

Szemle

---

MŰKÖDÉSI ZAVAROK ÉS MEGÚJULÁSI TÖREKVÉSEK  
A TUDOMÁNYOS PSZICHOLOGIÁBAN,  
A 21. SZÁZAD ELEJÉN

---

SZOKOLSZKY ÁGNES

Szegedi Tudományegyetem, Pszichológiai Intézet, Kognitív és Neuropszichológiai Tanszék

E-mail: szokolszky@gmail.com

*Beérkezett:* 2018. február 1. – *Elfogadva:* 2018. október 18.

*A 2010-es évek elején a tudományos pszichológiában kibontakozó működési zavarok és válságérzet (a tudományos csalások és visszavont publikációk számának megnövekedése, a vitatható kutatási eljárások, a gyenge vagy kétséges cikkek megjelenése vezet folyóiratokban, és a tudományos kommunikáció torzulásai) a Brian Nosek és munkatársai nevével jelzett „Megismétlési Projektben” (2011–2015) csúcspontot ki, egyfajta bizalmi válságot hozva létre. A pszichológiának szembe kellett néznie (más tudományágakkal együtt) a kutatási eredmények validitását és megbízhatóságát érintő problémákkal. Az alapos önvizsgálat sokféle – nem feltétlenül új keletű – hiányosságot tárt fel, amelyek különböző mértékben érintették a különböző kutatási területeket. Ilyen volt a pontos megismetések hiánya, a konceptuális megismetések visszásságai, a kutatói hipotézist nem igazoló negatív eredmények mostoha státusza, és az ezzel járó publikációs torzulás, a kutatói hipotézist alaptalanul megerősítő (hamis pozitív) eredmények elterjedtsége, és a nullhipotézis-teszteléssel és a statisztika alkalmazásával kapcsolatos korántsem új kritikák. A 2010-es évek vége felé már megfelelő rálátásunk van a feltárt problémákra, jobban értjük az okokat és háttértényezőket, és új megoldási módok és ajánlások is megfogalmazódtak. E tanulmányban a válság fő okainak, szimptomáinak és tanulságainak áttekintésére vállalkozom, és a kibontakozó reformmozgalom nyomán jelentkező új eljárásokra és ajánlásokra hívom fel a figyelmet. A pszichológia e válságperiódusban bizonyította önkorrekciós képességét. A kutatási és publikációs gyakorlatot a jobb működés irányába mutató impulzusok érik, de ezek hatása csak úgy tud érvényesülni, ha a kutatók mint egyének és mint kutatói közösségek tudatosabban viszonyulnak a problémák megoldásához.*

**Kulcsszavak:** adatmanipuláció, Megismétlési Projekt, megismétlési válság, metodológiai reformmozgalom a pszichológiában, publikációs gyakorlat, tudományos csalás, elfogadhatatlan kutatói viselkedés, nullhipotézis-tesztelés

A pszichológia történetében nem ismeretlen jelenség a krízis és az önvizsgálat. A válság érzete lényegében a tudományos pszichológia megalakulása óta kísért, amikor is egymással vitázó nézőpontok fogalmazódtak meg az új tudomány tárgyára, módszertanára és emberképére nézve. Nem egyértelmű azonban az, hogy mit tekintünk válságnak. Egy nézet szerint a pszichológia alapvetően válságban levő tudomány, mert széttöredezett elméleti keretek és módszertani irányzatok jellemzik, és kevés jele van a humán jelenségek átfogó, belső ellentmondásoktól mentes megértésének (vö. Dilworth, 2008). Más megközelítésben az egységes és konzisztens pszichológia ideálja irreális, és sokkal gyakorlatiasabb kritériumokat kell felállítani: akkor áll fenn válság, ha elméleti és módszertani eszközeinkben és eljárásainkban olyan hibák fordulnak elő, amelyek súlyosan hátráltatják vagy akadályozzák a feladatok ellátását (Zittoun, Gillespie és Cornish, 2009).

Nem idealizmus azt állítani, hogy minden tudomány alapvető feladata az, hogy érvényes és megbízható tudást állítson elő, és megkérdőjelezhetetlen legyen etikai integritása. Ha ezekkel a kritériumokkal szemben merülnek fel kétségek, az feltétlenül kimeríti a válság pragmatikus feltételeit, és kiemelt figyelmet érdemel.

A 2010-es évek közepén több jel szerint egy olyan tényleges válság alakult ki, amelyik a tudományos pszichológia működésének alapjait érintette. A jelzések többféle irányból érkeztek. Az ezredfordulót követően feltűnő módon megszaporodtak az etikailag megkérdőjelezhető eljárásokról és tudatos kutatási csalásokról (adathamisításokról, adatmanipulációkról) szóló beszámolók. Az évtized elején legismertebbé vált „*elfogadhatatlan kutatói viselkedés*” („*scientific misconduct*”) eset Diederik Stapelt, a holland Tilburg Egyetem kutatóját érintette. Az évtized vége felé Brian Wansink, a nagyhírű amerikai Cornell Egyetem kutatója körül keletkezett botrány keltett nagy figyelmet (van der Zee, 2017), de a két kutatón kívül számos esetre derült fény, és újabb és újabb esetek derülnek ki (vö. Randall és Welser, 2018).

A gyakran vezető kutatókat érintő esetekkel kapcsolatos vizsgálódások mentén felmerült: lehetséges, hogy ezek csupán a jéghegy csúcsát jelentik. Neves szaklapokban megjelent nehezen hihető, illetve megismételhetetlennek mutatózó eredmények jelezték a problémák sokrétűségét. A baj érzetét növelte annak tudatosulása, hogy a pszichológiai kutatásban és a publikációs gyakorlatban széles körben elterjedtek vitatható gyakorlatok. Ehhez váratlanul hozzáadódott még az is, hogy a kutatói integritást is érintő kritika bontakozott ki a pszichológia egyik klasszikus kísérletével: a stanfordi börtönkísérlettel szemben is (lásd Szokolszky, 2018).

A vizsgálódások és kétségek egyfajta bizalmi válság megfogalmazódásához vezettek az empirikus kutatások egy részének érvényességével és megbízhatóságával kapcsolatban. Bizonytalanság alakult ki a tekintetben, hogy a pszichológiai folyóiratokban megjelent eredmények mennyire hitelesek (Pashler és Wagenmakers, 2012). A feltárt problémák elsősorban a szociálpszichológiát érintették, de kiterjedtek más tudományterületekre is (pl. idegtudomány: Lisberger, 2013; orvostudomány: Bozzo, Bali, Evaniew és Ghert, 2017). A feltárt súlyos zavarok a Brian Nosek nevével fémjelzett Megismétlési Projekthez (Open Science Collaboration, 2015) vezettek, amely megerősítette a válságérzetet, mivel alacsony közvetlen megismételhetőségi arányt mutatott ki (ismét csak nem egyedülálló módon a pszichológiában). A projekt eredményeinek értelmezése azonban összetett feladat, és az elmélyülő önvizsgálat a metodológiai

tudatosság erősödésének forrásává is vált. A 2010-es évek második felében reformmozgalom bontakozott ki a hibák kiküszöbölésére. Új ajánlások és új gyakorlatok születtek, ilyen például a publikációk előregisztrációja (Nosek és Lindsay, 2018).

Tanulmányom célja ennek az összetett helyzetnek az áttekintése. Érinteni kívánom a válság okait, szimptomáit és tanulságait. Elsőként kitérek a megszorított csalások mibenlétére és hátterére, és az ennél jelentősebb volumenű problémára: a vitatható kutatási eljárások elterjedésére. Ezt követően kitérek a kiélezett verseny, az ösztönzési rendszer és a publikációs környezet hatásaira, és arra a jelenségre, hogy magas presztízsű folyóiratokban megkérdőjelezhető minőségű kutatások jelenhettek meg. A 2010-es évek közepétől nagy figyelmet kapott cikkek sorozata, majd a Megismétlési Projekt hívta fel a figyelmet a gondokra. Írásom második felében a problémák elemzésére és a következtetések levonására fókuszálok.

Mint ahogyan a problémák súlya sem egyaránt érinti a pszichológia különböző területeit, úgy az önvizsgálatnak és a jobb kutatási gyakorlatok kialakulásának is területspecifikusnak kell lenniük. A válság egyben hosszú távú növekedési lehetőség, így a 2010-es évek metodológiai krízise is lehetőséget nyújt a fejlődésre, az önreflexión keresztül. Írásom abból a megfontolásból született, hogy bár a Megismétlési Projekt eredményeiről széles körben lehetett értesülni, a válságjelenségek részletes és szisztematikus bemutatása, valamint az abból származó tanulságok összefogott áttekintése kevésbé elérhető a szakirodalomban.

## TUDOMÁNYOS CSALÁSOK ÉS VITATHATÓ KUTATÁSI ELJÁRÁSOK

A tudományos csalás hivatalos definíciójának és eljárási szabályzásának kidolgozása az 1990-es években kezdődött el az Egyesült Államokban, elsősorban a biomedikális kutatások területére koncentrálván (Lock, Wells és Farthing, 2001). A sok vita jelezte, hogy körültekintően kell eljárni, mivel mind a csalások rejtve maradása, mind pedig az igaztalan vád jelentős kárt okoz. A tudományos integritást felügyelő kormányhivatal (Office of Research Integrity) koordinálásával 1989-ben elfogadott definíció megállapítja, hogy a tudományos csalás megvalósulhat „kreált adatok előállításában, a felvett adatok megmásításában, plagizálásban vagy a megtévesztés más formáiban a kutatás pályázati, megvalósítási vagy publikálási fázisában” (Rennie és Gunsalus, 2001, 16). A szabályzás iránti igény jelezte, hogy akadémiai csalások korábban is előfordultak. A 2010-es években kiderített esetek viszont azt mutatták, hogy a hivatalos lépések nem voltak elégségesek a megelőzést illetően.

A 2010-es években megrázó erejű volt a feltárt esetek mélysége és gyakorisága. Marc D. Hauser, a Harvard vezető kutatóját (kutatási területe a kogníció és a moralitás evolúciója) az egyetem etikai bizottsága 3 éves belső vizsgálat után, 2010-ben találta bűnösnek adathamisítás miatt (Newcomer és Spitzer, 2010). Hauser és másik két munkatársa egyik jelentős publikációját (Hauser, Weiss és Marcus, 2002) visszavonták, de ezt a tanulmányt (amelyre azóta az adatbázisok „retracted” jelzővel adnak ki találatot) addigra már 151 esetben idézték. 2011-ben Diederik Stapel, egy holland egyetem elismert szociálpszichológus kutatója ellen indult vizsgálat (kutatási területe a hatalommal és a sztereotipizálással kapcsolatos szociális viselkedés), amely több éven

át tartó adatmanipulációt és kreált adatok használatát állapította meg. A Retraction Watch internetes oldal Stapel esetében 58 visszavonásról számol be 2017 októberében (Palus, 2015). Stapel elismerte a csalást, és egy könyvben megírta saját történetét (Stapel, 2014).

További esetekre is fény derült, vezető kutatók érintettségével: Karen Ruggiero, Harvard University, 2 visszavont cikk szociális igazságosság témában; Dirk Smeester, Erasmus University Rotterdam, 7 visszavont cikk szociális priming témában; Lawrence Sanna, University of Michigan, 8 visszavont cikk különböző szociálpszichológiai témákban. A leleplezések tovább folytatódtak (Van Kolschooten, 2014; O’Grady, 2017). 2017-ben Brian Wansink, a Cornell Egyetem vezető táplálkozáspszichológusa körül alakult ki botrányos helyzet. Wansink három évtizedes tevékenysége alatt több száz tanulmányt publikált, több mint húszeszeres idézettséget ért el, több könyve bestseller lett. Emellett az amerikai kormány tanácsadója volt, és „Smarter Lunchrooms” nevű kezdeményezése keretében több ezer iskolai menzán az ő kutatásai alapján futtatnak egészséges táplálkozás programokat. A problémákra 2017-ben derült fény, számos cikket rohamos gyorsasággal vonták vissza (van der Zee, 2017). Ezek a neves szerzők sok kollégával együttműködésben dolgoztak, így a társszerzők és a kezük alá dolgozó doktori diákok szélesebb körének szakmai integritása és jövője is kérdésessé vált.

Az akadémiai csalások terén korántsem a pszichológiára korlátozódó problémáról van szó. 2011-ben például a UCLA egyik politikatudomány szakos PhD hallgatója hamisított adatokat egy érdekes eredmény megkonstruálása érdekében, és a kutatást egy neves professzorral együtt közölte le, aki csak a nevét adta a kutatáshoz (Singal, 2015). Ugyanabban az évben a Boston University egyik rákkutató professzora ellen merült fel megalapozottan a csalás vádja (Johnson, 2011). 2012-ben a biológiai területen derült ki, hogy szerzők hamis néven saját cikkükről írtak lektori véleményt (Akst, 2012), majd 2017-ben a Tumor Biology c. folyóirat vont vissza 107 közleményt lektori vélemények meghamisítása miatt (McCook, 2017). A csalások mértékét tekintve tízszeres növekedést becsülnék meg egyes elemzések az 1970-es évekhez képest (Fang, Steen és Casadevall, 2012).

Hamarosan vizsgálódások indultak a pszichológia berkein belül annak felderítésére, hogy mennyiben beszélhetünk a jéghegy csúcsáról az említett esetekben. Egy kutatás 2009-ben becsülte meg korábbi felmérések metaelemzésére alapozva, hogy milyen gyakori lehet a tudományos csalás (Fanelli 2009). A kutatók 2 százaléka ismerte el saját magára nézve, anonim módon ezt a legsúlyosabb vétséget, viszont 14% állította, hogy személyes környezetében más kollégák elkövettek ilyent.

A vizsgálatok és elemzések (pl. Gadbury és Allison, 2012; Head, Holman, Lanfear, Kahn és Jennions, 2015; John, Loewenstein és Prelec, 2012; Fanelli, 2009, 2010) a csalások mellett egyéb, sokkal szélesebb kört érintő negatív jelenségekre is rávilágítottak. Ezek a kifogásolható, mégis gyakran előforduló eljárások az utóbbi időben a „*vitatható kutatási eljárások*” („*questionable research practices*”) gyűjtőnevet kapták (vö. Banks, Rogelberg, Woznyj, Landis és Rupp, 2016). Ilyen a szignifikancia manipulálása „rugalmas” adatkezeléssel (pl. a csoportszintű adatok összevonása vagy szétbontása), a statisztikai eljárások próbálgatása annak érdekében, hogy az eredmény elérje a szignifikancia-szintet („*p hacking*”); az elemzés megállítása egy kedvező pontnál; a meghosszabbított adatgyűjtés a jobb eredmény reményében, a post-hoc hipotézis állítás; a tendenciózusan szelektív adat- és eredményközlés; a kísérleti kondíciók és függő változók szelektív

bemutatása; egy váratlan eredmény várt eredményként való beállítása; az eredmények felszabdalt publikálása a publikációk számának növelése érdekében; a minta korlátozottságát figyelmen kívül hagyó jogosulatlan általánosítás; a negatív eredmények nem közlése; és a nyers adatok rövid ideig történő tárolása.

