

**POT LA PSICOLOGIA RESCATAR-SE A SI MATEIXA?  
INCENTIUS, BIAIX I REPLICABILITAT**  
**¿PUEDE LA PSICOLOGÍA RESCATARSE A SÍ MISMA?  
INCENTIVOS, SEGOS Y REPLICABILIDAD**  
**CAN PSYCHOLOGY RESCUE ITSELF?  
INCENTIVES, BIASES, AND REPRODUCIBILITY**

*Fernando Blanco\* \*\*, José C. Perales\* \*\*\*,  
Miguel A. Vadillo\* \*\*\*\**

---

DOI: 10.7203/anuari.psicologia.18.2.231

---

**Resum**

En els últims anys la Psicologia està patint internament i externa una crisi important de credibilitat, a la qual tampoc han estat alienes altres ciències, com ara la Medicina o la Biologia. Diversos projectes de ciència de col·laboració suggereixen que una gran part dels resultats de la investigació són difícils de reproduir. Això ve acompanyat d'un conjunt de simulacions i anàlisis quantitatives que situen la quantitat de falsos positius per damunt del 50% del total de dades publicades en la literatura psicològica actual.

En aquest breu treball realitzarem una anàlisi actualitzada de la situació i intentarem identificar les causes psicològiques que hi han contribuït. Entre aquestes destaquen, primer, un sistema d'incentius alineat més amb els interessos de promoció de l'investigador que no amb el descobriment i l'extensió d'una ciència transparent, fiable i reproduïble, i, segon, els biaixos individu-

\* Sociedad para el Estudio de los Juicios y las Decisiones (SEJyD), @Sociedad EJyD.

\*\* Departamento de Fundamentos y Métodos de la Psicología, Facultad de Psicología y Educación, Universidad de Deusto. Correspondència: <fernandoblanc@deusto.es>.

\*\*\* Departamento de Psicología Experimental, Centro de Investigación Mente, Cerebro y Comportamiento (CIMCYC), Universidad de Granada. Correspondència: <jcesar@ugr.es>.

\*\*\*\* Departamento de Psicología Básica, Universidad Autónoma de Madrid. Correspondència: <miguel.vadillo@uam.es>.

als en l'elaboració de judicis i en la presa de decisions que afecten de forma generalitzada els éssers humans –investigadors o no–.

Acabarem sotmetent a discussió com aquests biaixos, en concordança amb la investigació clàssica sobre biaixos i heurístics, s'acumulen a través dels individus i acaben afectant de forma substancial el projecte col·lectiu de la ciència psicològica. Enumerem també les estratègies que podrien contribuir, si es posaren en pràctica amb èxit, a corregir –en part– la situació i fins i tot estendre's a altres àrees científiques.

*Paraules clau: replicabilitat, biaixos, mètodes en Psicologia, incentius, ciència oberta.*

### **Resumen**

*En los últimos años la Psicología está sufriendo, interna y externamente, una importante crisis de credibilidad, a la que tampoco han sido ajenas otras ciencias como la Medicina o la Biología. Varios proyectos de ciencia colaborativa sugieren que gran parte de los resultados de la investigación son difíciles de reproducir. Esto viene acompañado de un conjunto de simulaciones y análisis cuantitativos que sitúan la cantidad de falsos positivos por encima del 50% del total de datos publicados, en la literatura psicológica actual.*

*En este breve trabajo realizaremos un análisis actualizado de esta situación, e intentaremos identificar las causas psicológicas que han contribuido a la misma. Entre éstas destacan, primero, un sistema de incentivos alineado más con los intereses de promoción del investigador que con el descubrimiento y diseminación de una ciencia transparente, fiable y reproducible y, segundo, los sesgos individuales en la elaboración de juicios y toma de decisiones, que afectan de forma generalizada a los seres humanos –investigadores o no–.*

*Terminamos discutiendo cómo estos sesgos, en concordancia con la investigación clásica sobre sesgos y heurísticos, se acumulan a través de individuos y acaban afectando de forma sustancial a la empresa colectiva de la ciencia psicológica. Enumeramos también las estrategias que podrían contribuir, de implementarse con éxito, a corregir en parte la situación, e incluso extenderse a otras áreas científicas.*

Palabras clave: replicabilidad, sesgos, métodos en Psicología, incentivos, ciencia abierta.

### **Abstract**

*In the last years, the credibility of psychological science –as well as the one of other disciplines as biology or medicine– has been called into question. Results from a number of collaborative science projects suggest that a large portion of results in published psychological research is not reproducible. In parallel, simulation and quantitative inference methods estimate the*

*proportion of false positives in current psychological literature to be well above 50%.*

*In this essay an up-to-date analysis of the current situation and its psychological underpinnings is showed. Among these psychological factors, the importance of (1) a career promotion-centred incentive system that disregards scientific transparency and dissemination of reliable and reproducible research are highlighted, but strongly motivates researchers to strive for publication, and (2) individual cognitive biases and distortions –common to laypeople and scientists– that affect research-related judgment and decision making.*

*The paper concludes discussing how these cognitive distortions, in line with available evidence on heuristics and biases, accumulate across researchers and end up distorting evidence accumulation, and thus making the enterprise of collective science derail. Awareness of these consequences and their roots should help implement strategies designed to correct the current distortions and even make psychological science a role model for other disciplines experiencing similar problems.*

Key words: reproducibility, biases, Psychology methods, incentives, pen science.

## **Introducció**

En un article de la revista *Nature*, sir Francis Galton (1907a) relata un famós estudi de camp en el qual va recollir les estimacions que van fer sense cap instrument de mesura 787 persones sobre el pes d'un bou. Entre aquestes persones hi havia ramaders i carnisers, però també un nombre considerable de ciutadans del carrer sense experiència en tals qüestions. L'estimació mitjana va ser de 1.207 lliures, una xifra molt propera a les 1.198 lliures que va resultar finalment que pesava el bou.

El mateix Galton (1907b) va utilitzar aquesta propietat dels judicis col·lectius per a defensar els principis de la democràcia, i posteriorment s'ha utilitzat per a defensar la naturalesa acumulativa i autocorrectiva de la ciència. Per exemple, Surowiecki (2004) relata sobre la cerca de la causa de la malaltia respiratòria SARS:

A principis de febrer de 2003 [...], es va encarregar treballar al mateix temps a diversos laboratoris sobre les mateixes mostres i multiplicaren la seua velocitat i la seua eficàcia. Després d'uns dies d'esforç, els laboratoris van considerar i van descartar un cert nombre de causes [...]. El 21 de març de 2003, uns científics de la Universitat de Hong Kong aïllaren un candidat. [...] Durant la setmana següent, els laboratoris de la xarxa van detectar el coronavirus en una àmplia varietat de mostres (p. 159).

