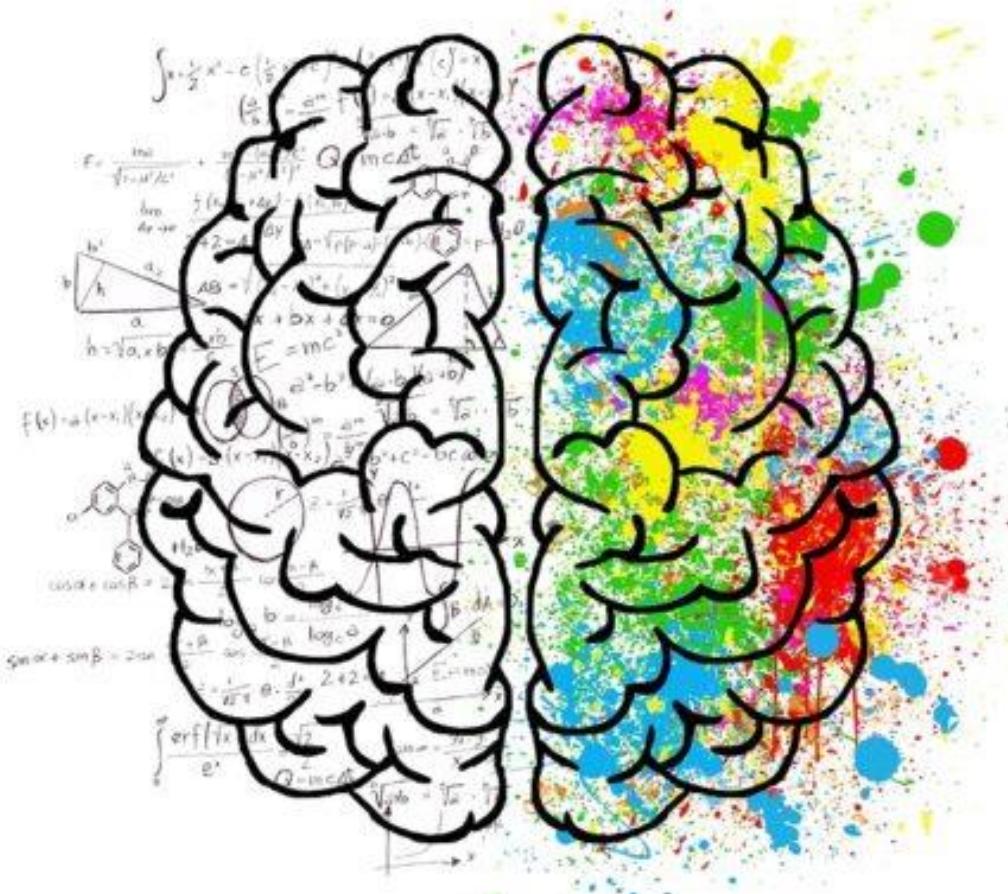


Les méthodes de recherche en psychologie sociale

Fabrice GABARROT & Thomas CARREL

Université de Bourgogne et Unidistance Suisse.

Introduction : Producteurs et consommateurs de recherches scientifiques



Beaucoup de gens croient que les hémisphères du cerveau ont des rôles très différents, et que des personnes peuvent privilégier l'un ou l'autres de ces hémisphères. On dit souvent que le "cerveau gauche" est responsable de la logique et des mathématiques, tandis que le "cerveau droit" est responsable de la créativité et des émotions. Cependant, les recherches scientifiques montrent que les deux hémisphères travaillent de concert pour l'ensemble les tâches que nous faisons quotidiennement, il n'y a pas de compartiment séparé pour les émotions ou les mathématiques par exemple, et encore moins des gens qui utiliseraient de façon privilégiée leur "cerveau gauche" ou leur "cerveau droit".

Certains étudiant·e·s en psychologie sont fascinés par le processus de recherche et ont l'intention de devenir des *producteurs* de recherche scientifique. Ils se destinent peut-être à travailler comme chercheurs·euses ou professeur·e·s. Cependant, la majorité des étudiant·e·s en psychologie ne souhaitent pas travailler dans un laboratoire. Ils se destinent plus probablement à poursuivre une carrière de thérapeute, d'enseignant, de conseiller d'orientation, d'entrepreneur, ou même d'agent de police, et ils s'attendent à ce que les cours de psychologie les aident dans ces rôles. Ces étudiants considèrent souvent les enseignements

de méthodologie (et de statistiques !!) comme un sujet ennuyeux, aride et difficile, qu'ils subissent plutôt qu'ils ne les apprécient. Ils préfèrent de loin leurs autres cours magistraux sur des sujets passionnants tels que le développement de l'enfant, les maladies mentales, les enquêtes médico-légales, la structure du cerveau et la pensée. Ce qu'ils oublient, c'est que la psychologie et les méthodes de recherche sont intimement liées - la psychologie et les moyens de développer des idées psychologiques par la recherche ne peuvent être différenciés. Il serait mal venu pour eux d'ignorer les méthodes de recherche en psychologie, au prétexte qu'ils ne se destinent pas à être chercheurs·euses. Les psychologues cliniciens, les psychothérapeutes, et les travailleurs sociaux doivent eux aussi lire la littérature scientifique pour savoir quelles prises en charge sont les plus efficaces. Même le monde des affaires fonctionne grâce à des données issues de la recherche, sur le comportement des consommateurs, par exemple. Quelle que soit la carrière qu'ils choisissent, il leur faudra lire sur la recherche en psychologie afin de pouvoir appliquer ces connaissances à votre travail, à vos loisirs, etc. Par exemple, vous avez déjà probablement entendu l'une ou plusieurs des affirmations suivantes :

- « Les êtres humains n'utilisent que 10% de leur cerveau »,
- « Les personnes créatives utilisent principalement leur cerveau droit alors que celles qui ont l'esprit analytique et logique utilisent majoritairement leur cerveau gauche »,
- « Faire écouter de la musique classique à un bébé permet d'augmenter son intelligence »,
- « Les élèves ont une mémoire visuelle, auditive ou kinesthésique, et ils apprennent mieux quand le style d'enseignement concorde avec leur style d'apprentissage »,
- « Les hommes viennent de Mars et les femmes viennent de Vénus, et c'est pour cette raison qu'ils ont parfois du mal à communiquer dans le couple »,
- « Le vaccin ROR (contre la rougeole, les oreillons et la rubéole) cause l'autisme »

Ces affirmations sont toutes extraits de Lilienfeld et collaborateur·ice·s (2010) et ont toutes été falsifiées par des recherches effectuées en psychologie. Ce type d'affirmations, dont nous sommes constamment bombardés, ont un impact direct sur notre façon de penser et de résoudre les problèmes de la société et de notre vie personnelle. Depuis maintenant quelques décennies, grâce à l'Internet, nous avons accès à plus d'explications et de réponses qu'à aucun autre moment de l'histoire. Mais, bien que la quantité d'informations ne cesse d'augmenter, il est nécessaire de juger de la qualité de ce que vous lisez ou regardez : ces informations ne sont pas toutes dignes de confiance. La pandémie de Covid-19 que nous a bien montré que la fiabilité des informations est particulièrement importante à une époque où les "fake news", les titres trompeurs des réseaux sociaux, ou encore les contenus conspirationnistes se disputent notre attention au détriment de conclusions bien informées et fondées sur des preuves scientifiques. Parfois, les conséquences sont désastreuses : la croyance selon laquelle le vaccin ROR serait responsable de troubles autistiques est liée à une diminution de vaccination des enfants, et à une augmentation de 556% des cas de rougeole dans le monde entre 2016 et 2019 (OMS, 2019). Déterminer quelles informations sont fiables et lesquelles ne le sont pas est impératif pour notre quotidien et pour la société en général.

Par conséquent, les étudiants en psychologie, qu'ils se destinent à *produire* ou *consommer* de la recherche, se doivent de trier les informations. La compréhension des méthodes de recherche permet de poser les questions appropriées afin d'évaluer correctement ces informations. Les compétences en matière de méthodes de recherche

s'appliquent non seulement aux études scientifiques, mais également à la plupart des autres types d'informations que vous êtes susceptible de rencontrer. L'objectif de ce chapitre est de vous familiariser avec la pensée scientifique et les méthodes de recherche. Dans un premier temps, ce chapitre présentera comment la pensée scientifique diffère du raisonnement quotidien et comment les méthodes scientifiques permettent de répondre efficacement aux questions que l'on se pose. Ensuite, il présentera un tour d'horizon des différentes méthodes de recherche que les psychologues utilisent pour collecter des données, de façon systématique, sur nos comportements. Enfin, nous verrons comment les psychologues peuvent utiliser plusieurs méthodes pour étudier comment notre environnement social et nos propres pensées et affects interagissent pour déterminer nos comportements.

La psychologie (sociale) est-elle une science ?

Savoir profane vs. Pensée scientifique

Un aspect excitant de la recherche en psychologie sociale est qu'elle explore des sujets pertinents pour notre vie quotidienne. Les psychologues sociaux étudient l'altruisme, l'impact de nos émotions, les conflits ou encore la discrimination. Tout comme les chercheurs·euses, la plupart d'entre nous se pose des questions sur notre monde et cherche des réponses. Nous voulons savoir quand et pourquoi les choses se produisent, si elles peuvent se reproduire, et comment les reproduire ou les modifier. Et, sans le savoir, nous recueillons même des données : il a été proposé que les gens agissent comme des « scientifiques naïfs », des « chercheurs·euses du quotidien », qui mènent des projets de recherche pour répondre à des questions sur le monde (Nisbett & Ross, 1980). Par exemple, lorsque nous obtenons de mauvais résultats à un examen important, nous essayons de comprendre ce qui a causé notre incapacité à nous souvenir ou à comprendre la matière et ce qui pourrait nous aider à mieux réussir la prochaine fois. Les résultats de ces « recherches du quotidien » peuvent nous enseigner de nombreux principes du comportement humain. Ces connaissances font partie intégrante de notre vie de tous les jours.

Cependant, ces observations du quotidien peuvent nous amener à tirer des conclusions erronées. Le problème avec la façon dont les gens recueillent et interprètent les données dans leur vie quotidienne est qu'ils ne sont pas toujours minutieux. Souvent, lorsqu'une explication d'un événement semble logique, cette explication devient vérité, même si d'autres explications sont possibles, voire plus précises. Par exemple, les témoins oculaires de crimes violents sont souvent extrêmement confiants dans leur identification des auteurs de ces crimes. Mais il a été montré que les témoins oculaires ne sont pas moins sûrs de leurs identifications lorsqu'elles sont incorrectes que lorsqu'elles sont correctes (Wells, Leippe, & Ostrom, 1979). Accepter des explications sans les vérifier peut parfois conduire les gens à penser qu'ils savent des choses qu'ils ne savent pas vraiment.

Cette première méthode de connaissance s'appelle *l'intuition*. Lorsque nous utilisons notre intuition, nous nous fions à nos émotions et/ou à nos instincts pour nous guider. Plutôt que d'examiner les faits ou d'utiliser la pensée rationnelle, l'intuition consiste à croire ce qui semble vrai. Le problème, c'est que nos intuitions peuvent être fausses, car elles sont guidées par des biais cognitifs et motivationnels plutôt que par un raisonnement logique ou des preuves scientifiques. Par exemple, nous pouvons être biaisés dans la sélection des

informations pertinentes, et interpréter les comportements d'autrui de façon complètement erronée, simplement parce que nous ne voyons que ce que nous voulons bien voir, et pas la réalité telle qu'elle est, simplement parce que nous ignorons les faits qui ne s'accordent pas avec nos croyances. Les spécialistes du comportement ont constaté qu'il existe une variété de biais cognitifs et motivationnels qui biaissent fréquemment nos perceptions et nous conduisent à tirer des conclusions erronées (Fiske & Taylor, 2013). Et comme ces biais agissent sans que nous en ayons conscience, il est très difficile pour nous de les corriger.

Il existe une autre façon par laquelle nos intuitions peuvent nous induire en erreur. Lorsque nous apprenons le résultat d'un événement donné (par exemple, lorsque nous lisons les résultats d'un projet de recherche), nous pensons souvent que nous aurions pu en prédire l'issue à l'avance. Ce *biais rétrospectif* peut même parfois amener le lecteur profane à penser qu'il n'y avait pas besoin de mener une recherche parfois coûteuse et qu'il aurait été en mesure de prédire ce résultat qui n'apparaît pour lui n'être que du « bon sens ».

Au-delà de l'intuition, nous utilisons d'autres méthodes pour acquérir des connaissances sur le monde. *L'autorité* est sans doute l'une des méthodes les plus courantes pour acquérir des connaissances. Cette méthode consiste à accepter de nouvelles idées parce qu'une figure d'autorité que nous percevons comme légitime affirme qu'elles sont vraies. Ces autorités comprennent les parents, les enseignants, des experts, les médecins, les médias, et même le gouvernement, qui nous enseignent à toutes et à tous nos premières connaissances sur le monde. Si, dans un monde idéal, nous devrions pouvoir faire confiance aux figures d'autorité, l'Histoire nous a enseigné le contraire : de nombreux cas d'atrocités contre l'humanité sont la conséquence du fait que des personnes ont suivi les autorités (par exemple, les procès des sorcières de Salem, les crimes de guerre nazis). Ces exemples montrent que le problème du recours à l'autorité pour obtenir des connaissances est que cette autorité peut se tromper, voire qu'elle peut avoir ses propres raisons de vous induire en erreur. Néanmoins, la plupart des informations que nous acquérons le sont par le biais de l'autorité, car nous n'avons pas le temps de questionner toutes les connaissances que nous possédons sur le monde, ni encore de faire nos propres recherches sur chacune de ces connaissances. Mais nous pouvons apprendre à évaluer les références des figures d'autorité, à évaluer les méthodes qu'elles ont utilisées pour parvenir à leurs conclusions, et à déterminer si elles sont en train de nous induire en erreur.