Egy több mint kétezer fős pszichológus kutató mintán végzett felmérés (John, Loewenstein és Prelec, 2012) megerősítette a rossz praxisok elterjedtségét. A válaszadók több mint 50%-a esetében előfordult például, hogy az adatgyűjtés során ellenőrizte a szignifikanciaértéket és ehhez igazította a további adatgyűjtést; több mint 60% szelektíven számolt be az alkalmazott függő változókról; és több mint 40% ismerte el, hogy egyes adatokat azután zárt ki az elemzésből, hogy megnézte, a kizárás hogyan befolyásolja a szignifikanciaszintet. (A megkérdőjelezhető eljárások elterjedtségét mutató további empirikus vizsgálatra lásd Banks és mtsai, 2016.)

A kutatók jó része nem tekinti súlyosnak ezeket a „szoft” módszertani vétségeket, amelyek pedig olyan tendenciózus rugalmasságot visznek a kutatási folyamatba, amely kipárnázza a kutatói hipotézist megerősítő, pozitív eredményekhez vezető utat. Ezek a szándékosan vagy átgondolatlan rutinból alkalmazott eljárások táptalajt jelentenek a kutatói elfogultság, és ezáltal a torzult eredményközlés érvényesülésének. Ennek eredménye az lehet, hogy a szakirodalomban elszaporodnak a *hamis pozitív*, azaz a kutatói hipotézist alaptalanul megerősítettnek tekintő eredmények. Másként: gyakorrivá válik a nullhipotézis helytelen visszautasítása – az a helyzet, amikor valósnak fogadunk el egy eredményt, pedig nem az (Simmons, Nelson és Simonsohn, 2011). Ha pedig egy-egy szakterületen megnövekszik a látens hamis pozitív eredmények jelenléte, amelyre újabb hamis pozitív eredmények épülhetnek, akkor hiába a produktivitás, nem bizonyos, hogy valóságos előrelépés történik a kutatási területen.

A hamis pozitív eredmények problémája a 2010-es évek válságának középpontjába került. A további részletezés előtt azonban vizsgáljuk meg, hogy milyen tényezők állhatnak az akadémiai csalások és rossz kutatási eljárások elterjedésének hátterében.

## AZ ÖSZTÖNZÉS, A VERSENY ÉS A PUBLIKÁCIÓS KÖRNYEZET

Kérdés, hogy miért követnek el jól képzett kutatók módszertani vétségeket, és egyesek miért mennek el az adathamisításig? A kutatók számának exponenciális növekedése (Bornmann és Mutz, 2015) bizonyosan magával hozza a negatív jelenségek megnövekedését is. Az egyéni felelősség megkérdőjelezhetetlen, de ha nem elszórtan és periférikusan jelentkezik a problémás jelenségek, akkor el kell gondolkodni azon, hogy vannak-e olyan rendszerszintű háttértényezők, amelyek nemkívánatos módon befolyásolják a kutatók gyakorlatát.

Indokoltnak tűnik a fő tényezőcsoportot a felerősödött versenyben és publikációs nyomásban látni (Editors, 2011).<sup>1</sup> A verseny mindig részét képezte a tudomány műkö-

<sup>1</sup> Ennek a kérdéskörnek igen nagy a szakirodalma. Szelektíven a következő referenciákat lehet említeni: Anderson és mtsai, 2007; Miller, Taylor és Bedeian, 2011; Sarewitz, 2016; Waaijer, Teelken, Wouters és van der Weijden, 2017. Hazai vonatkozásban: Csaba, Szentes és Zalai, 2014; Marton, Varró és Varró, 2004; Papp, 2004; Tóth, 2014.

désének. Az állásokért, pozíciókért és kutatási erőforrásokért folytatott jelenlegi kompetíció kiélezettebb, mint korábban volt. Versenyben állnak egymással az egyes kutatók (az állásokért és a karrierben való előmenetelért), a kutatócsoportok (a pályázati pénzekért), az egyetemek (a diákok megszerzéséért) és a folyóiratok (az olvasók és a finanszírozás megszerzéséért). A versenymezőny tipikusan magas színvonalú. A kiválóak a még kiválóbbakkal állnak versenyben, gyakran az is erős teljesítményt mutat fel, aki veszít, de vannak, akik még jobbak. A 21. század elején mindez egy folyamatos, élesedő egzisztenciális és pszichés „karriernyomás” részeként jelenik meg az akadémiai életben. Erik Poehlman, a University of Vermont kutatója, aki 12 hónapot töltött szövetségi börtönben adathamisításért, tettét részben annak tudta be, hogy úgy látta, pozícióját és önértékelését csak a folyamatos sikerrel tudja fenntartani (Interlandi, 2006).

Miközben a versenynek ismert a hatékonyságnövelő hatása, az is tudott, hogy a túl erős versengés csökkenti a kreativitást és gátolhatja a kollektív előrehaladást (Fang és Casadevall, 2015). Anderson, Ronning, De Vries és Martinson (2007) 51 karrierjének közepén tartó kutatóval készített fókuszcsoporthatár-interjú azzal kapcsolatban, hogy hogyan él meg a folyamatos versenyhelyzetet. A következő negatív tendenciák rajzolódtak ki: 1. a saját eredmény felnagyítása mások odavágó munkájának elhallgatásával, 2. adatvisszatartás a saját pozíció erősítése céljából; 3. a lektorálás felhasználása manipulatív célokra; 4. rossz kutatási gyakorlatok alkalmazása a pozitív eredmény és a publikálhatóság érdekében.

A tudományos teljesítmény folyamatos mérésére alapuló ösztönzési rendszer, a „publish or perish” normatíva a gyorsan és biztosan publikálható eredmények felé tereli a kutatókat, és ez szerepet játszik a csalások és problémás eljárások elszaporodásában. A publikációs nyomással más jelenségek is összefüggésbe hozhatóak, így például az egy cikkre eső szerzőszám növekedésének jelensége, ami egyrészt a teammunka szükségszerű következménye, de nehezen mérhető mértékben a „szívésségi” vagy „stratégiai” szerzőség elszaporodását is jelzi. Továbbá ide sorolható az eredmények többszörös vagy felszabdalt publikálása, amely elvezetett a „*legkisebb publikálható egység*” fogalmának megjelenéséhez (Wesel, 2016; Biswanger, 2014).

Az utóbbi időben a publikációk számának növelésével szemben előtérbe került a magas impaktfaktorú folyóiratokban való közlés fontossága (Franco, 2013). Lényeges megnézni, hogy mi jellemzi az ezen folyóiratok által kialakított publikációs környezetet, mert ez a kutatási gyakorlat fontos szabályzó eleme. Két tényező emelhető ki ebben az összefüggésben: 1. a folyóiratok nem publikálnak negatív eredményeket; 2. a folyóiratok szerkesztői előtérbe helyezik az újdonságértékű, érdekes, meglepő eredménnyel rendelkező kutatásokat (Fanelli, 2012; Head, Holman, Lanfear, Kahn és Jennions, 2015). E trendek mögött a szakmai és médiafigyelem megragadására irányuló verseny húzódik meg. A kutatók tehát különösen érdekeltek az újdonságértékű, meglepő és érdekes („szexi”) eredmények termelésében.

Természetesnek tűnhet a negatív vagy kevésbé újszerű eredményekkel szembeni érdektelenség. A negatív eredmények „fiókban maradása” (Rosenthal, 1979), csak úgy, mint a kutatói elfogultság a saját hipotézissel / elmélettel szemben, igencsak régi probléma (Sterling, 1959; Sterling, Rosenbaum és Weinkam, 1995). A probléma súlya és következményei iránti érzéketlenség azonban újra előtérbe került a 2010-es évek vitáiban. A negatív eredmények nem közlése összetett kérdés. Óhatatlan, hogy a tu-

dományos közlemények többsége pozitív eredményeket közöl. Világos azonban, hogy torzulásokhoz vezet, ha szélsőségesen érvényesül ez a tendencia. A negatív eredmények a tudományos kutatás szerves részét képezik, információs és falszifikációs értékük jelentős. Ha egy-egy kutatási terület fejlődésébe csak a pozitív eredmények épülnek be, akkor ez gyengíti az eredmények robusztusságának megítélhetőségét. A kutatás szintjén a negatív eredménnyel szembeni averzió a mindenáron való szignifikáns eredmény imperatívuszát hozza magával, a szakterület szintjén pedig a „*publikációs torzulást*” („*publication bias*”): a pozitív eredmények aránytalan túlsúlyát. Számszerű elemzések is alátámasztják, hogy a negatív eredmények gyakorlatilag hiányoznak a pszichológiai szakirodalomból; bár a tendencia más tudományágakban is jelen van, a pszichológia élen jár ezen a téren (Fanelli, 2010, 2011; Pautasso, 2010).

A pszichológiai szakirodalomban régóta erős a pozitív eredmények közzétételének túlsúlya (vö. Sterling, 1959; *Sterling, Rosenbaum és Weinkam, 1995*). Fanelli (2010) több mint kétezer publikációt átfogó elemzése szerint ez jelenleg a pszichológiában és pszichiátriában 95,1%, miközben az idegtudományokban 85%, és a legtöbb negatív eredményt közlő űrtudományban 70,2%. Ezzel az eredménnyel a pszichológia és a pszichiátria az első helyet foglalja el a pozitív eredmények szakirodalmi arányát tekintve a vizsgált 20 tudományág között.

A publikációs nyomás erősödésével a kutatók ellentmondásos helyzetbe kerültek: a tudományos kutatás természeténél fogva bizonytalan kimenetelű vállalkozás, ami kockázattal jár: a hipotézisek egy része (jó esetben) bukásra van ítélve. Ennek a ténynek az elismerését azonban az ösztönzési és közzételési rendszer nem tükrözi. A 2010-es évek önvizsgálata (amelynek részleteire a továbbiakban kitérek) azt mutatja, hogy a makroszintű rendszer a nemkívánatos eljárások felé terel sok kutatót, és olyan lokális szubkultúrák alakulhatnak ki, amelyben sok minden megengedett a publikálható cikkek megírása érdekében.

A problémákat emellett az is jelezte, hogy alapvető kétségeket ébresztő publikációk jelenhettek meg vezető folyóiratokban.

## DARYL BEM 2011-ES PARAPSZICHOLÓGIAI PUBLIKÁCIÓJA

A pszichológia területén Daryl Bem parapszichológiai témájú cikke váltott ki katalizáló hatású sokkot. A cikk az APA által kiadott *Journal of Personality and Social Psychology*-ban jelent meg (Bem, 2011). Bem cikke a jövőbe látás („*prekogníció*”) extraszenzoriális jelenségét vizsgálta kísérleti módszerrel, a Cornell University több mint ezer BA diákja részvételével. Bem 9 kísérletet közölt, amelyek közül 8 gyenge, de statisztikailag szignifikáns eredményt talált amelle, hogy bárki rendelkezik valamilyen szinten a prekogníció intuitív képességével.

Az egyik kísérletben a vizsgálati személy két függőnyt látott a képernyőn, és el kellett találnia, hogy melyik mögött van eltakart kép. Eltakart kép azonban a válaszadáskor még nem volt egyik függőny mögött sem, csak a válaszadás után dőlt el automatizált random módszerrel, hogy melyik mögé kerül kép. A vizsgálat logikája szerint, ha konzisztens módon magasabb a találati arány a véletlennél, akkor ez a prekogníció bizonyítékeként értékelhető. További kísérletekben a résztvevők azokra a szavakra jobban

emlékeztek, amelyeket csak a kísérlet egy további fázisában kellett leírniuk; félelmi reakciót mutattak félelemkeltő inger bemutatását megelőzően; és habituálódtak olyan kellemetlen ingerekre, amelyeket csak a habituáció után kaptak meg (Bem, 2011).

A publikáció közlése (amelyet erős médiafigyelem követett) a parapszichológia meghatározó pillanata volt, mivel példa nélkül állt egy ilyen témájú tanulmány megjelenése vezető tudományos pszichológiai folyóiratban. Mások számára viszont zavarba ejtő volt, hogy a lektorok (a szerkesztőség négy lektort kért fel a szokásos kettő helyett) nem találtak kifogásolnivalót sem a módszertanban, sem a statisztikában. Ez egy olyan helyzethez vezetett, hogy vagy a tudományos józan ész és az ismert fizikai törvények érvényességét kellett az olvasónak felfüggesztenie, vagy az abba vetett hitet, hogy a pszichológia konvencionális játékszabályai megfelelően működnek.

A megjelenés után többen kritikai vizsgálat alá vetették Bem kutatásait. Bem nyitottan, sőt, ösztönzően állt a rigorózus ellenőrzések elé – számolt ezzel, amikor a publikáció megjelentetésére készült. Módszertani problémákat a vizsgálódások valóban feltárták: így például előfordult, hogy a kísérleti protokollt menet közben megváltoztatták, a statisztikai feldolgozás során különböző adatszoportokat ad hoc módon vontak össze, és nem lehetett tudni, hogy hány személy adata maradt ki a végső elemzésből (Alcock, 2003). Ezeknek a nem túl jelentős elemzési döntéseknek a feltárása (amelyeket maga Bem is elismert) azonban nem volt perdöntő, mivel lényegében minden, amit Bem csinált, bevett gyakorlat volt a pszichológiai kutatásban. Sőt, Bem sok tekintetben az átlagnál is körültekintőbben járt el a publikált kísérletek során (LeBel és Peters, 2011).

Alcock (2003) meglátása szerint alapvető probléma, hogy a parapszichológusok valójában nem adnak esélyt a nullhipotézisnek; annyira hisznek a vizsgált psi jelenség létezésében, hogy tudatosan vagy tudattalanul az eredményesség irányába terelik a vizsgálat folyamatát, ha pedig nem jön ki pozitív eredmény, akkor vagy addig keresgélnek, amíg az ki nem jön, vagy pedig hajlékonyan, egyéb lehetséges okkal magyarázzák a sikertelenséget. 2012-ben a *Journal of Personality and Social Psychology* leközölt egy kutatást, amelyik megismételte Bem kísérleteit, de nem mutatott ki prekogníciót (Galak, LeBoeuf, Nelson és Simmons, 2012; a kutatást előzetesen több folyóirat visszautasította arra hivatkozva, hogy negatív eredményt nem közölnek, vö. Yong, 2012a). Más megismétlések is negatív eredménnyel zárultak, a vizsgált psi jelenségekbe vetett pro és kontra hitek azonban zavartalanul fennmaradtak.