En altres paraules, el principi de saviesa de les masses prediu que quan diversos agents independents aporten informació a un conjunt, allò que els agents aporten de cert tendeix a acumular-se, mentre que els errors tendeixen a eliminar-se. Aquesta eliminació de l'error per mitjana opera en molts contextos diferents i explica, per exemple, l'eficiència dels mercats de decisió –als quals tornarem al final d'aquest article–. En l'àmbit que ens incumbeix, aquest mateix principi està a la base de la principal font d'informació necessària per a construir una Psicologia basada en la prova –les metaanàlisis i les revisions sistemàtiques–. Aquests treballs acumulatius es basen en estudis individuals que realitzen investigadors i laboratoris a tot el món (Higgins i Green, 2008; Cumming, 2014).

Tanmateix, sovint s'oblida que la correcta operació d'aquest principi requereix que es complisquen diverses condicions. La primera és que el mostreig dels agents individuals i de la informació que aporten no siga selectiu. La segona, que els errors que cometem cadascun d'aquests actors siguen aleatoris –que no tendisquen a estar sistemàticament desviats en la mateixa direcció–. I la tercera, que els actors i els seus judicis siguen independents entre ells (Lorenz, Rauhut, Schweitzer i Heldbing, 2011). Per desgràcia, cap d'aquestes tres condicions es compleix actualment ni en la Psicologia científica ni en la ciència en general, per la qual cosa podem sospitar que ambdues estan en gran manera fracassant com a empreses acumulatives i autocorrectives (Ioannidis, 2005, 2012). Aquest breu assaig pretén il·lustrar com els factors psicològics i ambientals que afecten de forma concreta les decisions d'investigadors individuals són en última instància els responsables d'aquesta situació. A més, apuntarem alguns elements que podrien contribuir a corregir-la si es posaren en pràctica de forma eficaç. És a dir, pretenem que la Psicologia s'embarque en la missió de descriure, explicar i intervenir en la correcció dels seus propis biaixos i punts cecs.

## **La crisi de la Psicologia científica i les seues causes immediates**

En bona mesura, la credibilitat d'una ciència depèn de la possibilitat de reproduir els efectes que reporta en els seus estudis. Diem replicació metodològica de l'èxit en la reproducció d'un efecte exactament en les mateixes condicions en les quals es va descriure l'efecte original. La replicació conceptual, per contra, suposa la reproducció en condicions diferents de les originals –en una altra població, una altra tasca o un altre context, per exemple–, però en les quals se suposa que opera el mateix procés psicològic subjacent. Un fracàs en la replicació conceptual limita l'abast teòric d'un resultat i, per tant, també la seua possible utilitat pràctica, però no té per què afectar la veracitat del resultat original. Un fracàs en una rèplica metodològica tampoc és inherentment nociu:

pot indicar que el resultat original o el de l'intent de rèplica van ser causats per atzar –una possibilitat que és inherent a l'establiment de llimars de significativitat estadística–. Tanmateix, un fracàs sistemàtic en la replicació metodològica pot ser fatal per a la credibilitat d'una disciplina, mentre que pot indicar una tendència també sistemàtica en els resultats originals a ser falsos.

En els últims anys, diversos intents de reproduir efectes que eren centrals en determinats models teòrics han fracassat. Entre aquests efectes estan, per exemple, el de la posa de poder i altres efectes similars de corporització (Ranehill, Dreber, Johannesson, Leiberg, Sul i Weber, 2015), els de depleció de l'ego (Lurquin i al., 2016), molts dels efectes de facilitació social (Rohrer, Pashler i Harris, 2015; Shanks i al., 2015) i una bona part de les intervencions psicoterapèutiques en diferents contextos d'aplicació (Van der Gucht i al., 2016; Coyne, Thombs i Hagedoorn, 2010).

El problema de la fiabilitat de la literatura psicològica no és nou i ha estat assenyalat en moltes ocasions, de forma més o menys aïllada, al llarg de les últimes dècades (Meehl, 1978). Tanmateix, la comunitat científica en conjunt ha començat a prendre consciència del problema en els últims anys. Per exemple, en 2012, Daniel Kahneman –premi Nobel i probablement el psicòleg viu més influent en l'actualitat– va enviar una carta oberta en la qual advertia de la fragilitat de pràcticament la totalitat de la recerca sobre facilitació social i advertia els joves científics dels problemes per a la seua futura carrera que podia comportar investigar en aquesta àrea (Young, 2012).

Davant aquests problemes, de vegades s'ha argumentat, primer, que afectaven només certs temes de recerca, com la facilitació social i la literatura circumdant en judicis i presa de decisions. I, segon, que sobretot es tractava d'un fracàs en la rèplica conceptual i que, per tant, no necessàriament afectava la veracitat dels descobriments originals. L'intent més seriós, a gran escala, de quantificar els problemes de replicabilitat metodològica en el conjunt de la Psicologia ha estat fruit de la col·laboració entre múltiples investigadors i laboratoris, sota el paraigua de l'*Open Science Collaboration* (2015). En aquest projecte es van intentar cent rèpliques d'estudis extrets aleatòriament de tres de les revistes més prestigioses utilitzant dissenys de gran potència i amb la col·laboració dels autors dels estudis originals en el procediment i l'elaboració de materials. Dels cent estudis originals, noranta-set donaren resultats estadísticament significatius ( $p < 0,05$ ) i, no obstant això, en només un 39% d'ells l'intent de rèplica va aconseguir aquest nivell de significació estadística –és a dir, no es va replicar l'efecte–. I el que és més greu, la grandària mitjana de l'efecte observat en les rèpliques va ser menys de la meitat de l'observat en els estudis originals, i en només un 47% dels casos la grandària observada de l'efecte va estar dins de l'interval de confiança (95% IC) estimat en l'estudi original. Diem que és especialment greu perquè la grandària de l'efecte i no la  $p$ , malgrat el que molts

científics pensen, és el resultat més rellevant d'un estudi i l'únic a partir del qual es pot valorar la rellevància teòrica i pràctica d'aquest efecte i acumular proves a través d'estudis individuals (Cohen, 1990).

A grans trets, aquests resultats coincideixen amb un altres projectes de col·laboració (Klein i al., 2014). Encara que els han interpretat de forma més o menys optimista diferents autors (Baker, 2015; Van Bavel, Mende-Siedlecki, Bradi i Reinero, 2016), sembla difícil negar que una porció molt important dels resultats en Psicologia són falsos positius. La raó immediata d'això resulta evident en les xifres que s'aporten més amunt: un 97% dels efectes originals de la mostra de l'*Open Science Collaboration* es basaven en diferències estadísticament significatives –positius–. Això malgrat el fet que la potència mitjana dels estudis en Psicologia no assoleix el 50% (Bakker, Van Dijk i Wicherts, 2012; Button i al., 2013; Cohen, 1990). Això implica que més del 50% dels estudis en els quals s'intenta demostrar una hipòtesi alternativa veritable no tenen èxit i no ixen mai a la llum. Aquest ocultament dels resultats nuls (Franco, Malhotra i Simonovits, 2014; Sterling, Rosenbaum i Weikram, 1995) de vegades afecta estudis complets. Altres vegades afecta porcions d'estudis en els quals s'utilitzen diverses variables, de les quals només es reconeixen i reporten les que donen lloc a diferències significatives. Finalment, aquest ocultament pot referir-se a estratègies d'anàlisi d'entre un menú disponible, de les quals només es reconeix haver-ne dut a terme una.

