, la théorie s'affine au contact des données.

Mais ce faisant, les choix se cumulent et se multiplient, tant et si bien que le résultat *positif* présenté est un résultat parmi des dizaines, des centaines, des milliers de résultats *négatifs* que l'équipe aurait pu produire si elle avait fait d'autres choix. Ces résultats négatifs, d'ailleurs, tout au long de son travail, elle les a produits – et rejetés comme non satisfaisants. Si elle s'était arrêtée à la première analyse, elle aurait conclu qu'elle ne trouvait pas de lien statistique entre

diversité et performance (ce qui ne prouve pas qu'il n'en existe pas). Mais justement, elle ne s'est pas arrêtée là.

En quoi est-ce une faute, demanderez-vous ? N'est-ce pas une vertu que de chercher jusqu'à ce que l'on trouve ? Dans la vie courante, si, bien sûr. Mais dans une analyse statistique, c'est une erreur de raisonnement. Car si beaucoup de corrélations reflètent des réalités, il existe aussi des corrélations fictives qui ne sont produites que par le hasard. Il existe par exemple une corrélation quasi parfaite entre le nombre de diplômés en études interdisciplinaires et la production d'électricité en Angola, mais personne ne conclurait que l'un des phénomènes influe sur l'autre¹. Statistiquement parlant, si l'on cherche assez longtemps, dans assez de données, et avec assez de créativité, on finit *toujours* par trouver ce qu'on cherche. Cela prouve, certes, qu'on a cherché longtemps... mais pas qu'on peut tirer une conclusion. Quand on cherche l'aiguille dans la botte de foin et qu'on finit par la trouver, on n'en conclut pas que la botte est entièrement composée d'aiguilles.

Il s'agit bien d'une erreur de raisonnement, et non simplement de technique. Certes, l'étude décrite plus haut ne brille pas par la rigueur ni par la sophistication des méthodes analytiques qu'elle emploie. Mais même des chercheurs qui utilisent des méthodes statistiques autrement plus avancées sont souvent tombés, involontairement ou intentionnellement, dans le même piège. Cette pratique est même l'une des causes principales de la « crise de répliquabilité » en sciences, comme va le montrer le chapitre suivant.

**Même écrites par des auteurs honnêtes et sérieux,
les études qui visent à confirmer
une hypothèse séduisante sont souvent fausses.**

¹ Le site www.tylervigen.com répertorie des milliers de corrélations de ce type. Et il n'est pas entièrement exact que « personne » n'y voit de lien causal, puisque le site propose une « explication » du mécanisme à l'œuvre (générée, naturellement, par l'IA). Par exemple : « À mesure que de plus en plus d'étudiants se plongeaient dans le monde vaste et varié des études multi/interdisciplinaires, ils ont déclenché une montée en puissance intellectuelle. Cette tempête intellectuelle a conduit à un afflux d'idées innovantes pour améliorer l'efficacité de la production d'électricité. »

© Olivier Sibony, 2025

2

Devenez plus puissante en quelques minutes

Peut-être avez-vous déjà vu le *Ted Talk* dans lequel Amy Cuddy, chercheuse à Harvard, parle des « poses de pouvoir ». Douze ans après sa mise en ligne, cette courte conférence a été visionnée 71 millions de fois sur la plateforme TED, ce qui en fait le deuxième plus grand succès parmi les milliers de conférences que présente ce site. La chercheuse y explique que « votre langage corporel forge qui vous êtes » : adopter une « pose de pouvoir » modifie le mix d'hormones qui circule dans votre sang, augmente votre sentiment de confiance, et favorise donc votre réussite à des moments critiques.

Concrètement, explique Amy Cuddy, dans les situations d'évaluation, par exemple avant un entretien d'embauche, ne vous repliez pas sur vous-même (ni sur votre téléphone !). Au contraire, redressez-vous, étirez-vous, remplissez l'espace autour de vous. Si vous faites cela pendant quelques minutes (éventuellement en vous cachant aux toilettes, suggère la pragmatique conférencière), vous vous sentirez mieux. Faites cela même (et surtout) quand vous ne vous sentez pas confiant(e), et le succès suivra.

Voilà un conseil simple, que beaucoup adoptent avec enthousiasme, et notamment beaucoup de femmes, puisque c'est à elles que l'autrice s'adresse avec sincérité et émotion³. Elle souligne que les poses de pouvoir sont une application concrète du principe « *fake it till you make it* » (« simulez jusqu'à ce que ça marche »). Mais attention, ce n'est pas un simple conseil de développement personnel ! Car Amy Cuddy présente toutes les garanties de crédibilité scientifique. Elle est titulaire d'un doctorat en psychologie de Princeton et chercheuse à Harvard. Ses travaux sur les « poses de pouvoir » ont fait l'objet d'un article dont la co-autrice principale, Dana Carney, chercheuse à Berkeley, a publié une cinquantaine de papiers. Et l'article lui-même est publié dans la très prestigieuse revue *Psychological Science*, ce qui signifie qu'il a franchi toutes les étapes du

© Olivier Sibony, 2025

processus de *peer review* (revue par les pairs) conçu pour débusquer toute erreur ou irrégularité.

Bref, même vulgarisé dans un *Ted Talk*, c'est du solide. L'article scientifique est catégorique : « Le fait qu'une personne puisse, en prenant deux poses simples pendant une minute, incarner la puissance et devenir instantanément plus puissante, a des implications exploitables dans le monde réel⁴. »

Pourtant, une polémique scientifique va se déclencher. Car, quelques années après la publication de cet article, plusieurs groupes de chercheurs ont tenté de reproduire de manière indépendante l'expérience des « poses de pouvoir ». Et, contrairement à Amy Cuddy, ils n'ont pas constaté les effets attendus. Les conclusions « ne répliquent pas ».

