

АНЂЕЛА Д. КЕЉАНОВИЋ¹

УНИВЕРЗИТЕТ У ПРИШТИНИ СА ПРИВРЕМЕНИМ СЕДИШТЕМ
У КОСОВСКОЈ МИТРОВИЦИ, ФИЛОЗОФСКИ ФАКУЛТЕТ
КАТЕДРА ЗА ПСИХОЛОГИЈУ

ПСИХОЛОШКА НАУКА ОД „ОБЈАВИ ИЛИ НЕСТАНИ“ ДО „ВЕРУЈ, АЛИ ПРОВЕРИ“

САЖЕТАК. У време када су социјални психолози веровали да могу да буду поносни на своју дисциплину дошло је до разорне вести да је Дидерик Стапел починио велику научну превару. Овај догађај се поклопио са почетком расправе о поверењу у психолошке налазе. Убрзо је уследио извештај серије од 9 студија које нису успеле да понове „професорску студију“. Ови резултати репликација били су запањујући због ранијих извештаја о успешним репликацијама. Услед постојања кризе поверења у резултате истраживања на терену, Open Science Collaboration спровео је након тога репликацију 100 корелационих и експерименталних студија објављених 2008. у часописима *Psychological Science*, *Journal of Personality and Social Psychology* и *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Од 97% оригиналних студија које су имале позитиван ефекат, реплицирано је 36%. Међутим, и њихови налази су доведени у питање, рачунањем бајесовог фактора. Осим превара, и упитне истраживачке праксе настале услед пристрасности у објављивању које резултирају лажно позитивним налазима подривају поверење у ваљаност резултата психолошких истраживања. Можда је најскупља грешка лажно позитивних налаза погрешно одбацавање нулте хипотезе. Међутим, да је Стапел (2011) потврдио нулту хипотезу или да је Барг (1996) открио да примовање учесника не утиче на брзину ходања или да су Дајкстерхојс и Ван Книпенберг (1998) објавили да учесници примовани речју „професор“ нису побољшали свој учинак на задатку, нико не би био заинтересован за њихове налазе. Нулти налази су занимљиви само

¹ andjela.keljanovic@pr.ac.rs

Рад је примљен 17. јануара 2020, а прихваћен за објављивање на састанку Редакције Зборника одржаном 24. марта 2021.

ако су у супротности са главном хипотезом изведеном из теорије или су у супротности са низом ранијих студија. Чињеница да се добро експериментално истраживање обично спроводи са циљем да се тестирају теорије, истраживачи никада не могу бити сигурни јесу ли одабрали оптималну операционализацију датог конструкта. Како истраживачи никада не могу бити сигурни да су правилно операционализовали теоријске конструкте које процењују и да ли су били успешни у контроли трећих варијабли које могу бити одговорне за њихове резултате, теорија никада не може бити доказана као истинита.

КЉУЧНЕ РЕЧИ: репликација; пристрасност објављивања; психологија.

Када људи говоре о „кризи“ у било којој области која постоји дуго, као што је експериментална психологија (и наука уопште), мера скепсе је вероватно врло разумна реакција (Pashler & Harris, 2012). Све више критика се упућује психолозима на начин на који спроводе истраживања. Критике, као и неуспех у реплицирању студија, коришћене су за проглашење кризе репликације у психологији. Постоји тренутна криза поверења у психолошку науку која одражава невиђени ниво сумње међу практичарима о поузданости резултата истраживања на терену (Pashler & Wagenmakers, 2012). Криза репликабилности се првенствено заснива на епистемолошком неспоразуму који наглашава феномен уместо његових темељних механизма (Stroebe & Strack, 2014). Криза поверења у психолошку науку јавља се серијом несрећних догађаја, најпре имамо случај преваре Стапел (Stroebe, Postmes, Spears, 2012), потом објављивање „доказа“ о екстрасензорној перцепцији (Vem, 2011), онда скретање пажње на проблеме истраживача у реплицирању познатих резултата (Bower, 2012), до пројекта *Reproducibility* (Open Science Collaboration, 2015) у који се полагало много наде, након чега су уследиле и замерке на пројекат (Etz & Vandekerckhove, 2016; Gilbert, King, Pettigrew, Wilson, 2016). У чланку (Neuroskeptic, 2012) о круговима научног пакла, група аутора на шаљив начин приказује каква судбина чека научнике који греше, а сваки круг пакла резервисан је за неку упитну истраживачку праксу. Од лимба, где су сви они који су „зажмурили“ на проблематичне научне праксе кроз подржавање, награђивање, промовисање истраживача који их чине преко „overselling“ – пренаглашавања важности налаза, ХАРКинга, лова на п вредност, до креативне употребе аутлајера, плагијаризама и не-

објављивања података (file drawer phenomenon), парцијалног објављивања података (cherry-picking) и њихових измишљања (Neuroskeptic, 2012).

Две врсте догађаја су подстакле перцепцију кризе у овој научној области. Најпре, тврдње да се неки психолошки истраживачи баве упитним истраживачким праксама које резултирају лажно позитивним налазима (John, Loewenstein, Prelec, 2012; Simmons, Nelson, Simonsohn, 2011), а потом и велики број неуспеха да се реплицирају познати експерименти (Doyen, Klein, Pichon, Cleere-mans, 2012; Shanks et al., 2013).