A Bem-cikk egyféle lakmuszpapír volt: fókuszált sokakban már jó ideje meglévő aggodalmakat. A gondok túlmutattak a parapszichológián; fontosabbnak mutatkoztak a tudományos pszichológia művelését érintő következtetések. Bem cikke azért volt jó kiindulópont az általánosabb tanulságokhoz, mert az abszurd eredményhez a bevett módszertan alapján jutott. Carl Sagan népszerűsítette azt a tudomány klasszikusai által megfogalmazott szabályt, miszerint rendkívüli állításokat rendkívül erős bizonyítékokkal kell alátámasztani. A Bem-féle kutatás ennek a követelménynek nem felelt meg, de az elfogadott tudományosság játékszabályai szerint játszott. Az aggasztó felismerés az volt, hogy a hibákat és torzításokat, amiket a rendszer nem szűr ki, tulajdonképpen bárki elkövetheti.



## TOVÁBBI KÉRDÉSES CIKKEK VEZETŐ FOLYÓIRATOKBAN

A 2010-es évek elején világossá vált, hogy kérdéses minőségű cikkek jelenhettek meg jó nevű tudományos folyóiratokban. Három szociálpszichológus: Uri Simonsohn, Joseph Simmons és Leif Nelson listázni kezdték azokat a tanulmányokat, amelyek hiteltelennek tűntek (Simmons, Nelson és Simonsohn, 2018; Shea, 2012). Az egyik ilyen publikáció Lawrence Sanna és munkatársainak a vizsgálata volt (Sanna, Chang, Miceli és Lundberg, 2011), amelyik a *Journal of Experimental Social Psychology*-ban jelent meg. A szerzők azt a hipotézist ellenőrizték egy kísérletsorozattal, hogy az emberek proszociálisabban viselkednek fizikailag magasabban levő pozícióban. Az eredmények igazolták az elvárásokat: azok, akik egy mozgólépcsőn felfelé utaztak, többet adományoztak, mint azok, akik lefelé utaztak, és hosszabb ideig segítettek másoknak azok, akik magasabban ültek. A kritikai átvilágítás statisztikai hibákat mutatott ki, és adatmanipulációt is vélelmezett. Később az első szerző cikkét (több más cikkével együtt) visszavonták (Yong, 2012b).

Sanna és munkatársai ahhoz a paradigmához kapcsolódtak, amely a Yale Egyetemen John Bargh munkájából nőtt ki. A „szociális priming” elnevezésű kutatási terület olyan vizsgálatokat foglal magában, amelyek a finom fizikai ingerekkel történő előfeszítés tudattalan viselkedésmódosító hatását igyekeznek kimutatni. Az első, sokat idézetté vált kutatásban (Bargh, Chen és Burrows, 1996) Barghék öregséggel kapcsolatos hívószavaknak tették ki a kísérletben résztvevőket, és azt találták, hogy az előfeszített követően a vizsgálati személyek lassabb ütemben hagyták el a vizsgálat helyszínét, mint azok, akik nem voltak kitéve ennek az előfeszítésnek. Miközben felkapottá vált a kutatási terület, kétségek merültek fel az eredmények megbízhatóságával kapcsolatosan. Néhányan közvetlen megisméltést kezdeményeztek, és ezek rendre negatív eredménnyel zárultak. 2012-ben Stephane Doyennek és munkatársainak sikerült a negatív eredményeket és a kutatásokkal kapcsolatos kételyeiket publikálniuk (Doyen, Klein, Pichon és Cleeremans, 2012). Még ebben az évben a Nobel-díjas pszichológus közgazdász Daniel Kahneman a *Nature*-ben nyilvános emailben kérte a terület kutatóit, hogy a különböző laborok független replikációval igyekezzenek mielőbb tisztázni a szociális priming hatások valóságát a kutatási terület hitelességének megőrzése, valamint a területen dolgozó fiatal kutatók jövője érdekében (vö. Yong, 2012c.; Kahnemann 2011-ben megjelent „Thinking fast and slow” című könyvében maga is épített a szóban forgó kutatásokra, így könyvét is érintették a kérdésesnek mutatkozó eredmények).

A kételyeket keltő publikációk közé sorolható a Harvard Business School szociálpszichológusának, Amy Cuddy-nak a kutatása is, amely a *Psychological Science* című folyóiratban jelent meg (Carney, Cuddy és Yap, 2010). A kísérleti kutatás a „hatalmi póz” (egyfajta kiterjesztett testtartás) hatását vizsgálta, és megerősítette azt a feltételezést, hogy a testtartás 1 perces gyakorlása azonnali pozitív fiziológiai és viselkedéses változásokat eredményez: a kísérleti csoportban levők tesztoszteronszintje megemelkedett, kortizolszintje csökkent, kompetenciaérzésük és rizikóvállalásuk megemelkedett a kontrollcsoporthoz képest, akik 1 percet összegörnyedt testtartásban töltöttek. A kutatás módszertanát később komoly kritika érte (Simmons és Simonsohn, 2015; Gelman és Fung, 2016). 2016-ban a szenior szerzőtárs, Carney elhatárolta magát a kutatástól: azt nyilatkozta, hogy nem hiszi, hogy a hatalmi póz hatása valós (Singal, 2016).

További példa lehet a Kristina Durante nevével fémjelzett vizsgálat (Durante, Rae és Griskevicius, 2013), amely a nők politikai szavazással kapcsolatos magatartását vizsgálta a 2012-es elnökválasztási kampányt követően; a publikáció szintén a *Psychological Science* c. folyóiratban jelent meg. Az evolúciós pszichológiával elkötelezett kutatás feltevései szerint ovulációs időszakban a nők nyitottá válnak a rövid távú szexuális kapcsolatokra, és olyan férfiakat preferálnak, akik a genetikai fitness jegyeit hordozzák – ilyen az attraktivitás és az arcszimmetria. Mivel ezekkel a jegyekkel Obama inkább rendelkezik, mint Romney, várható, hogy az ovuláló nők inkább Obamára szavaznak. Feltételezték továbbá, hogy mivel a rövid távú szexuális kapcsolatok nem illeszkednek a vallásossághoz, „az ovuláció liberálisabbá, egyben kevésbé vallásossá teheti a nőket” (i. m. 2.). A megjelenés utáni kritika az elméleti háttér spekulatív jellege mellett komoly módszertani-statisztikai problémákra mutatott rá (Gelman, 2013).

Melyek a közös vonásai az említett (és más hasonló) cikkeknek? 1. Érdekes és meglepő eredménnyel állnak elő; 2. Elméleti hátterük gyenge; 3. Kis mintán végzett néhány kísérletre alapoznak, majd túlzó és kétséges konklúziókat vonnak le gyenge bizonyítékok alapján; 4. Anonim szakmai bírálattal dolgozó vezető folyóiratokban jelennek meg, amelyek igen magas visszautasítási rátával működnek, így feltételezhetően csak a legjobb minőségű kutatásokat publikálják; 5. A publikálás után komoly problémákra derül fény.

Önmagában nem meglepő, hogy előfordulnak gyenge minőségű vagy kétes eredményű kutatások és óvatlan lektorok. A tudományterület diszfunkcionális működését mutatja viszont, ha a kétséges minőségű tudomány a szakmai szűrőkön átjutott legjobb tudományként kerül elénk. Növeli a problémát, ha erre a tudományos kommunikáció torzulásai is ráarakódnak.

## TORZULÁSOK A TUDOMÁNYOS KOMMUNIKÁCIÓBAN

A tudományos tömegtermelés és a tömeges tudományos hírfogyasztás időszakában éles verseny bontakozott ki a médiafigyelem megragadása érdekében. A tudományt nemcsak művelni kell, de el is kell tudni adni. Az „eladás” a kutatók és a kutatás mögött álló intézmények számára a médiafigyelem megragadásának feladatát jelenti. Mindez korunk természetes velejárója, azonban érzékelhető a „média impaktfaktor” negatív hatása. Ezek közül szembetűnő megnyilvánulás az ún. „overselling”: az eladásra irányuló túlzó igyekezet.

Tanulságos történet Amy Cuddy-é, akinek hatalmi pózzal kapcsolatos kutatásairól már volt szó. Cuddy a *Psychological Science*-ben megjelent, szerzőtársakkal írt cikke (Carney, Cuddy és Yap, 2010) nyomán vált széles körben ismertté. Az érdekes kutatási eredményt a média hamar felkapta, és ennek eredménye lett 2012-ben a TED történetének második legsikeresebb előadása, amit 2017-ig több mint 43 millió ember nézett meg. Amy Cuddy kiváló előadói képességének és inspiratív beszédének köszönhető a sikert, aminek fő hitelességét mégis legfőképpen a Harvard Egyetem keretében végzett tudományos kutatás, és az a felszított remény adta, hogy bárki néhányzori pár perces hatalmi póz gyakorlással jelentősen javítani tudja esélyeit, hogy nehéz helyzetekben helytálljon, sőt megváltoztassa az életét. Sikerkönyvében Cuddy

(2015) már azt is kijelentette, hogy a hatalmi póz gyakorlása eszköz lehet a nők és a kisebbségi csoportok helyzetének javítására. 2017-re azonban már egy tucat megismétlést végeztek, amelyek nem erősítették meg az eredeti kutatási eredményeket. Cuddy a kritika keresztútjába került, és 2017-ben elhagyta a Harvardot (vö. Engber, 2017, Dominus, 2017).

A klasszikus munkamegosztás szerint a jó minőségű, érvényes és megbízható kutatás végzése a tudományos kutatók dolga, a tudományos szűrőn átment publikálás a folyóiratoké, a fontos eredmények széleskörű közzététele pedig a hírügynökségeké és a médiáé. A rangsorolásért, finanszírozásért és hallgatókért folytatott versenyben azonban ez a modell módosult: a kutató egyetemek PR, illetve sajtórészlegeket hoztak létre, és gyakorlattá vált, hogy az egyetemen folyó kutatásokról a saját ügynökség tájékoztatja a médiát. Az egyetem vezetésének elvárása, hogy az intézményben folyó kutatások a lehető leggyorsabban, a lehető legszélesebb körben kapjanak nyilvánosságot. Ismert panasz, hogy a média torzítja a tudományos híreket. A mai helyzetben úgy tűnik viszont, hogy a felelősségben osztoznak a kutatóegyetemek PR-sajtó részlegei, és maguk a kutatók is (vö. Moore, 2006).

Rossz gyakorlattként kialakult az a szokás, hogy a tipikusan kis mintán végzett, első eredményeket hozó kutatásokat megbízható, áttörő eredményként mutatják be a sajtónak, még mielőtt a publikáció megjelenne szakmai bírálatot alkalmazó folyóiratban (vö. Brown, 2012; Brechman, Lee és Cappella, 2011). A torzító túlzások már az egyetemek közleményeiben megjelennek, pedig a sajtóosztályok általában jóváhagyatják a közleményeket magukkal a kutatókkal is. Íme egy példa a szenzációhajhász egyetemi sajtóközleményekre: A Hibbing, Smith és Alford (2014) „*Predisposed: liberals, conservatives and the biology of political differences*” c. publikációját bemutató közlemény így zárul: „*A politika feltehetően nem a lelkünkben van, hanem a DNS-ünkben – állítja John Hibbing és Kevin Smith (UNL) és John Alford (Rice University)*” (Office of University Communications, University of Nebraska-Lincoln, jul. 31, 2014).

## SZEMBENÉZÉS A PROBLÉMÁKKAL

2005-ben John Ioannidis epidemiológus (Stanford University) megjelentetett egy cikket a *PLoS Medicine* folyóiratban, a következő provokatív címmel: „*Why most research findings are false*” (Ioannidis, 2005). A cikk megállapítja, hogy az orvostudományban erősödő aggodalom van a tekintetben, hogy a publikált kutatási eredmények többsége nem valós. Ioannidis nem szándékos hamisításról beszélt, hanem arról, hogy az általa jól ismert kutatási gyakorlat hiányosságai, együttjárva az elterjedt statisztikai elemzések tárházával, hamis eredményekhez vezethetnek még jóhiszemű eljárás mellett is. A bajok forrásaként olyan jellemzőket említett meg, amelyek hosszú évek alatt beépültek a kutatási gyakorlatba és elnézettekké váltak; ilyen a kis mintán végzett adatgyűjtés, az alacsony hatásnagyság elfogadása, a statisztikai programok, amelyek sokféle elemzési lehetőséget kínálnak fel, de a felhasználó nem teljesen érti ezeket. Mindehhez hozzájárul a mindenáron való publikálás motivációja. Ioannidis azt a lehetőséget vetette fel, hogy a megalapozatlan kutatási eredmény az orvostudományi közleményekben nem ritka anomália, hanem tömeges jelenség.

Volt aki vitatta, volt aki megerősítette a drámai felvetést, a cikk mindenesetre nagy figyelmet váltott ki, és egy megismétlési hullámot indított el. Ennek során az Amgen Inc. gyógyszergyártó cég független megismétlést végeztetett a rákkutatás területén mérőföldkőnek tekintett, a *Nature*-ben megjelent eredményekre vonatkozóan. Ennek keretében 53-ból 47 esetben nem sikerült az eredményt reprodukálni (Begley és Ellis, 2012). A Bayer által indított hasonló projekt, amelyik kísérleti stádiumú gyógyszerekre vonatkozó publikált eredményeket ismételtetett meg, 65%-os eredménytelenséget hozott (Mullard, 2011). A rákkutatás területén azóta is több megismétlési projektet végeznek, és a terület komolyan keresi a rossz gyakorlatoktól való megszabadulás módját (vö. Errington, Iorns, Gunn, Tan, Lomax és Nosek, 2014).