Dana Carney, la co-autrice, ne tardera pas à se rétracter, dans des termes inhabituellement virulents : « je ne pense pas que ces effets soient réels » ; « les preuves de leur inexistence sont indéniables » ; « je n'étudie pas les effets des poses de pouvoir ; je déconseille de les étudier »⁵. Il est vrai que les raisons de douter se sont, entretemps, accumulées. Amy Cuddy, pour sa part, maintient ses conclusions, en les nuancant légèrement⁶. Forte du succès planétaire de son *Ted Talk*, elle poursuit une carrière de conférencière de haut vol : n'espérez pas l'entendre parler pour moins de 75 000 dollars, indique le site de son agent.

Au-delà de la question des poses de pouvoir, cette polémique est devenue le symbole d'une crise plus vaste : la « crise de la répliquabilité » (ou de la « reproductibilité ») en psychologie et plus largement en sciences sociales. Depuis 2011, en effet, un mouvement de fond remet en question un nombre croissant de travaux de recherche pourtant publiés dans les meilleures revues. En psychologie, plusieurs équipes ont tenté de répliquer systématiquement les articles les plus célèbres et les plus souvent cités : ils n'y sont parvenus que pour à peine plus de la moitié d'entre eux – et encore, avec des effets généralement bien moins marqués^{7,8}.

Si « la plupart des conclusions publiées sont fausses », comme l'affirme le titre d'un article de 2005, les études auxquelles nous faisons confiance sont suspectes – et c'est tout aussi vrai en économie, et même en médecine, qu'en management ou en

psychologie sociale ⁹. Il est donc essentiel pour chacun d'entre nous de comprendre ce qui sous-tend cette crise. Car elle révèle en réalité plusieurs problèmes de nature très différente.

Fraude

Le plus spectaculaire, mais heureusement pas le plus fréquent, est celui de la fraude pure et simple. Contrairement à Amy Cuddy et à ses collègues, dont personne ne remet en cause l'intégrité, plusieurs chercheurs de grande réputation ont été récemment accusés de fraude scientifique caractérisée ². L'un des exemples les plus connus est une étude co-signée par Dan Ariely, professeur à Duke et auteur de nombreux ouvrages à succès ¹⁰.

Ce qui fait tout le sel de ce cas de fraude est que l'étude en question portait, justement, sur la malhonnêteté ! Son objet est un simple formulaire, celui d'une compagnie d'assurances qui demande à ses clients d'indiquer le nombre de kilomètres parcourus. Pour minorer leur prime d'assurance, les clients peuvent être tentés de sous-estimer ce kilométrage. Ariely teste donc deux versions du formulaire. Dans la version originale, les clients certifient la sincérité de leur déclaration en la signant, comme c'est l'usage, en bas de page. Dans la version « améliorée », au contraire, ils signent au début, *avant* de noter leur kilométrage.

Cette idée repose sur une théorie : en soulignant l'importance de l'intégrité en amont, pense Ariely, les clients se comporteront de manière plus éthique. Et la théorie se vérifie, puisque les clients signant au début déclarent un kilométrage plus important que ceux signant à la fin, ce qui suggère qu'ils fraudent moins. D'où une recommandation simple et pleine de bon sens : faire signer les déclarants au début du formulaire, plutôt qu'à la fin. Cette idée sera d'ailleurs rapidement adoptée par de nombreuses entreprises, et même par des États pour leurs formulaires de déclarations fiscales.

² On parle ici de fraude portant sur les données et leur analyse. D'autres formes d'inconduite scientifique ont fait l'objet d'une attention nouvelle au cours des dernières années, en particulier les cas de plagiat. Ceux-ci, pour condamnables qu'ils soient, ne remettent en principe pas en cause la validité des conclusions.

© Olivier Sibony, 2025

L'histoire est belle... mais, hélas, fausse. Car les données de l'étude étaient grossièrement trafiquées. Une partie des chiffres présentait des caractéristiques statistiques très différentes des autres, suggérant qu'ils étaient fabriqués de toutes pièces par les expérimentateurs. Signe d'une étonnante absence de précautions, ils étaient même parfois saisis dans une police de caractères distincte... Ariely convient aujourd'hui que les données étaient trafiquées, mais nie être à l'origine de la fraude ¹¹⁻¹³. Et de nombreuses autres expériences ont conclu que signer au début d'un formulaire ne change pas le comportement des signataires. Cette accumulation de données contredisant l'étude initiale a conduit les auteurs de l'article à publier une rétractation officielle ¹⁴.

Une autre affaire de fraude présumée concerne justement l'une des co-autrices de l'article en question, Francesca Gino. Gino était une des stars de la Harvard Business School, réputée à la fois pour sa « productivité » étonnante (c'est-à-dire pour le nombre d'articles scientifiques qu'elle publiait) et pour sa capacité à vulgariser ses idées. Elle est accusée d'avoir falsifié des données dans au moins quatre articles, dont celui co-écrit avec Ariely ¹⁵. (Vous avez bien lu : apparemment sans se concerter, deux co-auteurs auraient falsifié les données de la même étude *consacrée à la malhonnêteté*.)