En aquest context, interessa saber quants dels resultats reportats són resultats veritables. El percentatge de veritables positius sobre el total d'efectes estadísticament significatius reportats en la literatura d'un determinat tema de recerca rep el nom de valor predictiu positiu (VPP). Segons Ioannidis (2005), és possible estimar el VPP en un camp determinat simulant adequadament les pràctiques de recerca en aquest camp i algunes de les seues característiques inherents. Entre elles, estan (1) la potencia habitual dels estudis –que depèn principalment de la grandària de les mostres i de la grandària de l'efecte que es persegueix–, (2) la novetat i el valor contraintuïtiu dels efectes que es pretenen demostrar, (3) la flexibilitat de possibilitats d'anàlisi i (4) l'excessiva llibertat en la definició i selecció de variables dependents.

Així, els assajos clínics de gran potència ( $1-\beta = 0,80$ ), amb variables de resultat i anàlisi preespecificades i amb hipòtesis ben informades, podrien assolir un VPP de 0,85, mentre que, en l'altre extrem, la gran massa d'anàlisis exploratòries, amb mostres petites ( $1-\beta = 0,25$ ), sobre efectes improbables a priori i anàlisis selectives, estaria prop d'un VPP de 0. Desgraciadament, també és fàcil demostrar que la major part dels treballs que es publiquen en l'actualitat estan més prop del segon extrem que del primer –el lector pot simular fàcilment el que passa en diferents situacions en l'aplicació en línia que han creat Zehetleitner i Schönbrodt, 2016–. El que implica això és que en la major part dels camps

d'estudi les metaanàlisis i les revisions sistemàtiques que haurien de servir per a basar la pràctica en les proves són virtualment irrelevantes, perquè aglutinen dades estadístiques profundament esbiaixades.

## **Causas intermèdies: mecanismes psicològics responsables de les pràctiques de recerca qüestionables**

### *a) Coneixement estadístic*

Encara que la causa més important de la poca fiabilitat de la recerca psicològica és la publicació gairebé exclusiva de resultats positius obtinguts en condicions subòptimes i propenses a la distorsió sistemàtica, les raons subjacents tenen a veure amb els nostres comportaments i decisions com a investigadors. Voluntàriament o incidentalment, empenyem de forma sistemàtica els resultats de la nostra recerca en la direcció que millor serveix als nostres interessos.

Un dels mecanismes que perpetuen aquests problemes és que els mateixos científics manquem de la formació o del criteri necessari per a jutjar la qualitat dels nostres mètodes. Per exemple, com hem assenyalat amb anterioritat, la potència estadística és un factor determinant del valor predictiu positiu de tota recerca. Els estudis que s'han realitzat fins avui suggereixen que la potència estadística de la recerca psicològica és ridículament baixa (Bakker i al., 2012; Button i al., 2013; Cohen, 1990). Aquesta situació no només no sembla millorar amb el pas dels anys, sinó que, de fet, tendeix a empitjorar amb cada nova avaluació. Com ja assenyalaven Sedlmeier i Gigerenzer (1989) fa gairebé tres dècades, els estudis sobre la potència estadística tenen poc o cap efecte sobre la potència estadística dels estudis.

Per què els científics insistim a realitzar estudis sense suficient potència? Com és possible que a un científic li resulte rendible realitzar un estudi que només té un 30% o un 40% de probabilitats de donar resultats significatius, suposant que la hipòtesi que el motiva siga certa? Aparentment, els investigadors no es plantegen la qüestió en aquests termes. Més aviat, tendim a sobreestimar la potència estadística que s'aconsegueix amb les grandàries mostrals i els efectes més habituals en la nostra disciplina (Bakker, Hartgerink, Wicherts i van der Maas, 2016).

La incapacitat per a comprendre la informació estadística essencial no es limita a la potència. El mateix concepte de significació de la hipòtesi nul·la és objecte de tot tipus de malentesos, tal com revela el fet que només una minoria d'investigadors coneix la definició correcta del valor  $p$  (Badenes-Ribera, Frias-Navarro, Iotti, Bonilla-Campos i Longobardi, 2016; Badenés-Ribera, Frias-Navarro i Pascual-Soler, 2015; Schervish, 1996). Tot suggereix que al-

guns aspectes crucials de la recerca, com la selecció de la grandària mostral o del procediment estadístic que cal utilitzar, no es basen en cap procés analític precís, sinó en el simple hàbit i en intuïcions vagues que no sempre es corresponen amb la realitat. En la major part dels casos, és probable que tant aquestes decisions com les interpretacions dels resultats publicats per altres investigadors es basen en heurístics senzills com «incloure almenys 25 participants per condició» o «un resultat estadísticament no significatiu indica que no hi ha diferències entre condicions» (Hoekstra, Finch, Kiers i Johnson, 2006; Nelson, Rosenthal i Rosnow, 1986). Com s'esdevé amb qualsevol altre heurístic, aquestes regles de decisió poden funcionar correctament en moltes ocasions, però la seua utilització mecànica i inconscient pot tenir importants conseqüències negatives (Gigerenzer, 2004).

### ***b) Biaixos individuals en la presa de decisions***

Amb tot, seria un error atribuir íntegrament l'actual crisi de replicabilitat a la simple ignorància estadística. Al cap i a la fi, la major part dels científics hem de bregar en la nostra activitat quotidiana amb conceptes i eines considerablement més complexos que els estadístics. Per contra, és possible imaginar que bona part d'aquests problemes es donen perquè els investigadors no som testimonis neutrals de la prova que recollim, sinó observadors interessats que percebem les nostres pròpies dades a través dels mateixos biaixos cognitius que poden trobar-se en qualsevol ciutadà del carrer. Com qualsevol altra persona, els científics som incapaços d'evitar veure patrons on només hi ha soroll i atzar. Encara que els mètodes de recerca estan dissenyats com a contramesures per a anul·lar l'efecte d'aquests biaixos, aquest objectiu rarament s'aconsegueix.

L'estadística inferencial que es basa en la significació de la hipòtesi nul·la requereix que l'experimentador establisca a priori quina hipòtesi desitja posar a prova abans de recollir qualsevol dada. Tanmateix, una vegada que acaba l'experiment i s'analitzen les dades, l'investigador inevitablement començarà a veure-hi patrons, alguns directament relacionats amb la seua hipòtesi inicial i d'altres totalment aliens a ella (Gelman i Loken, 2014). Per exemple, un simple ullada als estadístics descriptius pot suggerir que la hipòtesi inicial es compleix, però només per a una part de la mostra. Potser només per a les dones, o només per als participants més joves, o tal vegada per als qui puntuen més alt en neuroticisme. Possiblement cap d'aquests factors formava part del pla d'anàlisi que l'investigador s'havia proposat fer a priori. Però una vegada vistos els estadístics descriptius és difícil no pensar que puguen ser rellevants. En altres paraules, amb freqüència les hipòtesis experimentals no estan completament definides abans de recollir les dades, sinó que es van formant en la ment de l'investigador a mesura que s'exploren les dades.



