

Scientific Reform and Replication Crisis in Psychology

Flis, Ivan

Source / Izvornik: **Psihologijske teme, 2023, 32, 227 - 256**

Journal article, Published version

Rad u časopisu, Objavljena verzija rada (izdavačev PDF)

<https://doi.org/10.31820/pt.32.2.1>

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:186:083904>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-12-25**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Humanities and Social Sciences - FHSSRI Repository](#)



Znanstvena reforma i replikacijska kriza u psihologiji

Ivan Flis

Sveučilište u Rijeci, Filozofski fakultet, Odsjek za kulturalne studije, Rijeka, Hrvatska

Sažetak

Početak 2010-ih godina u psihologiji se ubrzano počela širiti rasprava o problemima s replikacijom već objavljenih istraživanja. Rasprave su se uskoro pretvorile u višeslojnu kritiku kvantitativnih metoda i metodologija koje psiholozi koriste, posebice u dijelu inferencijalne statistike. Problemi s replikacijom opisani su kao simptom dublje krize uvriježenih nacrtava istraživanja i statističkih postupaka, u kombinaciji s funkcioniranjem znanstvenoga objavljivanja rezultata i znanstvene komunikacije. U ovome ćemo radu ponuditi pregled rasprava o replikacijskoj krizi i reformskome pokretu u psihologiji, prvi put na hrvatskome jeziku.

Ključne riječi: replikacijska kriza, reformski pokret u psihologiji, metaznanost

Uvod

Tijekom posljednjih deset godina u mnogim se znanstvenim disciplinama proširila takozvana replikacijska kriza ili, kako se danas primjerenije naziva, kriza povjerenja u znanstvene rezultate (Baker, 2016; u daljnjemu tekstu: *kriza*). U ovome ćemo radu ponuditi pregled krize u okvirima znanstvene psihologije (Chambers, 2017; Nosek i sur., 2021; Pashler i Wagenmakers, 2012). Pregled će se fokusirati na dvije velike teme: (1) na okvir i uzroke krize te (2) na pokušaje psihologa okupljenih u reformskome pokretu da poboljšaju prakse znanstvenih istraživanja (ili kraće znanstvene prakse, engl. *scientific practices*) u psihologiji da bi riješili probleme koji su identificirani u diskusijama o krizi. Znanstvene prakse obuhvaćaju prikupljanje i analizu podataka, postavljanje nacrtava istraživanja, dijeljenje podataka i postupka analize te objavu i recenziju radova. Drugim riječima, uključuju sve metodološke i analitičke odluke te ponašanja samih istraživača pri provođenju istraživanja, od prikupljanja podataka do publikacije rezultata.

Ovaj pregledni rad pokušava ponuditi širok pregled izrazito žive interdisciplinarnе literature kojoj je, zbog specifičnoga sklopa povijesnih i društvenih

Ivan Flis <https://orcid.org/0000-0002-0493-8026>

✉ Ivan Flis, Filozofski fakultet u Rijeci, Odsjek za kulturalne studije, Sveučilišna avenija 4, 51000 Rijeka, Hrvatska. E-adresa: ivan.flis@uniri.hr

razloga, jedan od glavnih interesa reforma i poboljšanje kvalitete istraživanja u kvantitativnoj psihologiji. Pregled će biti sinoptički te zbog toga plitak – zahvatit će široke rasprave velikoga broja metaznanstvenika, metodologa, psihologa, filozofa, povjesničara i sociologa te naznačiti okvire rasprava o replikaciji u svojoj širini. Ovo nije pregled svih provedenih replikacija u posljednjih deset godina ni taksativan popis konstrukata koji su odbačeni ili kritizirani, već prikaz kompleksne, višegodišnje i multidisciplinarnе kritike kvantitativne metodologije u psihologiji. S obzirom na obilan žargon koji je nastao u reformskoj literaturi, u tekstu će se (u nekim slučajevima prvi put na hrvatski jezik) prevesti, koristiti i ponekad definirati veći broj novih pojmova. Gdje nisu detaljno razrađeni u tekstu, za pojmove će u Tablici 2. u *Prilogu* biti prevedena definicija iz rječnika FORRT-a (*Framework for Open and Reproducible Research Training*; <https://forrt.org/>). Kada se prvi put pojavljuju u tekstu, takvi će pojmovi biti podcrtani. Čitatelj može u tablici, prateći za svaki pojedini termin poveznicu na originalnu FORRT-ovu definiciju, pronaći daljnje poveznice na radove koji ih dublje razrađuju i usložnjavaju.

Literatura o replikacijskoj krizi već je izrazito obimna (za zastarjeli, ali kvalitetan pregled v. Shrout i Rodgers, 2018; za noviji pregled v. Nosek i sur., 2021). U empirijskome dijelu okuplja mnoge pokušaje replikacija istraživanja koji su veći i suradnički (najpoznatiji je primjer Open Science Collaboration, 2015) ili pak manji, na razini pojedinih istraživačkih grupa (npr. Lynott i sur., 2014). U metodološkome dijelu uključuje rasprave o najboljim praksama (npr. Navarro, 2021), polemiske diskusije znanstvenika na društvenim mrežama, kako u popularnoj, tako i u stručnoj periodici (v. Derksen i Field, 2022), radove iz novouspostavljenoga područja metaznanosti („znanosti o znanosti”, Devezer i sur., 2021; Schooler, 2014), reformske manifeste i slične tekstove upućene stručnoj publici (Munafò i sur., 2017; Nosek i Bar-Anan, 2012; Vazire i Holcombe, 2022). Rasprave o replikacijskoj krizi zaista dubinski zahvaćaju psihologijsku znanost: na najplićoj razini u metodološkome smislu, na nešto dubljoj u normativnome aspektu etike znanstvenoga rada te na dubinskoj razini kada dovode u pitanje epistemološke temelje discipline ili čak metafizičke postavke u koje sami empiričari rijetko kada zalaze. Širu se raspravu može pratiti u radovima filozofa (Feest, 2019; Maatman, 2021; Morawski, 2019; Romero, 2019), povjesničara (Derksen, 2019; Flis, 2019, 2022; Mülberger, 2022) i sociologa (Nelson i sur., 2022; Peterson i Panofsky, 2021, 2020).