Comme Ariely, Francesca Gino s'est distinguée par un intérêt prononcé pour la transgression : l'un de ses articles suspects est sous-titré « Comment la malhonnêteté peut rendre plus créatif » ; et son dernier livre s'intitule *Talents rebelles. Pourquoi il faut violer les règles au travail et dans la vie*. Il semble bien que l'autrice ait mis en pratique ses propres préceptes : les indices de fraude, rassemblés par Harvard dans un rapport de 1 200 pages, sont nombreux, concordants et, pris dans leur ensemble, plus que préoccupants ¹¹.

Gino clame elle aussi son innocence et, après avoir été suspendue par Harvard, a attaqué en justice non seulement son université, qu'elle accuse de discrimination, mais aussi les scientifiques qui ont lancé l'alerte ¹⁶. Poursuivre en diffamation les auteurs d'une critique scientifique n'est pas, c'est le moins qu'on puisse dire, conforme aux

usages académiques ³. C'est dire si la crise de la répliquabilité secoue la communauté universitaire.

p-hacking

D'où est partie cette petite révolution, et qui sont les dangereux agitateurs qui l'ont lancée ? Principalement, trois chercheurs de Wharton et Berkeley, qui ont publié en 2011 un article retentissant pour dénoncer les faiblesses méthodologiques de beaucoup de publications scientifiques ; et qui tiennent depuis un blog, *Data Colada*, consacré aux problèmes de non-répliquabilité et de fraude.

La thèse centrale de Joseph Simmons, Leif Nelson et Uri Simonsohn est énoncée dans le titre de leur article original : « La *flexibilité dissimulée* dans la collecte et l'analyse des données permet de présenter *n'importe quoi* comme significatif ¹⁷ » (c'est moi qui souligne). Pour l'illustrer, les chercheurs recourent à une étonnante démonstration par l'absurde. Après avoir recruté 20 étudiants, expliquent-ils, ils leur ont fait écouter soit la chanson des Beatles *When I'm Sixty-Four*, soit une autre chanson. Ils leur ont ensuite demandé d'indiquer leur âge. Résultat : ceux qui ont écouté *When I'm Sixty-Four* sont plus jeunes de 18 mois en moyenne que les autres. Non seulement on peut devenir instantanément plus puissant en se tenant bien droit deux minutes, mais on peut rajeunir d'un an et demi en écoutant une chanson des Beatles : décidément, les progrès de la science sont ébouriffants.

Comment des chercheurs sérieux peuvent-ils arriver à un résultat pareil ? Exactement de la même manière, expliquent-ils, que dans l'exemple fictif du chapitre précédent. À chaque étape de la collecte et de l'analyse de leurs données, les chercheurs ont des « degrés de liberté » dans la conduite de leur étude. Toute étude est un « jardin aux sentiers qui bifurquent ¹⁸ », où l'on décide des méthodes, des données et des présentations qu'on va utiliser. En empruntant les sentiers les plus compatibles avec ses hypothèses, on parvient au résultat recherché.

³ En septembre 2024, la justice fédérale américaine a rejeté la plainte en diffamation de Francesca Gino. La plainte pour discrimination, en revanche, suit son cours.
© Olivier Sibony, 2025

Il peut sembler étrange que les méthodes statistiques les plus modernes ne suffisent pas à prévenir ce problème. Car enfin, quand on fait une étude sérieuse, on ne se contente pas, comme des amateurs, d'aligner des points pour voir s'ils forment plus ou moins une droite. Toute étude sérieuse utilise des tests statistiques, qui permettent d'affirmer qu'un résultat est « significatif ». Ces tests ne servent-ils pas, précisément, à garantir que les résultats observés sont fiables ?

Pour répondre à cette question, il nous faut revenir à une idée essentielle, mais parfois perdue de vue par les scientifiques eux-mêmes : les tests de significativité mesurent des *probabilités*. Un résultat « significatif », ce n'est pas un résultat certain : c'est un résultat qui n'a qu'une *faible probabilité d'être dû au hasard*.

Illustrons ce principe par l'exemple des « poses de pouvoir ». L'article de Cuddy et de ses collègues affirme que la prise d'une pose de pouvoir fait baisser, dans la salive des sujets, le taux de cortisol (une hormone considérée comme un marqueur de stress). Mais le taux d'hormones varie d'un sujet à l'autre et d'un moment à l'autre, pour mille raisons connues ou non. Sa variation pendant l'expérience pourrait n'être due qu'au hasard.

Ce que nous indique l'article, c'est justement la probabilité, appelée *p-value*, de ce scénario où les variations ne sont dues qu'au hasard, scénario appelé *hypothèse nulle*. Pour croire qu'un effet est réel, et non le fruit du hasard, il faut donc que la *p-value* soit aussi faible que possible. En l'occurrence, elle est de moins de 2 %. Autrement dit, si la pose de pouvoir n'avait en réalité *aucun effet*, il y aurait moins de 2 % de chances qu'on observe une aussi forte variation des taux d'hormones. Cela vous semble sans doute suffisant pour inspirer confiance⁴. Et de fait, par convention, une *p-value* inférieure à 5 % permet d'affirmer que l'hypothèse nulle est assez peu probable pour qu'il y ait matière à publication. Voilà pourquoi cette étude a été publiée.