Une troisième façon d'acquérir de la connaissance s'appuie sur le *rationalisme*. Le rationalisme consiste à utiliser la logique et le raisonnement pour acquérir de nouvelles connaissances. Cette méthode permet d'énoncer des prémisses et de suivre des règles logiques pour parvenir à des conclusions solides. Par exemple, si l'on me donne comme prémisses que tous les cygnes sont blancs et comme prémisses que ceci est un cygne, je peux arriver à la conclusion rationnelle que ce cygne est blanc sans avoir réellement vu le cygne. Le problème de cette méthode est que si les prémisses sont fausses ou s'il y a une erreur de logique, la conclusion ne sera pas valide. Par exemple, la prémisses selon laquelle tous les cygnes sont blancs est incorrecte ; il existe des cygnes noirs en Australie. En outre, à moins d'avoir reçu une formation formelle sur les règles de la logique, il est facile de commettre une erreur. Néanmoins, si les prémisses sont correctes et que les règles de logique sont suivies de manière appropriée, il s'agit d'un bon moyen d'acquérir des connaissances.

Une dernière méthode de construction de savoir profane est *l'empirisme*. L'empirisme consiste à acquérir des connaissances par l'observation et l'expérience. Une fois encore, beaucoup d'entre vous ont peut-être cru que tous les cygnes sont blancs parce que vous avez vu que des cygnes blancs. Cet exemple et les nombreuses illusions visuelles qui trompent nos sens illustrent les problèmes que pose le fait de se fier uniquement à l'empirisme pour obtenir des connaissances. Nous sommes limités dans ce que nous pouvons expérimenter et observer, et nos sens peuvent nous tromper. En outre, nos expériences antérieures peuvent modifier la façon dont nous percevons les événements. Néanmoins, l'empirisme est au cœur de la méthode scientifique au travers des observations. Mais pas n'importe quelles observations, la science s'appuie sur des observations structurées, ce que l'on appelle l'empirisme systématique.

En résumé, bien que l'intuition, l'autorité, le rationalisme, et l'empirisme soient utiles pour se faire une idée, et bien que nos intuitions soient parfois correctes, elles ne sont pas infaillibles. Nos théories naïves sur la façon dont nous portons des jugements ne correspondent pas toujours bien à la façon dont nous prenons réellement des décisions. Et nous croyons souvent pouvoir prédire des événements que nous n'aurions pas prévus, ce qui donne l'impression que les résultats des recherches ne sont parfois que du bon sens. Cela ne signifie pas que ces méthodes de construction du savoir ne sont pas importantes - les scientifiques s'appuient eux aussi sur leur intuition, l'autorité de leurs pairs, des raisonnements logiques rationnels, et des observations empiriques pour les aider à résoudre des problèmes. Mais alors, qu'est-ce qui différencie le savoir profane de la pensée scientifique ?

La nature du raisonnement scientifique : induction, déduction, et abduction.

Le savoir profane et la pensée scientifique peuvent s'appuyer tous les deux sur les mêmes opérations mentales. Par exemple, ils font tous deux appels à *l'induction*, qui consiste à tirer une conclusion générale, une hypothèse, un principe théorique, ou une loi¹, à partir d'un sous-ensemble d'observations spécifiques. Par exemple, l'opinion d'une personne selon laquelle le bachotage augmente les résultats d'un examen peut être fondée sur le souvenir d'un examen réussi après une nuit d'étude. De même, les conclusions d'un·e chercheur·euse qui s'opposent au bachotage peuvent s'appuyer sur des études qui mettent en avant qu'espacer le temps des sessions d'études est plus efficace que le bachotage. Dans ces scénarios, les conclusions scientifiques et quotidiennes sont tirées d'un échantillon limité d'observations.

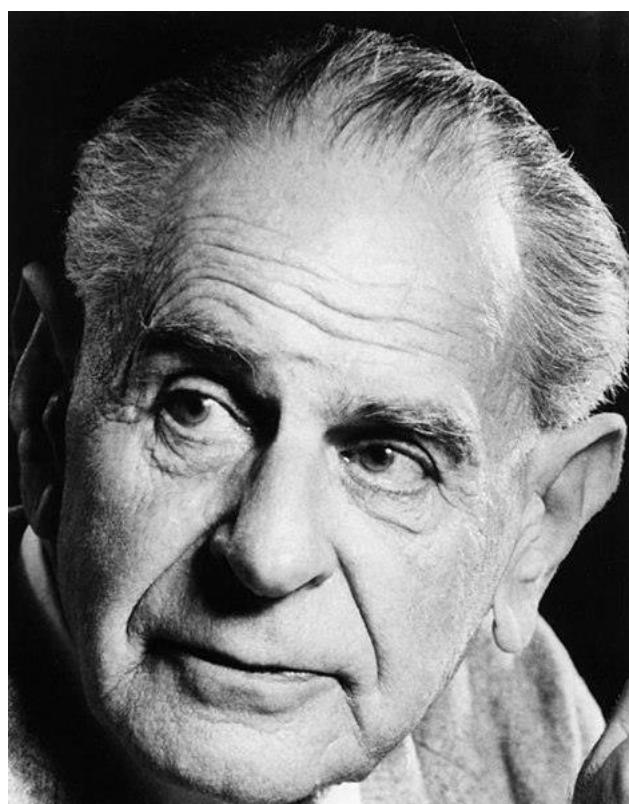
Que faire alors de ces observations contradictoires ? Le processus d'induction, à lui seul, ne semble pas suffisant pour fournir des informations dignes de confiance. Que doit faire un étudiant qui souhaite obtenir de bons résultats aux examens ? Une source d'information l'encourage à bachoter, tandis qu'une autre lui suggère de répartir son temps d'étude. Pour prendre la meilleure décision avec les informations dont nous disposons, nous devons

¹ En épistémologie, une hypothèse définit une proposition avancée, provisoirement, comme explication de faits ou phénomènes et qui doit être, ultérieurement, vérifiée par la déduction ou par l'expérience. Une loi définit une règle générale décrivant la relation entre des phénomènes naturels observés de façon répétée, fiable et précise.

apprécier les différences entre les opinions personnelles et les déclarations scientifiques, ce qui nécessite une compréhension de la science et de la nature du raisonnement scientifique.

La première différence entre le savoir profane et la pensée scientifique réside dans la nature des observations qui vont être utilisées pour acquérir une connaissance. En d'autres termes, plutôt que de différer du savoir profane par le contenu, la pensée scientifique se caractérise principalement par la méthode qu'elle emploie pour déterminer si ces connaissances reflètent de manière adéquate la réalité. Cette *méthode scientifique* est un processus de collecte et d'évaluation systématiques de preuves pour tester des idées et répondre à des questions. Si les scientifiques peuvent faire appel à l'intuition, à l'autorité, au rationalisme et à l'empirisme pour générer de nouvelles idées, ils ne s'arrêtent pas là. Les scientifiques vont plus loin en utilisant l'empirisme systématique pour collecter des observations minutieuses dans diverses conditions contrôlées afin de tester leurs idées, et ils utilisent le rationalisme pour arriver à des conclusions valides.

Cependant, pour Popper (1934/1973), l'empirisme systématique n'est pas suffisant. Une démarche exclusivement inductive, c'est-à-dire qui s'appuierait sur un nombre plus ou moins important de données d'observation pour en tirer une conclusion générale, ne peut suffire, pour lui, à établir une vérité scientifique. Par exemple, si, après avoir observé plusieurs milliers de cygnes blancs dans des circonstances variées, je tirais la conclusion que tous les cygnes sont blancs, je ferais une inférence inductive tout à fait légitime, basée sur l'observation empirique et sur un raisonnement rationnel (puisque tous les cygnes que j'ai observés sont blancs, j'en tire naturellement la conclusion que tous les cygnes sont blancs). Cependant, toutes systématiques que soient mes observations, cette conclusion serait mise à mal par l'observation d'un seul cygne noir (dont nous avons déjà établi l'existence).



Karl Popper

Une seconde différence entre le savoir profane et la pensée scientifique est donc, pour Popper, que cette dernière se doit d'être *réfutable*. Au-delà de son caractère naïf ou même parfois biaisé, le savoir profane, en comparaison à la pensée scientifique, est souvent constitué de croyances érigées comme des faits, emprunts de vérité et de certitude. La pensée scientifique, à l'opposé, est en constante évolution, et a besoin de s'actualiser. La *réfutabilité* (parfois nommée, par anglicisme, *falsifiabilité*) est donc une caractéristique essentielle de la méthode scientifique. Au début du 20e siècle, le philosophe Karl Popper a suggéré que la science peut être distinguée du savoir profane (mais également des pseudo-sciences) parce que les affirmations scientifiques peuvent être réfutées. Pour Popper, c'est le raisonnement *déductif* (c'est-à-dire le raisonnement à partir d'observations qui peuvent infirmer une théorie) plutôt que le raisonnement *inductif* (le raisonnement à partir d'un nombre quelconque d'observations qui semblent avoir confirmé une théorie) qui est la caractéristique d'une théorie scientifique. Autrement dit, une théorie scientifique doit faire des prédictions qui peuvent être testées, et la théorie doit être rejetée s'il s'avère que ces prédictions ne sont pas correctes. Ce qui distingue les énoncés scientifiques des autres, ce n'est donc pas qu'ils sont vérifiables empiriquement, mais qu'il doit être possible de démontrer par l'expérience qu'ils sont faux. Cependant, le fait que les hypothèses et la théorie soient réfutables ne signifie pas qu'elles soient nécessairement fausses, mais simplement qu'il doit être possible de montrer qu'elles le sont.

De même, une théorie n'a pas besoin d'être falsifiée – au sens de démontrée comme étant fausse – pour être abandonnée et remplacée par une autre théorie. Pour Thomas Kuhn (1962), une étude attentive de l'histoire des sciences révèle que le développement de tout domaine scientifique passe par une série de phases. La première phase est, pour Kuhn, *pré-paradigmatique*, caractérisée par un grand nombre d'hypothèses et de théories concurrentes, et par un manque de consensus sur les méthodes et les théories à utiliser pour expliquer un phénomène. Durant cette phase, les scientifiques sont confrontés à de nombreuses incohérences et à des contradictions, et ils doivent choisir parmi plusieurs options pour orienter leurs recherches. Puis, lorsqu'un des modèles offre un cadre de référence clair et cohérent pour orienter les recherches futures, il sera perçu comme étant capable d'expliquer de manière convaincante les observations et les expériences. Il peut alors faire consensus auprès de la communauté scientifique et devenir le paradigme dominant. Lorsque ce nouveau paradigme est adopté, il devient le cadre de référence pour la recherche et l'enseignement dans le domaine, et la phase *paradigmatique* commence.



Thomas S. Kuhn interviewé dans son bureau du MIT en Novembre 1989 (© Skúli Sigurdsson)

Cette deuxième phase correspond à ce que Kuhn baptise « science normale ». La science normale est donc une phase de consolidation et de développement de la connaissance. Les chercheurs·euses partagent alors un cadre intellectuel commun - appelé *paradigme* ou "matrice disciplinaire" -, travaillent sur des problèmes précis et des questions précises qui sont déjà bien compris dans le cadre du paradigme dominant. Cette phase est généralement relativement stable et productive, mais elle peut être perturbée par des données qui ne peuvent pas être expliquées par le paradigme dominant. La plupart du temps, ces données sont expliquées soit par des modifications progressives du paradigme, soit par un biais méthodologique. Le problème est que, sur de longues périodes, ces données non expliquées par le paradigme dominant s'accumulent et finissent par atteindre le point où certains scientifiques commencent à remettre en question le paradigme lui-même. À ce moment-là, la discipline entre dans une phase de crise.

Cette phase de crise est résolue par un changement paradigmatisant, dans lequel le paradigme désormais déficient est remplacé par un nouveau. Le domaine scientifique revient alors à une science normale, fondée sur ce nouveau cadre, jusqu'à la nouvelle crise. Pour Kuhn, la phase de crise est nécessaire pour permettre l'émergence de nouvelles idées et de nouvelles théories. C'est à travers la confrontation et le débat que les scientifiques peuvent évaluer les différentes options et choisir la direction à suivre pour la recherche future. Le développement de la pensée scientifique s'établirait ainsi par cycles mêlant science normale, crises, et révolutions.

Kuhn (1977) nous informe également sur les caractéristiques d'une bonne théorie. Si les observations scientifiques sont le carburant qui alimentent les découvertes scientifiques, les théories scientifiques en sont le moteur. Les théories permettent aux scientifiques d'organiser et de comprendre des observations antérieures, puis de prédire et de créer des observations futures. Comme vous le verrez dans les chapitres de ce traité, la psychologie sociale a donné naissance à de nombreuses théories scientifiques. Ces théories ont toutes des caractéristiques communes qui les différencient des idées non-scientifiques et des pseudo-sciences.

