



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Disponible en ligne sur www.sciencedirect.com

ScienceDirect

et également disponible sur www.em-consulte.com



Article original

Les pratiques de recherche ouvertes en psychologie[☆]



Open research practices in psychology

A. Lantian

Laboratoire parisien de psychologie sociale, UPL, université Paris Nanterre, 200, avenue de la République, 92100 Nanterre, France

INFO ARTICLE

Historique de l'article :

Reçu le 20 juillet 2019

Accepté le 9 septembre 2020

Mots clés :

Science ouverte

Méthodologie

Transparence

Répliquabilité

Pré-enregistrement

RÉSUMÉ

Cet article vise à offrir une vision d'ensemble des récentes évolutions des pratiques de recherche en psychologie. Un rappel des différents symptômes de la crise de la répliquabilité (et de confiance) ayant affecté la psychologie sera suivi par une discussion approfondie et nuancée des facteurs responsables de cette situation. Il s'agira ensuite, en s'appuyant sur des illustrations et des ressources, de démontrer le rôle crucial des pratiques de recherche ouvertes comme moyen de résoudre ces difficultés. La connaissance et l'adoption de ces pratiques de recherche popularisées par le mouvement de la science ouverte sont indispensables afin de contribuer, via la transparence et l'ouverture, à l'effort collectif d'amélioration de la fiabilité et de la répliquabilité des résultats en psychologie.

© 2020 Société Française de Psychologie. Publié par Elsevier Masson SAS. Tous droits réservés.

ABSTRACT

This article aims to provide an overview of the recent developments of research practices in psychology. A review of the different symptoms of the crisis of replicability (and of confidence) that affected

Keywords:

Open science

Methodology

Transparency

[☆] Thèmes de recherche : croyances, conspirationnisme, libre arbitre, jugement moral.

Adresse e-mail : anthony.lantian@parisnanterre.fr

psychology will be followed by an in-depth and nuanced discussion of the factors responsible for this situation. Then, the next step will be to demonstrate, using illustrations and resources, the crucial role of open research practices to address these challenges. Knowledge and adoption of these research practices popularized by the open science movement are essential to contribute, through transparency and openness, to the collective effort to improve the reliability and replicability of psychological results.

© 2020 Société Française de Psychologie. Published by Elsevier Masson SAS. All rights reserved.

L'objectif de cet article est de traiter des dites « pratiques de recherche ouvertes » (parfois connues sous le nom de « *best* » ou « *good research practices* »¹, Asendorpf et al., 2013 ; Finkel, Eastwick, & Reis, 2015 ; van't Veer & Giner-Sorolla, 2016). Ces dernières sont au cœur des recommandations mises en avant par le mouvement de la *science ouverte* en psychologie (*open science* en anglais ; Spellman, Gilbert, & Corker, 2018). Après un bref constat des difficultés identifiées au sein du système de production des connaissances en psychologie, il s'agira de recenser différents facteurs susceptibles d'en être à l'origine. Pour terminer, je donnerai un aperçu des récentes évolutions des pratiques de recherche ouvertes adoptées par les chercheurs afin d'améliorer la transparence, la fiabilité, et la répliquabilité des résultats issus des publications scientifiques.

1. La crise de la répliquabilité

La répliquabilité est une des pierres angulaires de la méthode scientifique (Zwaan, Etz, Lucas, & Donnellan, 2018). Les termes « répliquabilité » (« *replicability* ») et « reproductibilité » (« *reproducibility* ») sont souvent utilisés de manière interchangeable (Condon, Graham, & Mroczek, 2017 ; Fidler & Wilcox, 2018). Ils véhiculent un sens différent en fonction du domaine disciplinaire dans lequel ils sont employés (Kenett & Shmueli, 2015). Cependant, à l'instar de certains auteurs, pour des raisons de clarté, nous distinguerons les deux termes (Condon et al., 2017 ; Plesser, 2018). La reproductibilité désigne le fait de pouvoir reproduire les résultats d'une étude en utilisant les mêmes procédures analytiques et le même jeu de données que les investigateurs originaux : les chercheurs doivent être en mesure d'aboutir aux mêmes conclusions que l'étude originale. Quant à la répliquabilité, elle renvoie à l'emprunt d'une méthodologie plus ou moins proche de l'étude originale afin de récolter de nouvelles données pour tenter d'obtenir des résultats similaires (Bollen, Cacioppo, Kaplan, Krosnick, & Olds, 2015 ; Condon et al., 2017).

Les répliquations ne visent pas nécessairement toutes le même objectif (Crandall & Sherman, 2016 ; Zwaan, Etz et al., 2018). Il faut distinguer la répliquabilité directe (ou exacte) de la répliquabilité conceptuelle. D'une part, la répliquabilité directe vise à tester la ou les prédictions d'origine dans les conditions les plus proches possibles de celles de l'étude dont on cherche à répliquer l'effet, ou du moins, dans les conditions considérées comme nécessaires à l'observation de l'effet d'intérêt (Simons, 2014 ; Zwaan, Etz et al., 2018). D'autre part, la répliquabilité conceptuelle consiste à répliquer les résultats obtenus dans un contexte différent. Pour être plus précis, la répliquabilité conceptuelle cherche à mettre à l'épreuve la pertinence des hypothèses auxiliaires, à savoir, s'assurer que l'effet obtenu ne tenait pas seulement au type de matériel ou de mesures utilisés, à la population étudiée, etc. Ce type de répliquabilité permet de tester la généralisabilité et la robustesse des découvertes scientifiques (Crandall & Sherman, 2016 ; LeBel, Berger, Campbell, & Loving, 2017). Bien entendu, dans la réalité, il

¹ Je privilégie l'emploi de « pratiques de recherche ouvertes » plutôt que « bonnes » ou « meilleures » pratiques de recherche, car la première expression présente l'avantage d'être plus descriptive comparativement aux deux dernières, qui peuvent revêtir une connotation morale et véhiculer un jugement de valeur. L'usage de l'expression « nouvelles pratiques de recherche » est tout autant discutable, car doublement imprécis (que ce soit au sujet de la nature de ces pratiques de recherche tout autant que leur temporalité).

existe tout un éventail de situations intermédiaires entre ces deux pôles que sont la réplication directe et conceptuelle (LeBel et al., 2017)².

Ces dernières années, il était difficile pour les chercheurs en psychologie de passer à côté de nombreuses discussions et réactions à ladite crise de la réplication en psychologie (Asendorpf et al., 2013 ; Allard, 2018 ; Baker, 2016 ; Bastart, Klein, & Ijzerman, 2018 ; Bêgue, 2018 ; Chopik, Bremner, Defever, & Keller, 2018 ; Cursan, 2018 ; De Boeck & Jeon, 2018 ; Earp & Trafimow, 2015 ; Kruglanski, Chernikova, & Jasko, 2017 ; Maxwell, Lau, & Howard, 2015 ; Motyl et al., 2017 ; Pashler & Wagenmakers, 2012 ; Swiatkowski & Dompnier, 2017), c'est-à-dire la surprenante difficulté à réussir à répliquer des effets issus de publications scientifiques (Doyen, Klein, Pichon, & Cleeremans, 2012 ; Harris, Coburn, Rohrer, & Pashler, 2013 ; Klein et al., 2014 ; Nosek & Lakens, 2014). Certains parlent également de « crise de confiance » (Pashler & Wagenmakers, 2012). En 2015, un collectif de chercheurs sous l'appellation *Open Science Collaboration* délivre les résultats d'un projet de réplication massif (*Open Science Collaboration*, 2015). Sur 100 tentatives de réplifications de recherches en psychologie, seules 36 s'avèrent « réussies ». Il faut cependant bien garder à l'esprit que la simple attribution du qualificatif « échec » ou « réussite » lorsque l'on parle d'une réplication fait l'objet de très nombreux débats, amenant naturellement à des constats différents suivant les critères retenus, sans même évoquer la question complexe de la détermination des véritables causes de l'échec d'une réplication (Bryan, Yeager, & O'Brien, 2019 ; Gilbert, King, Pettigrew, & Wilson, 2016 ; Nelson, Simmons, & Simonsohn, 2018 ; Stanley, Carter, & Doucouliagos, 2018 ; Zwaan, Etz et al., 2018).

Cette crise³ est loin de ne se limiter qu'à la psychologie (Makel, Plucker, & Hegarty, 2012). Sans chercher à être exhaustif, diverses disciplines scientifiques sont concernées, que ce soit les recherches médicales (Ioannidis, 2005), notamment les recherches précliniques (Begley & Ellis, 2012 ; Freedman, Cockburn, & Simcoe, 2015), les sciences criminelles (Pridemore, Makel, & Plucker, 2018), les neurosciences (Button et al., 2013) et les sciences économiques (Camerer et al., 2016 ; Chang & Li, 2018). Toutefois, le simple fait d'utiliser le terme « crise » pour qualifier ces difficultés de réplifications reste source de débats parmi les chercheurs, comme le souligne à juste titre une enquête de la revue scientifique *Nature* (Baker, 2016). Approximativement la moitié des 1576 interrogés pensent que l'on peut parler d'une « crise de reproductibilité⁴ » sérieuse, tandis que l'autre moitié ne partage pas ce constat, ou alors de manière moins marquée, sans compter les 7 % qui admettent ne pas savoir ce qu'il en est. Le grand public est susceptible d'être informé de la situation de crise qui frappe la psychologie et il est indispensable de se mobiliser sur les moyens d'accroître ce taux de répliquabilité, car c'est en partie là-dessus que repose la valeur perçue et la confiance du grand public envers la communauté scientifique en psychologie (Wagen, Berkessel, & Englich, 2020). Avant de poursuivre, il est indispensable de prendre le temps de s'attarder sur les facteurs à l'origine de cette crise afin de bien comprendre la nécessité d'adopter ces pratiques de recherche ouvertes.

2. Les facteurs à l'origine de cette crise

La communauté scientifique identifie régulièrement plusieurs facteurs suspectés d'être à l'origine de cette crise (Bishop, 2019 ; Friese & Frankenbach, 2019). Commençons cette série de facteurs par

² Le flou qui découle de cette complexité motive des chercheurs (Machery, 2020 ; Nosek & Errington, 2020) à proposer d'abandonner l'usage des termes « réplication directe » et « réplication conceptuelle » jugés mal définis et sources de querelles verbales contreproductives. À la place, toujours d'après ces chercheurs (Machery, 2020 ; Nosek & Errington, 2020), il faudrait distinguer les unités expérimentales (e.g., les participants ou les groupes), les traitements (e.g., les variables indépendantes), les mesures (e.g., les variables dépendantes), et le cadre contextuel (e.g., le laboratoire, le dispositif expérimental, etc.).

³ D'un point de vue historique, l'usage du terme « crise » pour caractériser l'état actuel de la recherche en psychologie n'est en réalité pas nouveau. Bien que fondé sur des critères différents, le mot « crise » était déjà employé dans les années 1960–1970 (Giner-Sorolla, 2012 ; Morawski, 2020). Ainsi, notre discipline n'a pas été frappée par une, mais plusieurs crises successives de différentes natures.

⁴ Il faut remarquer que le terme « crise de reproductibilité » est employé dans la formulation de la question posée et non « crise de répliquabilité ». Compte tenu de l'occasionnel usage interchangeable des termes « reproductibilité » et « répliquabilité », le rapprochement entre les résultats de cette enquête et la reconnaissance des chercheurs de la communauté scientifique dans son ensemble de l'existence d'une crise de répliquabilité (comparable à celle qui concerne la psychologie) telle que définie précédemment ne peut pas se faire sans ambiguïté.

l'évocation de la facilité déconcertante avec laquelle on peut obtenir des faux positifs, à savoir, conclure à tort à l'existence de quelque chose qui n'existe pas en réalité (Simmons, Nelson, & Simonsohn, 2011). Dans un article publié dans *The American Scientists* (Gelman & Loken, 2014), des chercheurs illustrent cette idée à l'aide de la métaphore d'un jardin aux sentiers qui bifurquent. Celle-ci illustre l'arborescence insoupçonnée résultante de microdécisions successives prises a posteriori, qui peuvent radicalement changer les conclusions d'une étude. Une fois les données récoltées, il peut être tentant d'ajouter des tests statistiques à ceux initialement prévus en amont de l'étude. Par exemple, il peut s'agir de savoir si les résultats varient si l'on considère séparément les hommes et les femmes ou différentes catégories d'âge. Ou encore, de voir ce qu'il se passe si l'on enlève un item d'une échelle, si l'on regroupe deux items que l'on avait prévu d'analyser séparément, si l'on applique telle ou telle transformation sur une variable donnée, si l'on rajoute une variable en covariée, etc. En clair, sans fixer précisément au préalable les paramètres clés sur lesquels nous allons nous reposer à l'issue de l'expérience, il existe une infinité de manières de regrouper ou diviser des données en fonction d'une multitude de critères plus ou moins arbitraires. L'usage approprié d'un test statistique implique de fixer au préalable le seuil α correspondant au risque de faux positif que l'on est prêt à tolérer si l'hypothèse nulle est vraie (i.e., erreur de Type I), et ce pour un seul test. En multipliant le nombre de tests, le seuil α augmente considérablement par rapport à celui fixé au départ (dans l'éventualité où l'on ne prend pas la peine de le corriger en fonction du nombre de tests réalisés). Il s'ensuit qu'à force de répétition, on finira mécaniquement par tomber sur un résultat dit « significatif » avec un risque accru que ce dernier soit un faux positif. On nomme communément *p-hacking* la pratique consistant à faire émerger un effet significatif en multipliant déraisonnablement le nombre de tests statistiques sur une même base de données (Head, Holman, Lanfear, Kahn, & Jennions, 2015 ; Nelson et al., 2018). Cela traduit un détournement de l'usage prévu des tests statistiques et de nombreux signes indiquent que le *p-hacking* est pratiquée au sein de différentes disciplines scientifiques (Baum & Bromiley, 2019 ; Head et al., 2015).