„ОБЈАВИ ИЛИ НЕСТАНИ“ – СЛУЧАЈ СТАПЕЛ

Недолично научно понашање је дефинисано као измишљање (потпуно измишљање скупа података), фалсификовање (манипулисање опремом или промена података) или плагијаризам у предлагању, извођењу или преиспитивању истраживања или пријављивању резултата истраживања. Такво недолично понашање мора се починити намерно и свесно (National Science Foundation, 2001).

Истраживачка превара може имати четири врсте спољних ефеката (Stroebe et al., 2012). Прво, може наштетити каријери студената и колега који несвесно учествују у радовима који се заснивају на лажним истраживањима. Друго, у случајевима клиничких истраживања пацијенти могу претрпети штету. Треће, превара може одложити научни напредак јер истраживачи троше драгоцене ресурсе. И као крајњи продукт су вести о научним преварама које штете имиџу поља и смањују поверење у науку. Још увек не постоје тачне информације о раширености преваре у науци, па чак ни о броју случајева преваре које се откривају на годишњем нивоу (Stroebe et al., 2012). У једном истраживању (Fanelli, 2009) 1,97% научника је признало да је бар једном измислило, фалсификовало или модификовало податке или резултате истраживања, док је 33,7% признало друге упитне истраживачке праксе. Приликом процене понашања колега, стопа пријављивања износила је 14,12% за фалсификовање и до 72% за друге упитне истраживачке праксе (Fanelli, 2009). Вероватно да култура „објави или нестани“ доприноси да недолична понашања превладавају у науци (Drenth, 2013). Додатну препреку ствара и свест да су многи укључени у упитне истраживачке праксе (Keljanović,

2020), као и потреба неких истраживача да своје податке не учине јавно доступним.

Вест да је цењени холандски социјални психолог Дидерик Стапел оптужен за превару била је шок за научну заједницу (Stroebe et al., 2012). У његовом случају било је речи о фалсификовању и измишљању, а након детаљне истраге свих 137 публикација, три одбора су закључила да је у 25 чланка манипулисано подацима, а да су подаци измишљени у 30 чланака. За 10 чланака је закључено да је превара била вероватна (засновано на статистичкој анализи). У 10 од 18 докторских дисертација коришћени су измишљени подаци (Drenth, 2013). Приликом анализе публикација утврђено је да се служио и другим упитним истраживачким праксама, попут повећања величине узорка док се не утврде значајне разлике, брисањем резултата неких експеримената, уклањањем непожељних опажања пост-хоц оправдањима, пропустима у експерименталном тестирању, непотпуним или нетачним извештавањем (Drenth, 2013). Превару су открили његови студенти, а сумња се најпре појавила приликом читања рукописа истраживања спроведеног на узорку деце школског узраста чија је просечна старост била 19 година, а још уверљивији доказ био је идентичан низ резултата у другој студији (Stroebe et al., 2012). Постоје обрасци понашања који су карактеристични за научнике који чине преваре, они су или веома поштовани или објављују радове са старијим цењеним колегама (Stroebe et al., 2012).

ЗАШТО СЕ ОДНОС СТЕРЕОТИПА О СТАРИМ ОСОБАМА И БРЗИНЕ ХОДАЊА ПРОМЕНИО ОД ОРИГИНАЛНОГ ИСТРАЖИВАЊА ДО РЕПЛИКАЦИЈЕ?

Барг и сарадници (Bargh, Chen, Burrows, 1996) излагали су своје испитанике у експерименталној групи шифрованим реченицама које треба да изазову стереотип о старим особама. Након експеримента, други експериментатор је мерио колико брзо ће се учесници кретати низ ходник. Учесници који су били примовани речима које изазивају стереотипе о старијим особама кретали су се знатно спорије ходником од испитаника који су били у контролној групи.

Репликације од стране истих аутора, користећи различите учеснике, резултирале су готово идентичним резултатима, као и концептуалне репликације које су успешно понављале налаз (Macrae & Johnston, 1998; Dijksterhuis, Spears, Lepinasse, 2001; Kawakami, Young, Dovidio, 2002). Дајкстерхојс и сарадници (2001) по-

казују да примовање стереотипом о старим особама успорава реакције на задатку лексичке одлуке, а исти ефекат је добијен и у студији других аутора (Kawakami et al., 2002). Са друге стране, аутори су открили да учесници брже читају листу речи када су примовани именом возача формуле 1, него неутралним стимулусом (Macrae & Johnston, 1998).