Il est important de noter que le terme « théorie » n'a pas le même sens en science et dans le langage courant. Dans le langage courant, le terme « théorie » peut être utilisé pour

désigner une idée vague et floue, qui peut être vraie ou fausse, sur la façon dont les choses fonctionnent. En science, le terme « théorie » désigne un ensemble cohérent de propositions, portant sur un sujet précis, pouvant inclure des lois et des hypothèses, et induites par l'accumulation de faits provenant d'une observation systématique ou l'expérimentation. Une théorie doit également permettre de réaliser des *prédictions*, découlant d'un raisonnement logique, sur ce qui va être observé. Enfin, comme nous l'avons vu, pour qu'une théorie soit valide, elle doit être réfutable — mais pas falsifiée. Elle doit donc résister à l'expérience et être compatible avec les nouveaux faits qui peuvent s'ajouter au cours du temps, ou rester valide dans de nouveaux domaines non encore explorés lors de sa première élaboration. Si ce n'est pas le cas, la théorie doit être corrigée, voire invalidée.

Kuhn identifie 5 critères pour une bonne théorie : la précision, la consistance (ou cohérence), la portée (ou utilité), la simplicité (ou parcimonie), et la fécondité. Premièrement, une théorie doit être *précise* : dans son domaine d'application, les conséquences déductibles d'une théorie doivent être en accord avec le résultat de l'expérience. Deuxièmement, une théorie doit être *cohérente*, à la fois intrinsèquement, dans le sens où les propositions qui la composent ne peuvent pas amener à une contradiction logique, mais également extrinsèquement, c'est-à-dire qu'elle n'entre pas en contradiction logique avec d'autres théories acceptées et applicables, en psychologie ou dans d'autres disciplines scientifiques. Troisièmement, elle doit avoir une large *portée* : en particulier, les conséquences d'une théorie doivent s'étendre bien au-delà des observations, lois ou sous-théories particulières qu'elle a été initialement conçue pour expliquer, et doit se généraliser à d'autres aspects de la nature. Quatrièmement, elle doit être *parcimonieuse* : les propositions qu'elle formule ne doivent pas être plus nombreuses que nécessaire, apportant ainsi un ordre aux phénomènes qui, en l'absence de théorie, seraient isolés individuellement et, en tant qu'ensemble, confus. Cinquièmement - un élément un peu moins standard, mais d'une importance particulière pour les décisions scientifiques réelles - une théorie doit être *féconde* pour les nouvelles découvertes de la recherche. Elle doit, en d'autres termes, permettre de faire de nouvelles prédictions, révélant de nouveaux phénomènes ou des relations précédemment ignorées parmi ceux déjà connus. Ces cinq critères s'ajoutent à celui déjà mentionné de réfutabilité.

Bien que la démarche hypothético-déductive et l'idée de réfutabilité restent centrale dans l'élaboration des théories scientifiques et en particulier dans l'analyse des données, cette démarche n'est plus utilisée de nos jours de la manière dont Popper l'avait initialement envisagée. Pour commencer, les scientifiques ne sont pas uniquement intéressés par réfuter les théories. Les scientifiques sont également intéressés à fournir des descriptions et des explications des phénomènes. Nous voulons décrire différentes causes et les diverses conditions dans lesquelles ils se produisent. Ces explorations nous obligent à tirer des conclusions à partir d'échantillons limités de données. Dans certains cas, ces données semblent correspondre à nos hypothèses, dans d'autres cas elles paraissent plus surprenantes. Les scientifiques font alors parfois appel à *l'abduction*. Le raisonnement abductif est une forme d'inférence logique formulée et avancée par le philosophe américain Peirce (1998). Il part d'une observation ou d'un ensemble d'observations et cherche ensuite l'explication la plus simple et la plus probable à partir de ces observations. Ce processus, contrairement au raisonnement déductif, aboutit à une hypothèse de travail. Les conclusions abductives sont donc qualifiées comme ayant un reste d'incertitude ou de doute. On peut comprendre le raisonnement abductif comme une *inférence à la meilleure explication*. Alors

que le raisonnement déductif consiste à faire une proposition, ou hypothèse, à partir de prémisses largement acceptées, et organisées en théories, le raisonnement abductif consiste à tirer une conclusion probable à partir de ce que l'on sait de la théorie et des observations. Pour expliquer un fait surprenant par rapport à la théorie, les scientifiques vont chercher à en inférer la cause probable : dans un premier temps, en s'appuyant sur la théorie, puis dans une nouvelle recherche testant les nouvelles hypothèses. L'abduction diffère de l'induction car l'effet de surprise quant au phénomène étonnant observé par le·a chercheur·euse signale que celui-ci avait des préconceptions sur son objet, donc des attendus, fussent-ils naïfs. Elle diffère également de la déduction dans le sens où l'hypothèse générée ne l'est pas uniquement sur la base de la théorie, mais est informée par les faits observés.



Charles Sanders Peirce

Pour résumer, la science s'appuie principalement sur une méthode qui consiste à utiliser l'observation systématique afin de développer et tester des théories permettant d'acquérir des connaissances sur le fonctionnement du monde. Les scientifiques observent le monde, de manière systématique et organisée, afin que certains biais soient moins susceptibles d'entrer dans leurs conclusions. L'observation conduit à des hypothèses que nous pouvons tester et, le cas échéant, réfuter. Cette méthode permet d'élaborer un ensemble de connaissances, mais ces connaissances sont nécessairement temporaires. De plus, elle permet à la science de s'auto-corriger : il est probable que dans un futur plus ou moins éloigné, ce que nous pensions savoir s'avère faux, et que certaines de nos théories soient falsifiées. Par conséquent, pour répondre à la question du début de cette section, on peut en conclure que la psychologie (sociale) est, dans sa démarche, scientifique. Elle appuie ses conclusions sur des données d'observation systématique collectées à l'aide d'une méthodologie scientifique. Cependant, cela n'implique pas nécessairement que la totalité des études en psychologie (sociale) soient crédibles ou valides. Cela nous amène à deux nouvelles questions : Comment collecter des données sur ce qu'on ne peut pas observer directement ? Comment savoir si une recherche est crédible ?

L'observation systématique : Variables et méthodes de collecte des données

L'observation systématique est un des piliers de la méthode scientifique. Pour la psychologie, cela soulève une question particulière : comment peut-on observer ce qui se trouve dans la tête des gens et qui ne se voit pas ? La dépression, l'intelligence et les préjugés ne semblent pas être directement observables. Comme les pensées et les sentiments sont invisibles, de nombreux·ses chercheurs·euses en psychologie ont choisi de se concentrer sur le comportement, c'est un courant que l'on appelle le *béhaviorisme*. Par exemple, vous avez peut-être remarqué que certaines personnes se comportent de manière extravertie, d'autres de manière introvertie. Comme les premiers psychologues, vous utilisez le comportement pour tirer des conclusions sur différents types de personnalités. D'autres chercheurs·euses ont essayé de développer des outils permettant de mesurer de façon plus ou moins directe ce qu'ils ne peuvent voir. Par exemple, si vous êtes fan d'Harry Potter, vous avez peut-être répondu à un questionnaire pour savoir dans quelle maison de Poudlard vous auriez été placé·e. Dans ce cas, vous avez utilisé un outil visant à "mesurer" votre personnalité afin de vous assigner à Serpentard, Gryffondor, Serdaigle, ou Poufsouffle. En utilisant des mesures comportementales et des échelles d'évaluation, il est possible pour le psychologue de mesurer ce qu'il ne peut pas observer directement, comme des pensées et des sentiments.

L'utilisation de ces mesures, qu'elles soient comportementales ou déclaratives, démontre de façon quasi-systématique la variation du comportement humain. Après avoir répondu au même questionnaire (ou s'être soumis au jugement du Choixpeau), tous les aspirants élèves de Poudlard ne sont pas assignés à la même maison. La question qui se pose alors aux chercheurs·euses en psychologie sociale, c'est d'identifier dans quelle mesure des variations dans le contexte social, ou dans les pensées, sentiments ou comportements des individus sont associées à d'autres variations dans leur comportement ou celui d'autrui. Souvent, les affirmations ou les croyances concernant le comportement humain relèvent d'une association entre les variations d'une première caractéristique et les variations d'une deuxième caractéristique, avec, souvent, une supposition de causalité. Par exemple, lorsque les médias nous « informent » que faire écouter de la musique classique à son bébé permettrait de le rendre plus intelligent, cela implique que certains enfants écouteraient ce genre de musique et d'autres non, et que cette différence est liée, de façon causale, à leur intelligence. En d'autres termes, les chercheurs·euses en psychologie sociale vont se demander dans quelle mesure une *variable* peut être associée à une autre variable.

Variables mesurées ou manipulées

Les variables sont l'unité centrale de la recherche psychologique. Une variable, comme le mot l'indique, est quelque chose qui varie, elle doit donc avoir au moins deux niveaux, ou valeurs. Ces variables peuvent être soit *mesurées* (ou *invoquées*), soit *manipulées* (ou *provoquées*). Cette distinction est importante car certaines affirmations sont testées avec des variables mesurées, tandis que d'autres doivent être testées à la fois avec des variables mesurées et manipulées. Ensuite, une variable mesurée est une variable dont les différents niveaux sont simplement observés et enregistrés. Ces variables peuvent également être qualitatives ou catégorielles — nominales ou ordinaires —, ou quantitatives — discrètes ou continues, d'intervalles ou de ratio. Certaines variables, comme la taille et le QI, sont mesurées à l'aide d'outils familiers (une règle, un test). D'autres variables, comme le sexe biologique et

la couleur des cheveux, sont également dites « mesurées ». Pour mesurer des variables abstraites, comme la dépression et le stress, les chercheurs·euses peuvent concevoir un questionnaire pour représenter les différents niveaux. Dans chaque cas, la mesure d'une variable consiste à enregistrer une observation, une déclaration ou une valeur telle qu'elle se présente naturellement.

En revanche, une variable *manipulée* est une variable que le·a chercheur·euse contrôle, généralement en affectant les participants de l'étude aux différents niveaux de cette variable. Par exemple, un·e chercheur·euse peut donner à certains participants 10 milligrammes d'un médicament, à d'autres 20 mg et à d'autres encore rien du tout. Ou encore, un·e chercheur·euse peut demander à certaines personnes de passer un test soit seules, soit accompagnées d'autres personnes, et tester ainsi les effets de la coprésence (seul vs. à plusieurs) sur la performance au test. Dans ces deux exemples, ce sont les chercheurs·euses qui manipulent, qui assignent les participants à un niveau de la variable ou un autre.

Variables dépendantes, indépendantes, ou confondues

Certaines variables peuvent jouer le rôle de *mesures*, parfois appelées *variables dépendantes* (ou VD), qui représentent ce que le·a chercheur·euse veut expliquer. Ces variables sont nommées dépendantes parce que leurs valeurs sont censées dépendre des conditions des autres variables. Ces autres variables auront le rôle de *prédicteur*, parfois appelées *variable indépendante* (ou VI ; par opposition à variable dépendante). La VI représente donc la « cause potentielle » des variations dans la VD ; la VD « dépend » de la VI.

Certaines variables ne peuvent pas être manipulées - elles peuvent seulement être mesurées. L'âge ne peut pas être manipulé parce que les chercheurs·euses ne peuvent pas assigner les participants à un âge donné ; ils ne peuvent que mesurer l'âge qu'ils ont déjà. Le sexe biologique des participants, bien qu'étant traité comme une variable catégorielle, ne peut pas non plus être manipulé. Ces variables ne peuvent pas non plus prendre le rôle de variables dépendantes. Une étude qui viserait à montrer que, par exemple, écouter la chanson « When I'm Sixty-four » des Beatles rendrait les participants plus jeunes d'environ 1,5 an en comparaison à une autre chanson, n'aurait pas beaucoup de sens (Simmons et al., 2011).

A l'opposé, certaines variables peuvent être soit manipulées soit mesurées, en fonction des objectifs de l'étude. Si vous vous intéressez à l'effet de activités extrascolaires sur les performances académiques d'un enfant, vous pourriez manipuler cette variable en assignant certains enfants à des cours de musique, d'autres à des cours de théâtre, et un troisième groupe à aucune activité de ce type, et comparer après un certain temps les résultats scolaires de ces trois groupes. Vous pourriez, également, mesurer si les enfants prennent déjà des cours de musique ou de théâtre et mettre cette mesure en relation avec leurs résultats scolaires. Attention toutefois, dans ce dernier exemple, il est possible que vous soyez confronté à un problème de *variable confondue*. Il pourrait exister une troisième variable, qui pourrait expliquer à la fois le fait que les enfants participent à ces activités extra-scolaires et en même temps pourquoi ils ont également de meilleurs résultats scolaires, comme le statut socio-économique des parents (dont les ressources financières permettent la participation à ce type d'activités, ainsi que l'accès à plus de livres ou autres contenus culturels qui participent à la

réussite scolaire), ou une plus grande discipline personnelle (qui expliquerait à la fois la motivation à participer à ces activités extra-scolaires et l'assiduité dans le travail scolaire).

Opérationnaliser les variables ?