Une publication illustre formidablement bien qu'en adoptant beaucoup de flexibilité dans l'analyse des données, on peut conclure des aberrations telles que le fait d'écouter « *When I'm sixty-four* » des Beatles plutôt qu'une autre musique fait rajeunir les gens au sens propre (Simmons et al., 2011). La même année, une publication accréditant l'existence de la faculté de prédire l'avenir, ou autrement dit, du don de précognition, paraît dans la revue *Journal of Personality and Social Psychology* (Bem, 2011). Depuis lors, de nombreuses critiques ont été émises réfutant l'idée que ces résultats seraient une démonstration de l'existence de la précognition : ils seraient plutôt le fruit d'une trop grande flexibilité dans l'usage des statistiques (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011). Comme le rappelle l'expression populaire « Si vous torturez vos données suffisamment longtemps, elles vont finir par vous dire ce que vous avez envie d'entendre » Mills (1993, p. 1196).

Lorsqu'une exploration des données conduit à des découvertes ou conclusions qui n'étaient pas initialement prévues, cette grande flexibilité peut conduire à la tentation du « HARKing », à savoir, *Hypothesizing After the Results are Known*. Cela consiste à prétendre que des données corroborent une hypothèse a priori fondée sur un raisonnement théorique préalable qui a motivé la réalisation d'un travail empirique, en omettant de signaler que l'hypothèse présentée a en réalité été fabriquée (ou modifiée) a posteriori de façon à correspondre à des résultats qui n'ont pas été anticipés (Kerr, 1998). La pratique du HARKing est problématique à plus d'un titre, comme le refus de s'engager dans une démarche de réfutation d'une hypothèse ou encore le risque d'inflation des faux positifs (Kerr, 1998 ; voir aussi Rubin, 2019 pour une analyse critique et nuancée des effets supposément délétères du HARKing). Imaginons que des chercheurs publient un article démontrant l'effet d'une thérapie psychologique donnée sur la qualité du sommeil, tout en prétendant avoir conçu l'étude afin de précisément tester cette hypothèse. On admet aisément que le résultat paraît beaucoup plus douteux s'il a été porté à notre connaissance que ce résultat est extrait d'une étude dans laquelle cinq autres thérapies psychologiques concurrentes ont été testées, accompagnées d'une dizaine de mesures différentes, et sans que cette hypothèse spécifique portant sur cet effet en particulier ait été formulée au préalable. Plus préoccupant encore, l'hypothèse de départ motivant l'étude portait en réalité sur l'effet supposé d'une autre thérapie psychologique sur l'anxiété. Dans la même veine, la transformation de résultats initialement chaotiques en une version sublimée a été nommée « effet chrysalide », en

référence au processus de métamorphose de la chenille en paillon (O'Boyle, Banks, & Gonzalez-Mulé, 2017). En l'occurrence, dans cet article, les auteurs illustrent le décalage entre les hypothèses et les résultats rapportés dans des thèses de doctorat en management, et ces mêmes résultats concrétisés sous la forme de publications scientifiques (voir Cairo, Green, Forsyth, Behler, & Raldiris, 2020 pour un constat similaire en psychologie sociale). En un mot, les résultats dans la revue scientifique sont présentés comme plus conclusifs qu'ils ne l'étaient dans le manuscrit de thèse de doctorat.

Ces pratiques peuvent être assimilées à une catégorie plus englobante : les pratiques de recherches discutables (connu sous le terme de « *questionable research practices* » en anglais ; John, Loewenstein, & Prelec, 2012 ; Simmons et al., 2011). Selon certaines études, si l'on accorde du crédit au déclaratif, ces petits arrangements avec la réalité ne sont pas rares dans la pratique des chercheurs (Martinson, Anderson, & de Vries, 2005). Par exemple, dans une étude publiée en 2012, environ un tiers des plus de 2000 chercheurs états-uniens issus de différentes spécialités en psychologie admettent avoir eu au moins une fois recours à ce type de pratique, comme omettre de déclarer l'inclusion d'une mesure voire d'une condition expérimentale (John et al., 2012). Cette publication fréquemment citée (et dont les résultats ont été répliqués chez des chercheurs en psychologie italiens, Agnoli, Wicherts, Veldkamp, Albiero, & Cubelli, 2017 ; voir aussi Rabelo et al., 2019 pour les résultats obtenus chez des chercheurs brésiliens) doit toutefois être interprétée avec prudence. Ceci en raison de la formulation parfois problématique de ces questions, mais également plus fondamentalement de la portée très restrictive de ce qui est mesuré. En effet, il est question d'avoir au moins une fois recouru à un de ces comportements problématiques, ce qui ne permet pas de conclure quant à la prévalence effective de ces comportements (Fiedler & Schwarz, 2016). Ces mêmes auteurs ont par ailleurs conduit une nouvelle étude avec une méthodologie plus rigoureuse qui a donné lieu à des résultats beaucoup moins alarmistes (Fiedler & Schwarz, 2016). Ils prennent la peine de rajouter que le simple fait de suggérer sur la base de résultats déformés que ce type de conduite est fréquent, affaiblit non seulement les messages de prévention destinés à prévenir ces pratiques de recherche discutables, mais nuit également gravement à l'image de la recherche scientifique et par extension la confiance en ses recommandations (Fiedler & Schwarz, 2016). Là encore, ces pratiques de recherche discutables sont loin d'être uniquement propres à la psychologie, elles concernent par exemple l'écologie et la biologie de l'évolution (Fraser, Parker, Nakagawa, Barnett, & Fidler, 2018), les recherches biomédicales (Gardner, Lidz, & Hartwig, 2005), les sciences économiques (List, Bailey, Euzent, & Martin, 2001), les sciences de l'éducation (Makel, Hodges, Cook, & Plucker, 2019), le management (O'Boyle et al., 2017) ou encore l'ingénierie civile (Swazey, Anderson, Lewis, & Louis, 1993), etc. Si ces pratiques de recherche discutables paraissent à première vue inoffensives, elles sont de plus en plus montrées du doigt en raison de leur écart avec les nouveaux standards de recherche qui aspirent à une plus grande rigueur scientifique.

Il faut veiller à ne pas faire de procès d'intention aux chercheurs qui ont recours à ces pratiques, ces dernières ne sont pas nécessairement le fruit d'intentions malhonnêtes et peuvent très bien être expliquées en partie par de simples biais de confirmation (Bishop, 2020), sans même évoquer les causes situationnelles⁵. Si l'on revient sur la trop grande flexibilité dans l'analyse des données, c'est un travers très difficilement évitable, d'autant plus lorsqu'il s'agit d'étudier des questions de recherche complexes. Dans ce cas, il n'y a pas qu'une seule, mais une multitude de stratégies analytiques possibles basées sur des postulats défendables et justifiés. Ces différentes stratégies peuvent conduire à des conclusions divergentes. Une étude démontre cela en constatant l'étonnante hétérogénéité des conclusions lorsque l'on présente le même jeu de donnée et la même question de recherche à 29 équipes de recherche de différentes spécialités ayant travaillé indépendamment les unes des autres (Silberzahn et al., 2018). D'après cette recherche, ni l'expertise des chercheurs en statistiques ni leurs croyances initiales au sujet de la question de recherche ne peuvent expliquer de manière satisfaisante les différences observées. Plus récemment, les analyses de 70 équipes indépendantes portant sur le même jeu

⁵ Bien entendu, cela n'implique pas qu'il faille écarter pour autant les explications mettant en avant des déterminants plus dispositionnels. En ce sens, des données suggèrent que le machiavélisme, une appétence plus marquée en faveur de la publication dans les revues à haut facteur d'impact par rapport à l'adhésion aux standards scientifiques ou encore les buts d'approche de la performance prédisent positivement le recours aux pratiques de recherche discutables (Janke, Daumiller, & Rudert, 2019).

de données d'imagerie par résonance magnétique fonctionnelle débouchent là encore sur ces mêmes conclusions (Botvinik-Nezer et al., 2020). Cela souligne la complexité de l'analyse des données ainsi que la place importante occupée par l'arbitraire.

Un autre facteur expliquant ces difficultés à répliquer provient des biais de publication (voir Cursan, 2018 pour une discussion en français détaillée à ce sujet), désignant la tendance à conditionner la décision de soumettre un article scientifique dans une revue en fonction des résultats obtenus (Cooper, DeNeve, & Charlton, 1997 ; Francis, 2012). Plus précisément, cela revient à favoriser la publication de résultats significatifs corroborant une hypothèse par rapport aux résultats non significatifs.⁶ Ces derniers ont statistiquement plus de chances d'être rejetés des revues scientifiques pour des raisons de manque d'intérêt, d'absence de dimension conclusive ou de caractère suffisamment spectaculaire, traduisant ce faisant un problème structurel (Franco, Malhotra, & Simonovits, 2014). De surcroît, ces données non significatives du point de vue statistique ont une moindre probabilité d'être retranscrites sous forme de manuscrit et donc par extension de faire l'objet d'une communication sous une forme ou une autre (Franco et al., 2014 ; Greenwald, 1975).

Ces résultats non significatifs peuvent faire naître un sentiment de démotivation chez les chercheurs qui, via l'anticipation de divers obstacles et difficultés qui se profilent, conduisent à une forme d'autocensure (Franco et al., 2014). Cela n'exclut pas pour autant l'explication plus prosaïque d'un frein psychologique à l'idée de présenter des données ne permettant pas de conclure en faveur des hypothèses qui nous sont chères. De l'aveu même de certains d'entre eux, ces données non significatives perdent de leur intérêt (Franco et al., 2014). La tendance ne va pas en s'arrangeant, car les résultats significatifs sont surreprésentés dans les publications scientifiques : plus que ce que l'on devrait attendre compte tenu des lois du hasard. Entre 1990 et 2007, on constate une augmentation des résultats soutenant l'hypothèse des chercheurs (Fanelli, 2012). Pour être plus nuancé, à un niveau d'analyse plus fin, on observe des disparités entre les disciplines et les différentes origines géographiques des chercheurs (Fanelli, 2012). Dans une autre étude, comparativement aux sciences physiques, les sciences sociales rapportent une plus grande proportion de résultats qui vont dans le sens des hypothèses, des résultats dits « positifs » (Fanelli, 2010). Selon Fanelli (2010), une des raisons qui expliquerait ces différences proviendrait du fait que les sciences qui cherchent à expliquer des phénomènes de moindre complexité, généralement plus anciennes, posent des hypothèses plus contraignantes, donc plus facilement réfutables, que les sciences qui cherchent à expliquer des phénomènes d'une plus grande complexité, généralement plus récentes. En définitive, ces données qui ne sont pas statistiquement significatives et qui restent dans les tiroirs (d'où le nom « effet tiroir », Rosenthal, 1979), conduisant à une vision déformée de l'état de l'art sur un sujet donné. En effet, au moment de compiler les résultats lors des méta-analyses, les données non publiées qui n'iraient pas dans le sens des hypothèses entraînent une vision partielle (et partielle) de la réalité (Friese & Frankenbach, 2019).