Једина директна репликација која је реплицирала оригиналну студију (а није успела да понови резултате оригиналне студије) (Doyen et al., 2012), с једином разликом да брзина хода није била мерена штоперицом него инфрацрвеним сензором, није пронашла никакву разлику између контролне и експерименталне групе у првом експерименту упркос већем узорку и употреби аутоматизованих метода мерења брзине хода. У другом експерименту су манипулисали веровањима експериментатора, па је половина наведена да мисли да ће учесници ходати спорије када се изложе стереотипу о старима, а друга половина је наведена да очекује супротно. Тада је примећен ефекат брзине ходања и то у групи где су експериментатори били упућени да очекују да ће учесници из експерименталне групе ходати спорије због ефекта примовања (Doyen et al., 2012), што доводи до тога да су у изворном истраживању резултати можда узроковани очекивањем експериментатора (Stroebe & Strack, 2014). Међутим, у оригиналном раду (Bargh et al., 1996) наведено је да експериментатори нису били упознати са сврхом експеримента, па остаје нејасно како су очекивања експериментатора могла утицати на брзину кретања учесника. Чињеница да је примовање појединца стереотипом старим особама резултирало смањењем брзине ходања био је неочекиван налаз, иако је било конзистентно са постојећим теоријским знањем, није постојао консензус о процесима који посредују у утицају стереотипа на брзину ходања (Stroebe & Strack, 2014).

ДА ЛИ НАС ПРАЈМИНГ СПЕЦИФИЧНИМ ПОЈМОВИМА ЧИНИ УСПЕШНИЈИМ – „ПРОФЕСОРСКА СТУДИЈА“

Примовање учесника категоријом „особа која се обично сматра високо интелигентном“ (нпр. професор) или категоријом „особа која се сматра мање интелигентном“ (нпр. хулиган), а потом и проверавање општег знања, довело је до закључка да размишљање о интелигентној особи резултира бољим извођењем тривијалног задатка од размишљања о мање интелигентној особи (Di-

jksterhuis & van Knippenberg, 1998). Спровели су четири истраживања где се примарна манипулација састојала од тога да учесници мисле о типичном представнику категорије (професор/хулиган), затим описују његов стил живота, атрибуте, понашање, а у другом делу експеримента, наизглед неповезаном, одговарају на питања из опште културе.

Извештај (Shanks et al., 2013) директне репликације, серија од 9 експеримената са 475 учесника, није показао да манипулација има статистички значајан утицај на перформансе ни у једној студији. Ови налази нису само у нескладу са оригиналном студијом, већ су и неспојиви са другим налазима који показују да стереотипи могу утицати на индивидуалне перформансе интелектуалних задатака (Stroebe & Strack, 2014).

Неуспела понављања „професорске студије“, чак и ако се чини да су професори сматрани интелигентнијим од фудбалских хулигана од стране ученика који су служили као субјекти, показују да манипулација примовањем можда није успела повећати когнитивну репрезентацију концепта интелигенције. Могуће је да су ови налази у већини уџбеника за социјалну психологију и, како су широко познати, могли су да утичу на резултате. Други могући разлог би могао бити избор знања. Ако је ефекат примовања мотивациони, онда и предмет знања мора бити одабран да одражава ефекте мотивације (Stroebe & Strack, 2014).

ДА ЛИ „PSI“ ПОСТОЈИ И МОЖЕ ЛИ СЕ РЕПЛИЦИРАТИ?

У психологији се често порекло репликационе кризе везује за радове аутора који извештава да постоји „psi“, односно да постоји екстрасензорна перцепција (Вем, 2011). „Psi“ означава процес преноса информација или енергије који се не може објаснити физичким или биолошких механизмима, и аутор (Вем, 2011) разликује две врсте: прекогницију и предосећање, који представљају утицај неког будућег догађаја на тренутне одговоре појединца. Бем (Вем, 2011) покушава да докаже да будући догађаји ретроактивно утичу на одговоре људи. Чланак извештава о 9 експеримената са више од 1.000 учесника, а сви осим једног експеримента дали су статистички значајне резултате. Бем (Вем, 2011) наводи у самом раду да сумња у могућност репликације студије, јер се добијени ефекти тешко могу реплицирати нарочито на узорку који се значајно разликује од коришћеног. Сматра да је једна од главних препрека успешној репликацији утицај експе-

риментатора на резултате, што показују и три експеримента дизајнирана да истраже ефекат експериментатора, у једној групи су они који су промотери пси, а у другој скептици. У два од три експеримента, промотери екстрасензорне перцепције постижу значајан резултат (Wiseman & Schlitz, 1997; 1999; Schlitz, Wiseman, Radin, Watt, 2005). Зато Бем (Bem, 2011) ефекат експериментатора покушава да сведе на најмању могућу меру, препуштајући упутства и интеракције са учесником рачунару. Овај рад је прошао кроз стандардни поступак рецензије и објављен је у часопису високог утицаја – *Journal of Personality and Social Psychology*. Контроверзна природа налаза инспирисала је три независне студије репликације, од којих ниједна није успела да понови резултате. Ове студије су одбијене да буду објављене, уз образложење да репликације нису биле довољно оригиналне (Fidler & Wilcox, 2018). Поступак 8. и 9. експеримента о екстрасензорној перцепцији (Bem, 2011) поновљен је на узорку од 3.289 испитаника кроз 7 експеримената и истраживачи (Galak, LeBoeuf, Nelson, Simmons, 2012) нису успели да понове налаз. Метаанализом свих покушаја репликације овог експеримента открили су да се просечна величина ефекта не разликује од 0 (Galak et al., 2012). Поновљеним анализирањем добијених података Бајесовим т-тестом показано је да су докази за „psi“ слаби до непостојећи (Wagenmakers, Wetzels, Borsboom, van der Maas, 2011) уз закључак да „ \bar{u} “ вредности не доказују постојање прекогниције, него да указују на то да експериментални психолози треба да промене начин на који спроводе експерименте и анализирају податке.