2008-ban egy hasonlóan erős nyelvezettel megfogalmazott cikk jelent meg a szociális idegtudományi vizsgálatokban elterjedt kétséges elemzési eljárásokról, amely lehetetlenül magas korrelációs értékek közlésében nyilvánult meg (Vul, Harris, Winkielman és Pashler, 2009; A cikk eredetileg a „*Voodoo correlations in social neuroscience*” címet viselte, amit később a szerzők szerkesztői kérésre „*Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition*” címre változtattak). Vulék sok vizsgálat alapján arra figyelmeztettek, hogy az fMRI-vizsgálatokban a valamilyen szociális válasz és egyes agyterületek aktivitása közötti korrelációs értékek irreálisan magasak, és ezért túlbecsültek. Legfőképpen azért, mert a kutatók szelektíven azokat a mérési pontokat (voxeleket) választják ki a lehetséges választékból, amelyek maximális korrelációt mutatnak. Egy másik elemzés arra mutatott rá, hogy „voodoo” korrelációk nemcsak az idegtudományok terén bukkannak föl (Fiedler, 2011).

Meghatározó publikációk a pszichológia vonatkozásában is születtek, és nagyban hozzájárultak ahhoz, hogy a tudományterület szembenézzen a kialakult rossz eljárásokkal.

2011-ben (két hónappal a Bem-cikk megjelenése után) Joseph Simmons, Leif Nelson és Uri Simonsohn közöltek egy pszichológiai kutatásokat célba vevő cikket: „*False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant*” címmel (Simmons, Nelson és Simonsohn, 2011). A szerzők egy képtelen kérdésre épített szimulációs kísérlettel illusztrálták mondanivalójukat. A kísérleti feltételben virtuális egyetemi hallgatók a Beatles „When I’m Sixty-Four” című számát hallgatták meg, míg a kontrollcsoport másik zenét hallgatott. A szignifikáns eredmény igazolta az abszurd hipotézist: a szám meghallgatása több évvel fiatalabbá tette a vizsgálati személyeket. Az adatokat megszokott eljárásokkal, menet közben ad hoc elemzési döntéseket hozva dolgozták fel. A konklúzió: bármilyen értelmetlen hipotézis könnyen kihozható szignifikáns eredményre a bevett rugalmas gyakorlatnak megfelelő adatelemzéssel.

A cikk fő állítása az volt, hogy bár a pszichológiai kutatások névlegesen a  $\leq .05$  szignifikanciaértéket teljesítik, azaz  $.05$  szinten rögzítik a „hamis pozitív” következtetések valószínűségét, valójában jóval gyakoribb a hamis pozitív eredmények előfordulása. Ez elsősorban az adatgyűjtés és adatelemzés során helytelenül megengedett ad-hoc döntéseknek köszönhető, amelyek segítségével lényegében bármely hipotézis igazolható. A szerzők véleménye szerint a rossz eljárások elterjedtségének következtében akár 60%-ra is felmehet a hamis pozitív eredmények aránya és ezáltal a publikációs torzulás

egy-egy terület szakirodalmában. Ioannidishez (2005) hasonlóan tanulmányuk arra figyelmeztetett, hogy a publikált szakirodalomban messze magasabb lehet a hamis pozitív eredmények aránya, mint bárki is gondolta.

Végül említeni kell Andrew Gelman és Eric Loken 2013-as meghatározóvá vált cikkét (Gelman és Loken, 2013), amely pontos elemzéssel tovább tudatosította, hogy a kutatók akkor is belevihetnek torzítást az adatelemzésbe, ha nem szándékoznak ismert nemkívánatos eljárásokba (pl. p hacking) bocsátkozni.

## A MEGISMÉTLÉSI PROJEKT ÉS AZ ÖNIZSGÁLAT

A fent leírt előzmények összefüggésében érthető, hogy 2011-ben Brian Nosek, a University of Virginia szociálpszichológusa miért indított másokkal együtt kezdeményezést egy, a pszichológiát érintő kiterjedt megismétlési projekt érdekében. A kezdeményezés egyben a *Center for Open Science* létrehozását is eredményezte, melynek célja az, hogy bármely területen elősegítse a tudományos tevékenység nyitottságát, integritását és reprodukálhatóságát (Open Science Collaboration, 2012).

A pszichológiára irányuló projekt 2011 novemberében indult, és az eredményeket 2015-ben publikálták (Open Science Collaboration, 2015). A „*Reproducibility Project: Psychology*” (a továbbiakban: Megismétlési Projekt) keretében 270 önkéntes közreműködő kutató vállalta, hogy elvégzi 100, 2008-ban megjelent vizsgálat pontos megismétlését. A publikációkat a *Psychological Science*, a *Journal of Personality and Social Psychology* és a *Journal of Experimental Psychology: Learning and Memory* című vezető folyóiratokból válogatták ki. Az eredeti szerzők minden esetben együttműködtek a vizsgálat apró részleteinek megosztásában. Az érdeklődéssel övezett eredmények szerint csupán 36,1% volt a sikeres megismétlések száma, ezekben az esetekben is azonban tipikusan kevésbé erős eredményeket kaptak az ismétlésekben (Open Science Collaboration, 2015; Weir, 2015).

A sikertelen megismétlések nagy száma és a vezető laborok érintettsége kétségtelesen rossz hír volt, és az sem nyújthatott vigaszt, hogy más tudományterületeken (így a már említett rákkutatásban és idegtudományban, de a közgazdaságtanban és a kémiában is) hasonló problémák ütötték fel a fejüket. Az értékelés azonban megfontolt értelmezést kíván. Először is felvethető, hogy mit – vagy mit nem – bizonyít egy vizsgálat sikertelen megismétlése? Másrészt megválaszolendő kérdés, hogy mit jelent egy tudományterületre nézve a megismétlési kudarc magas aránya?

Hangsúlyozni kell, hogy a projekt nem azt mutatta ki, hogy a vizsgált kutatások 68,9 százaléka elvetendőnek bizonyult, hanem azt, hogy ennyi esetben nem sikerült, pontos megismétlésre törekedve, az eredetihez közel hasonló eredményt reprodukálni egyetlen független megismétlés alkalmával. Ugyanakkor a pontos közvetlen megismételhetőség a tudományosság egyik alapkritériuma. Nem véletlenül jelentette ki Karl Popper, hogy „*a nem megismételhető egyszeri eredményeknek nincs tudományos jelentőségük*” (Popper, 1959, 66). Ha a fizikában egy kutatócsoport közvetlen bizonyítékot talál a gravitációs hullámok létezésére Einstein száz évvel korábbi hipotézisét megerősítve, de azt más kutatócsoportok nem tudják megerősíteni, akkor kétségessé válik a gravitációs hullámok létezésébe vetett hitünk, még ha Einstein hipotéziséről van is szó.

Másrészt viszont: mi van akkor, ha az ismétlők tévednek, vagy ők az elfogultak? Egyetlen vizsgálatot sem lehet teljesen pontosan megismételni. Mi van akkor, ha az ismétlés megváltoztatja a feltételeket, és ezek az akár csak kismértékű eltérések az okai annak, hogy más eredmény jön ki? És mi van akkor, ha időközben megváltozott maga a jelenség? Többféle oka lehet a pontos megismétlés kudarcának. Felmerül a *konceptuális megismétlés* kérdése is, amely arra utal, hogy hogyan értelmezhető az a helyzet, amikor egy vizsgálatot más kutatások kisebb-nagyobb változtatásokkal ismételnék meg, ugyanarra a fő hipotézisre építve. Mennyiben tekinthető egy ilyen megismétlés pozitív eredménye az eredeti vizsgálat megerősítésének? Továbbá: mi a megismételhetőség elvárható aránya? 100% bizonyosan nem, hiszen még a legjobb minőségű tudományban is jelen van a bizonytalansági faktor. Ezek jogos kérdések és felvetések, a válaszok keresését azonban folytassuk azzal, hogy megvizsgáljuk, milyen meglátásokat kínált fel a problémák hátterét kutató kollektív gondolkodás.

### *1. A pontos megismétlések hiánya a pszichológiai szakirodalomban*

A pontos megismétlések hiánya a szakirodalmat ismerve bárkinek feltűnhet. Makel és munkatársai viszont szisztematikusan megvizsgálták 100 tudományos pszichológiai folyóirat teljes publikációs történetét, és 1900-ig visszamenőleg megállapították, hogy a reprodukálás céljával végzett publikációk 1,07%-ban fordultak elő a vizsgált anyagban (Makel, Plucker és Hegarty, 2012). Igen nagy ellentmondás van aközött, hogy a kutatómódszertan „arany sztenderdnek” tekinti a közvetlen megismétlést (lásd a fenti Popper-idezetet, vagy pl. Jasny, Chin, Chong és Vignitery, 2011 állásfoglalását), a valóságban azonban senki sem műveli. A szakmai közösségekben a megismétlésnek alacsony a presztízse, sőt többnyire a kreativitás hiányának tekintik, ha valaki mások kutatási eredményeit kívánja megismételni. Mint láttuk, a folyóiratok megerősítik ezt a hozzáállást azzal, hogy deklaráltnak nem publikálnak megismétléseket (Makel, Plucker és Hegarty, 2012).

Van, aki szerint ez nem baj. Jason Mitchell szociálpszichológus (Harvard Egyetem) például kifejtette, hogy a megismétlésnek nincs értelme, mert 1. a sikertelen kísérleteknek nincs értelmezhető tudományos jelentésük, ugyanis egy pszichológiai kísérletben mindig annyi rosszul kontrollált vagy kontrollálhatatlan ok játszik közre az eredményben, hogy sosincs kizárva, hogy ezek állnak a sikertelenség hátterében; 2. a megismétlést végzők sem tökéletesek, ők is elkövethetik ugyanazokat vagy más hibákat, sőt elfogultak, mert eleve feltételezik, hogy az eredeti eredmény megbízhatatlan, így viszont a megismétlés sem lehet magasabb mérce (Mitchell, 2014). Mások (pl. Lutus, 2017) elfogadhatatlannak tartják ezt az érvelést. A megismétlésellenes hozzáállás szerint ha már valaki kimutatott egy új eredményt/hatást, még ha kis mértékben és alacsony hatásmagysággal is, és az megjelent egy megfelelő impaktfaktorral rendelkező folyóiratban, akkor az megalapozott eredménynek tekinthető. Ha viszont feladjuk a megismételhetőség kritériumát, akkor feladjuk a tudomány egyik alapkövetelményét.

## *2. Visszás gyakorlat a konceptuális ismétlések terén*

Ha közvetlen megismétlés nem is, de konceptuális megismétlés nagy számban található a pszichológiai kutatásokban és a szakirodalomban. Ezek lényege az, hogy a kutató egy kezdeti vizsgálatban teszti a hipotézisét, és ha szignifikáns eredményt kap, akkor továbblép, hogy kicsit megváltoztatott feltételek mellett tesztelje a kérdést. Meggyőzőek és jól publikálhatóak az olyan kísérletsorozatok, amelyek ilyen módon szolgálnak az eredeti hipotézis kiterjesztésére. Egyes érvelések szerint az ilyen konceptuális megismétlés értékesebb, mint a közvetlen megismétlés, mert egyszerre teszteli az eredeti eredmény validitását és kiterjeszhetőségét, és így megalapozottabbá teszi a konklúziót (az érvet kritikailag elemzi Pashler és Harris, 2012).

A konceptuális ismétlés jó kutatási stratégia, de a kialakult gyakorlat rejt magában problémákat. Ugyanis, ha a konceptuális ismétlés megerősíti az eredeti eredményt, akkor ezt a kutató a hipotézis megerősítésének tekintheti. Ha viszont a konceptuális ismétlés nem erősíti meg az eredeti eredményt, akkor ritkán hajlik a kutató arra, hogy az eredeti hipotézisbe vetett hitét felülvizsgálja, mivel a módosított körülményeket és módszertant teheti felelőssé a kudarcért (vö. Meehl, 1967; Mitchell, 2014). Így a sikeres vizsgálat konceptuális megismétléssé válik, a sikertelen viszont módszertanilag hibás próbálkozásként végzi az ismeretlenség homályában. Amennyiben a konceptuális megismétlés az itt leírt gyakorlatnak megfelelően történik, úgy gyenge bizonyíték, amely nem tekinthető sem az eredeti eredmény megerősítésének, sem érvényes kiterjesztésnek (vö. LeBel és Peters, 2011).

## *3. A negatív eredmények mostoha státusza és a publikációs torzulás*

Egy módszertanilag erős bizonyítási rendszerben a negatív eredménynek tartalmi jelentése van: megkérdőjelezi a vizsgált hipotézis megalapozottságát. Ezzel összefüggésben beszélt Karl Popper arról, hogy egy elmélet csak akkor tekinthető tudományosnak, ha cáfolható (Popper, 1959/2005). Tény azonban, hogy a negatív eredmény jelentése mindig sokértelmű, ezért a negatív eredmény nehezebben értelmezhető, mint a pozitív. Továbbá, a tesztelt elméletbe/hipotézisbe vetett hit erősségével arányos a hajlandóság arra, hogy a negatív eredményt módszertani problémákkal magyarázzuk (LeBel és Peters, 2011).

Nagyobb baj akkor keletkezik, ha egy terület kutatói kultúrájában bevett gyakorlatá válnak a hipotézis megmentésére tett manőverek, miközben a kutatók alig törekszenek komoly kihívás elé állítani elméleti feltevéseiket és megalapozatlannak nyilvánítani azokat, ha nem állják ki a tesztelés próbáját (Fanelli, 2010). Ebben az irányba hat az a publikálási rendszer is, amely nem teszi lehetővé a negatív eredmények közzétételét. Mindezek hatására a kutatói gyakorlatban elsikkad az eredmények időt és erőforrást igénylő türelmes ellenőrzése, amelyben a megerősítésnek (verifikációnak) és a megkérdőjelezésnek (falszifikációnak) egyaránt szerepe van.

Egyes megállapítások szerint a pszichológiai kutatás egész rendszerét áthatja a negatív eredmények iránti averzió. Ez viszont megteremti a „halhatatlan elméletek teme-

tőjét” – a formálisan soha el nem utasított, de egy bizonyos futamidő után csendesen elhagyott elméletek és kutatási programok sorát (Ferguson és Heene, 2012, 555.; lásd még: Heene és Ferguson, 2017). A közvetlen megisméltés hiánya, a konceptuális megisméltések fent leírt rossz gyakorlata és a negatív eredmények mostoha helyzete oda vezet, hogy a szakirodalomba megkérdőjelezetlenül és tartósan beépülhetnek valótlan – tipikusan hamis pozitív – eredmények. Ezekre újabb megalapozatlan eredmények épülhetnek rá, és így akár egész kutatási területek virágozhatnak valódi megalapozottság nélkül (Pashler és Harris, 2012; LeBel és Peters, 2011; Fanelli, 2010).