Nous pouvons également identifier un facteur additionnel qui contribue à ces échecs de réplifications : le manque de puissance statistique. En statistiques, la puissance correspond à la probabilité sur le long terme de détecter une relation existante entre les variables étudiées (ou une différence existante entre des groupes) compte tenu de la taille réelle de cette relation (ou de cette différence), d'un effectif donné, et d'un seuil d'erreur fixé au préalable (Forstmeier, Wagenmakers, & Parker, 2017). Il est reconnu depuis des décennies que les études conçues en psychologie partent bien souvent avec une probabilité beaucoup trop faible de détecter un effet au regard de la taille présumée de cet effet ; on parle alors de sous-puissance statistique (Bakker, van Dijk, & Wicherts, 2012 ; Cohen, 1962 ; Maxwell, 2004 ; Stanley et al., 2018). Pour faire une analogie, cela reviendrait à essayer de détecter une planète d'un autre système solaire avec de simples jumelles (analogie dérivée de celle du « télescope » développée par Simonsohn, 2015). Ce problème est encore loin d'être résolu étant donné que les données accumulées depuis des décennies permettent d'avoir une idée assez précise de l'ordre de grandeur des phénomènes étudiés en psychologie (Richard, Bond, & Stokes-Zoota, 2003).

En règle générale, le recrutement de trop petits effectifs est problématique à plusieurs niveaux. Pour faire simple, cela augmente les chances de conclure à tort que ce que l'on cherche à observer

⁶ Par mesure de simplicité, je n'aborderai pas ici les répercussions faisant suite à des situations dans lesquelles un résultat est statistiquement significatif, mais dans le sens opposé à l'hypothèse (Friese & Frankenbach, 2019).

n'existe pas (erreur de Type II), mais également nous conduit à conclure à l'existence de quelque chose à tort (erreur de Type I, [Button et al., 2013](#)). Et ce, notamment en raison des grandes fluctuations d'échantillonnage qui surviennent lorsque l'effectif est trop faible ([Maxwell, 2004](#)). S'ajoute à cela un plus fort niveau d'incertitude concernant l'estimation de la taille de l'effet observé, pouvant donner lieu à une surestimation de la taille de l'effet de même qu'une erreur d'estimation de sa direction ([Gelman & Carlin, 2014](#)). Un cercle vicieux peut alors s'installer, car les études subséquentes sont basées sur une taille d'effet faussée, en raison des biais de publications évoqués précédemment, aggravant du même fait le problème initial. Néanmoins, il ne faut pas être fataliste pour autant, car rétrospectivement, des données indiquent qu'en psychologie sociale, la taille moyenne des effectifs recrutés dans les études augmente depuis 2011 (année à partir de laquelle la crise de confiance en psychologie a gagné du terrain), ce qui peut être interprété comme une prise de conscience de l'importance accordée aux tailles d'échantillons ([Sassenberg & Ditrich, 2019](#)). Cela étant, cette progression des effectifs ne va pas croître indéfiniment, et compte tenu des contraintes structurelles qui pèsent sur l'activité de recherche (e.g., manque de moyens, population d'intérêt difficile d'accès, etc.), il serait injuste d'interpréter le recrutement d'effectifs insuffisants comme le reflet d'un manque d'effort et/ou de méconnaissances statistiques des chercheurs.

Tout bien considéré, on peut légitimement se demander dans quelle mesure ces difficultés à répliquer sont dues à des fraudes scientifiques telles que des manipulations voire des fabrications de données (on pensera notamment à la tristement célèbre *affaire D. Stapel*, [Budzicz, 2015](#) ; [Levelt, Drenth, & Noort, 2012](#)). Tout en reconnaissant les limites inhérentes aux études portant sur ces questions épineuses de fraudes scientifiques ([Fanelli, 2009](#) ; [Gardner et al., 2005](#)), il faut garder à l'esprit que les cas avérés de fabrication de données sont relativement rares ([Fidler & Wilcox, 2018](#) ; [Stricker & Günther, 2019](#)) et devraient mécaniquement expliquer seulement un pourcentage marginal des difficultés de réplifications. En réalité, d'autres facteurs tels que l'erreur de mesure ([Loken & Gelman, 2017](#)) et l'erreur d'échantillonnage semblent beaucoup plus convaincants pour expliquer ces difficultés de réplifications ([Stanley & Spence, 2014](#)).

3. L'adoption des pratiques de recherche ouvertes comme piste de sortie de crise

Après avoir passé en revue un certain nombre de facteurs pouvant expliquer cette difficulté à répliquer, nous allons voir en quoi les pratiques de recherche ouvertes font office de solution afin d'y remédier. Pour concrétiser celles-ci, nous allons nous appuyer sur un exemple fictif. Par exemple, imaginons que l'on souhaite tester l'hypothèse que les personnes portant des vêtements de couleur rouge sont jugées plus attirantes que les personnes portant des vêtements de couleur bleue⁷.

3.1. Le pré-enregistrement

L'adoption des pratiques de recherche ouvertes nécessite d'intégrer le pré-enregistrement comme nouvelle corde à son arc. Certains chercheurs n'hésitent pas à qualifier le pré-enregistrement de « révolution » dans l'activité de recherche ([Nelson et al., 2018](#) ; [Nosek, Ebersole, DeHaven, & Mellor, 2018](#) ; [Spellman, 2015](#) ; [Vazire, 2018](#)). Celui-ci consiste à renseigner un certain nombre d'informations capitales avant de conduire une étude. Parmi ces informations, il est nécessaire de renseigner précisément les hypothèses qui vont être mises à l'épreuve (i.e., les prédictions théoriques), les tests statistiques et les paramètres associés qui vont être appliqués (e.g., seuil d'erreur α), le nombre de participants visé, la règle d'arrêt de la récolte des données ainsi que les critères d'exclusions établis au préalable. Ces informations doivent être enregistrées sur des sites internet spécifiquement conçus à cette fin (sur *Open Science Framework* [OSF], <https://osf.io>

⁷ Cet exemple, bien que semblant à première vue arbitraire, correspond à une hypothèse issue d'un programme de recherche existant. À ce sujet, les lecteurs intéressés sont invités à consulter les résultats témoignant de l'existence de l'effet positif de la couleur rouge sur l'attractivité ([Elliot et al., 2010](#) ; [Elliot & Niesta, 2008](#)) de même que les travaux n'ayant pas répliqué ces résultats ([Peperkoorn, Roberts, & Pollet, 2016](#) ; [Wagge et al., 2019](#)) ou seulement sous certaines conditions ([Lehmann & Calin-Jageman, 2017](#)). Une très récente méta-analyse ([Lehmann, Elliot, & Calin-Jageman, 2018](#)) rend compte de la difficulté de conclure avec certitude à l'existence de cet effet.

ou *AsPredicted*, <https://aspredicted.org/> pour ne citer que les plus notables). Ces sites internet disposent d'un horodatage permettant de faire foi sur la date du pré-enregistrement et rendent impossible la modification a posteriori du document. Il existe de nombreuses ressources sous forme de guides, exemples et modèles, permettant de guider la rédaction d'un pré-enregistrement (van't Veer & Giner-Sorolla, 2016 ; <https://osf.io/registries/discover?provider=OSF> ; <https://osf.io/e6auq/wiki/Example%20Preregistrations>).

Par exemple, nous pouvons déclarer que nous souhaitons recruter 150 participants (avant application d'éventuels critères d'exclusion) qui devront évaluer successivement l'attractivité physique de 250 personnes différentes à partir de photographies présentées dans un ordre aléatoire. Disons qu'une moitié des personnes présentées porteront des vêtements rouges et l'autre moitié des vêtements bleus. Il faudrait préciser qu'il est planifié de moyenniser l'ensemble des évaluations réalisées. En effet, dans le cadre de l'usage des statistiques fréquentistes, il est important de fixer par avance le nombre d'évaluations qui va être pris en compte, car en l'absence de précision sur ce critère, nous laissons une porte ouverte à une modification de ce critère. On pourra ainsi éviter la tentation de changer les critères de manière ad hoc, comme n'inclure que les 150 dernières évaluations de chaque participants, sous prétexte que les 100 premières évaluations sont des essais d'entraînement. Il faudrait de même déclarer par avance le test statistique réalisé ainsi que le seuil d'erreur de première espèce, classiquement 5 % ou parfois 1 %.

On notera au passage que le choix de la valeur du seuil d'erreur de première espèce fait actuellement l'objet de débats entre les tenants d'un seuil plus sévère à 5 sur 1000 (Benjamin et al., 2018) pour qualifier un résultat comme « découverte » contre les tenants de l'idée que ce changement de seuil est une fausse « bonne solution » qui ne réglera pas pour autant tous les problèmes connus (Trafimow et al., 2018). Sans oublier la multitude d'autres positions qu'il est possible de défendre (Lakens et al., 2018), voire l'abandon pur et simple de ces seuils propres au paradigme statistique inférentiel « classique » au profit d'un paradigme statistique bayésien (Kruschke & Liddell, 2018). Quoi qu'il en soit, cette question difficile du choix du seuil en ébullition la communauté scientifique, comme en témoigne un numéro spécial de la revue *The American Statistician* entièrement consacré à cette question (Wasserstein, Schirm, & Lazar, 2019).

Pour revenir au pré-enregistrement, il faut bien admettre que si cette pratique se généralise, les résultats obtenus seraient vraisemblablement beaucoup plus répliquables que les résultats issus d'études non pré-enregistrées. L'une des raisons pourrait être par exemple que le pré-enregistrement du plan d'analyse pourrait réduire les risques d'erreur de Type I (Ledgerwood, 2018) mais plus encore, « d'éviter que les chercheurs ne réduisent de manière non transparente la capacité d'un test à réfuter une prédiction » (Lakens, 2019, p. 223). De récents résultats corroborent cette hypothèse. Un examen de 127 rapports pré-enregistrés dans le domaine biomédical et en psychologie affiche un taux de 61 % d'études dont les conclusions ne viennent pas confirmer les hypothèses, tandis que ce taux est estimé entre 5 et 20 % pour les études non pré-enregistrées (Allen & Mehler, 2019 ; Warren, 2018). De même, sans grande surprise, les tailles d'effets obtenues dans les études pré-enregistrées sont significativement plus petites que dans les études non pré-enregistrées (Schäfer & Schwarz, 2019). Si ces écarts ne s'expliquent pas simplement par une variable confondue (e.g., une plus grande tendance à pré-enregistrer des hypothèses plus risquées), alors pré-enregistrer pourrait conduire ainsi à une réduction des biais de publication. Nous pouvons donc penser que la tendance suivie ce qui a été observé en sciences médicales, à savoir, une réduction drastique des résultats statistiquement significatifs à partir du moment où certains financeurs ont exigé des chercheurs de pré-enregistrer les essais cliniques sur un site conçu à cet effet (clinicaltrials.gov ; Kaplan & Irvin, 2015).

Cependant, le pré-enregistrement ne constitue pas pour autant une solution ultime. Outre le fait qu'un pré-enregistrement ne compensera jamais un cadre théorique défaillant (Lakens, 2019 ; Szollosi et al., 2019), il peut techniquement être falsifié. Par exemple si l'on dépose son pré-enregistrement après avoir analysé ses données ou en réitérant l'étude jusqu'à ce que les résultats corroborent l'hypothèse par simple hasard, en admettant que l'hypothèse nulle soit vraie (Yamada, 2018). Bien que possible, ce type de détournement reste plutôt dissuasif dans le sens où son coût de mise en œuvre est relativement élevé et est un cas de fraude aggravée. Pour autant, cela ne nous dispense pas de nous intéresser de près au contenu de ce pré-enregistrement afin de vérifier ses qualités (Veldkamp et al., 2018). Cet examen est d'autant plus justifié que récemment, une étude de 23 articles

publiés dans la revue *Psychological Science* pour lesquels un pré-enregistrement était rendu accessible montre que l'ensemble de ces articles commettaient un écart par rapport à ce qui était prévu : un écart entre le rêve et la réalité pour reprendre le titre de l'article (Claesen, Gomes, & Tuerlinckx, 2019). Concrètement, cet écart peut se manifester par une analyse annoncée dans le pré-enregistrement qui n'est pas réalisée et/ou rapportée dans l'article, un critère d'exclusion prévu qui n'est pas pris en compte ou le rajout d'un nouveau critère d'exclusion non anticipé, ou encore un effectif recruté qui diffère de l'effectif annoncé dans le pré-enregistrement, etc. Les écarts entre ce qui est consigné dans le pré-enregistrement et ce qui est réellement fait ne posent pas de problème en soi, du moment que ceux-ci sont signalés et justifiés (Nosek et al., 2019). Sur ces 23 articles examinés, seul un article a rapporté l'ensemble des écarts produit par rapport au pré-enregistrement (Claesen et al., 2019).