НЕКЕ СТВАРИ ЈЕ БОЉЕ ОСТАВИТИ НЕИЗГОВОРЕНИМ (ВЕРБАЛНО ЗАТАМЊИВАЊЕ ВИЗУЕЛНИХ СЕЋАЊА)

У школи су нас учили да вербална обрада углавном побољшава перформансе меморије, међутим, крајем прошлог века, кроз серију експеримената откривено је да вербализовање претходно виђених визуелних стимулуса слаби касније препознавање (Schooler & Engstler-Schooler, 1990). Експериментално је показано да је, када се испитаницима прикаже лице и затражи од њих да га опишу, много мање вероватно да ће препознати лице касније, од оних из контролне групе који су само погледали лице без задатка описивања. Феномен је назван вербално затамњивање (verbal overshadowing). Од првог објављивања рада, цитиран је више од 400 пута (Lehrer, 2010). Проширио је модел на разне друге задат-

ке, попут сећања укуса вина (Melcher & Schooler, 1996). Његов први покушај понављања студије из 1990. пет година касније је резултирао ефектом који је био 30% мањи (Fallshore & Schooler, 1995). Наредне године се величина ефекта смањила за још 30% (Melcher & Schooler, 1996). Проблем је почео да назива „космичком хабитуацијом“ аналогно смањењу реакције која се јавља када се појединци навикну на одређене стимулусе (Lehrer, 2010).

„ВЕРУЈ, АЛИ ПРОВЕРИ“ – REPRODUCIBILITY PROJECT

Поновљивост налаза није добро схваћена, јер подстицаји појединих научника дају предност новостима над репликацијом (*Open Science Collaboration*, 2015), а аутори пројекта наводе да иновација истиче путеве који су могући, репликација указује на путеве који су вероватни, док се напредак ослања на обоје. Научне тврдње не би требало да се заснивају на статусу аутора, већ на реплицирању доказа, а то је једино могуће директном репликацијом (*Open Science Collaboration*, 2015), иако она некада неће поновити налазе оригиналне студије из различитих разлога.

Истраживачи су спровели 100 репликација експерименталних и корелационих студија објављених у три водећа психолошка часописа (*Psychological Science*, *Journal of Personality and Social Psychology* и *Journal of Experimental Psychology: Learning, Memory, and Cognition*) током 2008. године. Постоји више разлога зашто су баш ови часописи изабрани. Прво, да би обезбедили оквир за узорковање, затим да би омогућили поређење међу врстама часописа и дисциплинама, уклапали су се у распон стручности тимова, били су довољно нови да дају оригиналне материјале и довољно стари да имају значајне показатеље утицаја цитирања и представљају психолошке дисциплине са високом учесталашћу студија које нису скупе за понављање.

Тимови су бирали из базе првих 20 чланака из сваког часописа почев од првог чланка који је објављен у првом броју за ту годину. Последњи експеримент у сваком чланку био је предмет репликације. Укупно 84 од 100 репликација биле су последње пријављене студије у чланку. Укупан број чланака у свим часописима био је 488, од тога је 158 ушло у избор за репликацију током трајања пројекта (*Open Science Collaboration*, 2015).

Позитивних резултата било је 97% од 100 ефеката оригиналних студија. На основу просечне снаге репликације од 97% значајних ефеката, очекивано је приближно 89% позитивних резултата у

репликацијама под условом да су сви оригинални ефекти тачни, међутим добијено је 35. Анализа сугерише стопу успеха од 47,4%. У области социјалне психологије 14 од 55 (25%) ефеката је реплицирано, док је за когнитивну психологију 21 од 42 (50%). Утврђено је да су изненађујући ефекти били мање понављани, као и ефекти за које је било изазовније спровести репликацију. Успех репликације био је више повезан са оригиналном снагом доказа (п вредност, величина ефекта) него са карактеристикама тимова и применом репликације (*Open Science Collaboration*, 2015).

У зависности од критеријума који се користи, само 36 до 47% оригиналних студија је успешно реплицирано, што је многе довело до закључка да постоји „репликациона криза“ у психологији, а један од најпознатијих приговора се осврнуо на то колико су репликационе студије заиста биле директне репликације (Gilbert et al., 2016). Гилберт и сарадници (2016) истичу конкретне примере где су се студије репликације директно разликовале од оригиналних студија. Једини начин на који су се репликационе студије разликовале од оригиналних, како се наводи у пројекту (*Open Science Collaboration*, 2015), јесте начин узорковања из популације, међутим Гилберт (Gilbert et al., 2016) наводи и друге начине. Почине од разлика у узорку на примеру истраживања где је оригинална студија мерила став Американаца према Афроамериканцима, док је поновљена студија радила са Италијанима који не деле исте стереотипе.