Si en el procés d'explorar les dades l'investigador es troba amb un resultat significatiu, encara que no formara part del pla d'anàlisi original, no li serà difícil canviar el curs real dels esdeveniments i convèncer-se a si mateix que, en realitat, aquesta hipòtesi s'havia tingut en compte des del primer moment –pràctica coneguda com HARKing, *Hypothesis After Results are Known*; Kerr, 1998–. L'extensa literatura sobre el biaix de retrospecció mostra que, una vegada que coneixem el resultat d'un esdeveniment, tendim a pensar que aquest resultat era perfectament predictable per endavant (Arkes, 2013; Fischhoff, 1975). És a dir, una vegada que sabem quelcom, ens costa retrotraure'ns a la situació en què estàvem quan encara no ho sabíem. És improbable que l'investigador que per accident es troba amb un resultat inesperat siga immune a aquest biaix. Segurament, una vegada vistes les dades, pensarem que aquests resultats eren predictibles i que d'alguna forma vaga formaven part de la hipòtesi inicial.

El fet que les hipòtesis vagen emergint a mesura que s'exploren les dades en lloc de ser clarament formulades a priori, convida a analitzar les dades de tantes formes com siga necessari fins que es trobe un resultat positiu (pràctica coneguda com a *p-hacking*; Wicherts i al., 2016). Les enquestes que s'han realitzat fins al moment confirmen que molts investigadors maquillen les dades amb pràctiques com, per exemple, registrar diverses variables dependents, però informar només de les que donen resultats significatius –*cherry picking*–, eliminar qualsevol al·lusió a condicions experimentals en què no es va trobar cap resultat significatiu o afegir covariables a una anàlisi estadística amb l'única finalitat de fer que una tendència no significativa assolisca la significació estadística.

En la literatura, aquestes pràctiques s'han donat a conèixer com a pràctiques de recerca qüestionables (John, Loewenstein i Prelec, 2012) i tenen com a conseqüència el increment en la taxa de falsos positius. En combinar diverses d'aquestes pràctiques és fàcil aconseguir taxes de falsos positius superiors al 60% (Simmons, Nelson i Simonsohn, 2011). Per descomptat, qualsevol d'aquestes estratègies constitueix mala praxi. Tanmateix, només ho són perquè en tots els casos es recorre a un pla d'anàlisi que s'allunya de la hipòtesi inicial de l'estudi. Si l'investigador és capaç de convèncer-se a si mateix que aquestes estratègies d'anàlisi s'havien considerat per endavant, llavors és perfectament factible posar-les en pràctica sense tenir en cap moment la impressió d'haver actuat malament. El biaix de retrospecció pot jugar un paper clau en aquest procés.

Al biaix de retrospecció s'afegeixen almenys dos biaixos cognitius més que porten qualsevol persona a percebre la realitat de la forma que millor s'acorde a les seues creences: els biaixos de confirmació i de distorsió de la informació (DeKay, 2015; Nickerson, 1998). Si un investigador ha realitzat dos experiments, dels quals un dóna suport a la seua hipòtesi i l'altre no, segurament donarà més pes a qualsevol prova que suggerisca que el primer resultat és més sòlid que el segon. Si l'estudi amb el resultat favorable té més participants,

s'atribuirà la diferència a la major potència estadística. Si l'estudi amb el resultat negatiu es va realitzar a una hora molt primerenca, tal vegada se suppose que els participants no estaven totalment desperts quan vingueren al laboratori. Qualsevol persona intel·ligent –i els científics ho som– serà capaç de trobar raons que expliquen que el resultat esperat no s'haja acomplert en totes les ocasions. En aquest sentit, és interessant recordar que el biaix de publicació de què hem parlat més amunt no es deu tant al fet que les revistes no publiquen resultats negatius, com al fet que els mateixos investigadors, en primera instància, estan poc inclinats a incloure aquests estudis en els seus manuscrits (Dickersin, 2005). Això podria estar revelant que els científics donen més pes als seus experiments quan aquests donen resultats significatius, encara que lògicament és possible que altres factors estiguen contribuint a aquesta decisió.

Com diem en la introducció, en un món ideal pot esperar-se que encara que aquests biaixos puguen afectar el criteri d'un investigador o altre de forma aïllada, col·lectivament es cancel·larien entre si, de manera que la comunitat científica en el seu conjunt seria immune a ells. Per exemple, si el científic A és partidari d'una teoria i el científic B n'és detractor, cal esperar que els biaixos cognitius de A i de B s'anul·len mútuament, de tal forma que el cos de proves que han recollit tots dos porte a una visió no esbiaixada de la realitat. Lamentablement, dècades de recerca en presa de decisions de grup mostren que el consens final al qual arriba un grup rares vegades es correspon amb la creença mitjana dels seus individus, sinó que tendeix a polaritzar-se cap a l'extrem dominant (Myers i Lamm, 1976). En el context de la ciència, això suggereix que una vegada que una teoria es converteix en dominant dins d'una comunitat d'investigadors, encara que siga sobre la base d'una prova feble i poc fiable, és probable que qualsevol dada discordant es reinterprete o revalue fins que encaixe amb la teoria –tal vegada amb l'ajuda d'hipòtesis auxiliars– o bé es desestime com a irrellevant (Chinn i Brewer, 1998). Podem trobar una prova indirecta d'aquest fet en les anàlisis que revelen que la mort o el retir d'un investigador de referència en un camp de recerca sol anar seguida d'un creixement dels intents de posar a prova hipòtesis més herètiques i menys complaents amb l'opinió de la majoria (Fons-Rosen i Azoulay, 2015).

## **Causes distals: el sistema de recompenses i càstigs de la Psicologia acadèmica**

Entre totes les característiques que defineixen la pràctica actual de la ciència, una de les més rellevants és el seu caràcter competitiu. Atès que els recursos amb els quals treballen els científics són limitats, amb freqüència se suposa que la competició per mèrits és una manera adequada d'assignar aquests

recursos –ja siguin els fons d'un projecte de recerca, l'estabilitat d'una plaça, l'accés a instrumental o instal·lacions o l'espai en una revista–. Tanmateix, el nivell de competició que impregna la pràctica científica també pot acabar implantant incentius perversos en el sistema i establint-hi dinàmiques no desitjables, amb conseqüències negatives per als investigadors i per a l'empresa de l'avanç científic.

Com tota competició, la carrera científica actual té les seues pròpies regles per a premiar els guanyadors. En aquest context s'han establert tres indicadors que determinen l'èxit professional d'un investigador: el nombre de publicacions, l'índex d'impacte de les revistes en què s'ha publicat la recerca i el nombre de cites que aquesta ha tingut (Van Dijk, Manor i Carey, 2014). Pràcticament tots els índexs de qualitat i de productivitat utilitzats en la gestió de la recerca es basen en aquests pilars (vegeu Hirsch, 2005). A partir d'ací és fàcil intuir com l'ús generalitzat d'aquestes mètriques pot introduir distorsions indesitjades en el comportament dels científics. Si es planteja la ciència com una competició, no és estrany que els investigadors acaben seleccionant les millors estratègies per a guanyar (Bakker i al., 2012).