⁴ Un point d'attention : la valeur *p* indique la *probabilité de trouver les résultats observés* (ou plus extrêmes) si l'hypothèse nulle est vraie, $p(R|H_0)$. Elle n'indique pas la *probabilité que l'hypothèse nulle soit vraie* sachant qu'on a trouvé les résultats observés, $p(H_0|R)$. Ici, il serait tentant de conclure qu'il y a 2 % de chances que l'hypothèse nulle soit vraie (ou 98 % de chances qu'elle soit fausse), mais ce ne serait pas exact.
© Olivier Sibony, 2025

Mais souvenons-nous un instant des innombrables choix que faisaient les analystes du chapitre précédent. Comme eux, les chercheuses qui étudient les poses de pouvoir ont une multitude de degrés de liberté. Le résultat qu'elles nous présentent n'est qu'un des très nombreux résultats qu'elles auraient pu présenter (et qu'elles ont peut-être testés). Le fait que la *p-value* soit une probabilité, et non une certitude, devient alors un très sérieux problème. Car même un phénomène qui n'a que 2 % de chances d'être observé par hasard sera observé par hasard... dans 2 % des cas.

Imaginons par exemple une urne qui contient 50 boules : 49 boules rouges et une seule boule verte, la boule gagnante. Vous n'avez que 2 % de chances de tirer la boule verte. Heureusement, vous avez une arme secrète : une formule magique à prononcer avant le tirage. Si vous tirez la boule verte du premier coup, il est certain que votre auditoire sera impressionné par l'effet de vos incantations ! Mais si vous remettez la boule dans l'urne et répétez l'opération cinq, dix, ou trente fois, jusqu'à ce que sorte la boule verte, votre démonstration sera nettement moins convaincante... alors, pourtant, que vous n'avez toujours que 2 % de chances de gagner à chaque tirage.

Les chercheurs ne puisent pas leurs résultats au hasard dans des urnes. Mais si leurs études sont entachées de « flexibilité dissimulée », comme le disent Simmons et ses collègues, le problème est le même.

Revenons au résultat absurde de l'étude sur les Beatles. Ce que vous ignoriez quand vous l'avez lu, c'est que les participants n'avaient pas seulement répondu à une question sur leur âge, mais aussi à une douzaine d'autres questions. Il s'est avéré – par hasard, bien sûr – que l'écart de réponse entre les deux sous-groupes était considérable sur la question de l'âge. C'est donc ce résultat que nous rapportent les auteurs. Mais si une différence importante entre les deux sous-groupes était apparue sur les opinions politiques ou sur la capacité à faire une division, les chercheurs auraient pu affirmer, de manière tout aussi absurde, qu'écouter une chanson des Beatles modifie votre vote ou affecte vos compétences mathématiques.

Le chercheur qui nous présente un seul résultat, et une seule *p-value*, alors qu'il a posé vingt questions, se comporte exactement comme le gagnant du jeu qui brandit fièrement la boule verte en

omettant de préciser qu'elle n'est sortie qu'après vingt tirages. Il nous fait prendre des vessies pour des lanternes, ou plutôt des hasards pour des effets. Simmons, Nelson et Simonsohn ont donné à cette pratique un nom devenu fameux : le *p-hacking*, c'est-à-dire le « piratage de *p-value* ».

Dans la démonstration par l'absurde de la chanson des Beatles, le *p-hacking* est délibéré : on cherche jusqu'à ce qu'on trouve. C'était aussi le cas dans l'exemple du lien diversité-performance au chapitre précédent. Mais, rappelons-le, délibéré ne veut pas dire mal intentionné. Comme on l'a vu, les managers qui multiplient les analyses n'ont pas le sentiment de commettre une quelconque irrégularité. De même, jusqu'à une époque récente, les chercheurs ne se formalisaient pas d'une petite dose de *p-hacking*. Comme l'ont écrit Simmons et ses co-auteurs, « tout le monde savait que c'était mal, mais on pensait que c'était un peu comme traverser en-dehors des clous. Ce dont on s'est rendu compte, c'est que c'est plutôt comme braquer une banque ¹⁹ ».

Le biais de publication

En plus de la fraude et du *p-hacking*, un troisième phénomène vient contribuer à la crise de la répliquabilité. Le problème est simple : personne n'a envie de lire, et encore moins de publier, une étude qui n'a rien trouvé d'intéressant ! Imaginons qu'Amy Cuddy et ses collègues aient « découvert » que prendre la pose de King Kong n'a strictement *aucun effet* sur le mix hormonal ni sur le sentiment de puissance des participants. Ou que Dan Ariely et ses co-auteurs aient conclu que signer au début du formulaire plutôt qu'à la fin ne change *absolument rien* à la propension à la fraude. Sans doute ces chercheurs seraient-ils passés à autre chose sans se donner la peine de rédiger les conclusions de leur étude. Les scientifiques, évalués sur le nombre et la qualité de leurs publications, n'ont pas de temps à perdre à enfoncer des portes ouvertes.

Les auteurs auraient donc classé leur étude dans un coin de leur bureau : on parle de ce phénomène comme de l'« effet tiroir », ou *file drawer problem*. Et s'ils avaient tout de même fait cet effort et soumis leurs résultats à une revue scientifique, le comité de lecture de celle-

© Olivier Sibony, 2025

ci aurait sans doute donné la priorité à des résultats « positifs », forcément plus intéressants pour les lecteurs que les résultats « négatifs ». Ces deux mécanismes concourent à ce que les résultats les plus surprenants et les plus contre-intuitifs soient surreprésentés dans la littérature. C'est ce qu'on appelle le *biais de publication* ⁵.