Open Science Collaboration (2015) наводи да је сваку репликацију поновио само једном, а тај покушај је дао резултате да је само 47% оригиналних студија успешно реплицирано. Са друге стране, Many Lab Project (Klein et al., 2014) реплицирао је сваку од својих студија више пута, а затим објединио податке, извештавајући о 85% реплицираних оригиналних студија. Намеће се и питање пристрасности. Пре спровођења репликационе студије питали су ауторе оригиналних студија да подрже методолошке протоколе и то је учинило 69% оригиналних аутора. Да су ограничили анализе на одобрене студије установили би да је 59,7% успешно реплицирано (Gilbert et al., 2016).

Андерсон и сарадници (Anderson et al., 2016) у одговору објашњавају да су у половини случајева аутори оригиналне студије потврдили да је репликација директна или блиска оригиналној и да независно оцењена сличност оригиналних и репликационих студија није успела да предвиди репликациони успех.

Са друге стране, аутори (Etz & Vandekerckhove, 2016) проверавају добијене резултате израчунавањем Бајесовог фактора. У рачунању узимају у обзир вероватни сценарио да је пристрасност објављивања искривила првобитно објављене резултате. Од укупног броја, 75% студија дало је квалитативно сличне резултате у погледу количине пружених доказа, али су ти докази често били слаби. Више од половине студија и оригиналних и реплицираних (64%) није пружило доказе ни за нулту ни за алтернативну хипотезу. У случајевима када је оригинални рад пружео снажне доказе, а репликација није (15%), величина узорка је била мања него у оригиналној студији, док тамо где је репликација пружила снажне доказе, а оригиналне студије нису (10%), величина узорка у репликационој студији је била већа, па је, сходно томе, закључак истраживача (Etz & Vandekerckhove, 2016) који су применили Бајесов метод за проверу добијених резултата да привидни неуспех пројекта да реплицира многе циљане ефекте може бити адекватно објашњен прецењеним величинама ефекта због малих величина узорака и пристрасности публикација у психолошкој литератури.

ШТА МОЖЕМО НАУЧИТИ ИЗ КРИЗЕ ПОВЕРЕЊА У ПСИХОЛОШКУ НАУКУ?

Како је социјална психологија последњих година у кризи насталој губитком поверења у поновољивост налаза истраживања, Стробе (Stroebe, 2019) прави процену осврћући се на догађаје који су се одиграли претходних година желећи да сазна, прво, да ли репликационе студије информишу о поновљивости, друго, да ли репликације могу открити превару, треће, да ли неуспех репликација указује на неисправност оригиналних студија и, као последње питање, да ли репликације помажу одрживости или оповргавању теорије. Догађај који је покренуо лавину сумње био је случај преваре коју је починио холандски социјални психолог Дидерик Стапел (Stroebe et al., 2012; Drenth, 2013), убрзо затим су уследили резултати истраживања упитних научних пракси (Simmons et al., 2011; John et al., 2012). У исто време, појављују се извештаји неуспелих репликација познатих студија прајминга (Doyen et al., 2012; Shanks et al., 2013). Сумње су мотивисале истраживаче да се прикључе великим колаборативним пројектима и да покушају да понове резултате добијене у оригиналним студијама (Stroebe, 2019). Први пројекат репликабилности спровео је Open Science Collaboration (2015) добивши резултате да је само

36% покушаја репликација било успешно. Пројекат *Many Labs* (Klein et al., 2014) наводи да су успешно реплицирали 10 од 13 студија. У пројекту *Many Labs 2* (Klein et al., 2018) проналазе да је 54% показало значајне ефекте који иду у истом правцу као оригиналне студије.

Како је уложено доста времена и труда у покушаје репликације, Стробе (Stroebe, 2019) сматра да је дошло време за процену онога шта смо научили из репликационих истраживања.

Репликациони пројекти резултирали су променама начина на који радимо истраживања (Stroebe, 2019) кроз побољшање методологије, што подразумева озбиљније схватање статистичке снаге истраживања, а то значи да би требало смањити упитне истраживачке праксе као што су селективно извештавање и недоследни налази. Са друге стране, олакшан је приступ подацима па ће бити лакше спроводити метаанализе.

Истраживање *Open Science Collaboration* (2015) донекле нам је показало један од разлога зашто директне репликације нису интересантне јер мало доприносе научном сазнању. Ако тачна репликација добије исти налаз као и оригинална студија тада сазнајемо да је исход био поновљив, али и даље не знамо да ли је оригинална студија добар тест теорије, јер и ако је експеримент лоше дизајниран, директна репликација може резултирати истим налазом, али додатне информације можемо добити ако подржимо теоријску хипотезу а притом операционализујемо варијабле различито од оригиналног истраживања (Stroebe & Strack, 2014). Стапел (Stapel & Lindenberg, 2011) пише чланак како поремећени контексти промовишу стереотипе и дискриминацију где објављују резултате који показују да смеће или сломљени плочник могу повећати социјалну дискриминацију. Ови налази су првобитно објављени у *Science*, а по открићу преваре су повучени. Да је потврдио нулту хипотезу, налаз не би био занимљив, овако је добио однос који није очекиван, а са друге стране није било претходних емпиријских доказа за такав однос. Слично томе, да Барг (Bargh, 1996) или Дајкстерхојс (Dijksterhuis, 1998) нису добили позитивне резултате у складу са својим хипотезама нико не би био заинтересован за њихове налазе јер су нулти налази занимљиви само ако су у супротности са централном хипотезом која је изведена из теорије или су у супротности са низом ранијих студија (Stroebe & Strack, 2014).