Chaque variable d'une étude peut être décrite à deux niveaux. Tout d'abord, les variables *conceptuelles* (ou théoriques) sont des concepts abstraits, tels que « l'intelligence », ou « la performance scolaire ». Une variable conceptuelle est parfois appelée un *construit*. Les variables conceptuelles doivent être soigneusement définies au niveau théorique, et ces définitions sont appelées définitions conceptuelles. Des assertions du type « Écouter de la musique classique rend plus intelligent » nécessite, pour être étudiées empiriquement, de définir conceptuellement chaque variable : Qu'est-ce que « la musique classique » ? Qu'est-ce que l'intelligence ? ». Ensuite, pour tester l'hypothèse selon laquelle faire écouter de la musique classique rendrait plus intelligent, les chercheurs-euses créent des définitions opérationnelles des variables, également appelées variables opérationnelles ou opérationnalisations. Opérationnaliser signifie transformer une variable théorique en une variable mesurée ou manipulée : Quel(s) morceau(x) de musique classique choisir ? Quel test utiliser pour mesurer l'intelligence ?

Les relations entre les variables

L'objectif de la démarche scientifique en psychologie (mais pas uniquement) peut se résumer à décrire — de façon conceptuelle ou opérationnelle, mais également statistique — des variables, et établir des liens entre elles, dans le but de comprendre et d'expliquer le monde qui nous entoure. Il existe trois affirmations que les scientifiques effectuent sur la base de leurs observations et de leurs variables : des affirmations concernant leur *fréquence*, des affirmations concernant leur *association* et des affirmations concernant leur relation de *causalité*.

Quand pendant la pandémie de Covid-19, nous étions informés tous les jours de l'évolution de la maladie, c'était au travers d'affirmations de fréquence : le nombre de cas positifs, le nombre de décès, etc., sont des informations qui ont été présentées au grand public, et utilisées par le gouvernement pour décider de la politique sanitaire. Les affirmations concernant la fréquence d'une observation particulière décrivent un taux ou un degré particulier d'une seule variable. Les études qui mentionnent ou décrivent simplement le pourcentage d'une variable, le nombre de personnes qui pratiquent une certaine activité ou le niveau d'un certain groupe pour une variable peuvent toutes être appelées des affirmations de fréquence. Dans ces études, les variables sont toujours mesurées, et ne peuvent pas être manipulées. Ces affirmations de fréquence ne se concentrent que sur une seule variable à la fois. Elles sont *descriptives*, dans le sens où elle se contente de décrire un état de fait, et ne permettent pas d'établir une relation causale.

Les affirmations d'*association* sont caractéristiques des études dites *corrélationnelles*. Ces affirmations d'association soutiennent qu'un niveau d'une variable est susceptible d'être associé à un niveau particulier d'une autre variable. Les variables qui sont associées sont parfois dites corrélées, ce qui signifie que lorsqu'une variable change, l'autre a tendance à changer aussi. Plus simplement, on peut dire qu'elles sont liées. Dans ce cas, ce type

d'affirmation présente nécessairement deux variables (ou plus). Le lien qui unit ces deux variables peut être une association positive — pour laquelle les deux variables changent dans le même sens : quand l'une augmente, l'autre augmente également, et quand l'une diminue, l'autre diminue aussi —, négative — dans ce cas, les deux variables évoluent dans des sens opposés : quand l'une augmente l'autre diminue, et vice versa —, ou nulle — auquel cas, les variations d'une variable semblent indépendantes des variations de l'autre variable. Ces associations sont utile pour *prédir* ce qui va se passer pour l'une des variables, quand je connais la valeur de l'autre variable. Notons que le mot « prédir », tel qu'il est utilisé ici, ne signifie pas nécessairement prédir l'avenir et ne sous-entend pas de relation de causalité.

Alors qu'une affirmation d'association ne fait que constater une relation entre deux variables, parfois les chercheurs·euses désirent aller plus loin, et affirmer l'existence d'un lien de *causalité*, en affirmant que l'une des variables est responsable du changement de l'autre. Notons que ces affirmations de causalité comportent deux variables, tout comme les allégations d'association. De plus, les affirmations de causalité suggèrent, comme les affirmations d'association, que les deux variables en question covariant. Cependant, les affirmations de causalité diffèrent des affirmations d'association en donnant à chacune des variables un rôle particulier, celui de *cause* et celui d'*effet*. Une assertion du type « écouter de la musique classique rend plus intelligent » sous-entend que l'écoute de musique classique *causerait* une augmentation de l'intelligence.

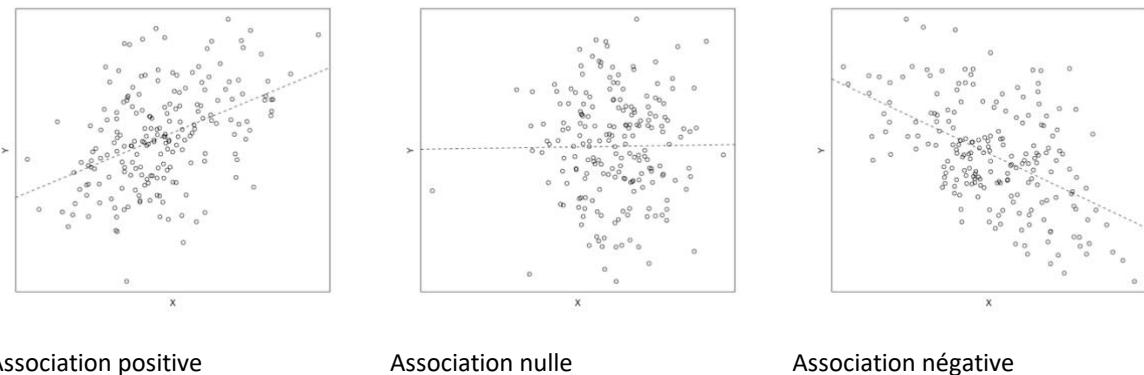


Figure 1. Association positive, nulle, ou négative entre les variables X et Y.

En matière d'explication, les affirmations de causalité sont un cran au-dessus des affirmations d'association. Parce qu'elles constituent une déclaration plus forte, nous les soumettons à des normes plus élevées. Pour permettre d'appuyer une affirmation de causalité, une étude doit faire trois critères (Kenny, 1979). Premièrement, elle doit établir que les deux variables sont *corrélées* ; la relation peut être positive, ou négative, mais elle ne peut être nulle. Deuxièmement, elle doit montrer que la variable présentée comme la cause précède temporellement celle présentée comme l'effet. Troisièmement, une affirmation de causalité doit établir qu'il n'existe aucune autre explication possible à cette relation. Un exemple de relation possible serait, comme noté plus haut, l'existence d'une variable

confondue, l'effet des activités extrascolaires pourrait être expliqué par une troisième variable : le statut socio-économique des parents, ou la discipline personnelle.

Au-delà d'une simple relation de confusion, la relation entre ces trois variables peut prendre d'autres formes. Elle peut par exemple prendre la forme d'une relation de *médiation*, la troisième variable venant expliquer — plutôt que causer — la relation entre les deux premières². On pourrait par exemple penser que la discipline personnelle explique la relation entre la participation aux activités extra-scolaires et la performance académique. La participation à des activités extra-scolaires, comme le sport ou la musique, pourrait permettre de développer chez ces enfants une discipline personnelle plus grande que chez ceux qui ne participent pas à ses activités, et en retour, cette discipline personnelle serait appliquée au travail scolaire, permettant de meilleures performances académiques.

Finalement, la relation entre ces trois variables peut prendre la forme d'une *modération* : la relation entre les deux premières variables est dépendante du niveau (ou de la modalité) de la troisième variable. Par exemple, il est possible que la discipline personnelle soit un facteur de personnalité qui vient modérer la relation entre participation à ces activités et performance scolaire. En d'autres termes, le lien entre participation à des activités extrascolaires et performance à l'école pourrait être plus fort chez les élèves ayant une forte discipline personnelle que chez ceux ayant une faible discipline. Ainsi beaucoup d'affirmations en apparence simple peuvent finalement cacher des relations entre les variables en réalité beaucoup plus complexes (voir la Figure 2. pour une illustration des différentes relations).

Comme nous l'avons déjà dit, le cœur de la démarche scientifique est de fournir des observations systématiques à partir desquelles nous allons élaborer des théories précises, cohérentes, parcimonieuses, utiles, et fructueuses. Pour obtenir ces observations, le·a chercheur·euse doit tout d'abord définir de façon conceptuelle et opérationnelle les variables qu'il veut observer, mais également s'interroger sur les relations que ces variables entretiennent. Pour que sa théorie, son étude, ou les conclusions qui en découlent aient du sens et traduisent quelque chose du comportement humain, le·a chercheur·euse doit également s'assurer de la *validité* de ses observations.

² Il existe une quatrième forme caractérisant la relation entre 3 variables, la relation de *collusion*. Dans ce type de relation, la troisième variable W est causée conjointement par X et Y.

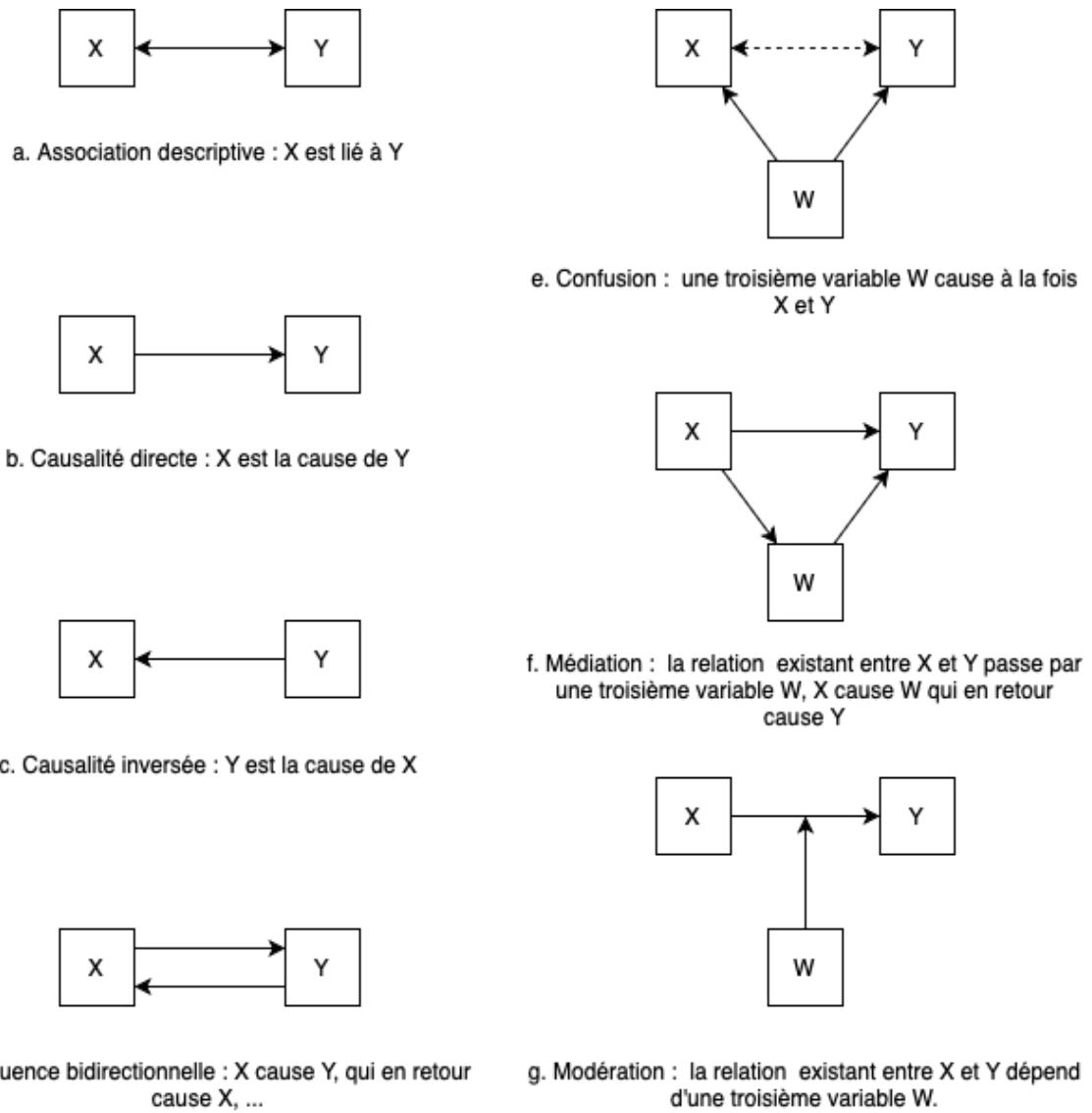


Figure 2. Représentation schématique des relations possibles entre des variables qui covarient.

Validité de construit, validité externe, validité interne et validité statistique.

La validité fait référence à la pertinence d'une conclusion ou d'une décision et, en général, une affirmation valide est raisonnable, précise et justifiable. Dans la recherche psychologique, cependant, nous ne disons pas qu'une affirmation est simplement "valide". Shadish, Cook & Campbell (2002) définissent quatre grands critères de validité auxquels le·a chercheur·euse doit prêter attention : la *validité de construit*, la *validité externe*, la *validité interne*, et la *validité statistique*, et il est parfois impossible de satisfaire les quatre critères dans la même étude. Par exemple, certaines méthodes de recherche plus écologiques, comme l'observation, mettront l'accent sur le critère de validité externe, parfois au détriment de la validité interne, alors que d'autres, comme l'expérimentation, se focaliseront sur la validité interne, le contrôle d'un ensemble d'autres variables rendant parfois la situation expérimentale moins écologique, diminuant ainsi la validité externe de la recherche. Par

conséquent, selon que les chercheurs·euses décident de tester une affirmation de fréquence, d'association ou de causalité, ils feront un choix sur le ou les critères de validité les plus pertinents.