#### 4. A mintanagysággal kapcsolatos problémák és a hamis pozitív eredmények

Logikus feltételezés, hogy a Megisméltési Projektben mutatkozó nagyarányú megisméltelhetetlenség mögött részben a hamis pozitív eredmények állnak. Mint láttuk, ezek, és velük összefüggésben a megkérdőjelezhető eljárások létre Simmons, Nelson és Simonsohn (2011), majd mások is (pl. Gelman és Loken, 2013; John, Loewenstein és Prelec, 2012) felhívták a figyelmet. A pontos megisméltéssel és a negatív eredményekkel szembeni averzió, háttérben a mindig új, jól publikálható eredmények produkálására való késztetéssel, szintén egyértelműen a hamis pozitív eredmények látens arányát növeli.

A hamis pozitív eredmény valószínűségét növeli az alacsony mintaszám és az ezzel járó alacsony statisztikai erő (Finkel, Eastwick és Reis, 2015). Egy 2011-es elemzés szerint a vizsgált APA folyóiratokban megjelent kutatások átlagos mintaszáma  $N = 40$  volt, ami messze elmarad a kívánatostól (Marszalek, Barber, Kohlhart és Holmes, 2011). A statisztikai erő utólagosan kiszámítható a kísérleti feltételekre jutó vizsgálati személyek száma (cell size), a tipikus hatásméret (effect size) és a statisztikai szignifikancia küszöbértéke (alpha) alapján. Ezzel a módszerrel, kiegészítve több száz kutatóhoz eljuttatott kérdőívvel, egy újabb vizsgálat is megerősítette, hogy a pszichológiai vizsgálatok statisztikai ereje nagyon sok esetben gyenge (Bakker, Hartgerink, Wicherts és van der Maas, 2016). Ez a vizsgálat azt mutatta ki, hogy miközben a kutatók tipikusan intuitív módon határozzák meg a mintanagyságot és azt gondolják, hogy megfelelő mintanagysággal dolgoznak, ez a feltevés formális számítások alapján nem megalapozott. Régóta ismert, hogy az alacsony mintaszám megnöveli a hamis negatív következtetések esélyét, de kevésbé köztudott, hogy az alacsony mintaszám (és ezzel járó alacsony statisztikai erő) a hamis pozitív következtetés esélyét is megnöveli (Button és mtsai, 2013; Gelman és Loken, 2013). Azaz, csak a mintanagyság növelésével csökkenthető egyidejűleg az első- és a másodfajú hiba.

Head és munkatársai (2015) a szakirodalomban közölt szignifikancia- és hatásméretértékeket alapul véve végeztek szisztematikus elemzést igen nagy adatmennyiség alapján adatbányászati módszerekkel, nemcsak a pszichológiára vonatkozóan. Ez az elemzés jelentős mértékű hamis pozitív jelenlétet valószínűsített a vizsgált szakirodalmakban. A hamis pozitív eredmények – ellenőrzés híján – könnyen gyökeret eresztenek, viszont nehezen korrigálhatóak. Jó példa erre a az MMR (bárányhimlő, mumsz, rubeola) oltás és a autizmus kapcsolatát felvető orvostudományi közlemény története. Andrew Wakefield 1998-ban a *The Lancet* nevű orvosi lapban publikálta 12 társszerző-



vel azt a kis mintán végzett kutatást, amelynek konklúziójaként az autizmus kialakulását az MMR vakcinához kapcsolták (Wakefield és mtsai, 1998). A bejelentés azonnal széleskörű nyilvánosságot kapott, és megjelent többek között a *Nature* és az *American Scientist* lapjain is. Ezzel párhuzamosan meredeken esett a vakcinációt kérő szülők száma, és megugrott a megbetegedett gyermekek száma, elhalálozásokat is beleértve. Később Wakefieldék kutatásával kapcsolatban súlyos aggályok merültek fel. A brit *General Medical Council* akadémiai csalásnak nyilvánította a kutatást, a cikket a *The Lancet* 2010-ben visszavonta (Flaherty, 2011). Az erőteljes cáfolatok dacára az oltásellenesség továbbra is makacsul jelen van.

A hamis pozitív eredmények erős jelenléte aggodalmat váltott ki a válságjelenségekkel kapcsolatos önvizsgálatok során, és a reformmozgalom egyik fő célpontjává vált. A túl sok hamis pozitív eredmény kétségtelenül tévútra viheti, akadályozhatja egy-egy tudományterület előrehaladását. Ugyanakkor nem reális elvárás, hogy a tudomány sose járjon tévutakon, és a becslési módszerek sem tökéletesek (vö. Jager és Leek, 2014). Az is nyilvánvaló, hogy a hamis negatív eredmények sem kívánatosak – ez az eset azonban nem sok figyelmet kapott a „válságdiskurzusok” folyamán, mivel kevés detektálható negatív eredményt hoznak a kutatások.

### 5. A nullhipotézis-teszteléssel kapcsolatos problémák, a statisztika fetisizálása

A válságjelenségek okainak keresése közben nagy hangsúlyt kapott a statisztika felhasználása a kutatási gyakorlatban. A nullhipotézis-tesztelés („NHST”: Null hypothesis significance testing) bizonyos előnnyel jár, de állandóan fellángoló viták kísérik (vö. Cohen, 1994; Dienes, 2007; Iacobucci, 2005). Az eljárást számos kritika érte az elmúlt évtizedek alatt, és ezek az utóbbi időben ismét felerősödtek (vö. Bárdits és Németh, 2017; Bárdits, Németh és Terplán, 2015; Morrison és Henkel, 2006; Szucs és Ionnidis, 2017).

Dienes (2007) részletesen kifejti, hogy a hipotézisvizsgálat Neyman és Pearson-féle hagyománya a valószínűség objektív, relatív gyakorisági értelmezésén nyugszik, és ez a statisztika nem képes arra, hogy megmondja: egy adott hipotézis milyen valószínűséggel igaz. Az alfa-értékre alapozó eljárás csupán annyit tesz, hogy döntési szabályt hoz arra nézve, hogy mikor fogadjunk el és mikor vessünk el egy hipotézist, így feltételezhető, hogy hosszú távon viszonylag keveset tévedünk. Az alfa-küszöb nem más, mint egy bizonyos fajta hibaaarány, a p-érték azonban nem mondja meg, hogy a hipotézis igaz-e vagy hamis, és valószínűséget sem rendel a hipotézishez. A kutatók viszont jellemzően úgy értelmezik a p-értéket, mintha az a nullhipotézis valószínűségét fedné fel (Dienes, 2007, ennek okait is kifejti). Az az elképzelés, miszerint a szignifikancia fogalmában bennefoglaltatik a hipotézis valószínűsége, makacs tévhit, sőt logikai hiba (Bárdits, Németh és Terplán, 2016; Dienes, 2007).

Bármennyire konszenzuson alapul a szignifikanciaszint küszöbértéke, mégis egy önkényes határról van szó, amelynek „mágikus” hatalma van eldönteni, hogy egy kutatási eredmény elfogadható-e vagy sem. A küszöbérték használata a komputert megelőző idők hagyatéka, amikor is a kézi számítás nehézségei miatt alakult ki ez a megoldás. Mai körülmények között már a pontos szignifikanciaérték közlése ajánlott a

publikációkban (Ziliak és McClosky, 2009), ez azonban nem oldja meg azt a problémát, hogy a megfelelő  $p$ -érték azt az azonnali jó érzést hozza magával, a fenti téves meggyőződésen alapulva, hogy a vizsgálat „sikeres” volt, és az eredmény „igaz”. Ezzel a publikálhatóság egyik feltétele is teljesül, abból a felfogásból kiindulva, hogy „a nem szignifikáns eredmény nem eredmény”. A statisztikai szignifikanciaszint tesztelés rituáléja (Kirk, 1996) azt az illúziót keltheti, hogy bizonyosságot nyertünk az eredmény igazságával kapcsolatban. Paradox módon a statisztika, amit a bizonytalanság tudományaként írnak le, a bizonyosság illúziójának eszközévé válik (Gelman, 2013).

A szignifikáns eredmény nem jelent védelmet vagy mentességet a rossz módszertantól: a nem valid mérőeszközök, a rossz mintavétel vagy a kontroll során elkövetett hibák hatásainak következményeitől (LeBel és Paunonen, 2011). A szakirodalom elemzése azt mutatja, hogy egy-egy területen tartósan alkalmaznak a kutatók alacsony pszichometriai minőségű mérőeszközöket. Ilyennek mutatkozik például több ToM (elmeolvasási) képességet mérő, a 7–13 éves egészséges populáció vonatkozásában használt teszt (Hayward és Homer, 2017). Az is félreértés, hogy a nullhipotézis-tesztelés biztosítékot nyújtana a hamis eredmények szakirodalomban való elszaporodása ellen. Ugyanis a konvencionális szignifikanciaszint nem szab határt a hamis eredmények terjedésének (Pashler és Harris, 2012, Ioannidis, 2005).

Széles körben elkövetett hiba továbbá a hipotézisvizsgálást exploratív jellegű kutatásoknál alkalmazni, vagy nem is tisztázni, hogy a kutatás exploratív vagy megerősítő jellegű, mert a statisztikai következtetés csak akkor érvényes, ha egy már előzetesen világosan megalapozott hipotézis tesztelésére épül. A kutatások fontos és szükségszerű fázisa az exploráció, amikor inkább hipotézisek generálása és előtesztelése folyik, semmint a már megalapozott hipotézisek tesztelése. Könnyű egy éppen generált hipotézist úgy tekinteni, mint ha az már megalapozott lenne – ez azonban sérti a statisztikai elemzés integritását és a szignifikancia értelmezését (a részletesebb kifejtésre lásd deGroot, 1969/2014; Kerr, 1998, Simmons és mtai, 2011; John és mtai, 2012).

A szignifikanciaszint teszteléssel való elégedetlenség újabb hulláma olyan radikális lépéseket is hozott, mint például a *Basic and Applied Social Psychology* 2015-ös szerkesztőségi állásfoglalása, amely kijelenti, hogy a nullhipotézis-tesztelést érvénytelen eljárásnak tekintik és nem követelik meg a továbbiakban a publikációkban (Trafimow és Marks, 2015). Az erősen kritikus állásfoglalások sora folytatható (pl. Cumming, 2014; Nuzzo, 2014). Szucs és Ioannidis (2017) is részletesen amellett érvelnek, hogy az NHST-nek nem szabadna dominálnia a biomedikális és pszichológiai kutatásokat. Ezzel együtt feltételezhető, hogy a nullhipotézis-tesztelés a pszichológiai kutatások eszköztárában fog maradni. A kritikák hulláma azonban megemelte a problémákkal kapcsolatos tudatosságot, átgondolásra és reformokra készítet a pszichológiai kutatásokban alkalmazott statisztikai eljárásokkal kapcsolatban (vö. Dienes, 2007; Klein, 2005; Vargha, 2016).

## 6. A „WEIRD” egyetemista mint „modell organizmus”

Nem javult az a régi probléma sem, hogy a pszichológiai kutatások túlnyomó többsége egyetemistákból álló kényelmi mintavételre támaszkodik. Az amerikai (vagy más nyugati országbeli) egyetemista „modell organizmusként” való alkalmazása tartósan

jellemző (becslések szerint tartósan 70–90% körüli) a pszichológiai és idegtudományi vizsgálatokban (Kimmel, 1998; Peterson, 2001). Az életkorban, szociális helyzetben és gondolkodásmódban is sajátos jellemzőkkel rendelkező egyetemista mintákon nyert eredmények más populációkra történő általánosítása erősen problematikus. E pont felett azonban rendre átsiklanak a kutatók, akik hajlamosak az általános populációra kiterjeszteni az egyetemista, tipikusan „WEIRD” (white, educated, industrialized, rich, democratic) mintán nyert eredményeket (Editorial, *Nature, Neuroscience*, 2010; Henrich, Heine, Norenzayan, 2010).

Mindez kérdéseket vet fel a megismételhetőséggel kapcsolatban is. A megismételhetőség állandóságot, azonosságot feltételez, holott a szociokulturális jelenségek erősen kontextusfüggőek, és a populációk sokban eltérőek egy kultúrán belül is, hát még kultúrák között. Az univerzalizmus még olyan jelenségekkel kapcsolatban sem feltétlenül tartható, mint a vizuális illúziók; a Müller–Lyer-illúziót például egyes afrikai populációk nem vagy kevésbé erősen észlelik, mint a nyugati világban élő általános populáció (Henrich, Heine, Norenzayan, 2010).

A fentiek érzékeltetik, hogy a 2010-es években alapvető kérdések merültek fel a tudományos pszichológia működésével kapcsolatban. Ilyen átfogó kérdés az, hogy mi számít értékes tudománynak? Mi a helye és szerepe az új eredményeknek, és mennyire fontos, hogy az eredmények szilárdak legyenek? Mit minősítsünk elfogadható módszertani rugalmasságnak és mit elfogadhatatlannak? Egyáltalában, hogyan tud az eddiginél jobban működni a pszichológiai kutatás? A válaszok keresése most is folyik, de már eddig is értékes lépések történtek a problémák megoldása érdekében.

## REFORMMOZGALOM A JÓ ELJÁRÁSOK MEGERŐSÍTÉSE ÉRDEKÉBEN

Van aki reformmozgalomként, van aki módszertani forradalomként értékeli a jó eljárások megerősítésére és a rendszerszintű hibák kiküszöbölésére történt kezdeményezések sorát. Biztos azonban, hogy sokféle változás indult el az elmúlt években. A bizalmi válság egyik oka abban állt, hogy úgy tűnt: a pszichológiai kutatások elfogadott gyakorlatában a kutatók sok mindent megtehettek annak érdekében, hogy bizonyítékot találjanak arra, amire bizonyítékot szerettek volna találni. Ezt a tágabb tudományos intézményesített rendszer akaratlanul is inkább elősegítette, semmint gátolta. A jelentős mértékben „grass root” reformmozgalom egyik fő célja a bizonyítási eljárás integritásának helyreállítása a pszichológiai kutatások széles spektrumában („*evidentiary value movement*”, Finkel, Eastwick és Reis, 2015).

A mozgalom az ismertetett előzményekből fokozatosan nőtt ki. Ide tartozik a *The Open Science Collaboration* (2012) és az ennek keretében létrejött *Open Science Framework* (OSF), amely a kutatások megosztására szolgáló nyitott platform. A mozgalom részének tekinthetőek azok a weboldalak és blogok is, amelyek a feltárt hiányosságokat segítenek orvosolni, így a *PsychFileDrawer.org* (megismérlésre irányuló kutatások archívuma); a *PsychDisclosure.org* (olyan platform, amelyen kutatók azokat a részleteket hozhatják nyilvánosságra, amelyek a publikációkból kimaradnak), vagy például a már említett *Retraction Watch* (amely a már publikált kutatások visszavonását tartja számon).