Fraude, *p-hacking* et biais de publication comptent parmi les principales causes de la crise de la répliquabilité. Ces problèmes distincts appellent des solutions distinctes, et la communauté scientifique s'attache à les mettre en place. Les meilleures revues déploient désormais une batterie de précautions techniques pour « solidifier » leurs publications ⁶. Il n'est pas praticable pour les lecteurs occasionnels d'études scientifiques sur le management (ou d'ailleurs sur tout autre sujet) de prendre eux-mêmes ces précautions. Et il faut bien sûr se garder de la réaction extrême qui consisterait à mettre en doute toutes les conclusions scientifiques au motif que certaines sont remises en question.

Mais, comme nous le verrons au chapitre suivant, il est possible de se poser quelques questions simples pour développer son esprit critique. Car même les meilleures sources, avec les meilleures intentions, peuvent induire en erreur.

**Quand une conclusion est vraiment très surprenante,
elle est souvent fausse.
Et il n'y a besoin ni de fraude, ni d'incompétence,
ni de conflits d'intérêts pour que ce soit le cas.**

⁵ À moins, bien sûr, que les résultats négatifs ne viennent invalider des résultats positifs déjà publiés et considérés comme acquis : c'est tout l'intérêt des campagnes de répliquabilité déjà mentionnées.

⁶ La description détaillée de ces techniques sort du champ de ce livre. Les lecteurs intéressés pourront se renseigner sur la *pre-registration*, qui contraint les auteurs d'une étude à indiquer à l'avance les données qu'ils vont collecter et la manière dont ils vont les analyser, et le *p-curve*, qui recense sur une courbe la distribution des *p-values* obtenues sur une question donnée.

© Olivier Sibony, 2025

3

Comment lire une étude ?

Nous avons vu que, sur la diversité, beaucoup d'idées répandues, et prétendument étayées par des études, sont fausses. Et ce problème n'est pas spécifique au sujet de la diversité : la crise de la réplicabilité touche toutes les sciences.

Cette observation conduit certains à une position extrême : ne plus faire confiance à aucune source scientifique. Vous les avez entendus, ces arguments : « Les chiffres, on leur fait dire ce qu'on veut » ; « des experts, il y en a toujours pour dire une chose et son contraire » ; « je m'en remets à mon bon sens ». Dans tous les domaines, et notamment pour les managers, une telle attitude revient non seulement à se priver d'informations importantes, mais condamne aussi celui qui l'adopte à commettre de lourdes erreurs.

Il est pourtant possible de faire un premier tri entre les sources fiables et celles qui le sont moins, même si l'on n'est pas soi-même chercheur. De même qu'il n'est pas nécessaire de savoir cuisiner pour être un gastronome averti, un lecteur de la presse managériale et des réseaux sociaux peut devenir un « consommateur éclairé » des travaux de recherche, et éviter la plupart des pièges que nous avons énumérés. Les questions qui suivent, ordonnées de la question la plus facile à la plus technique, ne sont pas exhaustives, mais peuvent servir de « check-list » pour évaluer, dans la presse généraliste, la fiabilité d'un article qui fait référence à des travaux de recherche.

1. Clarté : *Les affirmations sont-elles claires et précises ?*

La question peut sembler triviale, mais on ne compte plus les articles et les rapports où la seule formulation de la conclusion devrait suffire à la mettre en doute. Si vous lisez par exemple que la diversité « augmente l'innovation de 19 % » ou que les équipes diverses « ont une performance supérieure dans 87 % des cas », l'absence de définition des termes « innovation » et « performance » doit vous mettre la puce à l'oreille.

Inversement, il faut aussi se méfier d'une précision exagérée. Quand un rapport du Boston Consulting Group affirme qu'une entreprise qui compte 1 500 dirigeants devra, pour obtenir 1 % de chiffre d'affaires en plus, remplacer précisément 38 de ces dirigeants par des femmes, ou, à défaut, 23 par des personnes provenant d'un autre pays, une lecture critique s'impose ²⁰...

2. Émotion : Le compte-rendu fait-il appel à vos émotions ?

Une communication qui suscite l'émotion, c'est une évidence, gagne immédiatement en impact. Le *Ted Talk* d'Amy Cuddy est un modèle du genre : sur un sujet aride, où il est question de mesure des taux d'hormones dans la salive des volontaires, la conférencière, émue aux larmes, relie ses observations avec son histoire personnelle pour montrer que la volonté et le travail peuvent triompher de l'adversité. La *standing ovation* est garantie.

Mais, c'est une évidence aussi, l'émotion n'aide pas à évaluer froidement la qualité de l'information. Ce n'est pas par hasard que la désinformation et les *fake news* se caractérisent par leur propension à faire appel aux émotions, en particulier aux émotions négatives (colère, indignation...) ²¹. Même dans la presse généraliste, il n'est pas rare qu'un article sur les effets délétères des pesticides s'ouvre sur le récit des souffrances d'un agriculteur qui y a été trop longtemps exposé ; ou qu'on souligne la cupidité des multinationales du médicament avant de parler d'une étude sur l'efficacité des traitements qu'elles commercialisent. L'effet est, bien sûr, de susciter l'émotion pour anesthésier l'esprit critique (ce qui fonctionnera d'autant mieux chez le lecteur a priori disposé à se méfier de l'agrochimie et des laboratoires). Un effort conscient s'impose pour faire abstraction, autant que possible, du *pathos*, et se concentrer sur l'évaluation du *logos* – la logique des arguments – et de l'*ethos* – la crédibilité des sources.