La *validité de construit* fait référence à la façon dont une variable conceptuelle est opérationnalisée. Lorsque l'on évalue la validité de construction d'une affirmation de fréquence, la question est de savoir si les chercheurs·euses ont bien mesuré leurs variables. Pour conclure à une relation d'association ou de causalité, les chercheurs·euses vont devoir mesurer au moins deux variables, nous devons donc évaluer la validité de construit de chaque variable. Si vous estimatez qu'une des variables a été mal mesurée, vous ne pourrez pas faire confiance aux conclusions liées à cette variable. En revanche, si vous estimatez que la validité de l'étude était excellente, vous pouvez avoir davantage confiance dans l'allégation d'association rapportée. Pour garantir la validité de construit, les chercheurs·euses doivent établir deux critères : la *fiabilité* et la *sensibilité*. Chaque variable doit être mesurée de manière fiable, ce qui signifie que la mesure donne des résultats similaires lors de tests répétés. Une mesure qui donnerait des résultats différents dans des conditions de mesures identiques, et sur un même échantillon, ne serait donc pas fiable. La mesure doit également être sensible. Les différents niveaux d'une variable correspondent précisément à de véritables différences significatives, et inversement, que ces différences puissent être captées par la mesure.

La *validité externe* correspond à la possibilité de généraliser le résultat obtenu à d'autres populations, d'autres contextes, moments ou lieux. Il existe deux principaux types de validité externe : la validité de la population et la validité écologique. La *validité de la population* consiste à déterminer si vous pouvez raisonnablement généraliser les résultats de votre échantillon à un groupe de personnes plus important (la population). La validité de la population dépend du choix de la population et de la mesure dans laquelle l'échantillon étudié est représentatif de cette population. Des méthodes d'échantillonnage non probabilistes sont souvent utilisées pour des raisons de commodité (comme utiliser des étudiants de psychologie comme échantillon pour nos expériences). Avec ce type d'échantillonnage, la généralisation des résultats est limitée aux populations qui partagent des caractéristiques similaires à celles de l'échantillon. La *validité écologique* consiste à déterminer si l'on peut raisonnablement généraliser les résultats d'une étude à d'autres situations et contextes dans le "monde réel".

La *validité interne* est la mesure dans laquelle vous pouvez être sûr qu'une relation de cause à effet établie dans une étude ne peut être expliquée par d'autres facteurs. La validité interne rend les conclusions d'une relation causale crédibles et dignes de confiance. Sans une validité interne élevée, une expérience ne peut pas démontrer un lien de causalité entre deux variables. Comme nous l'avons évoqué plus haut, il existe trois conditions nécessaires à la validité interne : la covariance ou la corrélation entre la cause X et l'effet Y, la cause X précède l'effet Y, et l'absence d'une troisième variable W susceptible d'affecter la relation entre la cause X et l'effet Y. Ces trois conditions doivent être réunies pour établir la causalité entre une variable indépendante X et une variable dépendante Y.

Les chercheurs·euses utilisent les statistiques pour analyser leurs données. La *validité statistique*, est la mesure dans laquelle les conclusions statistiques d'une étude sont exactes et raisonnables. Dans quelle mesure les chiffres soutiennent-ils l'affirmation ? La validité

statistique implique deux types d'erreurs possibles. Premièrement, une étude peut conclure à tort, sur la base des résultats d'un échantillon de personnes, qu'il existe une association entre deux variables, alors qu'il n'y a en réalité aucune association dans l'ensemble de la population. Les chercheurs·euses prudents essaient de minimiser les risques de commettre ce type d'erreur, appelée "faux positif" ou **erreur de type I**. Pour ce faire, ils vont déterminer un seuil de significativité en dessous duquel ils pourront conclure à l'existence d'une association. Souvent en psychologie, ce seuil est fixé, par convention, à 5%. Cela signifie que, s'il n'existe pas d'association entre nos variables dans la réalité, seuls 5% des études montreraient un résultat significatif.

Deuxièmement, une étude peut conclure à tort, à partir d'un échantillon, qu'il n'y a pas d'association entre deux variables, alors qu'il existe réellement une association dans l'ensemble de la population. Les chercheurs·euses prudents essaient de réduire au minimum les risques de commettre ce type d'erreur ; c'est ce qu'on appelle une **erreur de type II**. Évidemment, ils veulent réduire les risques de manquer des associations qui existent réellement.

Comme vous pouvez l'imaginer, l'évaluation de la validité statistique peut être compliquée. Une formation complète sur la manière d'interroger la validité statistique nécessite... un cours de statistiques !

Menaces concernant la validité des conclusions d'une étude

Les résultats d'une étude scientifique ne pourront être crédibles que dans la mesure où ils sont valides. Les menaces à la validité sont des raisons spécifiques pour lesquelles nous pouvons nous tromper partiellement ou complètement lorsque nous faisons une inférence sur l'association, sur la causalité, sur les constructions, ou sur la question de savoir si la relation causale se maintient malgré les variations des personnes, des contextes, des traitements et des résultats. Identifier ces menaces est très utile. Cela permet aux chercheurs·euses d'anticiper les critiques probables de leurs études, afin de les écarter. Cela permet également aux lecteur·ice·s et autres consommateur·ice·s de recherche de déterminer si ces recherches sont dignes de confiance, et si il ou elle peut les employer dans sa pratique, pour développer des politiques publiques, ou encore pour ajuster ses actions au quotidien. Quelques exemples de ces menaces sont présentés dans le Tableau 2.

Tableau 1. *Les quatre grands types de validité d'après Shadish, Cook, & Campbell (2002).*

Type de validité	Définition	Exemple de questions que l'on peut se poser
Validité de construit	<p>La façon dont les variables d'une étude sont mesurées ou manipulées.</p> <p>La mesure dans laquelle les variables opérationnelles d'une étude sont une bonne approximation des variables conceptuelles.</p>	<p>Comment les chercheurs·euses ont-ils défini les variables qu'ils cherchent à mesurer ou manipuler ?</p> <p>Comment les chercheurs·euses ont-ils opérationnalisé les variables qu'ils cherchent à mesurer ou manipuler ?</p> <p>Les variables mesurées ont-elles une ou plusieurs dimensions sous-jacentes ?</p> <p>Est-ce que les outils utilisés pour mesurer les variables permettent de les mesurer telles qu'elles ont été définies ?</p>
Validité externe	<p>La mesure dans laquelle les résultats d'une étude sont généralisables à une population plus large, ainsi qu'à d'autres moments ou situations.</p>	<p>L'échantillon utilisé est-il représentatif de la population générale ?</p> <p>L'échantillon utilisé a-t-il été sélectionné de façon aléatoire ?</p> <p>Les mesures ou les manipulations utilisées sont-elles représentatives d'autres observations ?</p> <p>A quels autres problèmes ou situations les résultats observés peuvent-ils être généralisés ?</p> <p>Comment les résultats observés se généralisent-ils à d'autres échantillons ? A d'autres populations ? Dans d'autres cultures ? A d'autres moments ?</p>
Validité interne	<p>Dans une relation entre une variable (X) et une autre (Y), la mesure dans laquelle X, plutôt qu'une autre variable (W), est responsable des changements dans Y.</p>	<p>L'étude était-elle une expérience ?</p> <p>L'étude permet-elle d'établir la précédence temporelle de la cause sur l'effet ?</p> <p>L'étude contrôle-t-elle les explications alternatives en assignant aléatoirement les participants à des groupes ?</p> <p>Existe-t-il des variables confondues avec la mesure ou la manipulation ?</p> <p>Quelles mesures l'étude prend-elle pour éviter les biais ?</p>
Validité statistique	<p>La mesure dans laquelle les données appuient les conclusions. Parmi de nombreuses autres questions, il est important de s'interroger sur la force d'une association et sa significativité statistique (la probabilité que les résultats aient pu être obtenus par hasard s'il n'y a pas vraiment de relation).</p>	<p>Quelle est la taille de l'effet observé ?</p> <p>Le résultat est-il statistiquement significatif ?</p> <p>Si oui, quelle est la probabilité que le résultat soit un faux positif ?</p> <p>Si non, quelle est la probabilité que le résultat soit un faux négatif ?</p>

Hiérarchisation des types de validités

Lequel des quatre types de validités est la plus importante ? Cela dépend. Lorsque les chercheurs·euses planifient des études pour tester des hypothèses et étayer des affirmations, il leur est généralement impossible de mener une étude qui satisfasse les quatre types de validité à la fois. En fonction de leurs objectifs, les chercheurs·euses n'essaient parfois même pas de satisfaire certaines validités.

La validité externe, par exemple, n'est pas toujours possible à atteindre, et il arrive qu'elle ne soit pas la priorité. Pour pouvoir généraliser les résultats d'un échantillon à une large population, il faut un échantillon représentatif de cette population. Si des chercheurs·euses menaient une enquête téléphonique et voulaient en généraliser les résultats à l'ensemble de la population - pour maximiser la validité externe - ils devraient sélectionner au hasard des participants en fonction de leur zone géographique, ou de leur profession, de leurs niveaux de revenus, etc.

Lorsque les chercheurs·euses ont recours à des sondages formels et aléatoires, ils doivent souvent payer l'institut de sondage pour qu'il pose chaque question. Par conséquent, un·e chercheur·euse qui souhaite évaluer, par exemple, les niveaux de dépression dans une large population peut être contraint·e, pour des raisons économiques, d'utiliser un questionnaire ou un sondage court. Une mesure de 2 questions peut ne pas être aussi bonne qu'une mesure de 15 questions, mais cette dernière coûtera plus cher. Dans cet exemple, le·a chercheur·euse peut sacrifier une partie de la validité de construction afin de garantir la validité externe.

Par conséquent, au cours de la planification et de la réalisation d'une étude, les chercheurs·euses pèsent le pour et le contre des choix méthodologiques et décident des types de validité les plus importants, et vont choisir la méthode de recueil de données la plus appropriée par rapport à leur objectif. Lorsque vous lisez un article sur une étude, vous ne devez pas nécessairement en conclure qu'elle est invalide simplement parce qu'elle ne répond pas à l'un des critères de validité.

Tableau 2. Exemples de menace à la validité (d'après Shadish, Cook, & Campbell, 2002)

Menace	Définition
Biais d'échantillonnage	<p>Le ou les échantillons ne sont pas représentatifs de la population générale.</p> <p><i>Exemple :</i> L'échantillon ne comprend que des étudiants de licence de psychologie, blancs, éduqués, issus de pays industrialisés, riches, et démocratiques.</p> <p><i>(NB. Même si l'échantillonnage peut représenter une menace pour la validité externe et la généralisation des résultats, avoir des échantillons homogènes peut être avantageux pour la validité interne et l'établissement de relation causale entre des variables)</i></p>
Effet Hawthorne ou Effet de désirabilité sociale	<p>Le fait d'être interrogés modifie la réponse des participants.</p> <p><i>Exemples :</i> Les participants augmentent leur performance par rapport à d'habitude parce qu'ils se savent observés. Les participants répondent ce qu'ils pensent qui est attendu d'eux par les chercheurs-euses</p>
Effet d'apprentissage ou d'interférence	<p>Le fait d'avoir déjà passé l'étude, ou d'avoir répondu à certaines questions avant d'autres, modifie les réponses des participants.</p> <p><i>Exemples :</i> Les réponses au questionnaire au temps 1 influence les réponses au temps 2. Le fait d'avoir dû indiquer leur appartenance ethnique au début de l'étude diminue la performance des sujets Afro-Américains à un test d'intelligence.</p>
Effet d'attrition	<p>Les participants abandonnent l'étude avant d'être arrivés à son terme ou ne répondent pas à certaines questions d'une façon qui n'est pas aléatoire.</p> <p><i>Exemples :</i> Une partie des participants ayant un fort score d'orientation à la dominance sociale n'ont pas répondu à l'échelle de préjugés. 9% des participants de la condition placebo sont décédés avant la fin de l'étude (contre 0% de la condition traitement)</p>
Effet expérimentateur	Les caractéristiques ou le comportement de l'expérimentateur influence les réponses des participants de façon non intentionnelle.
Effet d'instrumentation	Un changement dans les mesures utilisées influence les résultats.
Effet de maturation	<p>Les résultats de l'étude varient naturellement avec le temps.</p> <p><i>Exemple :</i> Les participants ayant pris le traitement X ont vu leurs symptômes diminuer au bout de quelques jours, mais le cours naturel de la maladie fait que les symptômes disparaissent au bout de 3 ou 4 jours.</p>
Régression à la moyenne	Les résultats initialement les plus extrêmes ont tendances à se modérer avec le temps.
Effet situationnel	Des facteurs tels que le cadre, le moment de la journée, le lieu, les caractéristiques des chercheurs-euses, etc. limitent la généralisation des résultats.
Histoire	Un évènement indépendant influence la réponse des participants.