Репликације тешко да могу открити преваре, јер истраживање (Stroebe et al., 2012) показује да репликације нису играле никакву

улогу у откривању превара. Са друге стране, како се о преварама ретко говори у званичним извештајима, информације су сакупљали из новина, откривши да већину превара откривају блиске колеге и студенти. И метаанализе су слаби детектори превара (Stroebe et al., 2012; Stroebe, 2019). Метаанализа прајминг студија (Decoster & Claupool, 2004) која је истраживала како примовање утиче на формирање утисака кроз асимилацију, сидрење и корекцију, идентификовала је 47 чланака на основу 6.833 учесника и утврдила да су укупни ефекти статистички значајни за сва три случаја и извештава да су величине ефекта веће у студијама које су спроведене у земљама ван САД и Канаде, претпостављајући да ови ефекти одражавају специфичне разлике међу лабораторијама, а не разлике међу културама. Када је проверено ауторство европских студија, испоставило се да је већину спровео Стапел, а многе од тих студија су се касније испоставиле лажним (Stroebe, 2019). Стапел је био први аутор на 7 од 10 студија асимилације, 6 од 7 студија сидрења и на 2 од 3 студије контраста које су биле одговорне за ефекат националности (Stroebe et al., 2012). Закључци метаанализе (Decoster & Claupool, 2004) били су да је могуће да европски испитаници озбиљније схватају експерименте од америчких студената. Метаанализа указује да су резултати прајминг истраживања које је објавио Стапел били прилично конзистентни са вишеструким студијама у САД, а за случајеве као што је Стапел је типично да избегавају неочекиване налазе и да су њихови резултати у складу са очекивањима из те области (Stroebe et al., 2012).

Ако се крене од тога да је сврха понављања тестирање теоријске хипотезе, а не процена поузданости одређеног експеримента, тада проблем директне репликације постаје очигледан, јер понављање специфичне операционализације теоријског конструкта у другом времену или на различитој популацији можда неће одражавати исти теоријски конструкт као у оригиналној студији (Stroebe & Strack, 2014). Ово не представља проблем у студијама где варијабле нису културно или друштвено условљене. Међутим, у социјално психолошким студијама верна репликација операционализације теоријског конструкта у различитом тренутку и са различитом популацијом може бити другачија од оригиналне студије (Stroebe & Strack, 2014). Према теорији (Strack & Deutsch, 2004), прајминг активира концепте који проширују активацију на друге концепте који су епизодично или семантички повезани (професор-интелигенција). Суптилно при-

мовање може бити ефикасније од очигледног усмеравања пажње људи према неком садржају. Људи се разликују у својим уверењима о старим особама. Различита веровања су различито доступна у различитим контекстима. Исти прим може активирати различите концепте код различитих људи или под различитим условима (Stroebe & Strack, 2014). Приликом примовања стереотипа о старијим особама теоријски циљана когнитивна репрезентација је „лагано ходање“. Могуће је да репликациона студија (Doyen et al., 2012) није успела да понови ефекат иако је следила исту процедуру као у оригиналном истраживању (Bargh et al., 1996) јер би превод речи коришћених у првом истраживању могао имати другачије значење у другом. Осим тога, постоји објашњење да концепт „лаганог хода“ није централни део стереотипа о старим особама у Белгији (Stroebe & Strack, 2014).

Пашлер и Харис (Pashler & Harris, 2012) сматрају да штетна интеракција пристрасности објављивања и фокусирање на концептуалне, а не на директне репликације може осветлити неке од познатих случајева загонетки науке. На Гугл академику претрагом речи „failure to replicate, psychology“ проверили су првих 40 чланака. Медијана времена између оригиналног чланка и покушаја репликације била је 4 године, при чему се 10% покушаја репликације десило са закашњењем већим од 10 година (Pashler & Harris, 2012), што сугерише да се реплицирају недавно спроведена истраживања. Њихов је закључак да се одређене теме истражују у једном тренутку, а неколико година касније долазе друге теме на ред. Иако истраживачи одустају од теме када открију да се резултати не реплицирају, само је један од разлога зашто се интересовање мења.