Au-delà de ce type de pré-enregistrement, arrêtons-nous sur un autre type de pré-enregistrement qui pousse cette logique encore plus loin : les rapports pré-enregistrés (*registered reports* en anglais ; Chambers, 2013). Cela consiste à soumettre son article à l'évaluation par les pairs avant même de passer par la phase de récolte des données. Plus précisément, il s'agit ici de rédiger la partie théorique et la méthode avec le plus grand nombre possible de précisions et de décisions établies au préalable. Ce changement confère un grand nombre d'avantages. Par exemple, si l'étude présente un ou plusieurs défauts méthodologiques, il y a des chances qu'ils puissent être détectés et corrigés en amont. De plus, un des avantages majeurs est que si le protocole de l'étude est mené comme spécifié dans le document, et si les analyses statistiques correspondent bien à ce qui a été annoncé au départ, l'article sera publié, et ce, quels que soient les résultats. Ainsi, avoir la certitude que l'étude sera publiée même en l'absence de résultats statistiquement significatifs devrait entraîner moins de tentations d'enjoliver ses résultats pour augmenter ses chances de voir son article accepté dans une revue. Tout comme ce qui a été observé pour la première forme de pré-enregistrement abordée antérieurement, les premiers résultats sur les rapports pré-enregistrés font état d'un moindre taux de résultats positifs (i.e., rejetant l'hypothèse nulle) sur les études qui adoptent ce format (environ 44 %) en comparaison à celles qui ne l'adoptent pas (environ 96 % ; Scheel, Schijen, & Lakens, 2020 ; voir aussi Wiseman, Watt, & Kornbrot, 2019).

3.2. Donner accès aux ingrédients de ses recherches

Une autre pratique de recherche qui favoriserait cette fois-ci la reproductibilité est le fait de donner un libre accès aux données anonymisées et dans l'idéal, la procédure analytique suivie afin d'analyser ses données (e.g., le script ou le code statistique). De cette manière, tout un chacun peut reproduire les analyses, conduire des analyses alternatives, tester de nouvelles hypothèses, se servir de ces données pour conduire des méta-analyses ou bien détecter d'éventuels traitements sous-optimaux des données voire des erreurs statistiques. Certains chercheurs peuvent se poser la question de l'utilité de cette démarche étant donné que les données peuvent tout simplement être envoyées sur simple requête. Or, on sait justement que ce type de demandes par e-mail est loin de porter ses fruits. En 2005, des chercheurs en psychologie ayant publié dans 4 revues très sélectives ont été contactés par d'autres chercheurs afin de leur demander d'envoyer leurs données. Il en résulte que 73 % d'entre eux ne l'ont pas fait, ce qui viole le cadre éthique en vigueur dans lequel ces revues s'inscrivent (Wicherts, Borsboom, Kats, & Molenaar, 2006), en l'occurrence les principes éthiques de l'association américaine de psychologie (American Psychological Association, 2001). Plus récemment, il a été signalé que la majorité des données des articles les plus cités en psychologie et en psychiatrie publiés entre 2006 et 2016 sont à ce jour inaccessibles, malgré des démarches entreprises afin de pouvoir les récupérer (Hardwicke & Ioannidis, 2018).

Outre le simple aspect rebutant logistiquement parlant (Campbell et al., 2002), ainsi que des barrières légales et éthiques invoquées par les auteurs lorsque l'on demande d'expliquer leur refus de partager leurs données (Hardwicke & Ioannidis, 2018), on peut suspecter qu'une des raisons moins avouables qui pourrait expliquer cette appréhension de rendre accessibles ses données est la crainte qu'une analyse indépendante pointe des erreurs invalidant les conclusions de l'étude. Dans les publications scientifiques, les erreurs statistiques sont plus communes que ce que l'on peut imaginer de prime abord, et ce, dans différentes disciplines, que ce soit en psychologie (Bakker & Wicherts, 2011 ; Nuijten, Hartgerink, van Assen, Epskamp, & Wicherts, 2016 ; Rouder, Haaf, & Snyder, 2019), en

médecine (García-Berthou & Alcaraz, 2004 ; Murphy, 2004), en psychiatrie (Berle & Starcevic, 2007), en management (Bergh, Sharp, Aguinis, & Li, 2017) ou en philosophie expérimentale (Colombo, Duev, Nuijten, & Sprenger, 2018). Les erreurs étant humaines (et source d'embarras, Bishop, 2018), il serait extraordinairement étonnant de n'en observer aucune. Ces erreurs peuvent être de nature très diverse et varier en gravité. Elles peuvent provenir de simples erreurs de copies, intervenant entre le moment où le résultat du test statistique apparaît à l'écran et le moment où le résultat est retranscrit/copié-collé sur le document, ou une erreur plus amont, par exemple, un mauvais test statistique effectué. Certaines erreurs sont plus dommageables que d'autres, par exemple, lorsqu'elles invalident les conclusions d'une étude, faussant par la même occasion les méta-analyses subséquentes.

À titre d'illustration, il est estimé que 15 % d'une première sélection aléatoire d'articles publiés en psychologie contiennent des statistiques erronées conduisant à au moins une conclusion injustifiée (Bakker & Wicherts, 2011). Ces conclusions injustifiées vont plus souvent que ce que l'on devrait attendre du hasard dans un sens favorable aux hypothèses des chercheurs. À partir d'une nouvelle sélection aléatoire d'articles, Bakker et Wicherts (2011) ont contacté 21 auteurs d'articles présentant des incohérences statistiques de niveau de gravité variable. Fait notable : le temps que les chercheurs contactés ont mis à envoyer leurs données était relié à la gravité de ces erreurs, dans le sens où les délais les plus longs étaient associés à des erreurs plus sérieuses en moyenne. La rapidité de l'envoi des données peut être alors considérée comme un indice de la qualité de ces dernières. Outre le délai de réponse, le simple refus d'envoyer ses données est un autre signal négatif. Dans une autre étude, sur 49 chercheurs sollicités afin de fournir la base de données source de leur publication, seuls 21 d'entre eux ont répondu à cette requête (Wicherts, Bakker, & Molenaar, 2011). Les indices statistiques issus des articles des chercheurs qui n'ont pas fourni leurs bases de données présentent plus d'erreurs manifestes et des résultats moins convaincants que ceux ayant fourni leurs bases de données (Wicherts et al., 2011). Ces cas de refus d'envoi de données suite à des sollicitations s'observent également en sciences médicales (Reidpath & Allotey, 2001) et en génétique (Campbell et al., 2002).

Ainsi, dans un souci de conservation des données à long terme, il n'est pas justifié de ne pas donner un accès ouvert à celles-ci, à l'exception des données personnelles ne permettant pas de garantir l'anonymat des participants, dans le respect du cadre juridique en vigueur (Meyer, 2018). Il semblerait que le fait de considérer ses données comme un trésor qu'il ne faut pas partager est une conception archaïque qui tend à se perdre. En ce sens, une récente étude montre que dans le domaine de la biotélémétrie animale, 72 % de chercheurs en début de carrière ont partagé leurs données pour donner suite à une demande, tandis que seuls 11 % des chercheurs avec plus d'ancienneté ont partagé leurs données (Campbell, Micheli-Campbell, & Udyawer, 2019). Comme on l'observe parfois, les revues scientifiques pourraient démocratiser cette pratique en exigeant de donner un libre accès aux données lorsqu'un article est publié en leur sein. La barrière logistique ou financière n'est plus une excuse, car il est extrêmement aisé de déposer ses données sur une plateforme spécifiquement dédiée à cet effet (e.g., sur OSF ; Soderberg, 2018). Enfin, compte tenu des connaissances actuelles, il n'y a pas de preuves qu'annoncer que les données récoltées (et préalablement anonymisées) sont destinées à être librement accessibles sur un site internet public affecte négativement le taux de participations ou la teneur des réponses recueillies via un certain nombre d'échelles de mesure psychologiques (Eberlen, Nicaise, Leveaux, Mora, & Klein, 2019).

Il faut également par la même occasion s'attendre à ce qu'à l'avenir, les évolutions technologiques, informatiques, et statistiques facilitent la détection d'éventuelles erreurs dans les données d'articles publiés antérieurement. En réalité, ce processus a déjà débuté (Brown & Heathers, 2017 ; Heathers, Anaya, van der Zee, & Brown, 2018 ; Nuijten, van Assen, Hartgerink, Epskamp, & Wicherts, 2017). Des données librement accessibles (sous condition qu'elles soient complètes et réutilisables) rendront plus aisées la vérification et la mise à jour des résultats, et par extension leur exactitude. Notons au passage qu'en plus des données, il est d'autant plus souhaitable de partager le code/script utilisé pour les analyser (Goldacre, Morton, & DeVito, 2019). Une fois de plus, le partage du script ne garantit pas la reproductibilité des résultats tels qu'ils apparaissent dans un article, encore faut-il donner suffisamment d'information pour permettre à une personne extérieure d'exécuter le script et d'obtenir les mêmes résultats, sans avoir besoin de l'assistance de (ou des) auteur(s) de l'étude d'origine. À l'heure

actuelle, des données suggèrent que des progrès peuvent encore être faits sur cet aspect (Obels, Lakens, Coles, Gottfried, & Green, 2019).

Tout comme les données, le matériel de l'étude ainsi que la procédure complète devraient être partagés, ou a minima, ceux-ci devrait être suffisamment détaillé afin de permettre à d'autres équipes de reproduire l'étude (Grahe, 2017) et/ou de faciliter la détection de potentielles limites non détectables si un accès du matériel dans son entièreté n'est pas possible. Omettre ces détails revient à demander à une personne de réaliser une recette de cuisine sur la base d'un support incomplet, le résultat sera probablement approximatif. Certes, le nombre de mots limités par les revues ne permet pas toujours de développer cette partie, mais cela peut être compensé par l'ajout d'annexes. Il n'est pas difficile de trouver des situations où les chercheurs déplorent l'impossibilité de reproduire une étude en raison d'un manque de détail dans la publication originale (Kaiser, 2018).

Outre l'accès libre aux données et au matériel, le mouvement des Sciences ouvertes milite également pour un accès ouvert aux articles scientifiques. Les revues en libre accès (*open access*) permettent à quiconque d'accéder librement et gratuitement à leur contenu. Même pour les articles qui ne sont pas publiés dans des revues en libre accès, il est toujours possible de rendre accessible une version de prépublication (également connue sous le nom de *preprint* : version d'un article qui n'a pas encore été soumis pour publication dans une revue, ou qui n'a pas été encore relu par les pairs ou été accepté) ou une version de l'article accepté mais pas encore mis en forme par la revue (*postprint*). Cela peut se faire en proposant le manuscrit en téléchargement direct sur des sites d'archive (e.g., comme le propose le site <https://psyarxiv.com/>), ou en l'hébergement directement sur son site personnel ou institutionnel. Ayant gagné en popularité lors de la pandémie de Covid-19, cette initiative est la bienvenue, que ce soit pour les chercheurs travaillant au sein d'universités ne disposant pas des fonds suffisants pour prendre en charge les abonnements aux revues, en général très onéreux, et plus généralement, pour les citoyens désireux de se renseigner sur un sujet pointu.

3.3. Des transformations structurelles

Il faut bien avoir conscience que le recours à ces pratiques de recherche ouvertes à titre individuel n'atteindra pas un niveau de développement optimal tant que les différentes structures qui encadrent la recherche n'encourageront pas pleinement ces changements (Nosek, Spies, & Motyl, 2012). Si l'activité scientifique peut parfois être dépeinte comme un jeu (Bakker et al., 2012), alors il faut changer les règles du jeu. En clair, il est indispensable que les financeurs publics et privés, les institutions et organismes de recherches via les recrutements, promotions et prix, les sociétés savantes ainsi que les lignes éditoriales des revues scientifiques (Morey et al., 2016) mettent en avant ces critères d'ouverture et de transparence sous peine de ne pas voir une transformation radicale se réaliser (Gernsbacher, 2018). Cela revient à ne pas négliger les solutions dites distales, c'est-à-dire les changements qui s'inscrivent dans le contexte plus général de l'organisation de la communauté scientifique dans son ensemble (De Boeck & Jeon, 2018). La « feuille de route du CNRS pour la science ouverte » (2019) constitue un exemple concret d'engagement public d'une institution publique en faveur de la science ouverte.