En primer lloc, recompensar un investigador pel nombre de publicacions que té provoca que els equips de recerca opten per escriure articles més breus i menys elaborats, en compte de treballs complets –amb múltiples experiments, millors controls i mostres grans– que aporten proves més concloents. De vegades, la pressió de publicar porta a la pràctica de la publicació salami (Smolčić, 2013), que consisteix en la descomposició d'un estudi ampli en parts parcialment redundants –mateixa mostra, mateixes hipòtesis i similar metodologia– i la seua publicació separada en múltiples revistes, amb l'objectiu d'inflar el càlcul total de treballs publicats. Com a conseqüència, aquests resultats creen la falsa impressió de donar suport els uns als altres i inunden la literatura científica amb un cabal molt crescut de proves poc rellevants, la qual cosa dificulta traure conclusions clares de la seua revisió.

Quant a l'índex d'impacte i al recompte de cites que es fan d'un article, tots dos nombres es prenen com una aproximació indirecta a la qualitat de l'article, ja que la valoració directa d'aquesta és més complexa. Aquests índexs han estat criticats freqüentment (Alberts, 2013; Brembs, Buttn i Munafò, 2013; Kirschner, 2013). L'índex d'impacte d'una revista no sempre es correlaciona amb la repercussió de cada article individual, i fins i tot hi ha proves que les recerques publicades en les millors revistes podrien ser menys fiables –per exemple, les revistes de major impacte també tenen una major taxa de retraccions i d'errors en els valors  $p$ , Brembs i al., 2013; Szucs i Ioannidis, 2016–. Però a més l'ús d'aquestes mètriques com a índex de qualitat de la recerca introdueix artefactes en la pràctica diària del científic i crea incentius per a la manipulació d'aquests índexs.

En primer lloc, atès que les revistes competeixen entre si per obtenir un lloc avantatjat en les llistes de classificacions, el seu incentiu és sovint prioritzar la publicació dels resultats dels quals s'espere una repercussió gran, fora fins i tot del cercle de la premsa acadèmica. Molts dels efectes protagonistes en la recent crisi de replicació de la Psicologia –per exemple, que l'exposició a paraules com «jubilació» o «vell» alenteix la velocitat que du el participant quan camina per abandonar el laboratori– (Bargh, Chen i Burrows, 1996) tenen aquest ingredient de resultat sorprenent o contraintuïtiu al gust del públic i els editors de les revistes. Aquesta preferència pels resultats més vendibles estableix un altre biaix de publicació de facto: no només els resultats significatius, sinó també els que tenen més potencial mediàtic es publiquen amb major facilitat. Així, àrees amb molt atractiu com la neurociència adquireixen un avantatge quant a nombre d'investigadors i assignació de fons, mentre que altres àrees o enfocaments menys interessants per al públic es van despoblant a poc a poc. Al mateix temps, s'introdueix un biaix que penalitza la creativitat i el risc: els camins més transitats en la ciència actual són els que porten a la publicació en revistes de gran impacte i pocs investigadors joves arriscarien la seua carrera avui per prendre altres rutes més originals.

I en segon lloc, aquest sistema d'incentius basat en mètriques imperfectes és també un terreny adobat idoni per a l'aparició de les pràctiques científiques qüestionables. Una recerca sobre els efectes de la competició sobre la conducta d'investigadors joves (Anderson, Ronning, De Vries i Martinson, 2007) va revelar que la pressió de destacar en les mètriques els empenyia a adoptar les pràctiques problemàtiques següents: (1) publicació estratègica de resultats, que inclou aspectes com la descomposició d'una única línia de recerca en diversos articles, parcialment redundants entre si, o la publicació selectiva –únicament dels millors resultats, o els significatius–. (2) Tendència a ocultar o a no compartir informació rellevant sobre les recerques. Per exemple, molts investigadors eviten anticipar les seues idees o futures publicacions en congressos i reunions científiques, o fins i tot recelen dels revisors que examinen els seus treballs prèviament a la publicació. També se solen ocultar detalls del procediment per a evitar perdre l'avantatge respecte dels competidors. (3) Introducció de biaixos en el procés de revisió d'experts, de vegades per a dificultar la tasca dels competidors. (4) Ús descurat o poc meticulós de les tècniques estadístiques amb l'objectiu d'abreujar terminis i exagerar la importància dels resultats. Com a cas extrem, la pressió de sostenir un bon ritme de publicacions prestigioses com a condició necessària per a aconseguir i mantenir contractes, finançament i recursos pot fins i tot temptar una minoria dels investigadors a cometre frau o falsificació de dades (Zatonski i Witkowski, 2015), com es va fer palès en el cèlebre cas del psicòleg Diederick Stapel (Enserink, 2012). Pels motius esmentats i alguns altres, algunes associacions de científics, com la Societat Americana

de Biologia Cel·lular, han proposat que les mètriques basades en càlculs d'impacte i cites deixen d'utilitzar-se com a criteri per a contractacions, promocions i obtenció de finançament en la carrera científica (*American Society for Cell Biology*, 2013).

### *a) Simulació del sistema d'incentius en ciència*

Més enllà del fet que la pressió d'obtenir bones puntuacions en les mètriques pugui motivar pràctiques qüestionables, com hem vist, el sistema de competició en general crea incentius per a produir estudis de baixa qualitat i amb resultats poc fiables. Per exemple, Bakker i al. (2012) van realitzar simulacions per ordinador en les quals van comparar l'execució de diversos agents artificials –que en aquest cas feien el paper d'investigadors– amb diferents estratègies. Per exemple, una possible estratègia és fer un estudi a gran escala amb una mostra prou gran per assolir una potència estadística acceptable. D'altra banda, altres estratègies combinen la realització de múltiples estudis més petits –amb menor potència– amb l'ús de determinades pràctiques qüestionables, com la publicació selectiva –només s'escriuen els resultats significatius–. Les simulacions van mostrar com aquest segon tipus d'estratègia és molt més eficient a l'hora d'aconseguir el màxim de resultats significatius en el menor temps.

Més recentment, Higginson i Munafò (2016) van utilitzar una aproximació similar, però més completa en la seua recreació del sistema d'incentius, i trobaren que l'estratègia més racional per a maximitzar el nombre de publicacions consisteix a concentrar els esforços per a trobar resultats nous –la qual cosa deixa fora qualsevol treball de replicació d'estudis previs, així com estudis purament confirmatoris–, a força d'estudis amb molt poca potència –entre 0,1 i 0,4–. Com es va dir prèviament, els estudis amb baixa potència són també els que més freqüentment produiran resultats erronis –siga inflant la grandària d'efectes reals, o bé produint falsos positius–. El comportament d'aquest agent perfectament racional, en les simulacions, tal vegada no es distancia molt del de gran part dels investigadors reals, la pràctica diària dels quals està condicionada per un sistema de recompenses i càstigs similar, d'acord amb algunes recerques que s'han fet (Fanelli, 2010).