A reformmozgalom keretében a következő főbb ajánlások és újítások fogalmazódtak meg:

### 1. A kutatási tervek előregisztrációja

A kutatási gyakorlat új eleme az előregisztráció. A holland deGroot volt az első, aki ezzel a javaslattal élt (deGroot, 1969/2014), a hatás azonban igazán csak a 2010-es években felismert problémák nyomán kezd érvényesülni. Az előregisztráció azt jelenti, hogy a kutató/kutatócsoport önként vállalja, hogy az adatgyűjtés megkezdése előtt egy archívumban elhelyezi az adatgyűjtés és elemzés részletes tervét és protokollját. Fontos az alkalmazott mérőeszköz előregisztrációja is, valamint hogy a szerzők nyilatkozzanak arról, hogy a kutatás exploratív vagy megerősítő jellegű. A gyakorlatban az előregisztráció védett online elhelyezést jelent, amelyhez egy megjelölt dátum („embargo”) lejártáig csak korlátozott jogosultsággal lehet hozzáférni.

Az érvelés szerint az előregisztráció gátolja (vagy legalábbis csökkenti) a rossz kutatási szokásokat, növeli a megismételhetőség esélyét, és a publikációs torzulásra is kedvezően hat, mivel a negatív eredményeknek is nyomuk marad, feltételezve, hogy később az eredmények is bekerülnek az archívumba (Chambers, 2014; Wagenmakers és mtsai, 2012; Nosek és Lindsay, 2018). Az előregisztráció ma viszonylag könnyen megvalósítható, erre több platform áll rendelkezésre (például a már említett *Open Science Framework; OSF*). A *Psychological Science* c. folyóirat 2011-ben kezdeményezőként lépett fel és különböző ösztönzőket indított, hogy a szerzők kutatásukat előregisztrálják. A folyóirat vállalta a közvetlen megisméltéses vizsgálatok publikálását akár szignifikáns, akár nem szignifikáns azok eredménye (Association for Psychological Science, n.d.; Mission Statement).

Az előregisztráció egyik fő ellenérve az, hogy túlszabályozottá teszi a kutatás folyamatát, megköti a kutató kezét, és így akadályozza a kutatás közbeni felfedezések lehetőségét (Goldin-Meadow, 2016). Azt is figyelembe kell venni, hogy az előregisztráció nem minden típusú kutatás esetében alkalmazható egyforma módon (lásd pl. longitudinális vizsgálatok; vö. Finkel, Eastwick és Reis, 2015). Az átláthatóság azonban önmagában is érték. Az átláthatóság magában foglalja a kutatások módszertanának minél részletesebb ismertetését a publikációkban, a nyers adatok és a nem publikált elemzések megosztását, és az esetleges szponzorációk nyilvánosságra hozatalát.

### 2. Exploráció és megerősítés elválasztása, mintanagyság, statisztikai erő

A metodológus deGroot – másokkal együtt – hangsúlyozta az exploratív és bizonyító vizsgálatok különbségtételének fontosságát (deGroot, 1969/2014; Wagenmakers és mtsai, 2012). A klinikai kutatásokban különösen nagy jelentősége van (de nemcsak ott) a kettő megkülönböztetésének. A felfedező fázisban kis mintákon, különböző módszertani eljárások próbálgatásával lehet a felmerülő hipotéziseket elsődlegesen tesztelni. Ettől elkülönül a kutatásnak az a fázisa, amikor nagy bizonyító erőt kell alkalmazni, hogy az adott hipotézisről kiderüljön, mennyire igazolható. Fent említett hibás (de elterjedt) gyakorlat az, amikor a kutató a kis mintán végzett szignifikáns eredményt a hipotézisre nézve magas bizonyító erejűnek és általánosíthatónak tekinti.

A statisztikai irodalom hangsúlyozza, hogy csak a megerősítő jellegű vizsgálat nevezhető valódi hipotézistesztelésnek, ez az, amelyre vonatkozóan valódi értelme van

a statisztikai tesztek alkalmazásának (Button és mtsai, 2013; Vargha, 2000). Ennek érdekében hangsúlyozott ajánlás a kvantitatív kutatás „arany sztenderdje”: a kezdeti kis volumenű felfedező vizsgálatokra épülő, nagy statisztikai erővel bíró megerősítő célú vizsgálat, amely adott esetben falszifikáló erővel is bírhat. A statisztikai erő összetett feltételeket jelent: elegendően nagy minta szükséges a hatás detekciójához; magának a hatásnak is elegendően nagyoknak kell lennie; továbbá a mérőeszközöknek is érvényesnek és megbízhatónak kell lenniük (Dienes, 2007). Összességében tehát hangsúlyozott ajánlás a felfedező és a megerősítő kutatási kontextus elhatárolása, és a statisztikai erő növelése.

Ezzel együtt erősen ajánlott elhagyni a nullhipotézisre támaszkodást mint egyetlen módszert, és kiegészíteni azt a hatásméretre és a konfidencia-intervallumra vonatkozó információkkal, továbbá előtérbe helyezni a bayesiánus statisztika alkalmazását (erről részletesen lásd: Dienes, 2007; Henson és Smith, 2000). A kis mintákon ki lehet mutatni erős hatásokat, de a gyengébb (ettől még potenciálisan fontos) hatásokat csak nagy mintákon lehet kimutatni. Az eredményeket viszont a hatásméret mutatóival kell jellemezni, függetlenül attól, hogy kis vagy nagy mintán történt-e az adatfelvétel. Azon túl, hogy van-e hatás, az érdekes kérdés az, hogy ez a hatás mekkora.

### 3. Megismétlés, megismétlés

A Megismétlési Projekt középpontba állította a közvetlen megismétlés szerepét. Egyértelmű alakult ki arra nézve, hogy a megismétlés nem csodaszer, de ahhoz hogy hitelt nyerjen egy vizsgálati eredmény, megismételhetőnek kell lennie független kutatócsoportok által. A megismétlés a tudás megszilárdításának, és egyúttal a tudomány önkorrakciójának az eszköze. Ugyanakkor mások munkájának a megismétlése idő- és energiaigényes vállalkozás, amelyre nehéz rábírnai a kutatókat. Fontos lépések és javaslatok születtek ennek a kérdésnek az előmozdítására is.

Megjelent egy új publikációs műfaj, a „regisztrált megismétlési közlemény” („*registered replication report*”, RRR), amely szerint több független labor egységes protokoll alapján végez pontos megismétlést. 2011-ben a *Perspectives on Psychological Science* c. folyóirat bevezette az ilyen publikációk közzétételét (Simons, Holcombe és Spellman, 2014). A cél eléréséhez vinnének közelebb az „ismételd meg és terjeszd ki” típusú kutatások is, amelyek összekötnék a közvetlen és a koncepcionális megismétlést (Open Science Collaboration, 2015). Nagy lehetőségek rejlenek a kollektív (*crowdsourcing* alapú) nagy kutatási projekteknél, mint amilyen maga a Megismétlési Projekt is volt. Ezekben sok kutató egyenként viszonylag kevés munkát tesz bele a közös protokoll alapján folyó kutatásba, amely azonban a sok közreműködő folytán „megaprojektte” tud válni (Open Science Communication, 2015).

Javaslat született a „Megismétlési Index” („*Replicability Index*”, R-Index) fogalmának bevezetésére, amely egy kutatási területen született eredmények megismételhetőségét jósolja be (Schimmack, 2016). Más javaslatok a BA, MA és PhD diákok bevonását javasolják; eszerint sokkal nagyobb mértékben kellene bátorítani ebben a közegben a megismétlésre épülő kutatásokat, amelyek a szakmai szocializációt is segítenék (Frank és Saxe, 2012). Az Open Science Foundation keretében végzett *Collaborative Replications and Education Project* (<https://osf.io/wfc6u/>) például azonosítja azokat a kutató-

sokat, amelyeket BA oktatás keretében megismételésre ajánlanak, pontos instrukciókat adva a megismételéshez, és platformot biztosítva az eredmények feltöltéséhez.

Ugyanakkor a megismételés kapcsán is felmerül az a szempont, miszerint az nem feltétlenül alkalmazható mindenféle típusú pszichológiai vizsgálatra (Finkel, Eastwick és Reis, 2015). A szociokulturális jelenségek kontextusfüggő, változékony jellegéből adódó korlátokat is figyelembe kell venni. Továbbá azt is, hogy a megismételésnek is csak akkor van bizonyító ereje, ha megfelelő statisztikai erővel rendelkezik a vizsgálat (Francis, 2012).

#### 4. *Változtatni kell az ösztönzési és a kommunikációs rendszeren*

Az ösztönzési rendszer megváltoztatása heves vitákat kiváltó kérdés, ami messze nem csak a pszichológiát érinti. Érdekes fejlemény, hogy a *Journal Impakt Factor* (JIF) heves bírálata bontakozott ki az Amerikai Sejtbiológiai Társaság (American Society for Cell Biology, ASCB) 2012. decemberi San Franciscó-i ülésén, amely egy diszciplínákon átívelő manifesztum megfogalmazását eredményezte. A DORA néven ismertté vált nyilatkozatot (San Francisco Declaration on Research Assessment, é. n.) 82 tudományos szervezet jegyezte, köztük vezető folyóiratok szerkesztősegei, kemelkedő kutatóhelyek és tudománymetrikával foglalkozó szervezetek. A DORA részletesen elemzi a JIF hiányosságait, és elutasítja a JIF egyénekre vonatkoztatott használatát. Az ajánlások arra szólítanak fel, hogy minden tudományág gondolja át saját maga számára a minőségi kutatás koncepcióját, és a tudományos teljesítmények széles spektrumának értékét. A DORA ezzel összefüggésben ösztönzi az alternatív értékelési rendszerek kidolgozását, és a tudományos közösségek részvételét ebben a folyamatban. Megjelent az „*altmetrics*” fogalma: ez azokra a hatásmérő mutatókra terjed ki, amelyek az online publikációmegosztás ma egyre terjedő tudományos ökoszisztémáját jellemzik (pl. letöltések száma a Reserach Gate vagy az academia.edu oldalakon, vö. Priem, Taraborelli, Groth és Neylon, 2010).

## AZ ÖNKORREKCIÓ HAGYOMÁNYOS ÉS „HI-TECH” FORMÁI

Az önkorrekciónak a tudomány alapjellemezője (Merton, 1973). De mit jelent ez ténylegesen? A tudomány egy adott korszakban heterogén státuszú és validitású eljárások, módszerek és bizonyítékok halmaza, amelyben ismeretlen mértékben van egyszerre jelen időtálló tudás és ennek az ellenkezője, miközben mindez folytonosan változik. Az önkorrekciónak az a nagyobb időskiban kibontakozó folyamat, amelyik kirosszálja az érvénytelen tudást. Helytelen azonban azt gondolni, hogy az önkorrekciónak magától érvényesülő automatizmus. Ioannidis (2012) arra figyelmeztet, hogy a nem valós eredmények akár hosszú távon is beépülhetnek. Az önkorrekciónak az intézményes és az egyéni szereplők közös felelőssége, amiért aktív erőfeszítést kell tenni. Fel kell ismerni, ha egy adott időszakban az önkorrekciónak mechanizmusok legyengülnek, és tenni kell megerősítésük érdekében. Ennek hagyományos csatornája a tudományos közegre jellemző kritika és önkritika, az eredmények, elméletek és módszertani eljárások felülvizsgálatára való készenlét.

A reformmozgalom ajánlásai is ezen általános cél érdekében lépnek fel, azonban high-tech oldalról radikálisabb megoldások is születnek. Erre példa a publikációk programozott algoritmusok által való kritikai átvizsgálása. A holland Tilburg Egyetem Metakutatási Központjának (Meta-Research Center) egyik kutatócsoportja Michèle Nuijten vezetésével olyan programot fejlesztett ki, amelyik a világhálón megjelent publikációk statisztikai részeit (a szerző beleegyezése és tudta nélkül) kritikai elemzésnek veti alá. Az eljárás nagy felzúdulást keltett, amikor a világhálóra felkerült azon publikációk listája, amelyeket a program hibásnak talált (az érintett szerzők pedig egy utólagosan küldött emailből értesülhettek erről, vö. Buranyi, 2017). Ettől függetlenül is, ma többen feladatuknak érzik (többnyire fiatal kutatók, és néhány tudománnyal foglalkozó újságíró), hogy aktívan ellenőrizzék a megjelent publikációk megalapozottságát, s nyilvánosságra hozzák, ha gyanús jeleket észlelnek (Brian Wansink lebukását is ilyen „nyomozás” okozta).

## KONKLÚZIÓ

A 2010-es évek válsága tekinthető a régóta hurcolt problémák (statisztikai szignifikancia, publikációs torzulás, hamis pozitív eredmények) és az újabb körülmények között született problémák (csalások felszaporodása, hiteltelen publikációk megjelenése, tudományos kommunikáció problémái) egyvelegének. Több régi problémát az új körülmények tettek hangsúlyosabbá. Így például régebben is történtek tudományos csalások, és feltehetően jelen voltak az ideális sztenderdek el nem érő kutatói eljárások, ezek a jelenségek azonban felerősödtek, nem kis mértékben (de nem egyedül) a produktivitásra irányuló nyomás következtében.

Amennyire szerteágazóak a problémák, olyannyira nehéz a megoldások megtalálása.

A Nature 2017/551-es száma öt kiemelkedő statisztikust kért fel arra, hogy tegyenek javaslatot, miként lehetne jobbá tenni a tudományos kutatásokat (Leek, McShane, Gelman, Colquhoun, Nuijten, 2017). Jeff Leek, a Johns Hopkins School of Public Health kutatója az emberi tényezőt emelte ki: a komputeres és a programozott statisztikai eljárások hatalmas adat- és eredménytömeget dobnak ki, de mindig a tudományos kutató az, akinek értelmeznie kell ezeket. Jobban kell értenünk a kutatók gondolkodásmódját, és ennek figyelembevételével kell felvérteznünk őket a korszerű módszertani és adatelemzési technikákkal.