Полазећи од Поперовог гледишта да се теорија не може доказати као тачна, Стробе (2019) наводи да можемо повећати поверење у ваљаност теорије спровођењем строгих тестова, али никада неће постојати потпуна сигурност да је теорија валидна. Осим тога, сматра да се теорије не могу фалсификовати појединачним експериментима, зато покушај реплицирања појединачних, несистематски одабраних ефеката не допушта закључке о ваљаности теорија које су првобитно биле тестиране (Stroebe, 2019). Сходно томе, допринос репликација психолошком знању је ограничен и доводи се у питање да ли сазнања која они додају оправдавају потрошено време и ресурсе (Stroebe, 2019).

- ЛИТЕРАТУРА Anderson, C. J., Bahnik, S., Barnett-Cowan, M., Bosco, F. A., Chandler, J., Chartier, C. R., Zuni, K. (2016). Response to comment on „Estimating the reproducibility of psychological science“. *Science*, 351 (6277). <https://doi.org/10.1126/science.aad9163>
- 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 (2), 230–244. <https://doi.org/10.1037/0022-3514.71.2.230>
- 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. <https://doi.org/10.1037/a0021524>
- Bower, B. (2012). The hot and cold of priming: Psychologists are divided on whether unnoticed cues can influence behavior. *Science News*, 181 (10), 26–29. <https://doi.org/10.1002/scin.5591811025>
- Decoster, J. & Claypool, H. M. (2004). A Meta-Analysis of Priming Effects on Impression Formation Supporting a General Model of Informational Biases. *Personality and Social Psychology Review*, 8 (1), 2–27. https://doi.org/10.1207/S15327957PSPR0801_1
- Dijksterhuis, A., Spears, R., Lepinasse, V. (2001). Reflecting and deflecting stereotypes: Assimilation and contrast in impression formation and automatic behavior. *Journal of Experimental Social Psychology*, 37 (4), 286–299. <https://doi.org/10.1006/jesp.2000.1449>
- Dijksterhuis, A. & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology*, 74 (4), 865–877. <https://doi.org/10.1037/0022-3514.74.4.865>
- 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. <https://doi.org/10.1371/journal.pone.0029081>
- Etz, A. & Vandekerckhove, J. (2016). A Bayesian Perspective on the Reproducibility Project: Psychology. *PLoS One*, 11 (2), e0149794. <https://doi.org/10.1371/journal.pone.0149794>
- Fallshore, M. & Schooler, J. W. (1995). Verbal vulnerability of perceptual expertise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21 (6), 1608–1623. <https://doi.org/10.1037/0278-7393.21.6.1608>
- Fanelli, D. (2009). How Many Scientists Fabricate and Falsify Research? A Systematic Review and Meta-Analysis of Survey Data. *PLoS ONE*, 4 (5), e5738. <https://doi.org/10.1371/journal.pone.0005738>
- Fidler, F. & Wilcox, J. (2018). Reproducibility of Scientific Results. In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2018). Retrieved from:

- <https://plato.stanford.edu/archives/win2018/entries/scientific-reproducibility/>
- 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 (6), 933–948. <https://doi.org/10.1037/a0029709>
- Gilbert, D. T., King, G., Pettigrew, S., Wilson, T. D. (2016). Comment on „Estimating the reproducibility of psychological science“. *Science*, 351 (6277), 1037–1037. <https://doi.org/10.1126/science.aad7243>
- 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. <https://doi.org/10.1177/0956797611430953>
- Kawakami, K., Young, H., Dovidio, J. F. (2002). Automatic Stereotyping: Category, Trait, and Behavioral Activations. *Personality and Social Psychology Bulletin*, 28 (1), 3–15. <https://doi.org/10.1177/0146167202281001>
- Keljanović, A. (2020). Upitne istraživačke prakse u kliničkoj psihologiji. U: A. Dautbegović, N. Hadžiahmetović, M. Marković, K. Mišetić, S. Zvizdić (ur.), *Sarajevski dani psihologije* (223–237). Sarajevo: Filozofski fakultet.
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Nosek, B. A. (2014). Investigating Variation in Replicability: A “Many Labs” Replication Project. *Social Psychology*, 45 (3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Nosek, B. A. (2018). Many Labs 2: Investigating Variation in Replicability Across Samples and Settings. *Advances in Methods and Practices in Psychological Science*, 1 (4), 443–490. <https://doi.org/10.1177/2515245918810225>
- Lehrer, J. (2010, December 6). *The Truth Wears Off*. Retrieved from: <https://www.newyorker.com/magazine/2010/12/13/the-truth-wears-off>
- Macrae, C. N. & Johnston, L. (1998). Help, I Need Somebody: Automatic Action and Inaction. *Social Cognition*, 16 (4), 400–417. <https://doi.org/10.1521/soc.1998.16.4.400>
- Melcher, J. M. & Schooler, J. W. (1996). The Misremembrance of Wines Past: Verbal and Perceptual Expertise Differentially Mediate Verbal Overshadowing of Taste Memory. *Journal of Memory and Language*, 35 (2), 231–245. <https://doi.org/10.1006/jmla.1996.0013>
- Neuroskeptic. (2012). The Nine Circles of Scientific Hell. *Perspectives on Psychological Science*, 7 (6), 643–644. <https://doi.org/10.1177/1745691612459519>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349 (6251), aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>