## **Un nou sistema d'incentius per a una ciència millor**

Si en la secció anterior hem argumentat que el sistema d'incentius en ciència pot conduir a la proliferació de pràctiques qüestionables, estudis sense potència estadística i conclusions errònies, a continuació esmentarem algunes de les reaccions que hi ha hagut per a corregir els problemes detectats.

### ***a) Polítiques editorials diferents***

Les revistes científiques són en gran manera responsables de la situació actual de la ciència psicològica. Per aquest motiu, el sistema de recerca ha de ser sensible als canvis en les polítiques editorials de les revistes. En aquest sentit, estem assistint a alguns moviments l'objectiu dels quals és millorar la situació que hem descrit en seccions anteriors. En primer lloc, algunes revistes ja recullen explícitament la possibilitat de publicar rèpliques d'altres estudis o resultats nuls (PLoS Blogs, 2015). Unes poques han advocat expressament per deixar fora la novetat, la rellevància o l'impacte previst com a criteris per a acceptar un manuscrit i afirmen que es basen únicament en la qualitat de l'aspecte metodològic. Sense anar tan lluny, moltes de les revistes tradicionals en Psicologia, com *Psychological Science* (APS), estan ampliant l'espai que dediquen a descriure els mètodes i els procediments, de manera que aquests puguin ser coneguts amb detall, al mateix temps que fomenten noves perspectives estadístiques que eviten –o almenys compensen en certa mesura– els problemes associats als valors  $p$  i, en general, als tests de significació que es basen en la hipòtesi nul·la. Entre aquestes noves perspectives està l'ús adequat de l'anàlisi de potència, l'estudi de les grandàries de l'efecte amb intervals de confiança associats o, últimament, l'estadística bayesiana (Cumming, 2014; Etz, Gronau, Dablander, Edelsbrunner i Baribault, 2016; Wagenmakers, 2007).

### ***b) Premiar per compartir***

Més enllà d'oferir la possibilitat de pal·liar alguns problemes amb el sistema actual i de desincentivar algunes males pràctiques, algunes propostes s'estan centrant en el foment mitjançant incentius de les bones pràctiques en l'exercici de la ciència, i, particularment, pel que fa a una adequada transparència en la transmissió de les recerques. Una experiència en aquest sentit és l'ús d'insígnies –*badges*– per a assenyalar les publicacions que assolisquen uns determinats criteris, com per exemple oferir les dades brutes en obert –juntament amb l'article– o compartir els codis que s'utilitzen per a l'anàlisi de dades (Kidwell i al., 2016). En el cas de la revista *Psychological Science*, per exemple, després de la introducció d'aquestes insígnies es va donar un increment substancial en la taxa d'articles que s'acompanyaven amb dades o codi en obert, la qual cosa és un indicatiu que aquest tipus d'estratègies de ludificació poden ajudar els científics a superar les reticències inicials a la transparència.

### ***c) Els articles d'accés obert i els seus avantatges sobre l'impacte personal***

Actualment hi ha certa diversitat quant a les restriccions d'accés a la literatura científica publicada. Amb les revistes tradicionals –que cobren per l'accés

als articles– conviuen revistes d'accés obert –algunes carreguen el cost de la publicació a l'autor, d'altres, no– i revistes híbrides –habitualment, es tracta de revistes d'accés per subscripció que puntualment ofereixen als autors l'opció d'alliberar els seus articles–. Atès que les publicacions en obert les pot visitar qualsevol internauta interessat, sense necessitat d'accedir-hi a través d'una entitat que s'hi haja subscrit –com una universitat–, cal preguntar-se si publicar en aquest format podria incrementar de fet la difusió de les recerques. En aquest sentit, es fan nombroses recerques que semblen confirmar aquest avantatge de les publicacions en obert, almenys a curt-mig termini (Gargouri i al., 2010): els articles en obert es llegeixen més, es distribueixen en camps més heterogenis i per això acaben citant-se més. Es tracta, doncs, d'un nou incentiu que pot incrementar en el futur la preferència per publicar de forma oberta.

#### ***d) Incentius al preregistre: garantia prèvia de publicació***

En seccions anteriors hem explicat com el biaix de publicació –els resultats significatius que confirmen la hipòtesi plantejada es publiquen amb més probabilitat–, juntament amb algunes pràctiques qüestionables (*p-hacking*, *HARKing*, *cherry picking* i, en general, les derivades d'una excessiva flexibilitat de l'investigador), poden introduir elements distorsionadors en la pràctica del científic (Gelman i Loken, 2014). Una estratègia que pot alleujar en certa mesura aquests mals és el preregistre. Tradicionalment, la revisió de les revistes per experts es fa sobre l'article ja complet, amb els resultats i les conclusions. En un article preregistrat, el primer enviament consisteix en un pla detallat del mètode i anàlisi de dades, que és avaluat pels revisors. Una vegada passat aquest filtre de revisió per experts, se'n fa l'estudi i l'article es publica finalment independentment del resultat que haja obtingut (Chambers, 2013). La garantia de publicació, sense importar si les dades resulten ser les que s'esperen, és en si mateix un al·licient per a optar per aquesta modalitat d'article que algunes revistes prestigioses en el nostre camp ja estan implantant (*Psychological Science*, *Psychonomic Bulletin & Review*).

## **Discussió**

En aquest breu assaig presentem un panorama poc complaent, però creiem que realista, de l'estat actual de la credibilitat de la recerca en Psicologia. El conjunt de problemes analitzat no és exclusiu de la Psicologia, sinó que s'estén a altres branques de les ciències socials, biològiques i de la salut –economia, medicina, nutrició, recerca en càncer i neurociència, entre altres–, sobre

les quals també s'ha posat el focus de forma particularment intensa (Camerer i al., 2016; Errington i al., 2014; Prinz, Schlange i Asadullah, 2011). Tanmateix, el cas de la Psicologia ha tingut una repercussió i un impacte mediàtic major. Això, que es pot considerar injust, es deu en part al fet que la Psicologia és una ciència encara precària, particularment afectada per solapaments amb les pseudociències, però al mateix temps molt popular i de gran impacte en la política, l'economia i l'educació. La grandària de la controvèrsia és en gran manera el fruit del creixement de la importància de la Psicologia en la societat en les últimes dècades.

Des del nostre punt de vista, que el focus s'haja posat en la Psicologia pot servir d'esperó per a implantar les polítiques de millora que hem comentat anteriorment. En aquest sentit, podem considerar positiu que la Psicologia prenga cartes en l'assumpte abans que altres disciplines ni tan sols n'hagen pres consciència. En la mesura que les solucions siguin efectives en Psicologia, poden servir d'exemple per a les altres. A més, la Psicologia és la disciplina que pot aportar les eines aplicades necessàries, perquè és la ciència que es preocupa per la relació entre els sistemes d'incentius, la cognició i la conducta.

En el moment actual, no podem dir que l'autoescepticisme metodològic que defensem ací haja tingut un impacte generalitzat sobre el comportament dels científics. És més, aquest escepticisme s'ha trobat amb una forta resistència d'una bona part dels agents que han contribuït a l'estat actual (Baumeister, 2016; Fiske, 2016). Gràcies a l'Economia de la conducta, sabem de la importància del biaix de l'estatu quo i de la dificultat per a corregir-lo. També sabem de la importància d'una aproximació constructiva, en la qual l'escepticisme no es converteix en un assumpte de confrontació individual, sinó en una eina psicoeducativa i de millora de les estructures en les quals els agents individuals prenen les seues decisions, sempre contingents en unes circumstàncies materials determinades.