Blakeley McShane és Andrew Gelman (a Northwestern és a Columbia Egyetem kutatói) a statisztikai szignifikanciára épülő gondolkodásmód elhagyását említették. A lényeg nem a p-érték elvetése, hanem az, hogy a kutatók ezt csak egy mutatóként kezeljék, tekintetbe véve a hatásmagyságot, valamint a kutatási dizájn erejét és az adatok minőségét. A lényeg az, hogy a kutatók ne a hipotézismegtartás, illetve -elvetés dichotómiájában gondolkodjanak, hanem az eredmények komplexen értékelt megalapozottságában. David Colquhoun (University College London) és Michèle Nuijten (Tilburg University) a hamis pozitív eredmények kockázatát emelték ki, hangsúlyozva az előregisztráció és az átláthatóság fontosságát. Steven Goodman (Stanford University) a „tudományos kultúra” döntő szerepét említette: az egyének a tudományos kö-

zösszegek uralkodó normáihoz igazodnak. A megfelelő tudományos kultúra kialakítása és fenntartása közös feladat, amelyben részt kell venniük a folyóiratoknak, a finanszírozóknak és az egyes kutatóknak, különösképpen pedig a vezető kutatóknak.

A tudományos munka jellemzői jelentősen átfőrnélődnak a jelenkor feltételrendszerében. Ennek egyik szembetűnő következménye a tudományos szféra gyors növekedése, és a tudományos produktivitás felgyorsulása. A tudomány hatékonysága és hasznossága, csakúgy, mint a hatásmérés, elsődleges kérdések egy egyre inkább tudományra és digitális technológiákra épülő világban. Mindennél fontosabb azonban a tudomány hitelességének megtartása.

Brian Nosek és munkatársai (Nosek és Bar-Anan, 2012; Nosek, Spies és Motyl, 2012) egy olyan „utópiát” vázolnak fel a jövőre nézve, ahol megvalósul a tudományos eredmények nyílt hozzáférhetősége és átláthatósága, és a tudomány művelését elsősorban a valóság feltárásának szempontja, nem pedig a publikálhatóság vezérli. Lehet, hogy ez az utópia még várat magára, de annak szemtanúi vagyunk, hogy a tudomány átalakuláson megy keresztül és új utakat nyit meg (vö. Bartling és Friesike, 2014). A 2010-es évek tanulságai bizonyosan fontosak a jelen és a jövő számára.

## IRODALOM

- Akst, J. (2012). Scientists review own papers. *The Scientist*, Online, 2012 Oct.3. <https://www.the-scientist.com/?articles.view/articleNo/32810/title/Scientists-Review-Own-Papers/> Letöltve: 2017.12.20.
- Alcock, J. (2003). Give the null hypothesis a chance: Reasons to remain doubtful about the existence of psi. *Journal of Consciousness Studies*, 10(6–7), 29–50.
- Anderson, M. S., Ronning, E. A., De Vries, R., & Martinson B. C. (2007). The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics*, 13, 437–461.
- Bakker, M., Hartgerink, C. H., Wicherts, J. M., & van der Maas, H. L. (2016). Researchers' intuitions about power in psychological research. *Psychological Science*, 27(8), 1069–1077.
- Banks, G. C., Rogelberg, S. G., Woznyj, H. M., Landis, R. S., & Rupp, D. E. (2016). Evidence on questionable research practices: The good, the bad, and the ugly. Editorial. *Business and Psychology*, 31(3), 323–338.
- Barch, D. M., & Yarkoni, T. (2013). Introduction to the special issue on reliability and replication in cognitive and affective neuroscience research. *Cognitive, Affective, and Behavioral Neuroscience*, 13, 687–689.
- Bárdits A., & Németh R. (2017). A statisztikai szignifikanciateszt rítusa – kortárs kritikák; a rítus a szociológiában. *Szociológiai Szemle*, 27(1), 119–125.
- Bárdits A., Németh R., & Terplán G. (2016). Egy régi probléma újra előtérben: a nullhipotézis szignifikanciateszt téves gyakorlata. *Statisztikai Szemle*, 94(1), 52–75.
- Bargh, J. A., Chen M., & Burrows, L. (1996). Automaticity of social behavior: direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology*, 71, 230–244.
- Bartling, S., & Friesike, S. (2014). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Springer Open. Springer International Publishing.
- Begley, C. G., & Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391), 531–533.



- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, *100*(3), 407–425.
- Binswanger, M. (2014). Excellence by nonsense: The competition for publications in modern science. In S. Bartling & S. Friesike (Eds). *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing*. Springer Open (pp. 49–72). Springer International Publishing.
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, *66*(11), 2215–2222.
- Bozzo, A., Bali, K., Evaniew, N., & Ghert, M. (2017). Retractions in cancer research: a systematic survey. *Research Integrity and Peer Review*, *2*(1), 5.
- Brechman, J. M., Lee, C. J., & Cappella, J. N. (2011). Distorting genetic research about cancer: from bench science to press release to published news. *Journal of Communication*, *61*(3), 496–513.
- Brown, P. (2012). Nothing but the truth. Are the media as bad at communicating science as scientists fear? *EMBO Reports*, *13*(11), 964–967.
- Buranyi, S. (2017). The hi-tech war on science fraud. *The Guardian*, online: 2017.02.01. <https://www.theguardian.com/science/2017/feb/01/high-tech-war-on-science> Letöltve: 2018.05.19.
- Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376.
- Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science*, *21*(10), 1363–1368.
- Chambers, C. (2014). Psychology's 'Registration Revolution.'. *The Guardian*. Science section, online: 2014 May 2. <https://www.theguardian.com/science/head-quarters/2014/may/20/psychology-registration-revolution> Letöltve: 2017.12.22.
- Cohen, J. (1995). The earth is round ( $p < .05$ ): Rejoinder. *American Psychologist*, *50*, 997–1103.
- Cuddy, A. (2015). *Presence: Bringing your boldest self to your biggest challenge*. New York: Little, Brown and Company.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1) 7–29.
- Csaba L., Szentes T., & Zalai. E. (2014). Tudományos-e a tudománymérés? Megjegyzések a tudománymetria, az impakt faktor és az MTMT használatához. *Magyar Tudomány*, *175*(4), 442–466.
- De Groot, A. D. (1969/2014). *The meaning of "significance" for different types of research*. Translated and annotated by Eric-Jan Wagenmakers et. al. *Acta Psychologica*, *148*, 188–194.
- Dienes Z. (2007). *Mitől tudomány a pszichológia? A tudományos és statisztikai következtetés alapjai*. Budapest: Akadémiai Kiadó.
- Dilworth, C. (2008). *Scientific progress. A study concerning the nature of the relation between successive scientific theories*. Dordrecht: Kluwer Academic Publishers.
- Dominus, S. (2017) When the revolution came for Amy Cuddy. *The New York Times*, online: 18.10.2017. <https://www.nytimes.com/2017/10/18/magazine/when-the-revolution-came-for-amy-cuddy.html> Letöltve: 2017.12.20.
- DORA / Declaration on Research Assessment (2017). Annual Meeting of The American Society for Cell Biology (ASCB) in San Francisco, CA, (December 16, 2012), Online: <http://www.ascb.org/wp-content/uploads/2017/07/sfdora.pdf> Letöltve: 2017.09.23.
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind? *PloS One*, *7*(1), e29081.
- Durante, K. M., Rae, A., & Griskevicius, V. (2013). The fluctuating female vote: Politics, religion, and the ovulatory cycle. *Psychological Science*, *24*(6), 1007–1016.

- Editorial (2010). The university student as a model organism. *Nature, Neuroscience*, 13(5), 521. <https://www.nature.com/articles/nn0510-521.pdf> Letöltve: 2018.10.28.
- Editorial (2011). Combating scientific misconduct. *National Cell Biology*, 13(1), 1.
- Engber, Daniel (2017). The Trials of Amy Cuddy. A feminist psychologist was dragged through the mud for her mistakes. Did she deserve it? Online: 2017. Oct. 19. [http://www.slate.com/articles/health\\_and\\_science/science/2017/10/did\\_power\\_posing\\_guru\\_amy\\_cuddy\\_deserve\\_her\\_public\\_shaming.html](http://www.slate.com/articles/health_and_science/science/2017/10/did_power_posing_guru_amy_cuddy_deserve_her_public_shaming.html) Letöltve: 2017.12.20.
- Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *Elife*, 3, e04333.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One*, 4(5), 1–11.
- Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891–904.
- Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PloS One*, 5(4), e10271.
- Fanelli, D. (2010). Positive” results increase down the hierarchy of the sciences. *PloS One*, 5.4: e10068.
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90, 891–904.
- Fang, F. C., Steen, R. G., & Casadevall, A. (2012). Misconduct accounts for the majority of retracted scientific publications. *Proceedings of the National Academy of Sciences*, 109(42), 17028–17033.
- Fang, F. C., & Casadevall, A. (2015). Competitive science: is competition ruining science? *Infection and Immunity*, IAI-02939.
- Ferguson, C. J., & Heene, M. (2012). A vast graveyard of undead theories: Publication bias and psychological science's aversion to the null. *Perspectives on Psychological Science*, 7(6), 555–561.
- Fiedler, K. (2011). Voodoo correlations are everywhere – not only in neuroscience. *Perspectives on Psychological Science*, 6(2), 163–171.
- 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(2), 275.
- Flaherty, D. K. (2011). The vaccine-autism connection: a public health crisis caused by unethical medical practices and fraudulent science. *Annals of Pharmacotherapy*, 45(10), 1302–1304.
- Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review*, 19(2), 151–156.
- Franco, G. (2013). Research evaluation and competition for academic positions in occupational medicine. *Archives of Environmental & Occupational Health*, 68(2), 123–127.
- Frank, M. C., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, 7(6), 600–604.
- Gadbury GL, Allison DB (2012). Inappropriate fiddling with statistical analyses to obtain a desirable p-value: Tests to detect its presence in published literature. *PLoS One*, 7(10): e46363. <https://doi.org/10.1371/journal.pone.0046363> Letöltve: 2017.12.10.
- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103: 933–948.
- Gelman, A. (2013). How can statisticians help psychologists do their research better? *Statistical Modeling, Causal Inference, and Social Sciences Blog*. Online: 2013 May 17. <http://andrewgelman.com/2013/05/17/how-can-statisticians-help-psychologists-do-their-research-better/>

- Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition”. Online: 2013. Nov. 14. [http://www.stat.columbia.edu/~gelman/research/unpublished/p\\_hacking.pdf](http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf) Letöltve: 2017.09.20.
- Gelman, Andrew & Fung, K. (2016). The Power of the “Power Pose” Amy Cuddy’s famous finding is the latest example of scientific overreach. *Slate Online*: 2016.01.19. [http://www.slate.com/articles/health\\_and\\_science/science/2016/01/amy\\_cuddy\\_s\\_power\\_pose\\_research\\_is\\_the\\_latest\\_example\\_of\\_scientific\\_overreach.html](http://www.slate.com/articles/health_and_science/science/2016/01/amy_cuddy_s_power_pose_research_is_the_latest_example_of_scientific_overreach.html) Letöltve: 2017.10.01.
- Goldin-Meadow, S. (2016). Why preregistration makes me nervous. *APS Observer*, 29(7). <https://www.psychologicalscience.org/observer/why-preregistration-makes-me-nervous> Letöltve: 2018.12.06.
- Hauser, M. D., Weiss, D., & Marcus, G. (2002). RETRACTED: Rule learning by cotton-top tamarins. *Cognition*, 86(1), B15–B22.
- Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental Psychology*, 35(3), 454–462.
- 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(3), e1002106.
- Heene, M., & Ferguson, C. J. (2017). Psychological science’s aversion to the null, and why many of the things you think are true, aren’t. In S.O. Lilienfeld, & I.D. Waldman, (Eds). *Psychological science under scrutiny: Recent challenges and proposed solutions* (pp. 34–52). John Wiley & Sons.
- Henrich J., Heine S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavior and Brain Sciences*, 33(2–3), 61–83.
- Henson, R. K., & Smith, A. D. (2000). State of the art in statistical significance and effect size reporting: A review of the APA Task Force report and current trends. *Journal of Research & Development in Education*, 33, 285–296.
- Hibbing, J. R., Smith, K. B., & Alford, J. R. (2014). *Predisposed: Liberals, conservatives, and the biology of political differences*. New York and London: Routledge.
- Iacobucci, D. (2005). On p-values. *Journal of Consumer Research*, 32(1) 6–11.
- Interlandi, J. 2006. An unwelcome discovery. New York Times, New York, NY. <http://www.nytimes.com/2006/10/22/magazine/22sciencefraud.html?pagewanted=all&r=0>. Letöltve: 2018.10.28.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124.
- Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7(6), 645–654.
- Jager, L.R., & Leek, J.T. (2014). An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics*, 15: 1–12.
- Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Data replication and reproducibility. Again, and again, and again. *Science*, December 2., 334(6060):1225.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5) 524–532.
- Johnson, J. P. (2011). Cancer researcher fabricated data. *The Scientist*, 2011. Aug. 11. Online: 2011. August 11. <https://www.the-scientist.com/?articles.view/articleNo/31028/title/Cancer-Researcher-Fabricated-Data/> Letöltve: 2018.01.24.
- Johnson, C. S., Smeesters, D., & Wheeler, S. C. (2012). Retraction of Johnson, Smeesters, and Wheeler (2012). *Journal of Personality and Social Psychology*, 103: 605.
- Kahnemann, D. (2011). Thinking fast and slow. New York: Farrar, Straus and Giroux.