D'une part, les changements peuvent être accélérés par les politiques éditoriales des revues scientifiques. Ces dernières années, on peut constater des évolutions dans cette direction. À titre d'illustration, les revues scientifiques ont la possibilité d'adopter une charte de « promotion de la transparence et de l'ouverture » (« *Transparency and Openness Promotion* ») proposée par Nosek et al. (2015). Cette charte est conçue de manière à ce que les revues scientifiques puissent facilement intégrer des critères établis bien définis (e.g., transparence des données, du matériel, etc.) et ajuster leurs besoins en prenant pour repère différents niveaux d'exigence proposés pour chacun de ces critères (<https://www.osf.io/9f6gx/>). Tout laisse croire que les critères mis en avant dans les guides de soumission des revues ne sont pas sans effet (Giofrè, Cumming, Fresc, Boedker, & Tressoldi, 2017). En outre, certaines revues proposent des badges pour récompenser les articles qui se conforment aux pratiques de recherche ouvertes, à savoir les articles qui proposent un accès ouvert aux données, au matériel ainsi qu'un pré-enregistrement. À l'heure où ces lignes sont écrites, on peut recenser 68 revues qui proposent des badges (<https://www.cos.io/our-services/open-science-badges/>). Il n'est pas possible de prendre encore suffisamment de recul à ce sujet, mais les premières données tendent à démontrer l'efficacité

de ces badges (Kidwell et al., 2016). Également, de plus en plus de revues proposent de publier des rapports pré-enregistrés (que ce soit dans le cadre d'une publication ordinaire, dans une section dédiée ou encore dans des numéros spéciaux). Au 20 juillet 2020, 256 revues en sciences comportementales accordent une place à ce format de publication (<https://www.cos.io/our-services/registered-reports>).

D'autre part, le fonctionnement des revues scientifiques dépend des experts qui sont de précieux alliés dans l'évolution des pratiques de recherche. En signant le « Peer Reviewers' Openness Initiative » (<https://www.opennessinitiative.org>) les chercheurs s'engagent à refuser d'expertiser les manuscrits qui ne sont pas accompagnés de leur matière première publiquement accessible (e.g., données, matériel, etc.), ou a minima, qui ne justifient pas leur absence. Moins engageant, pour encourager la transparence au moment de la phase d'expertise, des listes de vérification à destination des experts (ou auteurs) recensant les points importants à vérifier ont vu le jour (Aczel et al., 2020 ; Parker et al., 2018). Les financeurs doivent également suivre cette tendance afin d'encourager des évolutions en ce sens. En France, saluons le « Plan national pour la science ouverte » (2018) qui entre autres « rend obligatoire l'accès ouvert pour les publications et pour les données issues de recherches financées sur projets ».

3.4. Des réformes statistiques

Des réformes de nature statistique peuvent s'ajouter aux réformes de nature méthodologique. Il faut également noter le récent effort de pédagogie et d'accessibilité fourni par certains chercheurs afin de faciliter la compréhension et la prise en main des statistiques Bayésiennes — dont la connaissance et l'usage était auparavant restreint à un cercle limité de chercheurs en psychologie — notamment via l'utilisation de JASP (<https://www.jasp-stats.org> ; Wagenmakers et al., 2018). C'est un logiciel gratuit et facile à prendre en main dont la popularité n'a cessé de croître ces dernières années. Dans la mesure où ce logiciel rend possible la quantification du niveau de preuve en faveur de l'hypothèse nulle (Quintana & Williams, 2018 ; Wagenmakers et al., 2018), cela pourrait encourager les chercheurs à publier une partie de leurs résultats qui auraient été jadis laissés dans leurs archives, sous réserve que l'étude en question atteigne un certain standard de qualité. Sans même aller jusqu'à la nécessité de se familiariser avec une autre philosophie (ou école de pensée) des statistiques, les tests d'équivalences (Lakens, 2017) permettent — tout en restant dans le cadre des statistiques fréquentistes — de tester si un effet obtenu est tellement petit que l'on puisse en conclure, par extrapolation, qu'il est équivalent à 0. Bien évidemment, l'usage d'une des techniques n'exclue pas l'autre, et il est judicieux d'en faire un usage complémentaire. Tout ceci a pour conséquence d'insuffler plus de crédibilité aux publications en psychologie.

3.5. Moins de spectaculaire et plus d'imperfections

En dehors de ces mesures relativement concrètes, une évolution de la philosophie sous-jacente du système de production des connaissances scientifiques pourrait, par l'intermédiaire des attentes normatives qui en découlent, encourager une transformation encore plus profonde. D'une part, cela implique d'assumer les imperfections relatives aux données recueillies et de les rendre banales (Cooper, 2016 ; Lakens & Etz, 2017 ; Nosek et al., 2012). La norme qui invite à raconter de belles histoires a pour défaut de passer sous silence les nuances et les contradictions afin de préserver coûte que coûte la dimension narrative (« Tell it Like it », 2020). D'autre part, si l'on y ajoute la pression à la publication et la valorisation de la nouveauté et du spectaculaire, cela installe les conditions idéales pour faire émerger des publications sensationnalistes de qualités douteuses (Higginson & Munafò, 2016 ; Kruglanski et al., 2017), cherchant à faire émerger un effet suscitant l'émerveillement (e.g., « WOW effect », Jussim, Crawford, Anglin, Stevens, & Duarte, 2016).

Pour briser cette course à la publication, une solution pourrait être de ralentir délibérément le rythme des publications (voir les suggestions pour une transition vers un mouvement de *Slow Science* ; Frith, 2020). Cette proposition peut paraître provocatrice compte tenu des injonctions d'excellence scientifique dont un des marqueurs est la quantité d'articles publiés. Il est clair que mécaniquement, toute chose étant égale par ailleurs, il est inconciliable de maintenir le même rythme de publication tout en mettant en pratique l'ensemble des modalités des pratiques de recherche ouvertes. Ces dernières

nécessitent en amont un temps de déploiement incompressible, ne serait-ce que pour s'assurer de vérifier et revérifier l'exactitude et le bon fonctionnement des contenus que l'on souhaite partager en libre accès. Il va sans dire que cette problématique concerne d'autant plus les jeunes chercheurs à la recherche d'un poste pour qui la pression à la publication est particulièrement forte compte tenu de l'importance accordée au nombre de publications dans les critères de recrutement. Ainsi, on comprend bien que si les financeurs ou les différentes commissions prenant des décisions relatives à la carrière scientifique ne mettent pas l'accent sur ces autres critères, non seulement l'adoption de ces pratiques de recherche ouvertes revient à s'infliger un sévère handicap, mais la généralisation de l'usage de ces pratiques de recherche ouvertes restera limitée. Un autre levier pour inciter à ralentir le rythme de publication serait d'augmenter le coût de soumission d'un article, de limiter le nombre d'articles que l'on peut soumettre sur une période donnée, voire plus radicalement de limiter le nombre de tentatives de soumission par manuscrit (Tiokhin et al., 2019)⁸. Enfin, cette fuite en avant n'encourage pas de prendre le temps de questionner les fondations de notre discipline, pouvant prendre par exemple la forme de projets de réplifications.

3.6. Des initiatives de grande ampleur

La place trop modeste occupée par les tentatives de réplification est un problème identifié depuis des décennies (Smith, 1970 ; Makel et al., 2012). Néanmoins, une issue est en train de voir le jour. Ces dernières années, des études de très grande ampleur ont été conçues dans l'idée de tester la répliquabilité des résultats en psychologie. Il faut remarquer l'initiative *Open Science Collaboration* (Open Science Collaboration, 2015) évoquée précédemment ainsi que les *Replication Registered Reports* (Simons, Holcombe, & Spellman, 2014) consistant à mobiliser de nombreux laboratoires qui, indépendamment et à partir d'un même protocole prédéterminé, vont conduire une réplification directe d'une étude spécifique (Alogna et al., 2014 ; Bouwmeester et al., 2017 ; O'Donnell et al., 2018 ; Wagenmakers et al., 2016). Ajoutons à cette liste les projets baptisés « Many Lab », fonctionnant sur le même principe, à l'exception que ce n'est pas une mais plusieurs études qui vont faire l'objet d'une tentative de réplification dans plusieurs laboratoires. À ce jour, trois d'entre eux ont été publiés et totalisent plus de 24 000 participants (Ebersole et al., 2016 ; Klein et al., 2014 ; Klein, Vianello et al., 2018). D'autres sont en préparation. Plus récemment, un article publié en 2018 présente les résultats d'une tentative de réplification de 21 expériences en sciences sociales publiées entre 2010 et 2015 au sein des deux revues à très haut facteur d'impact *Nature* et *Science* (Camerer et al., 2018). Parfois, ces projets de réplifications de grande ampleur délivrent une conclusion rassurante avec un taux de réplification relativement bon, comme on a pu le constater en psychologie cognitive (Zwaan, Pecher et al., 2018) de même qu'en psychologie de la personnalité (Soto, 2019). Il faut faire le pari que ces efforts collectifs permettent de faire le tri parmi les résultats auxquels on peut accorder notre confiance ou non.

Dans cette continuité, saluons le « *Psychological Science Accelerator* » (<https://www.psycsciacc.org/>). Cette initiative décentralisée consiste à fédérer un ensemble de laboratoires réparti sur toute la planète. Sur la base de volontariat, ces laboratoires proposent de mettre à disposition des ressources – par exemple un effectif de participants – autour d'un projet de recherche ayant été préalablement soumis à une présélection. En somme, comme son nom l'indique, l'objectif est d'accélérer l'accumulation de résultats dont on cherche à examiner la fiabilité et la généralisabilité (Moshontz et al., 2018). Cette initiative répond aux exigences de l'élévation des standards de l'activité de recherche en psychologie, qui pourront difficilement être atteints en l'absence d'une coordination collaborative à grande échelle (Uhlmann et al., 2019). Enfin, un facteur à ne surtout pas négliger est la formation des prochaines générations de chercheurs : les chercheurs de demain sont en ce moment même assis sur les bancs des universités. On peut mentionner des initiatives, par exemple le « Collaborative Education and Research Project » (CREP ; Grahe et al., 2017), qui permet aux étudiants de se former en s'impliquant dans des

⁸ À l'origine, ces propositions ont plutôt été pensées afin d'installer un mécanisme de contrôle de qualité au niveau scientifique, et plus spécifiquement de décourager la soumission de recherche de faible qualité dans des revues de haut niveau scientifique, pour garantir un « signal honnête », expression empruntée à la biologie évolutionniste. Mais rien n'interdit de généraliser ces préconisations à l'ensemble des revues afin de freiner le nombre de publications.

projets encadrés de réplication de très hauts niveaux (voir aussi Hawkins et al., 2018 ; Sarafoglou, Hoogeveen, Matzke, & Wagenmakers, 2020 pour des procédures comparables). Ces projets s'intègrent au sein des formations en méthodologie de la recherche et proposent de suivre une succession de procédures standardisées. Ce découpage en petites étapes permet non seulement de développer la rigueur et la maîtrise des pratiques de recherche ouvertes, mais représente également une très bonne manière d'apprendre par la pratique. Des comptes rendus de cours plus classiques délivrés à ce sujet à destination des étudiants en psychologie attestent de leur valeur pédagogique (Blincoe & Buchert, 2020 ; Chopik et al., 2018). Pour finir, il ne faut pas oublier le rôle important des encadrants de thèse dans la sensibilisation et la formation des doctorants sur ces questions (Krishna & Peter, 2018).

3.7. Communiquer pour dissiper les dernières hésitations

Bien que ces pratiques de recherche ouvertes soient encore source de réticences (Frankenhuis & Nettle, 2018 ; Hardwicke et al., 2020 ; Houtkoop et al., 2018 ; Spellman et al., 2018 ; Washburn et al., 2018), et nécessitent de passer par une phase d'apprentissage (Nosek et al., 2019 ; Veldkamp et al., 2018), la nature des contre-arguments laisse suggérer que davantage d'informations et une meilleure communication vis-à-vis de ces sujets permettraient d'abaisser certaines de ces barrières (Houtkoop et al., 2018 ; Washburn et al., 2018). Par exemple, une objection fréquente consiste à pointer la perte de créativité que provoquerait le pré-enregistrement (Frankenhuis & Nettle, 2018), renvoyant à l'image de l'édification d'une prison, un plan rigide et contraignant qui ne permet pas de laisser une place aux découvertes fortuites pouvant naître d'une exploration moins guidée des données. Cette représentation est incorrecte, car le pré-enregistrement n'est pas incompatible avec une démarche exploratoire : cette pratique implique seulement de signaler de manière transparente la démarcation entre les résultats découlant d'une démarche confirmatoire d'une démarche exploratoire (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). De plus, Frankenhuis et Nettle (2018) argumentent que le pré-enregistrement pousse au contraire à être plus créatif. En effet, le climat plus paisible et collaboratif de même que le temps de la réflexion et la nécessaire anticipation des éventuelles difficultés et imprévus sont un terrain favorable au développement de la créativité.