Més enllà de les mesures concretes, encara incipients, que ja s'estan implantant en el sistema editorial, un signe per a l'optimisme és la corroboració que la comunitat científica sembla no navegar totalment a cegues en la boira de soroll de dades que actualment sembla caracteritzar la Psicologia. En alguns estudis recents, s'han establert mercats de decisió per a predir si els efectes concrets que s'han trobat en els estudis científics aconseguiran replicar-se i sobreviure a un escrutini posterior. Aquests jocs, en els quals els agents inverteixen quantitats de diners reals –de tal manera que estan incentivats per a encetar en les seues decisions d'inversió individuals–, han demostrat capacitat per a predir resultats polítics, econòmics i socials amb bastant precisió (Surowiecki, 2004) en certes condicions d'operació, i sorprenentment, han demostrat una capacitat superior al 70% en la predicció de la fiabilitat d'un resultat científic (Dreber i al., 2015).



Dit d'una altra manera, lluny de reduir els debats a petits comitès de científics interessats de forma intensa en problemes concrets –sotmesos a incentius, fortament polaritzats i amb un important biaix d'estatu quo, amb els problemes que comporten–, les dades sobre mercats de decisió suggereixen que les prediccions són millors quan els agents són independents, estan directament incentivats per a prendre decisions objectivament encertades –i no pel sentit de la decisió de grup– i no són necessàriament experts en la matèria. Una bona part de la millora la ciència psicològica passa, doncs, per democratitzar-la i sotmetre-la de forma oberta a l'escrutini de la comunitat científica.

## Referències

- Alberts, B. (2013). Impact Factor Distortions. *Science*, 340(6134), pp. 787-787.
- American Society for Cell Biology (2013). *The San Francisco Declaration on Research Assessment (DORA)*. Recuperat el 26 de desembre de 2016 de <<http://am.ascb.org/dora/>>.
- Anderson, M. S.; Ronning, E. A.; De Vries, R. i Martinson, B. C. (2007). The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics*, 13(4), pp. 437-461.
- Arkes, H. R. (2013). The consequences of the hindsight bias in medical decision making. *Current Directions in Psychological Science*, 22, pp. 356-360.
- Badenes-Ribera, L.; Frias-Navarro, D.; Iotti, B.; Bonilla-Campos, A. i Longobardi, C. (2016). Misconceptions of the *p*-value among Chilean and Italian Academic Psychologists. *Frontiers in Psychology*, 7, a1247.
- Badenes-Ribera, L.; Frías-Navarro, D. i Pascual-Soler, M. (2015). Errors d'interpretació dels valors *p* en estudiants universitaris de Psicologia. *Anuari de Psicologia*, 16(2), pp. 15-31. doi: 10.7203/anuari.Psicologia.16.2.15.
- Baker, M. (2015, abril). First results from psychology's largest reproducibility test. *Nature News*. Recuperat el 26 de desembre de 2016 de <<http://www.nature.com/news/first-results-from-psychology-s-largest-reproducibility-test-1.17433>>.
- Bakker, M.; Hartgerink, C. H. J.; Wicherts, J. M. i van der Maas, H. L. J. (2016). Researchers' intuitions about power in psychological research. *Psychological Science*, 27(8), pp. 1069-1077.
- Bakker, M.; van Dijk, A. i Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, 7(6), pp. 543-554.
- Bargh, J. A.; Chen, M. i Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, 71(2), pp. 230-244.

- Baumeister, R. F. (2016). Charting the future of social psychology on stormy seas: Winners, losers, and recommendations. *Journal of Experimental Social Psychology*, *66*, pp. 153-158.
- Brembs, B.; Button, K. i Munafò, M. (2013). Deep impact: unintended consequences of journal rank. *Frontiers in Human Neuroscience*, *7*, a291.
- Button, K. S.; Ioannidis, J. P. A.; Mokrysz, C.; Nosek, B. A.; Flint, J.; Robinson, E. S. J. i Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews*, *14*, pp. 365-376.
- Camerer, C. F.; Dreber, A.; Forsell, E.; Ho, T. H.; Huber, J.; Johannesson, M. i Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*, pp. 1433-1436.
- Chambers, C. D. (2013). Registered reports: A new publishing initiative at Cortex. *Cortex*, *49*, pp. 609-610.
- Chinn, C. A. i Brewer, W. F. (1998). An empirical test of a taxonomy of responses to anomalous data in science. *Journal of Research in Science Teaching*, *35*, pp. 623-654.
- Cohen, J. (1990). Things I have learned (thus far). *American Psychologist*, *45*, pp. 1304-1312.
- Coyne, J. C.; Thombs, B. D. i Hagedoorn, M. (2010). Ain't necessarily so: Review and critique of recent meta-analyses of behavioral medicine interventions in health psychology. *Health Psychology*, *29*(2), pp. 107-116.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), pp. 7-29.
- DeKay, M. L. (2015). Predecisional information distortion and the self-fulfilling prophecy of early preferences in choice. *Current Directions in Psychological Science*, *24*, pp. 405-411.
- Dickersin, K. (2005). Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm. En Rothstein, H. R.; Sutton, A. J. i Borenstein, M. (eds.), *Publication bias and meta-analysis: Prevention, assessment and adjustments* (pp. 11-33). Nova York: John Wiley and Sons.
- Dreber, A.; Pfeiffer, T.; Almenberg, J.; Isaksson, S.; Wilson, B.; Chen, Y. i Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, *112*(50), pp. 15343-15347.
- Enserink, M. (2012, November). Final Report: Stapel Affair Points to Bigger Problems in Social Psychology. *Science Insider*, Recuperat el 26 de desembre de 2016 de <<http://www.sciencemag.org/news/2012/11/final-report-stapel-affair-points-bigger-problems-social-psychology>>.
- Errington, T. M.; Iorns, E.; Gunn, W.; Tan, F. E.; Lomax, J. i Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *eLife*, *3*, e04333.