3. Source : La source originale est-elle mentionnée et crédible ?

Un « rapport » mentionné sur LinkedIn n'est pas une étude scientifique ; et les assertions du type « un expert affirme » ou « une étude a démontré » doivent alerter le lecteur. Un renvoi à une source

© Olivier Sibony, 2025

scientifique *peer-reviewed*, où les publications sont passées au crible critique de l'examen par leurs pairs, est au contraire un indice (sinon une garantie) de qualité. Quand un article scientifique est cité, encore faut-il, bien sûr, qu'il ait été correctement analysé et résumé. Mais l'expérience suggère que, dans leur grande majorité, les publications douteuses ne se donnent pas la peine de citer une source originale. Dans le doute, il est facile de vérifier la réputation des auteurs en recherchant sur Google Scholar des indicateurs (imparfaits mais utiles) comme le nombre de citations obtenues par les travaux d'un chercheur, ou son *h-index*, qui mesure leur impact ⁷.

4. Ton : *L'article laisse-t-il une place raisonnable au doute ?*

Les charlatans sont toujours catégoriques, quand les véritables experts sont prudents. La polémique sur l'hydroxychloroquine au début de la pandémie de Covid-19 en fournit un exemple frappant. Ce ne sont pas les qualifications scientifiques du Pr Raoult qui pouvaient alerter : elles étaient impressionnantes. Mais le ton péremptoire qu'employait le professeur pour présenter ses conclusions (et tenter de discréditer ses adversaires) était en lui-même un signal d'alarme. Quand une conclusion importante et non triviale est présentée comme ne souffrant pas le débat, c'est en soi une raison pour en douter.

5. Parti pris : *Si la conclusion était opposée, la même source l'aurait-elle publiée ?*

Il est devenu habituel de prêter attention aux conflits d'intérêts des auteurs – et c'est bien sûr une saine précaution. Mais même en l'absence de conflits d'intérêts objectifs, chaque source a sa ligne éditoriale, sa culture, ou ses partis pris. À votre avis, McKinsey, Catalyst ou Crédit Suisse auraient-ils publié une étude intitulée « Diversité ou pas, cela ne changera rien à vos résultats ? » Un « TED Talk » aurait-il accumulé les millions de vues s'il clamait : « inutile de

⁷ Le *h-index* est défini comme le plus grand nombre *h* tel que le chercheur a publié au moins *h* articles ayant reçu chacun au moins *h* citations. Par exemple, un chercheur a un *h-index* de 10 s'il a au moins 10 articles qui ont été cités au moins 10 fois chacun.

simuler, ça ne sert à rien » ? Le journal où vous lisez que les banques dirigées par des femmes sont mieux gérées aurait-il publié un article affirmant le contraire ? À vous d'en juger, bien sûr. Mais si vous pensez que la réponse est non, ne perdez pas votre temps à lire des articles dont vous connaissez déjà la conclusion.

Notons au passage que cette prudence s'applique aussi, sur certains sujets, aux sources scientifiques les plus sérieuses. Alice Eagly note ainsi que « la plupart des chercheurs universitaires en sciences sociales [...] espèrent vraisemblablement que leurs recherches soutiendront des programmes progressistes visant à accroître la diversité et l'inclusion », ce qui explique que « les interprétations simplistes des résultats scientifiques sur la diversité restent populaires » parmi eux.

6. Allégations : Les affirmations sont-elles exagérément surprenantes ?

Bien sûr, vous ne liriez jamais la presse, et encore moins des articles de management, si vous n'y trouviez que la confirmation de ce que vous pensez déjà. Et plus fondamentalement, les progrès de la science consistent, par définition, à surprendre, en contredisant les théories existantes. Comme l'écrivait Daniel Kahneman à propos de la crise de la répliquabilité, « l'implausibilité ne suffit pas à justifier l'incrédulité, et nous n'avons pas l'option de ne pas croire à des conclusions scientifiques bien étayées ²² ».

Mais plus une conclusion est improbable, et plus il faut, justement, qu'elle soit « bien étayée ». C'est le principe de Laplace : « le poids des preuves doit être proportionné à l'étrangeté des faits ». Même publiées par des sources crédibles, donc, « les affirmations extraordinaires exigent des preuves extraordinaires⁸ ». Une affirmation extraordinaire, c'est par exemple de suggérer, comme le faisaient Amy Cuddy et ses collègues, qu'il est possible de « devenir instantanément plus puissant »... Une telle assertion devrait d'abord éveiller notre esprit critique, avant de susciter notre émerveillement. Quand c'est trop beau pour être vrai, généralement, c'est faux !

⁸ Formulation attribuée à Carl Sagan. Le « principe de Laplace » est lui-même une reformulation d'une idée qu'il développe dans son *Essai philosophique sur les probabilités* (1814).

© Olivier Sibony, 2025

7. Échantillon suffisant : *La taille de l'échantillon est-elle suffisante ?*

Il n'y a pas besoin d'être un expert en statistiques pour se demander s'il est raisonnable de tirer des conclusions aussi extraordinaires que celles d'Amy Cuddy à partir d'un échantillon de 42 sujets. Il serait sans doute impossible aujourd'hui de publier dans une revue scientifique une étude aussi sous-dimensionnée par rapport à son objectif. Mais dans le monde des affaires, l'attention aux tailles d'échantillon est loin d'être aussi systématique.