En conclusion, il faut voir cet épisode de crise et de remise en question de la crédibilité des résultats scientifiques comme une excellente occasion de prendre du recul sur ses propres pratiques et penser à de nouvelles solutions pour améliorer les faiblesses bien identifiées. La documentation au sujet des procédures disponibles afin de rendre le plus transparent possible les différents ingrédients d'une recherche sous forme de produit fini ne manque pas (Klein, Hardwicke et al., 2018 ; Soderberg, 2018). C'est l'occasion de voir se dérouler sous nos yeux le caractère autocorrectif propre à la démarche scientifique (Białek, 2018), participant à la renaissance de la psychologie (Nelson et al., 2018).

Déclaration de liens d'intérêts

L'auteur déclare ne pas avoir de liens d'intérêts.

Références

- Aczel, B., Szasz, B., Sarafoglou, A., Kekecs, Z., Kucharsky, S., Benjamin, D., et al. (2020). A consensus-based transparency checklist. *Nature Human Behaviour*, 4, 4–6. <http://dx.doi.org/10.1038/s41562-019-0772-6>
- Agnoli, F., Wicherts, J. M., Veldkamp, C. L., Albiero, P., & Cubelli, R. (2017). Questionable research practices among Italian research psychologists. *PLoS ONE*, 12. <http://dx.doi.org/10.1371/journal.pone.0172792>
- Allard, A. (2018). *La crise de la réplicabilité*. <https://www.laviedesidees.fr/IMG/pdf/20180320.allardreplicabilite-2.2..pdf>
- Allen, C., & Mehler, D. M. (2019). Open Science challenges, benefits and tips in early career and beyond. *PLoS Biology*, 17. <http://dx.doi.org/10.1371/journal.pbio.3000246>
- Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., et al. (2014). Registered replication report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science*, 9, 556–578. <http://dx.doi.org/10.1177/1745691614545653>
- American Psychological Association. (2001). Ethical standards for the reporting and publishing of scientific information. In *Publication manual of the American Psychological Association* (5th ed., pp. 387–396). Washington, DC: Author.
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., et al. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108–119. <http://dx.doi.org/10.1002/per.1919>
- Baker, M. (2016). Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the crisis rocking science and what they think will help. *Nature*, 533, 452–455.

- Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science*, 7, 543–554. <http://dx.doi.org/10.1177/1745691612459060>
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, 43, 666–678. <http://dx.doi.org/10.3758/s13428-011-0089-5>
- Bastart, J., Klein, R. A., & IJzerman, H. (2018). *Replication initiatives in psychology*. Oxford Bibliographies. <https://www.oxfordbibliographies.com/view/document/obo-9780199828340/obo-9780199828340-0212.xml>
- Baum, J. A., & Bromiley, P. (2019). P-hacking in top-tier management journals. In *Academy of Management Proceedings*. Briarcliff Manor, NY 10510: Academy of Management. <http://dx.doi.org/10.5465/AMBPP.2019.10810Abstract> (Vol. 2019, No. 1, p. 10810)
- Begley, C. G., & Ellis, L. M. (2012). Drug development: raise standards for preclinical cancer research. *Nature*, 483, 531–533. <http://dx.doi.org/10.1038/483531a>
- Bègue, L. (2018). Malscience et renouveau en psychologie sociale. In M. Silberstein (Ed.), *Qu'est-ce que la science... pour vous? Tome 2: 51 scientifiques et philosophes répondent* (pp. 61–68). Paris: Éditions Matériologiques.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407–425. <http://dx.doi.org/10.1037/a0021524>
- Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., et al. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2, 6–10. <http://dx.doi.org/10.1038/s41562-017-0189-z>
- Bergh, D. D., Sharp, B. M., Aguinis, H., & Li, M. (2017). Is there a credibility crisis in strategic management research? Evidence on the reproducibility of study findings. *Strategic Organization*, 15, 423–436. <http://dx.doi.org/10.1177/1476127017701076>
- Berle, D., & Starcevic, V. (2007). Inconsistencies between reported test statistics and p-values in two psychiatry journals. *International Journal of Methods in Psychiatric Research*, 16, 202–207. <http://dx.doi.org/10.1002/mp.225>
- Białek, M. (2018). Replications can cause distorted belief in scientific progress. *Behavioral and Brain Sciences*, 41. <http://dx.doi.org/10.1017/S0140525X18000584>
- Bishop, D. V. M. (2018). Fallibility in science: Responding to errors in the work of oneself and others. *Advances in Methods and Practices in Psychological Science*, 1, 432–438. <http://dx.doi.org/10.1177/2515245918776632>
- Bishop, D. (2019). Rein in the four horsemen of irreproducibility. *Nature*, 568, 435. <http://dx.doi.org/10.1038/d41586-019-01307-2>
- Bishop, D. V. (2020). The psychology of experimental psychologists: Overcoming cognitive constraints to improve research: The 47th Sir Frederic Bartlett Lecture. *Quarterly Journal of Experimental Psychology*, 73, 1–19. <http://dx.doi.org/10.1177/1747021819886519>
- Blincoe, S., & Buchert, S. (2020). Research preregistration as a teaching and learning tool in undergraduate psychology courses. *Psychology Learning & Teaching*, 19, 107–115. <http://dx.doi.org/10.1177/1475725719875844>
- Bollen, K., Cacioppo, J. T., Kaplan, R., Krosnick, J., & Olds, J. L. (2015). *Social, behavioral, and economic sciences perspectives on robust and reliable science*. Arlington, VA: National Science Foundation. <https://www.nsf.gov/sbe/SBE.Spring.2015.AC.Meeting.Presentations/Bollen.Report.on.Replicability.Subcommittee.May.2015.pdf>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., et al. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*. <http://dx.doi.org/10.1038/s41586-020-2314-9>
- Bouwmeester, S., Verkoeijen, P. P., Aczel, B., Barbosa, F., Bègue, L., Brañas-Garza, P., et al. (2017). Registered replication report: Rand, Greene, and Nowak (2012). *Perspectives on Psychological Science*, 12, 527–542. <http://dx.doi.org/10.1177/1745691617693624>
- Brown, N. J., & Heathers, J. A. (2017). The GRIM test: A simple technique detects numerous anomalies in the reporting of results in psychology. *Social Psychological and Personality Science*, 8, 363–369. <http://dx.doi.org/10.1177/1948550616673876>
- Bryan, C. J., Yeager, D. S., & O'Brien, J. M. (2019). Replicator degrees of freedom allow publication of misleading failures to replicate. *Proceedings of the National Academy of Sciences*, 116, 25535–25545. <http://dx.doi.org/10.1073/pnas.1910951116>
- Budzicz, Ł. (2015). Post-Stapelian psychology. Discussions on the reliability of data and publications in psychology. *Roczniki Psychologiczne*, 18, 25–40.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, 365–376. <http://dx.doi.org/10.1038/nrn3475>
- Cairo, A. H., Green, J. D., Forsyth, D. R., Behler, A. M. C., & Raldiris, T. L. (2020). Gray (literature) matters: Evidence of selective hypothesis reporting in social psychological research. *Personality and Social Psychology Bulletin*. <http://dx.doi.org/10.1177/0146167220903896>
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351, 1433–1436. <http://dx.doi.org/10.1126/science.aaf0918>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2, 637–644. <http://dx.doi.org/10.1038/s41562-018-0399-z>
- Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A., et al. (2002). Data withholding in academic genetics: Evidence from a national survey. *Journal of the American Medical Association*, 287, 473–480. <http://dx.doi.org/10.1001/jama.287.4.473>
- Campbell, H. A., Micheli-Campbell, M. A., & Udyawer, V. (2019). Early career researchers embrace data sharing. *Trends in Ecology & Evolution*, 34, 95–98. <http://dx.doi.org/10.1016/j.tree.2018.11.010>
- Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, 49, 609–610. <http://dx.doi.org/10.1016/j.cortex.2012.12.016>
- Chang, A. C., & Li, P. (2018). Is economics research replicable? Sixty published papers from thirteen journals say “often not”. *Critical Finance Review*, 7. <http://dx.doi.org/10.1561/104.00000053>
- Chopik, W. J., Bremner, R. H., Defever, A. M., & Keller, V. N. (2018). How (and whether) to teach undergraduates about the replication crisis in psychological science. *Teaching of Psychology*, 45, 158–163. <http://dx.doi.org/10.1177/0098628318762900>
- Claesen, A., Gomes, S. L. B. T., & Tuerlinckx, F. (2019). *Preregistration: Comparing dream to reality*. <http://dx.doi.org/10.31234/osf.io/d8wex>

- Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology*, 65, 145–153. <http://dx.doi.org/10.1037/h0045186>
- Colombo, M., Duev, G., Nuijten, M. B., & Sprenger, J. (2018). Statistical reporting inconsistencies in experimental philosophy. *PLoS ONE*, 13. <http://dx.doi.org/10.1371/journal.pone.0194360>
- Condon, D. M., Graham, E., & Mroczek, D. K. (2017). On replication research. In *The Wiley-Blackwell encyclopedia of personality and individual differences: Vol II. Research methods and assessment techniques*. Hoboken, NJ: John Wiley & Sons.
- Cooper, H., DeNeve, K., & Charlton, K. (1997). Finding the missing science: The fate of studies submitted for review by a human subjects committee. *Psychological Methods*, 2, 447–452. <http://dx.doi.org/10.1037/1082-989X.2.4.447>
- Cooper, M. L. (2016). Editorial. *Journal of Personality and Social Psychology*, 110, 431–434. <http://dx.doi.org/10.1037/pspp0000033>
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99. <http://dx.doi.org/10.1016/j.jesp.2015.10.002>
- Cursan, A. (2018). Un chercheur sachant chercher : De l'importance scientifique des résultats « nuls » et négatifs en psychologie. *Pratiques Psychologiques*, 24, 309–324. <http://dx.doi.org/10.1016/j.prps.2018.03.001>
- De Boeck, P., & Jeon, M. (2018). Perceived crisis and reforms: Issues, explanations, and remedies. *Psychological Bulletin*, 144, 757–777. <http://dx.doi.org/10.1037/bul0000154>
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE*, 7. <http://dx.doi.org/10.1371/journal.pone.0029081>
- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6. <http://dx.doi.org/10.3389/fpsyg.2015.00621>
- Eberlen, J. C., Nicaise, E., Leveaux, S., Mora, Y. L., & Klein, O. (2019). Psychometrics anonymous: Does a transparent data sharing policy affect data collection? *Psychologica Belgica*, 59, 373–392. <http://dx.doi.org/10.5334/pb.503>
- Ebersole, K. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., et al. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68–82. <http://dx.doi.org/10.1016/j.jesp.2015.10.012>
- Elliot, A. J., & Niesta, D. (2008). Romantic red: Red enhances men's attraction to women. *Journal of Personality and Social Psychology*, 95, 1150–1164. <http://dx.doi.org/10.1037/0022-3514.95.5.1150>
- Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., et al. (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General*, 139, 399–417. <http://dx.doi.org/10.1037/a0019689>
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, 4. <http://dx.doi.org/10.1371/journal.pone.0005738>
- Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, 5. <http://dx.doi.org/10.1371/journal.pone.0010068>
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891–904. <http://dx.doi.org/10.1007/s11192-011-0494-7>
- (2019). *Feuille de route du CNRS pour la science ouverte*. https://www.cnrs.fr/sites/default/files/press_info/2019-11/Plaqueette_ScienceOuverte.pdf
- Fidler, F., & Wilcox, J. (2018). Reproducibility of scientific results. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Winter 2018 ed.)*. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/win2018/entries/scientific-reproducibility/>
- Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science*, 7, 45–52. <http://dx.doi.org/10.1177/1948550615612150>
- Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology*, 108, 275–297. <http://dx.doi.org/10.1037/pspi0000007>
- Forstmeier, W., Wagenmakers, E. J., & Parker, T. H. (2017). Detecting and avoiding likely false-positive findings—a practical guide. *Biological Reviews*, 92, 1941–1968. <http://dx.doi.org/10.1111/brv.12315>
- Francis, G. (2012). Publication bias and the failure of replication in experimental psychology. *Psychonomic Bulletin & Review*, 19, 975–991. <http://dx.doi.org/10.3758/s13423-012-0322-y>
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345, 1502–1505. <http://dx.doi.org/10.1126/science.1255484>
- Frankenhuis, W. E., & Nettle, D. (2018). Open science is liberating and can foster creativity. *Perspectives on Psychological Science*, 13, 439–447. <http://dx.doi.org/10.1177/1745691618767878>
- Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PLoS ONE*, 13. <http://dx.doi.org/10.1371/journal.pone.0200303>
- Freedman, L. P., Cockburn, I. M., & Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biology*, 13. <http://dx.doi.org/10.1371/journal.pbio.1002165>
- Friese, M., & Frankenbach, J. (2020). p-Hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*. Advance online publication. *Psychological Methods*, 25, 456–471. <http://dx.doi.org/10.1037/met0000246>
- Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, 24, 1–2. <http://dx.doi.org/10.1016/j.tics.2019.10.007>
- García-Berthou, E., & Alcaraz, C. (2004). Incongruence between test statistics and P values in medical papers. *BMC Medical Research Methodology*, 4. <http://dx.doi.org/10.1186/1471-2288-4-13>
- Gardner, W., Lidz, C. W., & Hartwig, K. C. (2005). Authors' reports about research integrity problems in clinical trials. *Contemporary Clinical Trials*, 26, 244–251. <http://dx.doi.org/10.1016/j.cct.2004.11.013>
- Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspectives on Psychological Science*, 9, 641–651. <http://dx.doi.org/10.1177/1745691614551642>
- Gelman, A., & Loken, E. (2014). The statistical crisis in science. *American Scientist*, 102, 460–465. <http://dx.doi.org/10.1511/2014.111.460>
- Gernsbacher, M. A. (2018). Rewarding research transparency. *Trends in Cognitive Sciences*, 22, 953–956. <http://dx.doi.org/10.1016/j.tics.2018.07.002>