- Kennedy, J. E. (2015). Critique of Cumming's „new statistics” for psychological research: A perspective from outside psychology. Online: [http://jeksite.org/psi/critique\\_new\\_stat.pdf](http://jeksite.org/psi/critique_new_stat.pdf) and [http://jeksite.org/psi/critique\\_new\\_stat.htm](http://jeksite.org/psi/critique_new_stat.htm) Letöltve: 2017.03.04.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217.
- Kimmel, A. J. (1998). In defence of deception. *American Psychologist*, 53, 803–804.
- Klein, D. F. (2005). Beyond significance testing: Reforming data analysis methods in behavioral research. *American Journal of Psychiatry*, 162(3), 643-a.
- LeBel, E. P., & Peters, K. R. (2011). Fearing the future of empirical psychology: Bem's (2011) evidence of psi as a case study of deficiencies in modal research practice. *Review of General Psychology*, 15(4), 371–379.
- LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, 37(4), 570–583.
- Leek J., McShane B., Gelman A., Colquhoun D., Nuijten, M. (2017). Five ways to fix statistics: Share analysis plans and results. *Nature*, 551, 557–559.
- Lisberger, S. G. (2013). Sound the alarm: Fraud in neuroscience. In *Cerebrum: the Dana Forum on Brain Science*, (May-June). Dana Foundation. Online: May. 02. [http://dana.org/Cerebrum/2013/Sound\\_the\\_Alarm\\_\\_Fraud\\_in\\_Neuroscience/](http://dana.org/Cerebrum/2013/Sound_the_Alarm__Fraud_in_Neuroscience/) Letöltve: 2018.01.24.
- Lock, S., Wells, F. O. & Farthing M. J. (Eds) (2001). *Fraud and misconduct in biomedical research*. London: BMJ Books.
- Lutus, P. (2017). Psychology and neuroscience. Ont he transition from psychology to neuroscience. *Arachnoid.com blog*, Online: Oct. 18. [https://arachnoid.com/psychology\\_and\\_neuroscience/](https://arachnoid.com/psychology_and_neuroscience/) Letöltve: 2018.01.21.
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7(6), 537–542.
- Marszalek, J. M., Barber, C., Kohlhart, J., & Cooper, B. H. (2011). Sample size in psychological research over the past 30 years. *Perceptual and Motor Skills*, 112(2), 331–348.
- Marton J., Varró A., & Varró V. (2004). Impaktfaktor és tudományos teljesítmény. *Magyar Tudomány*, 49(12), 1395–1403.
- McCook, Alison (2017). When a paper retracts 107 papers for fake reviews, it pays a price. *Retraction Watch*. Online: 2017. August 16. <http://retractionwatch.com/2017/08/16/journal-retracts-107-papers-fake-reviews-pays-price/#more-51443> Letöltve: 2017.09.20.
- Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34(2), 103–115.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press.
- Michigan State University. 'Power poses' don't work, eleven new studies suggest. *ScienceDaily*. Online: 2017 September 11. [www.sciencedaily.com/releases/2017/09/170911095932.htm](http://www.sciencedaily.com/releases/2017/09/170911095932.htm)
- Miller, A. N., Taylor, S. G., & Bedeian, A. G. (2011). Publish or perish: academic life as management faculty live it. *Career Development International*, 16(5), 422–445.
- Mitchell, J. (2014). *On the evidentiary emptiness of failed replications*. Working paper. [http://jason-mitchell.fas.harvard.edu/Papers/Mitchell\\_failed\\_science\\_2014.pdf](http://jason-mitchell.fas.harvard.edu/Papers/Mitchell_failed_science_2014.pdf) Letöltve: 2018.01.22.
- Moore, A. (2006). Bad science in the headlines. Who takes responsibility when science is distorted in the mass media? *EMBO Reports*, 7(12), 1193–1196.
- Morrison, D. E. & Henkel, R. E. (Eds) (2006). *The Significance Test Controversy: A Reader*. New Brunswick: Aldine Transaction Publishers.
- Mullard, A. (2011). Reliability of 'new drug target' claims called into question *Nature Reviews Drug Discovery*, 10, 643–644

- Neuroskeptic. (2012). The nine circles of scientific hell. *Perspectives on Psychological Science*, 7, 643–644.
- Newcomer, E. P. & Spitzer, E. A. L. (2010). Marc Hauser's fall from grace. *The Harvard Crimson* www.thecrimson.com/article/2010/9/14/hauser-lab-research-professor/ Letöltve: 2017. 12.20.
- Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, 23(3), 217–243.
- 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(6), 615–631.
- Nosek, B. A., & Lindsay, D. S. (2018). Preregistration becoming the norm in psychological science. *APS Observer*, 31(30/3). <https://www.psychologicalscience.org/observer/preregistration-becoming-the-norm-in-psychological-science> Letöltve: 2018.12.16.
- Nuzzo, R. (2014). Scientific method: Statistical errors: P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume. *Nature*, 506,150–152.
- O'Grady, C. (2017). New papers were found through investigations into previous fraud. *Ars Technica* , <https://arstechnica.com/science/2017/04/107-cancer-papers-retracted-due-to-peer-review-fraud/> Letöltve: 2017.09.20.
- Office of University Communications, Univeristy of Nebraska-Lincoln (2014). *Scientists find growing consensus: Political attitudes derive from body and mind*. Online: 2014.06.31. <http://newsroom.unl.edu/releases/2014/07/31/Scientists+find+growing+consensus%3A+Political+attitudes+derive+from+body+and+mind> Letöltve: 2017.09.20.
- Open Science Collaboration (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, (6) 657–660.
- Open Science Collaboration (2015). Estimating the reproducibility of Psychological Science. *Science*, 349 (6251), aac4716
- Palus, Shannon (2015). Diderick Stapel now has 58 retractions. *Retraction Watch*, 2015.12.08. <http://retractionwatch.com/2015/12/08/diederik-stapel-now-has-58-retractions/> Letöltve: 2017.09.20.
- Papp Z. (2004). A tudományos teljesítmény mérésének problémáiról. *Magyar Tudomány*, 49(2), 232–240.
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531–536.
- 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.
- Pautasso, M. (2010). A review of the worsening file drawer problem. *Scientometrics*, 85, 193–202.
- Peterson, R. A. (2001). On the Use of College Students in Social Science Research: Insights from a Second-Order Meta-analysis. *Journal of Consumer Research*, 28, 450–461.
- Popper, K. R. (1959). *The logic of scientific discovery* New York, Routledge.
- Popper, K. (1959/2005). *The logic of scientific discovery*. London: Routledge.
- Priem, J. D., Taraborelli, P., Groth, C. Neylon (2010). Altmetrics: A manifesto, 26 October 2010. <http://altmetrics.org/manifesto> Letöltve: 2018.05.22.
- Randall, D., & Welsler, C. (2018). *The irreproducibility crisis of modern science. Causes, Consequences, and the Road to Reform, April 2018*. Report. National Association of Scholars. Online: [https://www.nas.org/images/documents/irreproducibility\\_report/NAS\\_irreproducibilityReport.pdf](https://www.nas.org/images/documents/irreproducibility_report/NAS_irreproducibilityReport.pdf) Letöltve: 2018.09.21.

- Rennie, D. & Gunsalus, C. K. (2001). Regulations on scientific misconduct: Lessons from the US experience. In S. Lock, O. Wells, & M. J. Farthing (Eds) *Fraud and misconduct in biomedical research* (13–31). London: BMJ Books.
- Rosenthal, R. (1979). An introduction to the file drawer problem. *Psychological Bulletin*, 86, 638–641.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3) 638.
- Sanna, L. J., Chang, E. C., Miceli, P. M., & Lundberg, K. B. (2011). RETRACTED: Rising up to higher virtues: Experiencing elevated physical height uplifts prosocial actions. *Journal of Experimental Social Psychology*, 47(2), 472–476.
- Sarewitz, D. (2016). The pressure to publish pushes down quality. *Nature*, 533(7602). Online: 2016. May 11. <https://www.nature.com/news/the-pressure-to-publish-pushes-down-quality-1.19887> Letöltve: 2017.09.20.
- Schimmack, U. (2016). The Replicability-Index: Quantifying Statistical Research Integrity. <https://wordpress.com/post/replication-index.wordpress.com/920> Letöltve: 2017. 09.20.
- Schimmack, U. (2016). A revised introduction to the R-Index. *Replicability-Index: Improving the Replicability of Empirical Research*. Online: 2016.01.31. <https://replicationindex.wordpress.com/2016/01/31/a-revised-introduction-to-the-r-index> Letöltve: 2017.09.20.
- Schimmack, U., Heene, M., & Kesavan, K. (2017). Reconstruction of a train wreck: How priming research went off the rails. *Replicability-Index: Improving the Replicability of Empirical Research*. Online: 2017 February 2. <https://replicationindex.wordpress.com/2017/02/02/reconstruction-of-a-train-wreck-how-priming-research-went-off-the-rails/> Letöltve:2017.09.27.
- Shea, C. (2012). The data vigilante. *The Atlantic*, Online: 2012 December <https://www.theatlantic.com/magazine/archive/2012/12/the-data-vigilante/309172/> Letöltve: 2017.09.20.
- Simmons, J. P. & Simonsohn, U. (2015). Power Posing: Reassessing The Evidence Behind The Most Popular TED Talk. Blog post Data Colada, Online: 2015.05.08. <http://datacolada.org/37> Letöltve: 2017.09.19.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11) 1359–1366.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science*, 13(2), 255–259.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science*, 13(2), 255–259.
- 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(5) 552–555.
- Singal, J. (2015). The Case of the Amazing Gay-Marriage Data: How a graduate student reluctantly uncovered a huge scientific fraud. *Science of US*. <http://nymag.com/scienceofus/2015/05/how-a-grad-student-uncovered-a-huge-fraud.html> Letöltve: 2017.09.20.
- Singal, J. (2016). Power Posing' Co-author: 'I Do Not Believe That 'Power Pose' Effects Are Real' *New York Magazine*, 2016.09.16. <https://www.thecut.com/2016/09/power-poses-co-author-i-dont-think-power-poses-are-real.html> Letöltve: 2017.09.20.
- Smith, C. T. (2013). PsychDisclosure. org: Grassroots support for reforming reporting standards in psychology. *Perspectives on Psychological Science*, 8(4), 424–432.
- Stapel, D. (2014). Faking science: A true story of academic fraud. Translated by Nicholas J. L. Brown. Online: 2014.12.14. <https://errorstatistics.files.wordpress.com/2014/12/faking-science-20141214.pdf> Letöltve: 2017.09.20.

- Stapel, D. A., & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science*, 332(6026) 251–253. RETRACTED
- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance: Or vice versa. *Journal of the American Statistical Association*, 54(285), 30–34.
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice-versa. [Editorial Material]. *American Statistician*, 49(1) 108–112.
- Szokolszky Á. (2018). Ezután mit gondoljunk a stanfordi börtönkísérletről? *Mindset*, I. rész, online: 2018.11.04. <https://mindsetpszichologia.hu/2018/11/04/szokolszky-agnes-ezutan-mit-gondoljunk-a-stanfordi-bortonkiserletrol-i-resz/>; II. rész, online: 2018.11.09. <https://mindsetpszichologia.hu/2018/11/09/szokolszky-agnes-ezutan-mit-gondoljunk-a-stanfordi-bortonkiserletrol-ii-resz/>; III. rész, online: 2018.11.11. <https://mindsetpszichologia.hu/2018/11/11/szokolszky-agnes-ezutan-mit-gondoljunk-a-stanfordi-bortonkiserletrol-iii-resz/>
- Szűcs, D., & Ioannidis, J. (2017). When null hypothesis significance testing is unsuitable for research: a reassessment. *Frontiers in Human Neuroscience*, 11, 390.
- Tóth J. (2014). Akadémiai függőség és hamisítványok: az impaktfaktor-fetiszmus egy kísérő jelenségéről. *BUKSZ*, 26(3) 203–205.
- Trafimow, D., & Marks, M. (2015). Editorial //On the invalidity of statistical significance testing // *Basic and Applied Social Psychology*, 37(1–2) 1–2.
- van der Zee, T. (2017). The Wansink Dossier: An Overview. *The Skeptical Scientist* blog, Online: 2017.03.21. <http://www.timvanderzee.com/the-wansink-dossier-an-overview/> Letöltve: 2018.01.21.
- Van Kolschooten, F. (2014). Fresh misconduct charges hit Dutch social psychology. *Science*, 344(6184) 566–567.
- Vargha A. (2016). Szignifikanciaszűrés – negyven éve hibás elemzéseket végzek és téveszméket tanítok? *Statistikai Szemle*, 94(4) 445–451.
- Vargha, A. (2000). *Matematikai statisztikai, nyelvészeti és biológiai alkalmazásokkal: felsőoktatási tankönyv*. Budapest: Pólya Kiadó.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science*, 4, 274–290.
- Waaiker, C. J., Teelken, C., Wouters, P. F., & van der Weijden, I. C. (2017). Competition in science: links between publication pressure, grant pressure and the academic job market. *Higher Education Policy*, 1–19.
- 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(6), 632–638.
- Wakefield, A. J., Murch, S. H., Anthony, A., Linnell, J., Casson, D. M., Malik, M., & Valentine, A. (1998). Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet*, 351(9103) 637–641. RETRACTED.
- Weir, K. (2015). A reproducibility crisis? *American Psychological Association. Monitor on psychology*, 46(9). Online: <http://www.apa.org/monitor/2015/10/share-reproducibility.aspx> Letöltve: 2017.09.20.
- Wesol, M. van (2016). Evaluation by citation: Trends in publication behavior, evaluation criteria, and the strive for high impact publications. *Science and Engineering Ethics*, 22(1) 199–225.
- Yong, E. (2012a). Replication studies: Bad copy. *Nature*, News, 485(7398) Online: 2012 May 16. <https://www.nature.com/news/replication-studies-bad-copy-1.10634> Letöltve: 2018.01.20.

- Yong, E. (2012b). Uncertainty shrouds psychologist's resignation. *Nature, News*. Online: 2012 July 12. <https://www.nature.com/news/uncertainty-shrouds-psychologist-s-resignation-1.10968> Letöltve: 2017.09.20.
- Yong, E. (2012c). Nobel laureate challenges psychologists to clean up their act. *Nature, News*, 485(7398) Online: 212 May 16. <https://www.nature.com/news/replication-studies-bad-copy-1.10634> Letöltve: 2018.01.20.
- Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance*. Ann Arbor: University of Michigan Press.
- Zittoun, T., Gillespie, A., & Cornish, F. (2009). Fragmentation or differentiation: questioning the crisis in psychology. *Integrative Psychological and Behavioral Science*, 43(2) 104–115.

## DYSFUNCTIONS AND REFORM MOVEMENT IN PSYCHOLOGICAL SCIENCE, IN THE BEGINNING OF THE 21ST CENTURY

SZOKOLSZKY, ÁGNES

*Negative tendencies have surfaced in psychological research in the early 2010's, although other scientific fields have also registered these problems. Such was the exposition of multiple academic frauds, the increase of publication retractions, and evidence for the extensive presence of questionable research practices. Prestigious academic journals published studies of dubious quality, and new reasons for distortions in the communication of scientific results have surfaced. The feeling of a confidentiality crisis culminated as findings of the Replication Project (2011–2015), led by Brian Nosek and others, were published. Beyond the ethical issues, awareness of the following methodological problems have stood out: lack of direct replication of results; issues with conceptual replications; neglect of negative results and related publication bias; wide spread presence of false positive results; problems with null hypothesis testing and interpretation of statistics. By the end of the 2010's psychology has developed a better understanding of the crisis. A reform movement is on the way, as a consequence, affecting research methodology, research ethics, and publication practices. In this paper I review these critical developments which have an impact on the advancement of psychological science. Individual scientists, as well as research communities need to be aware of the new recommendations and critically evaluate their own research practices.*

**Keywords:** *data manipulation, evidentiary value movement, methodological reform in psychological research, NHST, questionable research practices, replication crisis in psychology, Repliation Project, research fraud, scientific misconduct*