Dans le domaine des études de marché, et malgré tous les avertissements des professionnels, il n'est pas rare qu'on entende un manager s'exclamer : « Je sais bien que ce n'est qu'une étude qualitative, mais quand même, 20 clients, ce n'est pas rien ! » Pour formuler des hypothèses, générer des idées, identifier des questions, non, bien sûr, ce n'est pas rien. Mais pour conclure que les hypothèses sont justes, que les idées sont bonnes, et qu'on a les réponses aux questions, il faut un échantillon plus grand.

8. Causalité : *Le lien causal allégué est-il crédible ?*

Nous avons déjà souligné que corrélation n'égale pas causalité, mais il est essentiel de le rappeler ici, car c'est sans doute la plus grande source d'erreurs dans le domaine du management. Pour une raison simple : la grande majorité des études sont *observationnelles*. Tandis que dans une expérimentation contrôlée, les chercheurs manipulent activement une variable pour observer son effet sur une autre variable, dans une étude observationnelle, ils se contentent d'enregistrer des associations entre les variables. Et ces associations peuvent souvent être diversement interprétées.

Un exemple parmi mille, lu sous la plume d'un cabinet spécialisé : « Présent dans plus de 100 pays, McDonald's a réussi à s'imposer sur la scène internationale en adaptant son menu aux goûts locaux et aux préférences culturelles de chaque région. » Ce qui en fait un « excellent exemple d'expansion internationale réussie *grâce à la diversité de ses équipes* ». Évidemment, la diversité des équipes de McDonald's est au moins autant la conséquence de l'expansion internationale que sa cause... Dans la même veine, l'association américaine des agents de voyage souligne que les salariés qui prennent

© Olivier Sibony, 2025

plus de vacances que leurs collègues ont reçu plus d'augmentations de salaire que ceux qui restent au bureau ! La causalité est clairement suggérée... mais où est la cause, et où est l'effet ?

9. Amplitude : L'amplitude d'effet est-elle importante ?

Des effets peuvent être réels, mais d'amplitude si faible qu'ils sont négligeables. Nous en avons vu des exemples à propos des différences de personnalité entre les genres, ou de l'effet du sexe du PDG sur la performance de l'entreprise. Le principal piège, ici, tient à un adjectif que vous avez sans doute l'habitude de lire, à propos de tel ou tel résultat scientifique : *significatif*. Dans le langage courant, significatif est synonyme *d'important*. Mais pas en statistiques ! Un effet peut être *très significatif* et néanmoins *insignifiant en pratique*.

Cette confusion, intentionnelle ou pas, peut être lourde de conséquences. En 2019, le ministère de l'éducation nationale a ainsi présenté comme « très significatifs » les résultats d'une expérimentation sur la réduction de la taille des classes de CP en zone prioritaire. L'amplitude de leur effet, mesurée avec une précision inhabituelle (et louable), était hélas très faible ($d = 0,1$ environ). Le fait que la conclusion soit « significative » signifiait simplement qu'on pouvait avoir une grande confiance, justement, dans le fait que le résultat était décevant. Ce qui n'a pas empêché, sur la base de cette expérimentation « réussie », de généraliser le dédoublement des classes de CP...

10. Littérature : La littérature scientifique contient-elle des répliques de la conclusion, ou, mieux encore, une méta-analyse ?

Une hirondelle ne fait pas le printemps. Une étude peut être techniquement parfaite, et pourtant ne pas « répliquer ». C'est pourquoi une grande partie des sources citées dans ce livre font référence à des méta-analyses. Car le plus sûr moyen de connaître la réponse à une question reste de compiler une série d'études qui s'efforcent, selon des modalités légèrement différentes, d'y répondre.

Paradoxalement, les études qui captent l'imagination des chercheurs, séduisent la presse et font mouche auprès du grand public sont souvent celles dont la méthodologie inhabituelle et la

© Olivier Sibony, 2025

conclusion sensationnelle facilitent le *storytelling*. Si vous avez entendu parler d'une expérimentation où des enfants doivent résister à la tentation de manger un marshmallow, ou d'une étude archéologique sur le rôle des femmes chez les chasseurs-cueilleurs il y a dix mille ans, vous vous en souvenez forcément. Pourtant, ces études sont sujettes à caution. Inversement, rien n'est moins « narratif » que de rendre compte d'une méta-analyse. Et pourtant, une conclusion issue d'une méta-analyse, surtout si elle ne nous surprend pas (par exemple, « pas de différence de performance entre les entreprises les plus diverses et les autres ») devrait normalement avoir beaucoup plus de poids que la conclusion d'une seule étude, surtout si elle est sensationnelle (par exemple, « l'entreprise X a diversifié son management et ses résultats se sont miraculeusement améliorés »).

La checklist C'EST PASCAL Quand on lit que « des études affirment... », on peut demander :	
Clarté :	<i>Les affirmations sont-elles claires et précises ?</i>
Émotion :	<i>Le compte-rendu fait-il appel à vos émotions ?</i>
Source :	<i>La source originale est-elle mentionnée et crédible ?</i>
Ton :	<i>L'article laisse-t-il une place raisonnable au doute ?</i>
Parti pris :	<i>Cette source aurait-elle publié la conclusion opposée ?</i>
Allégations :	<i>Les affirmations sont-elles exagérément surprenantes ?</i>
Suffisant :	<i>La taille de l'échantillon est-elle suffisante ?</i>
Causalité :	<i>Le lien causal allégué est-il crédible ?</i>
Amplitude :	<i>L'amplitude d'effet est-elle importante ?</i>
Littérature :	<i>Existe-t-il des répliques ou une méta-analyse ?</i>

Cette check-list n'éliminera pas tous les problèmes. Naturellement, elle ne débusquera pas les fraudes intentionnelles, qui échappent souvent à des relecteurs professionnels. Elle ne rentre pas

non plus dans les nombreuses questions de méthode qui préoccupent à juste titre les scientifiques – notamment pour comprendre les bases théoriques qui sous-tendent une étude, ou pour déterminer si ses conclusions peuvent être généralisées à d'autres contextes que celui où elles ont été obtenues.