- Gilbert, D. T., King, G., Pettigrew, S., & Wilson, T. D. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science*, 351, 1037. <http://dx.doi.org/10.1126/science.aad7243>
- Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science*, 7, 562–571. <http://dx.doi.org/10.1177/1745691612457576>
- Giofrè, D., Cumming, G., Fresc, L., Boedker, I., & Tressoldi, P. (2017). The influence of journal submission guidelines on authors' reporting of statistics and use of open research practices. *PLoS One*, 12. <http://dx.doi.org/10.1371/journal.pone.0175583>
- Goldacre, B., Morton, C. E., & DeVito, N. J. (2019). Why researchers should share their analytic code. *BMJ*, 16365. <http://dx.doi.org/10.1136/bmj.l6365>
- Grahe, J. (2017). Another step towards scientific transparency: Requiring research materials for publication. *The Journal of Social Psychology*, 158, 1–6. <http://dx.doi.org/10.1080/00224545.2018.1416272>
- Grahe, J. E., Brandt, M. J., Ijzerman, H., Cohoon, J., Peng, C., Detweiler-Bedell, B., et al. (2017). *Collaborative replications and education project (CREP)*. <http://dx.doi.org/10.17605/OSF.IO/WFC6U>
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, 82, 1–20. <http://dx.doi.org/10.1037/h0076157>
- Hardwicke, T. E., & Ioannidis, J. P. (2018). Populating the Data Ark: An attempt to retrieve, preserve, and liberate data from the most highly-cited psychology and psychiatry articles. *PLoS ONE*, 13. <http://dx.doi.org/10.1371/journal.pone.0201856>
- Hardwicke, T. E., Thibault, R. T., Kosie, J., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. (2020). *Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017)*. <http://dx.doi.org/10.31222/osf.io/9sz2y>
- Harris, C. R., Coburn, N., Rohrer, D., & Pashler, H. (2013). Two failures to replicate high-performance-goal priming effects. *PLoS ONE*, 8. <http://dx.doi.org/10.1371/journal.pone.0072467>
- Hawkins, R. X., Smith, E. N., Au, C., Arias, J. M., Catapano, R., Hermann, E., et al. (2018). Improving the replicability of psychological science through pedagogy. *Advances in Methods and Practices in Psychological Science*, 1, 7–18. <http://dx.doi.org/10.1177/2515245917740427>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology*, 13. <http://dx.doi.org/10.1371/journal.pbio.1002106>
- Heathers, J. A., Anaya, J., van der Zee, T., & Brown, N. J. (2018). *Recovering data from summary statistics: Sample Parameter Reconstruction via Iterative Techniques (SPRITE)*. *PeerJ Preprints* 6. <http://dx.doi.org/10.7287/peerj.preprints.26968v1>
- Higginson, A. D., & Munafò, M. R. (2016). Current incentives for scientists lead to underpowered studies with erroneous conclusions. *PLoS Biology*, 14. <http://dx.doi.org/10.1371/journal.pbio.2000995>
- Houtkoop, B. L., Chambers, C., Macleod, M., Bishop, D. V., Nichols, T. E., & Wagenmakers, E. J. (2018). Data sharing in psychology: A survey on barriers and preconditions. *Advances in Methods and Practices in Psychological Science*, 1, 70–85. <http://dx.doi.org/10.1177/2515245917751886>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *Plos Medicine*, 2, 696–701. <http://dx.doi.org/10.1371/journal.pmed.0020124>
- Janke, S., Daumiller, M., & Rudert, S. C. (2019). Dark pathways to achievement in science: Researchers' achievement goals predict engagement in questionable research practices. *Social Psychological and Personality Science*, 10, 783–791. <http://dx.doi.org/10.1177/1948550618790227>
- 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. <http://dx.doi.org/10.1177/0956797611430953>
- Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. T., & Duarte, J. L. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. *Journal of Experimental Social Psychology*, 66, 116–133. <http://dx.doi.org/10.1016/j.jesp.2015.10.003>
- Kaiser, J. (2018). *Plan to replicate 50 high-impact cancer papers shrinks to just 18*. <https://www.sciencemag.org/news/2018/07/plan-replicate-50-high-impact-cancer-papers-shrinks-just-18>
- Kaplan, R. M., & Irvin, V. L. (2015). Likelihood of null effects of large NHLBI clinical trials has increased over time. *PLoS ONE*, 10. <http://dx.doi.org/10.1371/journal.pone.0132382>
- Kenett, R. S., & Shmueli, G. (2015). Clarifying the terminology that describes scientific reproducibility. *Nature Methods*, 12, 699. <http://dx.doi.org/10.1038/nmeth.3489>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196–217. http://dx.doi.org/10.1207/s15327957pspr0203_4
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., et al. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14. <http://dx.doi.org/10.1371/journal.pbio.1002456>
- Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Hofelich Mohr, A., et al. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, 4, 20. <http://dx.doi.org/10.1525/collabra.158>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Jr., Bahník, Š., Bernstein, M. J., et al. (2014). Investigating variation in replicability: A “Many Labs” replication project. *Social Psychology*, 45, 142–152. <http://dx.doi.org/10.1027/1864-9335/a000178>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Jr., Alper, S., et al. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1, 443–490. <http://dx.doi.org/10.1177/2515245918810225>
- Krishna, A., & Peter, S. M. (2018). Questionable research practices in student final theses—Prevalence, attitudes, and the role of the supervisor's perceived attitudes. *PLoS ONE*, 13. <http://dx.doi.org/10.1371/journal.pone.0203470>
- Kruglanski, A. W., Chernikova, M., & Jasko, K. (2017). Social psychology circa 2016: A field on steroids. *European Journal of Social Psychology*, 47, 1–10. <http://dx.doi.org/10.1002/ejsp.2285>
- Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25, 178–206. <http://dx.doi.org/10.3758/s13423-016-1221-4>
- Lakens, D. (2017). Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8, 355–362. <http://dx.doi.org/10.1177/1948550617697177>

- Lakens, D. (2019). The value of preregistration for psychological science: A conceptual analysis. *Japanese Psychological Review*, 62, 221–230. <http://dx.doi.org/10.31234/osf.io/jbh4w>
- Lakens, D., Adolfs, F. G., Albers, C. J., Anvari, F., Apps, M. A., Argamon, S. E., et al. (2018). Justify your alpha. *Nature Human Behaviour*, 2, 168–171. <http://dx.doi.org/10.1038/s41562-018-0311-x>
- Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science*, 8, 875–881. <http://dx.doi.org/10.1177/1948550617693058>
- LeBel, E. P., Berger, D., Campbell, L., & Loving, T. J. (2017). Falsifiability is not optional. *Journal of Personality and Social Psychology*, 113, 254–261. <http://dx.doi.org/10.1037/pspi0000106>
- Ledgerwood, A. (2018). The preregistration revolution needs to distinguish between predictions and analyses. *Proceedings of the National Academy of Sciences*, 115, E10516–E10517. <http://dx.doi.org/10.1073/pnas.1812592115>
- Lehmann, G. K., & Calin-Jageman, R. J. (2017). Is red really romantic? *Social Psychology*, 48, 174–183. <http://dx.doi.org/10.1027/1864-9335/a000296>
- Lehmann, G. K., Elliot, A. J., & Calin-Jageman, R. J. (2018). Meta-analysis of the effect of red on perceived attractiveness. *Evolutionary Psychology*, 16. <http://dx.doi.org/10.1177/1474704918802412>
- Levelt, W. J., Drenth, P. J. D., & Noort, E. (2012). *Flawed science: The fraudulent research practices of social psychologist Diederik Stapel*. <https://poolux.psychopool.tu-dresden.de/mdcfiles/gwp/Reale%20FC3%A4lle/Stapel%20-%20Final%20Report.pdf>
- List, J. A., Bailey, C. D., Euzent, P. J., & Martin, T. L. (2001). Academic economists behaving badly? A survey on three areas of unethical behavior. *Economic Inquiry*, 39, 162–170. <http://dx.doi.org/10.1111/j.1465-7295.2001.tb00058.x>
- Loken, E., & Gelman, A. (2017). Measurement error and the replication crisis. *Science*, 355, 584–585. <http://dx.doi.org/10.1126/science.aal3618>
- Machery, E. (2020). What is a replication? *Philosophy of Science*, 87, 545–567. <http://dx.doi.org/10.1086/709701>
- Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. (2019). *Questionable and open research practices in education research*. <http://dx.doi.org/10.35542/osf.io/f7srb>
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537–542. <http://dx.doi.org/10.1177/1745691612460688>
- Martinson, B. C., Anderson, M. S., & de Vries, R. (2005). Scientists behaving badly. *Nature*, 435, 737–738. <http://dx.doi.org/10.1038/435737a>
- Maxwell, S. E. (2004). The persistence of underpowered studies in psychological research: Causes, consequences, and remedies. *Psychological Methods*, 9, 147–163. <http://dx.doi.org/10.1037/1082-989X.9.2.147>
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, 70, 487–498. <http://dx.doi.org/10.1037/a0039400>
- Meyer, M. N. (2018). Practical tips for ethical data sharing. *Advances in Methods and Practices in Psychological Science*, 1, 131–144. <http://dx.doi.org/10.1177/2515245917747656>
- Mills, J. L. (1993). Data torturing. *New England Journal of Medicine*, 329, 1196–1199. <http://dx.doi.org/10.1056/NEJM199310143291613>
- Morawski, J. (2020). Psychologists’ psychologies of psychologists in a time of crisis. *History of Psychology*, 23, 176–198. <http://dx.doi.org/10.1037/hop0000140>
- Morey, R. D., Chambers, C. D., Etschells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., et al. (2016). The Peer Reviewers’ Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, 3, 150547. <http://dx.doi.org/10.1098/rsos.150547>
- Moshontz, H., Campbell, L., Ebersole, C. R., IJzerman, H., Urry, H. L., Forscher, P. S., et al. (2018). The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network. *Advances in Methods and Practices in Psychological Science*, 1, 501–515. <http://dx.doi.org/10.1177/2515245918797607>
- Motyl, M., Demos, A. P., Carsel, T. S., Hanson, B. E., Melton, Z. J., Mueller, A. B., et al. (2017). The state of social and personality science: Rotten to the core, not so bad, getting better, or getting worse? *Journal of Personality and Social Psychology*, 113, 34–56. <http://dx.doi.org/10.1037/pspa0000084>
- Murphy, J. R. (2004). Statistical errors in immunologic research. *Journal of Allergy and Clinical Immunology*, 114, 1259–1263. <http://dx.doi.org/10.1016/j.jaci.2004.09.023>
- Nelson, L. D., Simmons, J., & Simonsohn, U. (2018). Psychology’s renaissance. *Annual Review of Psychology*, 69, 511–534. <http://dx.doi.org/10.1146/annurev-psych-122216-011836>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science*, 348, 1422–1425. <http://dx.doi.org/10.1126/science.aab2374>
- Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., et al. (2019). Preregistration is hard, and worthwhile. *Trends in Cognitive Sciences*, 23, 815–818. <http://dx.doi.org/10.1016/j.tics.2019.07.009>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115, 2600–2606. <http://dx.doi.org/10.1073/pnas.1708274114>
- Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, 18, e3000691. <http://dx.doi.org/10.1371/journal.pbio.3000691>
- Nosek, B. A., & Lakens, D. (2014). Registered Reports. *Social Psychology*, 45, 137–141. <http://dx.doi.org/10.1027/1864-9335/a000192>
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, 7, 615–631. <http://dx.doi.org/10.1177/1745691612459058>
- Nuijten, M. B., Hartgerink, C. H., van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, 48, 1205–1226. <http://dx.doi.org/10.3758/s13428-015-0664-2>
- Nuijten, M. B., van Assen, M. A. L. M., Hartgerink, C. H. J., Epskamp, S., & Wicherts, J. M. (2017). The validity of the tool “statcheck” in discovering statistical reporting inconsistencies. <http://dx.doi.org/10.31234/osf.io/tcxaj>
- O’Boyle, E. H., Jr., Banks, G. C., & Gonzalez-Mulé, E. (2017). The chrysalis effect: How ugly initial results metamorphose into beautiful articles. *Journal of Management*, 43, 376–399. <http://dx.doi.org/10.1177/0149206314527133>