- Pashler, H. & Harris, C. R. (2012). Is the Replicability Crisis Overblown? Three Arguments Examined. *Perspectives on Psychological Science*, 7 (6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Pashler, H. & Wagenmakers, E. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?. *Perspectives on Psychological Science*, 7 (6), 528–530. <https://doi.org/10.1177/1745691612465253>
- Drenth, P. (2013, May 20). What lessons can we learn from the Stapel case? Retrieved September 25, 2019, from: Articles, papers from Pieter Drenth website: <https://pieterdrenth.wordpress.com/2013/05/20/what-lessons-can-we-learn-from-the-stapel-case/>
- Schlitz, M., Wiseman, R., Watt, C., Radin, D. (2006). Of two minds: Sceptic-proponent collaboration within parapsychology. *British Journal of Psychology (London, England : 1953)*, 97, 313–322. <https://doi.org/10.1348/000712605X80704>
- Schooler, J. W. & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22 (1), 36–71. [https://doi.org/10.1016/0010-0285\(90\)90003-M](https://doi.org/10.1016/0010-0285(90)90003-M)
- Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., Moore, C. (2013). Priming Intelligent Behavior: An Elusive Phenomenon. *PLOS ONE*, 8 (4), e56515. <https://doi.org/10.1371/journal.pone.0056515>
- Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22 (11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Stapel, D. A. & Lindenberg, S. (2011). Coping with chaos: How disordered contexts promote stereotyping and discrimination. *Science (New York, N.Y.)*, 332 (6026), 251–253. <https://doi.org/10.1126/science.1201068>
- Strack, F. & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, 8 (3), 220–247. https://doi.org/10.1207/s15327957pspr0803_1
- Stroebe, W. (2019). What Can We Learn from Many Labs Replications?. *Basic and Applied Social Psychology*, 41 (2), 91–103. <https://doi.org/10.1080/01973533.2019.1577736>
- Stroebe, W., Postmes, T., Spears, R. (2012). Scientific Misconduct and the Myth of Self-Correction in Science. *Perspectives on Psychological Science*, 7 (6), 670–688. <https://doi.org/10.1177/1745691612460687>
- Stroebe, W. & Strack, F. (2014). The Alleged Crisis and the Illusion of Exact Replication. *Perspectives on Psychological Science*, 9 (1), 59–71. <https://doi.org/10.1177/1745691613514450>

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, 100 (3), 426–432. <https://doi.org/10.1037/a0022790>

Wiseman, R. & Schlitz, M. (1997). Experimenter effects and the remote detection of staring. *Journal of Parapsychology*, 61 (3), 197–208.

Wiseman, R. & Schlitz, M. (1999). Replication of experimenter effect and the remote detection of staring. *Proceedings of the 42nd Annual Convention of the Parapsychological Association*, 471–479.

ANĐELA D. KELJANOVIĆ

UNIVERSITY OF PRIŠTINA IN KOSOVSKA MITROVICA
FACULTY OF PHILOSOPHY
DEPARTMENT OF PSYCHOLOGY

SUMMARY

PSYCHOLOGICAL SCIENCE FROM ‘PUBLISH OR PERISH’
TO ‘TRUST BUT VERIFY’

At the time when social psychologists believed they could be proud of their discipline, there was the devastating news that Diederik Stapel had committed a major scientific fraud. This event coincided with the start of the discussion on trust in psychological findings. It was soon followed by the report of a series of nine studies that failed to replicate the ‘professor's study’. These replication results were astounding due to earlier reports of successful replications. Due to the crisis of confidence in the results of field research, the Open Science Collaboration subsequently replicated 100 correlation and experimental studies published in 2008 in *Psychological Science*, *Journal of Personality and Social Psychology*, and *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Of the 97% of the original studies that had a positive effect, 36% were replicated. However, their findings have also been called into question by calculating the Bayesian factor. In addition to fraud, questionable research practices resulting from publication bias that results in false positives undermine confidence in the validity of psychological research findings. Perhaps the most costly mistake of false-positive findings is to erroneously reject the null hypothesis. However, that Stapel (2011) confirmed the null hypothesis, or that Bargh (1996) found that admission of participants did not affect walking speed, or that Dijksterhuis and van Knipenbergh (1998) reported that participants received with the word ‘professor’ did not improve their

performance on task, no one would be interested in their findings. Zero findings are only interesting if they contradict the main hypothesis derived from the theory or contradict a number of previous studies. The fact that good experimental research is usually conducted in order to test theories, researchers can never be sure whether they have chosen the optimal operationalization of a given construct. As researchers can never be sure that they have properly operationalized the theoretical constructs they are evaluating and whether they have been successful in controlling the third variables that may be responsible for their results, the theory can never be proven true.

KEYWORDS: replication; publication bias; psychology.



Овај чланак је објављен и дистрибуира се под лиценцом Creative Commons Ауторство-Некомерцијално Међународна 4.0 (CC BY-NC 4.0 | <https://creativecommons.org/licenses/by-nc/4.0/>).

This paper is published and distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial International 4.0 licence (CC BY-NC 4.0 | <https://creativecommons.org/licenses/by-nc/4.0/>).