Mais se poser ces dix questions – ou même quelques-unes d'entre elles seulement – est déjà une bonne discipline pour le lecteur occasionnel d'articles de management. Avec la pratique, nul doute qu'on prendra l'habitude de les lire avec un peu plus de prudence.

Les affirmations surprenantes, même quand vous voudriez y croire, exigent des preuves solides (et multiples).

Références

1. J. Joecks, K. Pull et K. Vetter, « Gender Diversity in the Boardroom and Firm Performance : What Exactly Constitutes a “Critical Mass ?” », *Journal of Business Ethics*, 118, 2013, p. 61-72.
2. R. M. Kanter, « Some Effects of Proportions on Group Life : Skewed Sex Ratios and Responses to Token Women », *American Journal of Sociology*, 82, 1977, p. 965-990.
3. D. Hochman, « Amy Cuddy Takes a Stand », *The New York Times*, 19 septembre 2014.
4. D. R. Carney, A. J. C. Cuddy et A. J. Yap, « Power posing : brief nonverbal displays affect neuroendocrine levels and risk tolerance. », *Psychological Science*, 21, 2010, p. 1363-1368.
5. R. Carney, « My position on “Power Poses” », 2015, p. 8-9.
6. J. Singal et M. Dahl, « Here is Amy Cuddy’s response to critiques of her power-posing research », *The Cut*, 2016. Disponible sur : <https://www.thecut.com/2016/09/read-amy-cuddys-response-to-power-posing-critiques.html>. [Consulté le 29 août 2024.]
7. C. F. Camerer, A. Dreber, E. Forsell, T. H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen et H. Wu, « Evaluating replicability of laboratory experiments in economics », *Science*, 351, 2016, p. 1433-1436.
8. R. A. Klein, M. Vianello, F. Hasselman, B. G. Adams, R. B. Adams, S. Alper, M. Aveyard, J. R. Axt, M. T. Babalola, Š. Bahník, R. Batra, M. Berkics, M. J. Bernstein, D. R. Berry, O. Bialobrzeska, E. D. Binan, K. Bocian, M. J. Brandt, R. Busching, A. C. Rédei, H. Cai, F. Cambier, K. Cantarero, C. L. Carmichael, F. Ceric, J. Chandler, J.-H. Chang, A. Chatard, E. E. Chen, W. Cheong, et al., « Many Labs 2 : Investigating Variation in Replicability Across Samples and Settings », *Advances in Methods and Practices in Psychological Science*, 1, 2018, p. 443-490.
9. J. P. A. Ioannidis, « Why most published research findings are false », *PLoS Medicine*, 2, 2005, p. 0696-0701.
10. L. L. Shu, N. Mazar, F. Gino, D. Ariely et M. H. Bazerman, « Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end », *Proceedings of the National Academy of Sciences*, 109, 2012, p. 15197-15200.
11. G. Lewis-Kraus, « They Studied Dishonesty. Was Their Work a Lie ? », *The New Yorker*, 30 septembre 2023.
12. « A study on dishonesty was based on fraudulent data », *The Economist*, 20 août 2021.

13. S. Dubner, « Why Is There So Much Fraud in Academia ? », s. d.
14. A. S. Kristal, A. V. Whillans, M. H. Bazerman, F. Gino, L. L. Shu, N. Mazar et D. Ariely, « Signing at the beginning versus at the end does not decrease dishonesty », *Proceedings of the National Academy of Sciences*, 117, 2020, p. 7103-7107.
15. N. Scheiber, « The Harvard Professor and the Bloggers », *The New York Times*, 30 septembre 2023.
16. K. Piper, « A Harvard dishonesty researcher was accused of fraud. Her defense is troubling », *Vox*, 2024. Disponible sur : <https://www.vox.com/future-perfect/24107889/francesca-gino-lawsuit-harvard-dishonesty-researcher-academic-fraud>. [Consulté le 29 août 2024.]
17. J. P. Simmons, L. D. Nelson et U. Simonsohn, « False-positive psychology : Undisclosed flexibility in data collection and analysis allows presenting anything as significant », *Psychological Science*, 22, 2011, p. 1359-1366.
18. A. Gelman et E. Loken, « The Statistical Crisis in Science », in *Best Writ. Math. 2015*, éd. M. Pitici Princeton University Press, 2016, p. 305-318.
19. J. P. Simmons, L. D. Nelson et U. Simonsohn, « False-Positive Citations », *Perspectives on Psychological Science*, 13, 2018, p. 255-259.
20. R. Lorenzo, N. Voigt, M. Tsusaka, M. Krentz et K. Abouzahr, « How Diverse Leadership Teams Boost Innovation », 2018.
21. The fingerprints of misinformation : how deceptive content differs from reliable sources in terms of cognitive effort and appeal to emotions | Humanities and Social Sciences Communications, s. d. Disponible sur : <https://www.nature.com/articles/s41599-022-01174-9>. [Consulté le 6 décembre 2023].
22. U. Schimmack, « Reconstruction of a Train Wreck : How Priming Research Went off the Rails », *Replicability-Index*, 2017.