- O'Donnell, M., Nelson, L. D., Ackermann, E., Aczel, B., Akhtar, A., Aldrovandi, S., et al. (2018). Registered replication report: Dijksterhuis and van Knippenberg (1998). *Perspectives on Psychological Science*, 13, 268–294. <http://dx.doi.org/10.1177/1745691618755704>
- Obels, P., Lakens, D., Coles, N. A., Gottfried, J., & Green, S. A. (2019). Analysis of open data and computational reproducibility in registered reports in psychology. *Advances in Methods and Practices in Psychological Science*, 3, 229–237. <http://dx.doi.org/10.1177/2515245920918872>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349. <http://dx.doi.org/10.1126/science.aac4716>
- Parker, T. H., Griffith, S. C., Bronstein, J. L., Fidler, F., Foster, S., Fraser, H., et al. (2018). Empowering peer reviewers with a checklist to improve transparency. *Nature Ecology & Evolution*, 2, 929–935. <http://dx.doi.org/10.1038/s41559-018-0545-z>
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528–530. <http://dx.doi.org/10.1177/1745691612465253>
- Peperkoorn, L. S., Roberts, S. C., & Pollet, T. V. (2016). Revisiting the red effect on attractiveness and sexual receptivity: No effect of the color red on human mate preferences. *Evolutionary Psychology*, 14. <http://dx.doi.org/10.1177/1474704916673841>
- (2018). *Plan national pour la science ouverte*. <http://m.enseignementsup-recherche.gouv.fr/cid/132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html>
- Plesser, H. E. (2018). Reproducibility vs. replicability: A brief history of a confused terminology. *Frontiers in Neuroinformatics*, 11. <http://dx.doi.org/10.3389/fninf.2017.00076>
- Pridemore, W. A., Makel, M. C., & Plucker, J. A. (2018). Replication in criminology and the social sciences. *Annual Review of Criminology*, 1, 19–38. <http://dx.doi.org/10.1146/annurev-criminol-032317-091849>
- Quintana, D. S., & Williams, D. R. (2018). Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC psychiatry*, 18, 178. <http://dx.doi.org/10.1186/s12888-018-1761-4>
- Rabelo, A. L., Farias, J. E., Sarmet, M. M., Joaquim, T. C., Hoersting, R. C., Victorino, L., et al. (2020). Questionable research practices among Brazilian psychological researchers: Results from a replication study and an international comparison. *International Journal of Psychology*, 55, 674–683. <http://dx.doi.org/10.1002/ijop.12632>
- Reidpath, D. D., & Allotey, P. A. (2001). Data sharing in medical research: An empirical investigation. *Bioethics*, 15, 125–134. <http://dx.doi.org/10.1111/1467-8519.00220>
- Richard, F. D., Bond, C. F., Jr., & Stokes-Zoota, J. J. (2003). One hundred years of social psychology quantitatively described. *Review of General Psychology*, 7, 331–363. <http://dx.doi.org/10.1037/1089-2680.7.4.331>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638–641. <http://dx.doi.org/10.1037/0033-2909.86.3.638>
- Rouder, J. N., Haaf, J. M., & Snyder, H. K. (2019). Minimizing mistakes in psychological science. *Advances in Methods and Practices in Psychological Science*, 2, 3–11. <http://dx.doi.org/10.1177/2515245918801915>
- Rubin, M. (2019). The costs of harking. *The British Journal for the Philosophy of Science*, axz050. <http://dx.doi.org/10.1093/bjps/axz050>
- Sarafoglou, A., Hoogveen, S., Matzke, D., & Wagenmakers, E. J. (2020). Teaching good research practices: Protocol of a research master course. *Psychology Learning & Teaching*, 19, 46–59. <http://dx.doi.org/10.1177/1475725719858807>
- Sassenberg, K., & Ditrich, L. (2019). Research in social psychology changed between 2011 and 2016: Larger sample sizes, more self-report measures, and more online studies. *Advances in Methods and Practices in Psychological Science*, 2, 107–114. <http://dx.doi.org/10.1177/2515245919838781>
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: Differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*, 10, 813. <http://dx.doi.org/10.3389/fpsyg.2019.00813>
- Scheel, A. M., Schijen, M., & Lakens, D. (2020). *An excess of positive results: Comparing the standard psychology literature with registered reports*. <https://psyarxiv.com/p6e9c/download?format=pdf>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtry, E., et al. (2018). Many analysts, one data set: Making transparent how variations in analytic choices affect results. *Advances in Methods and Practices in Psychological Science*, 1, 337–356. <http://dx.doi.org/10.1177/2515245917747646>
- 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, 1359–1366. <http://dx.doi.org/10.1177/0956797611417632>
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, 9, 76–80. <http://dx.doi.org/10.1177/1745691613514755>
- Simons, D. J., Holcombe, A. O., & Spellman, B. A. (2014). An introduction to registered replication reports at perspectives on psychological science. *Perspectives on Psychological Science*, 9, 552–555. <http://dx.doi.org/10.1177/1745691614543974>
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26, 559–569. <http://dx.doi.org/10.1177/0956797614567341>
- Smith, N. C. (1970). Replication studies: A neglected aspect of psychological research. *American Psychologist*, 25, 970–975. <http://dx.doi.org/10.1037/h0029774>
- Soderberg, C. K. (2018). Using OSF to share data: A step-by-step guide. *Advances in Methods and Practices in Psychological Science*, 1, 115–120. <http://dx.doi.org/10.1177/2515245918757689>
- Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, 30, 711–727. <http://dx.doi.org/10.1177/0956797619831612>
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*, 10, 886–899. <http://dx.doi.org/10.1177/1745691615609918>
- Spellman, B. A., Gilbert, E. A., & Corker, K. S. (2018). Open science: What, why, and how. In J. Wixted, & E.-J. Wagenmakers (Eds.), *Stevens' handbook of experimental psychology and cognitive neuroscience* (5) (4th ed., 5, pp. 729–776). New York, NY: Wiley. <http://dx.doi.org/10.1002/9781119170174.epcn519>
- Stanley, D. J., & Spence, J. R. (2014). Expectations for replications: Are yours realistic? *Perspectives on Psychological Science*, 9, 305–318. <http://dx.doi.org/10.1177/1745691614528518>

- Stanley, T. D., Carter, E. C., & Doucouliagos, H. (2018). What meta-analyses reveal about the replicability of psychological research. *Psychological Bulletin*, 144, 1325–1346. <http://dx.doi.org/10.1037/bul0000169>
- Stricker, J., & Günther, A. (2019). Scientific misconduct in psychology: A systematic review of prevalence estimates and new empirical data. *Zeitschrift für Psychologie*, 227, 53–63. <http://dx.doi.org/10.1027/2151-2604/a000356>
- Swazey, J. P., Anderson, M. S., Lewis, K. S., & Louis, K. S. (1993). Ethical problems in academic research. *American Scientist*, 81, 542–553. <https://www.jstor.org/stable/29775057>
- Swiatkowski, W., & Dompnier, B. (2017). Replicability crisis in social psychology: Looking at the past to find new pathways for the future. *International Review of Social Psychology*, 30, 111–124. <http://dx.doi.org/10.5334/irsp.66>
- Szollosi, A., Kellen, D., Navarro, D. J., Shiffrrin, R., van Rooij, I., Van Zandt, T., et al. (2019). Is preregistration worthwhile. *Trends in Cognitive Sciences*, 24, 94–95. <http://dx.doi.org/10.1016/j.tics.2019.11.009>
- (2020). Tell it like it is. *Nature Human Behaviour*, 4, 1. <http://dx.doi.org/10.1038/s41562-020-0818-9>
- Tiokhin, L., Panchanathan, K., Lakens, D., Vazire, S., Morgan, T., & Zollman, K. (2019). *Honest signaling in academic publishing*. <http://dx.doi.org/10.31219/osf.io/gyeh8>
- Trafimow, D., Amrhein, V., Areshenkoff, C. N., Barrera-Causil, C. J., Beh, E. J., Bilgic, Y. K., et al. (2018). Manipulating the alpha level cannot cure significance testing. *Frontiers in Psychology*, 9. <http://dx.doi.org/10.3389/fpsyg.2018.00699>
- Uhlmann, E. L., Ebersole, C. R., Chartier, C. R., Errington, T. M., Kidwell, M. C., Lai, C. K., et al. (2020). Scientific Utopia III: Crowdsourcing science. Perspectives on Psychological Science. Advance online publication. *Perspectives on Psychological Science*, 14, 711–733. <http://dx.doi.org/10.1177/1745691619850561>
- van't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology*, 67, 2–12. <http://dx.doi.org/10.1016/j.jesp.2016.03.004>
- Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, 13, 411–417. <http://dx.doi.org/10.1177/1745691617751884>
- Veldkamp, C. L. S., Bakker, M., van Assen, M. A. L. M., Cromptvoets, E. A. V., Ong, H. H., Nosek, B. A., et al. (2018). *Ensuring the quality and specificity of preregistrations*. <http://dx.doi.org/10.31234/osf.io/cdgyh>
- Wagenmakers, E. J., Beek, T., Dijkhoff, L., Gronau, Q. F., Acosta, A., Adams, R. B., Jr., et al. (2016). Registered replication report: Strack, Martin, & Stepper (1988). *Perspectives on Psychological Science*, 11, 917–928. <http://dx.doi.org/10.1177/1745691616674458>
- Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., et al. (2018). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin & Review*, 25, 58–76. <http://dx.doi.org/10.3758/s13423-017-1323-7>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. (2011). Why psychologists must change the way they analyze their data: the case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100, 426–432. <http://dx.doi.org/10.1037/a0022790>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 632–638. <http://dx.doi.org/10.1177/1745691612463078>
- Wagge, J. R., Baciú, C., Banas, K., Nadler, J. T., Schwarz, S., Weisberg, Y., et al. (2019). A demonstration of the collaborative replication and education project: Replication attempts of the red-romance effect. *Collabra: Psychology*, 5. <http://dx.doi.org/10.1525/collabra.177>
- Warren, M. (2018). First analysis of pre-registered studies shows sharp rise in null findings. *Nature*, <http://dx.doi.org/10.1038/d41586-018-07118-1> (d41586-018-07118-1)
- Washburn, A. N., Hanson, B. E., Motyl, M., Skitka, L. J., Yantis, C., Wong, K. M., et al. (2018). Why do some psychology researchers resist adopting proposed reforms to research practices? A description of researchers' rationales. *Advances in Methods and Practices in Psychological Science*, 1, 166–173. <http://dx.doi.org/10.1177/2515245918757427>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “p < 0.05”. *The American Statistician*, 73(Suppl. 1), 1–19. <http://dx.doi.org/10.1080/00031305.2019.1583913>
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6. <http://dx.doi.org/10.1371/journal.pone.0026828>
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61, 726–728. <http://dx.doi.org/10.1037/0003-066X.61.7.726>
- Wingen, T., Berkessel, J. B., & Englich, B. (2020). No replication, no trust? How low replicability influences trust in psychology. *Social Psychological and Personality Science*, 11, 454–463. <http://dx.doi.org/10.1177/1948550619877412>
- Wiseman, R., Watt, C., & Kornbrod, D. (2019). Registered reports: an early example and analysis. *PeerJ*, 7, e6232. <http://dx.doi.org/10.7717/peerj.6232>
- Yamada, Y. (2018). How to crack pre-registration: Toward transparent and open science. *Frontiers in Psychology*, 9. <http://dx.doi.org/10.3389/fpsyg.2018.01831>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, 41, e120. <http://dx.doi.org/10.1017/S0140525X17001972>
- Zwaan, R. A., Pecher, D., Paoilacci, G., Bouwmeester, S., Verkoeijen, P., Dijkstra, K., et al. (2018). Participant nonnaïveté and the reproducibility of cognitive psychology. *Psychonomic Bulletin & Review*, <http://dx.doi.org/10.3758/s13423-017-1348-y> (1968–1972)