Méthodes de recueil des données

Recherches qualitatives et quantitatives

Lors de la collecte et de l'analyse des données, les méthodes qualitatives traitent des mots et des significations, alors que les méthodes quantitatives traitent des chiffres et des statistiques. Les deux sont importantes pour acquérir différents types de connaissances. Les méthodes qualitatives et quantitatives diffèrent les unes des autres par deux aspects : la nature du raisonnement qu'elles utilisent (inductif vs. déductif), et le type de validité qu'elles privilégient (validité externe vs. validité interne). Même si les méthodes quantitatives, parce qu'elles utilisent des données chiffrées et des outils statistiques, peuvent paraître plus scientifiques, les méthodes qualitatives sont essentielles à l'étude du comportement humain. Pour la plupart des sujets de recherche, les chercheurs·euses en psychologie peuvent choisir une approche qualitative, quantitative ou mixte. Le type de méthode qu'ils choisissent dépendra, entre autres, de l'approche inductive ou déductive qu'ils adoptent, de leurs questions de recherche, de la nature expérimentale, corrélationnelle ou descriptive de leur recherche et de considérations pratiques (temps, coût, etc.).

Les méthodes qualitatives s'expriment par des mots. Elles sont utilisées pour comprendre des concepts, des pensées ou des expériences. Ce type de recherche vous permet de recueillir des informations approfondies sur des sujets qui ne sont pas bien compris. Les méthodes qualitatives courantes comprennent les entretiens (directive, semi-directive ou non directive), les observations décrites en mots et les analyses documentaires qui explorent les concepts et les théories. Les recherches qualitatives relèvent généralement d'une démarche inductive, et mettent l'accent sur la formulation de théories ou d'hypothèses *a posteriori*. Les méthodes de recherche qualitatives permettent d'approfondir les questions posées aux personnes interrogées sur la base de leurs réponses, l'enquêteur·rice/chercheur·euse essayant également de comprendre leurs motivations et leurs sentiments. Les méthodes qualitatives sont aussi généralement descriptives, et s'appuient sur des méthodes comme l'entretien, les focus groups, mais également les études de cas ou les études d'archives.

Les méthodes quantitatives quant à elles s'expriment par des chiffres et des graphiques. Elles relèvent le plus souvent d'un raisonnement hypothético-déductif, et sont utilisées pour tester des théories et des hypothèses. Les méthodes quantitatives mettent l'accent sur la corroboration ou la réfutation d'hypothèses ou de théories, et utilisent des outils statistiques pour analyser les données. Ce type de recherche peut être utilisé pour établir des faits généralisables sur un sujet. Le type de résultats obtenus avec ces méthodes peut soutenir chacune des trois affirmations présentées ci-dessus : une affirmation de fréquence, une affirmation d'association ou une affirmation de causalité. Comme noté plus haut, elles peuvent varier également sur le type de validité qu'elles mettent en avant, notamment sur la priorité donnée à la validité externe ou à la validité interne. Elles s'appuient sur des méthodes plus structurées de collecte de données, comme les expérimentations ou les questionnaires, mais également sur de l'observation systématique.

Les méthodes de recherche qualitatives

Les données qualitatives sont par essence des données non numériques qui nous aident à explorer la manière dont la situation est interprétée par les participants, et nous fournissent un aperçu détaillé des processus mentaux des participants tels qu'ils peuvent les verbaliser. Pour parvenir à de telles conclusions, les données collectées doivent être holistiques, riches et nuancées et les résultats doivent émerger d'une analyse minutieuse.

Il existe de nombreuses approches de la recherche qualitative. Toutes ont l'avantage d'être flexibles et à se concentrer sur la conservation d'une signification riche lors de l'interprétation des données. Les approches les plus courantes sont la recherche ethnographique, la recherche-action, et l'enquête narrative. Elles partagent certaines similitudes, mais mettent l'accent sur des objectifs et des perspectives différents.

La recherche ethnographique s'appuie sur une enquête de terrain composée de méthodes de collecte d'informations et de retranscription. Souvent les chercheurs·euses recourent à une période d'observation participative lors de laquelle ils seront partie prenante de leur objet d'étude. L'ethnographie en tant que première phase de collecte et de description a pour but d'enregistrer des données pertinentes pour l'analyse d'un peuple, d'une communauté ou d'une société. Les chercheurs·euses s'immangent dans des organisations ou des groupes afin d'en comprendre et décrire la culture.

Dans le cas de la recherche-action (Lewin, 1946), les chercheurs·euses et les participant·e·s établissent en collaboration un lien entre la théorie et la pratique afin de favoriser le changement social. La recherche-action a pour but un changement par le biais d'un processus simultané d'action et de recherche, qui sont liés par une réflexion critique. Dans son article de 1946, Lewin décrit la recherche-action comme « une recherche comparative sur les conditions et les effets de diverses formes d'action sociale et de recherche menant à l'action sociale » qui utilise « une spirale d'étapes, chacune d'entre elles étant composée d'un cercle de planification, d'action et de recherche de faits sur le résultat de l'action » (p. XX). L'objectif de la recherche-action est donc le développement d'une théorie qui fait participer les acteurs.

L'enquête narrative utilise des textes, tels que des histoires, des biographies, des articles de journaux, des notes de terrain, des lettres, des entretiens, des photos, etc., comme unités d'analyse pour rechercher et comprendre la façon dont les gens créent du sens dans leur vie en tant que récits. Ces recherches examinent la façon dont les histoires sont racontées pour comprendre comment les participant·e·s perçoivent et donnent un sens à leurs expériences de vie.

Quelles que soient les approches ou les méthodes choisies par un·e chercheur·euse pour collecter des données qualitatives, le processus générera une grande quantité de données. En plus de la variété des méthodes disponibles, il existe également différentes méthodes de collecte et d'enregistrement des données. Par exemple, si les données qualitatives sont recueillies par le biais d'un groupe de discussion ou d'entretiens, il y aura des notes informelles (avec des commentaires, des contextes environnementaux, l'analyse de l'environnement, les indices non verbaux), mais aussi des enregistrements audios ou vidéos. S'il y a des enregistrements, ils devront être transcrits avant que le processus d'analyse des

données puisse commencer. Les notes et les enregistrements vidéo et/ou audio devront être conservés.

Pour collecter des données qualitatives de façon systématique, les chercheurs·euses doivent se doter d'outils qui permettent de collecter toutes les données pertinentes, mais également éviter les biais. Iels vont pouvoir utiliser, le cas échéant, des grilles d'entretiens qui les aideront à collecter les données pertinentes pour leur recherche. Par conséquent, lorsqu'iels mettent en place leur méthodologie, il est important qu'iels réfléchissent en détails à leur approche et qu'iels justifient les choix effectués lors de la création de leurs outils, de la collecte et de l'analyse des données.

Les méthodes qualitatives visent à préserver la voix et le point de vue des participant·e·s et peuvent être adaptée en fonction des nouvelles questions de recherche. Ces méthodes ont plusieurs avantages. Tout d'abord, elles sont flexibles. Le processus de collecte et d'analyse des données peut être adapté lorsque de nouvelles idées ou de nouveaux modèles apparaissent. Cette flexibilité peut représenter un avantage pour la validité de construit du phénomène que l'on cherche à étudier. Les descriptions détaillées des expériences, des sentiments et des perceptions des personnes peuvent être utilisées pour préciser le phénomène étudié. Ensuite, elles prennent place dans le « milieu naturel » (lieu de travail, classe d'école, etc.) des participant·e·s, augmentant ainsi la validité écologique. Enfin, grâce aux réponses ouvertes, les chercheurs·euses peuvent découvrir de nouveaux problèmes ou opportunités auxquels ils n'auraient pas pensé autrement.

Cependant, ces méthodes ont également les inconvénients de leurs avantages. Par exemple, alors que leur ancrage dans le monde réel est bénéfique à leur validité écologique, les recherches qualitatives peuvent souffrir d'un manque de fiabilité en raison des facteurs non contrôlés susceptibles d'affecter les données. De plus, la subjectivité des chercheurs·euses peut jouer un rôle important dans ce type de recherche, et en affecter les résultats. Comme ce sont les chercheurs·euses qui décident de ce qui est pertinent et de ce qui ne l'est pas dans l'analyse des données, les interprétations des mêmes données peuvent varier considérablement d'un·e scientifique à l'autre. De plus, la faible taille des échantillons fait qu'il est parfois difficile de tirer des conclusions généralisables car les données peuvent ne pas être représentatives de la population au sens large. Finalement, d'un point de vue pragmatique, ces méthodes requièrent un travail long et intense de la part des chercheurs·euses qui peut venir en impacter la faisabilité. Bien que des logiciels puissent être utilisés pour gérer et enregistrer de grandes quantités de texte, l'analyse des données doit souvent être vérifiée ou effectuée manuellement, ce qui est couteux en temps et en ressources humaines.

Les méthodes de recherche quantitatives

Les méthodes de recherche quantitatives consistent en la collecte et d'analyse de données numériques. Comme noté plus haut, elles peuvent être utilisées pour trouver et décrire un état de fait, observer des associations et faire des prédictions, tester des relations causales et généraliser les résultats à des populations plus larges.

Pour recueillir des données quantitatives, les chercheurs·euses devront utiliser des définitions opérationnelles qui traduisent des concepts abstraits (par exemple, l'humeur) en mesures observables et quantifiables (par exemple, l'auto-évaluation des sentiments). Une fois les données collectées, les données devront être traitées avant de pouvoir les analyser. Par exemple, les données d'observations ou d'enquêtes devront probablement transformer des mots en chiffres. Ensuite, les chercheurs·euses utiliseront des outils statistiques pour répondre à leurs questions de recherche.

Les méthodes quantitatives se caractérisent par un recueil systématique de données chiffrées. Elles ont plusieurs avantages qu'il est utile de mettre en avant. Tout d'abord, elles permettent de tester des hypothèses. En opérationnalisant et en quantifiant certaines variables pourtant inobservables — comme l'humeur, la personnalité, ou l'idéologie —, les méthodes quantitatives permettent de décrire, d'étudier les liens, et de faire des hypothèses concernant les relations entre ces variables. Par ailleurs, la nature des méthodes quantitatives, qui simplifient sous forme numérique ces variables, fait qu'il est possible et plus facile de collecter un grand nombre de données, rendant de fait les conclusions de ces études plus généralisables et réplicables. L'opérationnalisation et la quantification des variables psychologique permet également de reproduire et comparés statistiquement l'étude et ses conclusions dans d'autres contextes culturels, à d'autres époques ou avec différents groupes de participants.

Cependant, tout comme les méthodes qualitatives, les méthodes quantitatives ont les inconvénients de leurs avantages. Tout d'abord, en se concentrant sur des choses mesurables, elles ne permettent d'avoir qu'une approche limitée du comportement humain et amène souvent à réduire les phénomènes complexes à des variables simples, ce qui peut entraîner une perte d'informations, et exclure des aspects importants de l'expérience humaine qui ne peuvent pas être facilement quantifiés, en limitant ainsi la validité externe. Par ailleurs, la qualité des données recueillies est contingente de la qualité des outils de mesure, ceux-ci peuvent présenter des biais venant fausser les résultats, limitant de ce fait la validité interne.

La subjectivité des chercheurs·euses peut influencer les résultats des recherches quantitatives ou expérimentales. Tout d'abord, dans leur choix des sujets d'étude et des méthodes, les chercheurs·euses peuvent choisir des sujets ou des méthodes spécifiques, qui correspondent à leurs propres intérêts ou à leurs propres hypothèses, ce qui peut biaiser les résultats. Les chercheurs·euses peuvent également inconsciemment donner plus de poids aux résultats qui confortent leur hypothèse, ignorer ceux qui lui sont contradictoires, ou encore interpréter les résultats d'une façon qui est conforme à leur hypothèses ou leur préconçus. Finalement, les chercheurs·euses peuvent être plus enclins à publier certains résultats qui soutiennent leurs hypothèses ou leur vision du monde plutôt que d'autres qui les contredisent. Il est donc important pour les chercheurs·euses d'être conscient·e·s de ces biais et de prendre des mesures pour les minimiser.

L'observation

Les chercheurs·euses peuvent étudier les gens en « situation naturelle » en les observant simplement en action, de la manière la plus discrète possible, les chercheurs·euses n'interagissant en aucune façon avec les participant·e·s. Lorsqu'un·e chercheur·euse observe

des personnes ou des animaux et enregistre systématiquement la manière dont ils se comportent ou ce qu'ils font, on parle de recherche par observation. Certains scientifiques pensent que l'observation du comportement peut être plus utile que l'utilisation de questionnaires, car les gens ne peuvent pas toujours faire état de leur comportement ou d'événements passés avec précision. La recherche par observation peut servir de base aux affirmations de fréquence, d'association ou de causalité. Par exemple, les chercheurs·euses peuvent observer les conducteur·ice·s, en comptant combien d'entre eux s'arrêteront à une intersection ou à un passage pour piétons, en fonction de la marque de la voiture qu'ils conduisent (Piff et al., 2012, Études 1 et 2). L'observation peut avoir lieu « sur le terrain », en situation naturelle, ou en situation contrôlée. Les chercheurs·euses peuvent agir comme un·e observateur·rice extérieur, caché ou visible, ou encore prendre part à la situation qu'iel désire observer.