- Etz, A.; Gronau, Q. F.; Dablander, F.; Edelsbrunner, P. A. i Baribault, B. (2016). *Understanding Bayes: How to become a Bayesian in eight easy steps: An annotated reading list*. Recuperat el 26 de desembre de 2016 de <<https://alexanderetz.com/2016/02/07/understanding-bayes-how-to-become-a-bayesian-in-eight-easy-steps/>>.
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS One*, 5(4), e10271.
- Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, pp. 288-299.
- Fiske, S. T. (2016). A call to change science's culture of shaming. *APS Observer*, 29(9), pp. 5-6.
- Fons-Rosen, C. i Azoulay, P. (2015). *Does Science Advance One Funeral at a Time?* National Bureau of Economic Research, *Working Paper*, 21788. Recuperat el 26 de desembre de 2016 de <<http://www.nber.org/papers/w21788>>.
- Franco, A.; Malhotra, N. i Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), pp. 1502-1505.
- Galton, F. (1907a). Vox populi (the wisdom of crowds). *Nature*, 75, pp. 450-451.
- Galton, F. (1907b). One vote, one value. *Nature*, 75, pp. 414-414.
- Gargouri, Y.; Hajjem, C.; Larivière, V.; Gingras, Y.; Carr, L.; Brody, T. i Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLoS ONE*, 5(10), e13636.
- Gelman, A. i Loken, E. (2014). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no «fishing expedition» or «p-hacking» and the research hypothesis was posited ahead of time. *Psychological Bulletin*, 140(5), pp. 1272-1280.
- Gigerenzer, G. (2004). Mindless statistics. *Journal of Socio-Economics*, 33(5), pp. 587-606.
- Higgins, J. P. i Green, S. (eds.) (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley-Blackwell.
- Higginson, A. D. i Munafò, M. R. (2016). Current Incentives for Scientists Lead to Underpowered Studies with Erroneous Conclusions. *PLOS Biology*, 14(11), e2000995.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), pp. 16569-16572.
- Hoekstra, R.; Finch, S.; Kiers, H. A. L. i Johnson, A. (2006). Probability as certainty: Dichotomous thinking and the misuse of p values. *Psychonomic Bulletin & Review*, 13, pp. 1033-1037.

- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124.
- Ioannidis, J. P. A. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, pp. 645-654.
- John, L. K.; Loewenstein, G. i Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, pp. 524-532.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, pp. 196-217.
- Kidwell, M. C.; Lazarević, L. B.; Baranski, E.; Hardwicke, T. E.; Piechowski, S.; Falkenberg, L. S. i Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLoS Biology*, 14(5), e1002456. doi: 10.1371/journal.pbio.1002456.
- Kirschner, M. (2013). A Perverted View of «Impact». *Science*, 340(6138), pp. 1265-1265.
- Klein, R.; Ratliff, K.; Vianello, M.; Adams Jr, R.; Bahník, S.; Bernstein, M. i Cemailcar, Z. (2014). Data from investigating variation in replicability: A «Many Labs» Replication Project. *Journal of Open Psychology Data*, 2(1), e4.
- Lorenz, J.; Rauhut, H.; Schweitzer, F. i Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), pp. 9020-9025.
- Lurquin, J. H.; Michaelson, L. E.; Barker, J. E.; Gustavson, D. E.; Von Bastian, C. C.; Carruth, N. P. i Miyake, A. (2016). No Evidence of the Ego-Depletion Effect across Task Characteristics and Individual Differences: A Pre-Registered Study. *PLoS One*, 11(2), e0147770.
- Meehl, P. E. (1978). Theoretical Risks and Tabular Asterisks: Sir Karl, Sir Ronald, and the Slow Progress of Soft Psychology. *Journal of Consulting and Clinical Psychology*, 46, pp. 806-834.
- Myers, D. G. i Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin*, 83(4), pp. 602-627.
- Nelson, N.; Rosenthal, R. i Rosnow, R. L. (1986). Interpretation of significance levels and effect sizes by psychological researchers. *American Psychologist*, 41, pp. 1299-1301.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2, pp. 175-220.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716.
- PLoS Blogs. (2015, February 25). *Positively Negative: A New PLOS ONE Collection focusing on Negative, Null and Inconclusive Results*. Recuperat el 26 de desembre de 2016 de <<http://blogs.plos.org/everyone/2015/02/25/>>

positively-negative-new-plos-one-collection-focusing-negative-null-inconclusive-results/>.

- Prinz, F.; Schlange, T. i Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10, p. 712.
- Ranehill, E.; Dreber, A.; Johannesson, M.; Leiberg, S.; Sul, S. i Weber, R. A. (2015). Assessing the Robustness of Power Posing No Effect on Hormones and Risk Tolerance in a Large Sample of Men and Women. *Psychological Science*, 26(5), pp. 653-656.
- Rohrer, D.; Pashler, H. i Harris, C. R. (2015). Do subtle reminders of money change people's political views? *Journal of Experimental Psychology: General*, 144, e73-e85.
- Schervish, M. J. (1996). P-Values: What They Are and What They Are Not. *The American Statistician*, 50(3), pp. 203-206.
- Sedlmeier, P. i Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, pp. 309-316.
- Shanks, D. R.; Vadillo, M. A.; Riedel, B.; Clymo, A.; Govind, S.; Hickin, N.; ... i Puhmann, L. (2015). Romance, risk, and replication: Can consumer choices and risk-taking be primed by mating motives? *Journal of Experimental Psychology: General*, 144(6), e142-e158.
- Simmons, J. P.; Nelson, L. D. i Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, pp. 1359-1366.
- Smolčić, V. Š. (2013). Salami publication: definitions and examples. *Biochemia Medica*, 23(3), pp. 137-141.
- Sterling, T. D.; Rosenbaum, W. L. i Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *American Statistician*, 49, pp. 108-112.
- Surowiecki, J. (2004). *The Wisdom of Crowds*. Nova York: Anchor.
- Szucs, D. i Ioannidis, J. P. (2016). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *bioRxiv*. doi: doi.org/10.1101/071530.
- Van Bavel, J. J.; Mende-Siedlecki, P.; Brady, W. J. i Reinero, D. A. (2016). Contextual sensitivity in scientific reproducibility. *Proceedings of the National Academy of Sciences*, 113, pp. 6454-6459.
- Van der Gucht, K.; Griffith, J. W.; Hellems, R.; Bockstaele, M.; Pascal-Claes, F. i Raes, F. (2016). Acceptance and Commitment Therapy (ACT) for Adolescents: Outcomes of a Large-Sample, School-Based, Cluster-Randomized Controlled Trial. *Mindfulness*. Advance online publication. doi: 10.1007/s12671-016-0612-y.

- Van Dijk, D.; Manor, O. i Carey, L. B. (2014). Publication metrics and success on the academic job market. *Current Biology*, 24(11), R516-R517.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin y Review*, 14(5), pp. 779-804.
- Wicherts, J. M.; Veldkamp, C. L. S.; Augusteijn, H. E. M.; Bakker, M.; van Aert, R. C. M. i van Assen, M. A. L. M. (2016). Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking. *Frontiers in Psychology*, 7, p. 1832. doi: 10.3389/fpsyg.2016.01832.
- Young, E. (2012, October). Nobel laureate challenges psychologists to clean up their act. *Nature News*. Recuperat el 26 de desembre de 2016 de <<http://www.nature.com/news/nobel-laureate-challenges-psychologists-to-clean-up-their-act-1.11535>>.
- Zatonski, T. i Witkowski, T. (2015). *Psychology Gone Wrong: The Dark Sides of Science and Therapy*. Boca Raton, FL, US: Brown Walker Publishers.
- Zehetleitner, M. i Schönbrodt, F. (2016). *When does a significant p-value indicate a true effect? Understanding the Positive Predictive Value (PPV) of a p-value*. <<http://shinyapps.org/apps/PPV/>>.

Data de recepció: 28-01-2017  
Data d'acceptació: 20-03-2017