Dans le cas de l'observation participante, les chercheurs·euses participent activement à l'étude elle-même. En plus d'observer les comportements, les chercheurs·euses peuvent mener des entretiens, prendre des notes, examiner des documents et prendre des photos. Par exemple, Festinger, Riecken et Shacter (1956) se sont beaucoup intéressés à la psychologie d'une secte. Cependant, cette secte était très secrète et n'accordait pas d'interviews aux membres extérieurs. Ainsi, afin d'étudier ces personnes, Festinger et ses collègues se sont fait passer pour des membres de la secte, ce qui leur a permis d'avoir accès au comportement et à la psychologie de la secte. Malgré cet exemple, il convient de noter que, dans la plupart des cas, les personnes observées dans le cadre d'une étude par observation participante savent que les chercheurs·euses sont là pour les étudier.

Questionnaires

La recherche par questionnaires est utile pour les psychologues sociaux afin d'étudier les différences individuelles et collectives dans les attitudes, comportements et émotions des gens. De plus, ils permettent d'avoir un grand nombre de personnes sondées en peu de temps. L'obtention d'échantillons importants, interculturels et représentatifs, a été facilitée depuis l'avènement d'Internet et la prolifération des plateformes d'enquête en ligne et des plateformes de recrutement de participant·e·s. Par exemple, la World Values Survey II, qui a inclus de grands échantillons représentatifs de 19 pays, a permis de montrer que l'influence du revenu sur le bien-être subjectif est plus importante dans les pays les moins riches (Diener & Oishi, 2000).

Les inconvénients de l'utilisation de questionnaires pour mener des recherches comprennent la validité basée sur des réponses honnêtes, les choix de réponses pourraient ne pas refléter les véritables opinions — le biais de désirabilité sociale — et une réponse particulière pourrait être comprise différemment par les sujets de l'étude, fournissant ainsi des résultats moins que précis.

Certains chercheurs·euses ont également des doutes sur la représentativité ou l'honnêteté des échantillons en ligne. Cependant, des études ont montré que les échantillons en ligne sont à bien des égards plus diversifiés et plus représentatifs que les échantillons recrutés à partir de pools de sujets humains (par exemple, Gosling et al., 2004). Les

échantillons en ligne se comparent également favorablement aux échantillons traditionnels en ce qui concerne l'attention portée au questionnaire, la fiabilité des données et la proportion de non-répondants (Paolacci et al., 2010 ; Paolacci & Chandler, 2014).

L'utilisation de plateformes de recrutement de participants soulève de nouveaux problèmes. Une étude récente (Chmielewski & Kucker, 2020) sur le recrutement via la plateforme MTurk met en évidence une augmentation du pourcentage données de faible qualité, d'un impact négatif important sur les résultats des études et de l'incapacité de ces études à reproduire des résultats bien établis. Un article encore plus récent de Webb & Tangney (2022) illustre l'expérience des chercheuses avec les données MTurk. D'un échantillon de départ de 529 participant·e·s, les chercheuses n'ont pu conserver que 14 participant·e·s (2.6%) qu'elles ont considéré valides, après avoir vérifié les conditions d'inclusion des participant·e·s (leur pays d'origine, leur âge et leur langue maternelle), leur déclaration de consentement, le fait qu'iels aient réalisé la totalité du questionnaire, leurs réponses aux contrôles d'attention³, les temps mis pour compléter le questionnaire et la qualité des réponses individuelles. Toutefois, ces problèmes peuvent être atténués, dans une certaine mesure, en prenant certaines précautions dans l'éthique et l'administration des études, en contrôlant et filtrant régulièrement les données, et en incluant des indicateurs de validité dans le questionnaire (voir Aguinis et al., 2021, ou Kennedy et al. 2020, pour une liste de recommandations).

Les questions d'un questionnaire peuvent suivre quatre formats de base. Tout d'abord, les chercheurs·euses peuvent poser des questions ouvertes qui permettent aux répondant·e·s de répondre comme iels le souhaitent. Iels peuvent demander aux gens de nommer la personnalité publique qu'ils admirent le plus, ou demander à un échantillon de personnes de décrire un évènement de leur quotidien. Les réponses aux questions ouvertes fournissent aux chercheurs·euses des données qualitatives spontanées et riches. L'inconvénient, comme pour toutes les recherches qualitatives, est que les réponses doivent être codées et catégorisées, un processus souvent difficile et long.

Une deuxième façon spécifique de poser des questions d'enquête consiste à utiliser des questions à choix forcé, dans lesquelles les personnes donnent leur opinion en choisissant la meilleure de deux options ou plus. Les questions à choix forcé sont souvent utilisées dans les sondages politiques, par exemple pour demander pour lequel des deux ou trois candidats les personnes interrogées ont l'intention de voter. Il est également possible de mesurer des construits psychologique de cette façon. Le questionnaire de personnalité professionnelle (*Occupational Personality Questionnaire*, Saville & Holdsworth, 1999), dans sa version 4.2, mesure la personnalité des participant·e·s en leur demandant de choisir une parmi quatre options d'une centaine de propositions (par exemple, « Je suis le genre de personne qui (a) joue pour gagner, (b) implique les autres dans les décisions, (c) analyse le comportement des gens, ou (d) se sent coupable de ses erreurs.»).

³ Les contrôles d'attention sont un moyen simple de savoir qui fait ou ne fait pas attention aux questions d'un questionnaire. Ces questions sont conçues pour s'assurer que l'attention des participant·e·s à l'enquête est maintenue à un niveau élevé pendant toute la durée de l'enquête - sinon, iels sont disqualifié·e·s. Il est recommandé d'utiliser ce type de contrôles pour les questionnaires en ligne.

Une troisième possibilité est de présenter aux gens une affirmation et on leur demande d'utiliser une échelle d'évaluation pour indiquer leur degré d'accord. Cette échelle contient plus d'un élément et chaque valeur de réponse est étiquetée avec les termes spécifiques suivants : tout à fait d'accord, d'accord, ni d'accord ni en désaccord, en désaccord et tout à fait en désaccord. Elle est souvent appelée échelle de Likert (Likert, 1932). Si elle ne suit pas exactement ce format (par exemple, si elle ne comporte qu'un seul élément, ou si ses étiquettes de réponse sont un peu différentes des étiquettes de Likert originales), elle peut être appelée échelle de type Likert. Le questionnaire de personnalité professionnelle (Saville & Holdsworth, 1999), dans sa version 5.2, mesure la personnalité des participants en utilisant une échelle de type Likert. Les échelles de Likert peuvent être unipolaires ou bipolaires. Les échelles unipolaires font référence à la présence ou à l'absence d'un attribut (par exemple, pour un degré d'accord, l'échelle pourrait varier de 1 « Pas du tout d'accord » à 7 : « Tout à fait d'accord »). Les échelles unipolaires n'ont pas de point neutre. Les échelles bipolaires, quant à elles, reposent, comme leur nom l'indique, sur deux pôles opposés. Elles demandent au répondant d'équilibrer ces attributs opposés. Elles sont généralement codées en sept points (par exemple de -3 « tout à fait défavorable » à +3 « tout à fait favorable »), et le point médian (0) est réservé à l'absence de l'une ou l'autre qualité (« ni favorable ni défavorable » par exemple).

Finalement, au lieu du degré d'accord, on peut demander aux répondant·e·s d'évaluer un objet cible à l'aide d'une échelle numérique ancrée par des adjectifs ; c'est ce qu'on appelle un format différentiateur sémantique. Ce type de questionnaires contient généralement une liste d'échelles à sept points (-3 ; 0 ; +3) entre des adjectifs contrastés (par exemple, bon-mauvais, clair-obscur, social-insocial) et un point zéro neutre. Les échelles de ce type sont nécessairement bipolaires.

Le nombre de points d'une échelle de type Likert — 4 points, 5 points, 7 points, 11 points, ou plus — est sujet à débat. D'une part, d'un point de vue psychométrique, une échelle qui dispose de plus de points sera plus sensible et donc, de meilleure qualité. Toutefois, cet avantage est mitigé au-delà de 11 points dans l'échelle (Nunally, 1978). D'un autre côté, un nombre trop élevé de points peut poser un problème parce que cela induit une fatigue cognitive supplémentaire pour les participants.

Recherches expérimentales et quasi-expérimentales

Dans une recherche expérimentale, les chercheurs·euses manipulent, ou provoquent des changements, dans la variable indépendante (VI), et observent ou mesurent l'impact de ces changements sur la variable dépendante (VD). Les expériences sont conçues pour établir des relations de cause à effet. En manipulant une ou plusieurs variables et en mesurant l'effet sur une autre variable, les expériences peuvent fournir des preuves solides de l'existence d'une relation de cause à effet.

Tout d'abord, comme les variables indépendantes sont sous le contrôle de l'expérimentateur·ice·, ce·tte dernièr·e assigne aléatoirement les participant·e·s aux conditions expérimentales. Cette assignation aléatoire est l'élément le plus important des méthodes expérimentales. L'assignation aléatoire fait en sorte que les groupes, en moyenne, sont similaires sur toutes les caractéristiques, faisant ainsi en sorte que la seule les modalités

de la VI soient différentes entre les deux groupes, et permettant ainsi déduire que seule cette VI peut être la cause des variations dans la VD.

De plus, l'expérimentateur·ice contrôle également l'environnement dans lequel les participant·e·s passent l'expérience. Iel s'assure qu'aucune autre variable ne peut expliquer la relation entre la VI et la VD, et donc, qu'il n'y a pas de variable confondue. Les facteurs de confusion sont des éléments qui peuvent nuire à la capacité à tirer des conclusions causales. Par exemple, si vous vouliez tester si une nouvelle pilule rend les gens plus intelligents, vous pouvez assigner aléatoirement les participant·e·s à prendre ou non la pilule (la VI) et comparer ces deux groupes sur un test d'intelligence (la VD). Cependant, si certains participant·e·s savent qu'iel vont prendre la pilule qui rend intelligent, iels peuvent développer des attentes qui pourront influencer le test. C'est ce qu'on appelle parfois l'effet placebo. Parfois, le simple fait qu'une personne sache qu'elle bénéficie d'un traitement spécial ou de quelque chose de nouveau suffit à provoquer des changements de comportement ou de perception. En d'autres termes, même si les participant·e·s de la condition "traitement" obtenaient de meilleurs résultats au test d'intelligence, nous ne saurions pas si la substance contenue dans la pilule les rend réellement plus intelligent ou s'il s'agit d'un effet placebo. Il y aurait alors ici une variable confondue.

Même les attentes de l'expérimentateur·trice peuvent influencer le résultat d'une étude. Par exemple, si l'expérimentateur·trice sait qui a pris la pilule et qui ne l'a pas prise, son comportement pourrait être légèrement différent en fonction du groupe de participant, et cela pourrait induire, sans qu'iel en ait conscience, des modifications dans les résultats. L'expérience pourrait alors montrer des améliorations dans la performance du groupe expérimental qui ne sont pas réelles. D'où l'importance d'utiliser une procédure en double aveugle, dans laquelle ni les participant·e·s ni l'expérimentateur·trice ne savent dans quelle condition se trouve les participant·e·s. Par conséquent, comme les deux parties sont « aveugles » à la condition, aucune d'entre elles ne pourra se comporter de manière à introduire une confusion. Finalement, la seule différence entre les groupes sera les pilules que les participant·e·s ont reçues, ce qui permettra aux chercheurs·euses de déterminer si la pilule a effectivement rendu les gens plus intelligents.

Et si un·e chercheur·euse voulait étudier les effets du sexe des participant·e·s sur le bonheur ? Pourrait-iel assigner aléatoirement le sexe masculin ou le sexe féminin aux participant·e·s ? Bien sûr que non. Iel se verrait obligé·e d'utiliser un plan quasi-expérimental. Une recherche quasi-expérimentale est similaire à la recherche expérimentale, à l'exception du fait que l'assignation aléatoire des participant·e·s aux différentes conditions de l'expérience n'est pas possible. Au lieu de cela, les chercheurs·euses en psychologie sociale sont parfois obligés de s'appuyer sur les appartenances de groupe existantes (par exemple, homme ou femme). Ces appartenances sont traitées comme des VIs, même si personne n'affecte les participant·e·s aux différentes conditions et ne manipule les variables. Par conséquent, avec les plans quasi-expérimentaux, l'inférence causale est plus difficile. Par exemple, les hommes et les femmes diffèrent sur un grand nombre d'autres caractéristiques que les différences biologiques : leurs rôles de genre, les stéréotypes, des socialisations et des attentes différentes, etc. autant de variables qui sont tout autant susceptibles d'expliquer des différences en termes de compétences, de personnalité, ou de comportements entre les hommes et les femmes que les différences génétiques, ou hormonales.

L'analyse documentaire

L'analyse documentaire, s'appuyant sur des archives, des statistiques et d'autres documents tels que des discours, des lettres ou même des tweets, offre une autre perspective sur le comportement humain. Bien que cette méthode soit généralement utilisée comme un type de recherche corrélationnelle - en raison du manque de contrôle sur les variables pertinentes - la recherche documentaire partage la validité écologique supérieure de l'observation naturaliste. En d'autres termes, les observations sont effectuées en dehors du laboratoire et représentent des comportements du monde réel. De plus, comme les archives examinées peuvent être collectées à tout moment et auprès de nombreuses sources, cette technique est particulièrement flexible et implique souvent une moindre dépense de temps et d'autres ressources lors de la collecte des données.

Les psychologues sociaux ont utilisé l'analyse documentaire pour tester une grande variété d'hypothèses en utilisant des données du monde réel. Par exemple, une étude a comparé les registres de lynchages raciaux aux États-Unis entre 1882 et 1930 au prix du coton ajusté à l'inflation pendant cette période (un indicateur clé de la santé économique du Sud), démontre une corrélation négative significative entre ces variables. En d'autres termes, il y a eu plus de lynchages lorsque le prix du coton est resté stable, et moins de lynchages lorsque le prix du coton a augmenté (Hovland & Sears, 1940). Cela suggère que la violence fondée sur l'ethnie est associée à la santé de l'économie, conformément aux propositions de la théorie des conflits réels (Sherif, 1966).

Les analyses des messages sur les médias sociaux comme Twitter fournissent également aux psychologues sociaux des ensembles de données extrêmement importants pour tester des hypothèses créatives. Dans un exemple de recherche sur les attitudes à l'égard des vaccinations, Mitra et ses collègues (2016) ont collecté plus de 3 millions de tweets sur quatre ans. Ils ont constaté que ceux qui avaient des attitudes anti-vaccination dans leur tweet étaient également plus susceptibles de tweeter sur leur méfiance à l'égard du gouvernement et leurs croyances dans les conspirations gouvernementales.

Les méta-analyses et les revues systématiques sont des cas particuliers d'analyses documentaires qui consistent en une synthèse et une analyse de recherches scientifiques précédemment menées. Une méta-analyse est une analyse statistique qui combine les résultats de plusieurs études quantitatives afin d'estimer l'effet global d'un traitement ou d'une intervention. Les méta-analyses sont généralement utilisées pour fournir une estimation plus complète et impartiale des effets d'un traitement ou d'une intervention particulière que ne peut l'être une étude seule. Le but d'une revue systématique est de fournir un résumé détaillé et impartial des données disponibles sur un sujet particulier. Les revues systématiques peuvent être plus vastes en termes de portée que les méta-analyses, car elles peuvent inclure non seulement des recherches quantitatives, mais également des recherches qualitatives. Ces études peuvent être utiles pour résumer et synthétiser la recherche existante sur un sujet particulier et pour identifier les lacunes de la littérature qui pourraient nécessiter d'être abordées dans des études futures. Les revues systématiques et les méta-analyses sont considérées comme précieuses car elles évaluent le poids des preuves dans une littérature scientifique. Elles permettent de savoir si, dans un certain nombre d'études, il existe une

relation entre deux variables - et si oui, quelle est sa force. Cependant, la puissance d'une méta-analyse dépend des données qui la composent. Si les chercheurs·euses ne s'efforcent pas d'inclure des résultats non publiés ou si les études qu'ils incluent ont suivi des pratiques questionnables de recherche⁴, la méta-analyse aboutira à une conclusion biaisée. De nouvelles méthodes permettent de s'assurer que les études incluses dans une méta-analyse ont été bien menées.

Conclusion : Des méthodes multiples pour une compréhension globale

Ce chapitre a présenté un aperçu des avantages et des limites des différentes méthodes de recherche en psychologie sociale, en fonction des différents types d'affirmation — de fréquence, d'association, ou de causalité — qu'elles permettent de faire, et du type de validité — de construit, externe, interne, ou statistique — qu'elles permettent d'accomplir. Nous avons également présenté les fondements épistémologiques de la méthode scientifique en psychologie. L'objectif de ce chapitre était de montrer que chaque méthode de recherche a ses propres forces et faiblesses, mais chacune d'entre elles est utile de différentes manières et s'adapte mieux à différents processus de raisonnement logique. Le choix de la méthode dépendra des questions de recherche, des objectifs et des ressources disponibles. Des méthodes de recherche quantitatives et qualitatives peuvent être utilisé conjointement pour parvenir à une compréhension plus complète d'un phénomène, en triangulant les résultats, ce qui augmente la validité et la fiabilité de la recherche (Diener et al., 2022).

Toutes les méthodes sont fondées sur des postulats. Pour que les conclusions d'un projet de recherche soient acceptées, ces postulats doivent être reconnus et vérifiés dans la mesure du possible. Tout d'abord, il est important de comprendre dans quelle démarche épistémologique les différentes méthodes de recherche s'inscrivent. Les méthodes de recherche qualitative, telles que l'ethnographie, l'étude de cas et l'analyse narrative, sont plus proches du raisonnement inductif. Ces méthodes impliquent la collecte et l'analyse de données riches et détaillées, telles que des entretiens, des observations et des textes, afin de générer des hypothèses et de comprendre des phénomènes complexes. Les méthodes qualitatives sont particulièrement utiles pour générer de nouvelles connaissances et une meilleure compréhension dans des domaines peu connus, ou pour explorer des phénomènes difficiles à quantifier. Les méthodes qualitatives impliquent souvent une compréhension approfondie du contexte et des perspectives des participants, ce qui permet une représentation plus précise des expériences et des perspectives étudiées dans leur contexte, favorisant ainsi la validité externe, mais diminuant la validité interne. En outre, la nature ouverte de nombreuses méthodes qualitatives permet d'explorer des résultats et des thèmes

⁴ L'expression "pratiques questionnables de recherche" (en anglais *Questionable Research Practices* ou QRP, cf. John et al., 2012) est utilisée pour décrire des pratiques visant à favoriser la « découverte » de résultats significatifs, telles que la publication sélective de résultats, ne pas décrire toutes les variables dépendantes ou toutes les conditions expérimentales de l'étude parce qu'elles n'ont pas fonctionné comme attendu, collecter davantage de données après avoir cherché à savoir si les résultats seront significatifs, présenter un résultat inattendu comme ayant été prédict dès le début (*HARKing : Hypothesizing After Results are Known*, Kerr, 1998), arrondir une p-value pour la faire paraître significative alors qu'elle ne l'est pas tout à fait, ne présenter que les études qui ont fonctionné, ... Une pratique questionnable n'est pas nécessairement frauduleuse ou répréhensible, mais elle "soulève des questions", et est susceptible de porter atteinte à la crédibilité et à la réplicabilité des résultats de recherche. La fraude se caractérise généralement par la falsification de données, c'est-à-dire la création de données de toute pièce, ou la modification de données pour appuyer les résultats.

inattendus qui n'auraient peut-être pas été identifiés par des méthodes quantitatives plus structurées. Cependant, elles s'appuient souvent sur des échantillons de petite taille et à des méthodes d'échantillonnage non probabilistes, ce qui peut rendre difficile la généralisation des résultats à l'ensemble de la population.

Les méthodes de recherche quantitatives, telles que les expériences et les questionnaires, sont généralement plus alignées avec un raisonnement déductif. Elles sont conçues pour identifier des associations entre les différentes variables, et pour contrôler les variables parasites et reposent habituellement sur des échantillons de grande taille, ce qui permet une meilleure généralisation des résultats. Elles favorisent donc la validité interne, mais souvent au détriment de la validité externe. Par conséquent, ces méthodes peuvent conduire à une vision trop simpliste de la réalité. Les expérimentations sont conçues dans un but précis, et elles se concentrent souvent sur un ensemble limité de variables ou supposent des relations linéaires entre les variables alors que ce n'est pas forcément le cas. Cela signifie que les expériences ne saisissent pas toujours tous les aspects des situations de la vie réelle qui pourraient avoir des implications importantes pour la compréhension des relations de cause à effet, et qui pourraient rendre plus difficile la réplication des résultats (van Bavel, et al., 2016). Cela signifie également que les expérimentations sont à privilégier quand l'objectif des chercheurs-euses est de vérifier une hypothèse explicative d'un comportement particulier et à corroborer une théorie. Néanmoins, l'expérimentation doit être utilisée avec précaution lorsqu'elle est conduite à des fins de découvertes. Utilisée en l'absence d'hypothèses fondées théoriquement, cette méthodologie comporte un risque élevé de mettre en évidence des résultats qui en réalité seraient des faux positifs (Oberauer & Lewandowsky, 2019).

La réplication est une étape absolument essentielle, mais ce n'est qu'une étape (Diener et al. 2022). Dans la mesure du possible, les chercheurs-euses doivent également effectuer des réplications conceptuelles dans lesquelles la variable indépendante théorique présumée est manipulée de plusieurs façons différentes et la variable dépendante théorique est mesurée de plusieurs façons différentes. De telles réplications conceptuelles réduisent le risque qu'un résultat soit le résultat d'un type particulier de demande expérimentale. Les réplications directes, étant donné qu'elles reposent sur une même opérationnalisation des variables en jeu, peuvent conduire à une compréhension déformée du comportement humain. Parallèlement, le recours exclusif à une seule méthode de recherche est susceptible de résulter en une vision fausse du comportement humain.

Pour conclure, même si l'expérimentation joue un rôle particulier dans la recherche en psychologie, elle doit être utilisée dans le cadre d'un programme de recherche cumulatif qui les combine avec d'autres méthodes (Diener et al., 2022). En raison de leur correspondance avec certaines postures épistémologiques, et de la nécessaire négociation entre différentes formes de validité qui peuvent parfois s'opposer, il est nécessaire de combiner différentes méthodologies disponibles pour permettre aux chercheurs-euses en psychologie d'acquérir une compréhension plus complète des phénomènes qu'ils étudient.

Bibliographie

- Aguinis, H., Villamor, I., & Ramani, R. S. (2021). Mturk research : Review and recommendations. *Journal of Management*, 47(4), 823-837.
<https://doi.org/10.1177/0149206320969787>
- Chmielewski, M., & Kucker, S. C. (2020). An mturk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464-473.
<https://doi.org/10.1177/1948550619875149>
- Diener, E., Northcott, R., Zyphur, M. J., & West, S. G. (2022). Beyond experiments. *Perspectives on Psychological Science*, 17(4), 1101-1119.
<https://doi.org/10.1177/17456916211037670>
- Diener, E., & Oishi, S. (2000). Money and happiness: Income and subjective well-being across nations. In E. Diener & E. M. Suh (Eds.), *Culture and subjective well-being* (pp. 185–218). The MIT Press.
- Festinger, L., Riecken, H., & Schachter, S. (1956). *When prophecy fails*. Minneapolis, University of Minnesota Press.
- Fiske, Susan T., Taylor, Shelley E. (2011). *Cognition sociale: Des neurones à la culture*. Mardaga.
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should We Trust Web-Based Studies? A Comparative Analysis of Six Preconceptions About Internet Questionnaires. *American Psychologist*, 59(2), 93–104. <https://doi.org/10.1037/0003-066X.59.2.93>
- Hovland, C. I., & Sears, R. R. (1940). Minor studies of aggression: Correlation of lynchings with economic indices. *Journal of Psychology*, 9, 301–310.
- John L. K., Loewenstein G., & Prelec D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524–532. doi:10.1177/0956797611430953
- Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P., Jewell, R., & Winter, N. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, 8(4), 614-629. doi:10.1017/psrm.2020.6
- Kenny D. A. (1979). *Correlation and Causality*. New York: John Wiley & Sons, Inc.
- Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, 2(3), 196–217.
https://doi.org/10.1207/s15327957pspr0203_4
- Kuhn, T.S. (1962). *La structure des révolutions scientifiques*. Paris, Flammarion.

- Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice, in T. S. Kuhn (Ed.), *The essential tension: Selected studies in scientific tradition and change* (pp. 320-339). Chicago: University of Chicago Press.
- Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues*, 2(4), 34-46.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 22, 140, 55.
- Mitra, T., Counts, S., & Pennebaker, J. (2016). Understanding Anti-Vaccination Attitudes in Social Media. *International AAAI Conference on Web and Social Media*, North America. Available at: <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13073/12747>.
- Nisbett, R. E. & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs, NJ, USA: Prentice-Hall.
- Nunnally, J. C. (1978). An Overview of Psychological Measurement. In B. B. Wolman (Ed.), *Clinical Diagnosis of Mental Disorders* (p. 97-146). Springer US. https://doi.org/10.1007/978-1-4684-2490-4_4
- Oberauer, K., Lewandowsky, S. Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26, 1596–1618 (2019). <https://doi.org/10.3758/s13423-019-01645-2>
- Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, 23(3), 184–188. <https://doi.org/10.1177/0963721414531598>
- Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5), 411-419. <https://doi.org/10.1017/S1930297500002205>
- Peirce, C.S. (1998). *The essential Peirce. Selected Philosophical Writings*. Volume 2 (1893-1913). Bloomington: Indiana University Press.
- Piff, P. K., Stancato, D. M., Côté, S., Mendoza-Denton, R., & Keltner, D. (2012). Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences*, 109(11), 4086-4091. <https://doi.org/10.1073/pnas.1118373109>
- Popper, K. R. (1973) *La logique de la découverte scientifique*. Payot.
- Saville & Holdsworth Ltd. (1999). Occupational Personality Questionnaires (OPQ 32) manual. Esher, Surrey: Saville & Holdsworth.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Sherif, M. (1966). *In common predicament: Social psychology of intergroup conflict and cooperation*. Boston: Houghton Mifflin

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366.
<https://doi.org/10.1177/0956797611417632>

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J., & Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, 113(23), 6454-6459. <https://doi.org/10.1073/pnas.1521897113>

Wells, G.L., Leippe, M.R. & Ostrom, T.M. (1979). Guidelines for empirically assessing the fairness of a lineup. *Law & Human Behavior*, 3, 285–293 .
<https://doi.org/10.1007/BF01039807>