Okvir i uzroci replikacijske krize

Replikacijska je kriza, najjednostavnije rečeno, stanje uznemirenosti određenoga dijela znanstvene zajednice koji se kroz niz empirijskih istraživanja i rasprava uvjerio da velik dio njihove literature sadrži rezultate koji nisu replikabilni (Wiggins i Christopherson, 2019). Na osnovno pitanje – možemo li ponoviti istraživačke postupke iz već objavljenih istraživanja i dobiti jednake rezultate kao u prvotnome istraživanju – odgovor je dovoljno često bio *ne* da bi potaknuo razvoj diskursa o krizi replikacija. Uglavnom je neupitan zahtjev pozitivističke i

postpozitivističke znanosti da je ponovljivost eksperimenta (ili drugoga istraživačkog nacrt) i dobivanje jednakoga rezultata jedan od minimalnih kriterija koje empirijska znanost mora ispunjavati (Romero, 2019). Shodno tomu, razumljivo je da će dovođenje u pitanje replikabilnosti istraživanja uzrokovati velik nemir ili čak krizu unutar znanstvenih zajednica. No, pitanje kako procijeniti što znači dobiti „jednake” rezultate te do koje je razine uistinu moguće ponavljanje već provedenih istraživanja u raspravama o replikacijskoj krizi preraslo je iz filozofskoga pitanja o znanstvenoj metodi u praktični problem empiričara.

Kroz te su rasprave istraživači došli do cijeloga niza finih razlika između različitih oblika ponovljivosti istraživanja. Tako je danas uvriježeno razlikovati replikabilnost (engl. *replicability*) i reproduktibilnost (engl. *reproducibility*). Replikabilnost je svojstvo istraživanja da pri ponovljenome prikupljanju podataka različiti istraživači dobiju jednake rezultate kao u originalnome istraživanju. Reproductibilnost je svojstvo istraživanja da pri ponovljenoj analizi prvotnih podataka istraživači dobivaju jednake rezultate (za više v. National Academies of Sciences, 2019). Dodatno je uveden pojam robusnost (engl. *robustness*) koji predstavlja svojevrsnu otpornost zaključaka pri testiranju podataka na kojima su temeljeni drugačijim analizama (Nosek i sur., 2021). Fine razlike u tim pojmovima uvode distinkciju između ponovljenih prvotnih analiza i potpuno novih analiza te već prikupljenih ili ponovno prikupljenih podataka po istome nacrtu (v. Tablicu 1. koja shematski prikazuje definicije tih pojmova u međuodnosu). Ako povijesno razmišljamo o razvoju psihologijskih metodologija, možemo razumjeti taj fokus na utjecaj istraživača na analize kao proširenje metodoloških uputa o artefaktima i efektima eksperimentatora. Što su efekti eksperimentatora na nacrt i metodologije bili od 1970-ih godina naovamo (Faye, 2012), u posljednjih su deset godina nadograđeni efektima analitičara na podatke (Flis, 2019). Naravno, slika je to složenija kada istodobno razmišljamo o istovjetnosti/različitosti metodoloških nacrti pri replikaciji/reprodukciji/ispitivanju robusnosti, bok uz bok s pitanjem o istovjetnosti i ponovljivosti podataka. Ponovljivost eksperimenta (a još više drugih, korelacijskih nacrt) ne treba uzeti zdravo za gotovo, pa se tako u literaturi čak problematizira kao relevantno narušavanje zahtjeva za istovjetnost nacrti pri replikaciji i činjenica da gotovo nijedna replikacija ne može biti izvedena na zaista istome npr. uzorku (za ozbiljnu kritičnu razradu v. Feest, 2019 i ideju „individuacijskih sudova”, engl. *individuation judgements*, kao ključnih metodoloških odluka koje imaju dublje implikacije od same valjanosti nacrti).

Ovdje postoji još jedna definicijska razlika, ona između izravne i konceptualne replikacije, kao nasljeđe ranih protukritika eksperimentalnih socijalnih psihologa koji su dovodili u pitanje valjanost replikacijskih istraživanja (npr. Stroebe i Strack, 2014) tako što su postavljali pitanje govore li nam replikacije nešto o samome efektu ili su neuspjele replikacije češće proizvod raznih suptilnih razlika između prvotnoga istraživanja i replikacije (za takve perspektive koje se danas uglavnom ne shvaćaju ozbiljno u reformskome pokretu v. Nosek i sur., 2021, str. 722; „alternativne perspektive o vrijednosti replikacija”).

Tablica 1.

Shematski prikaz međuodnosa definicija pojmova s obzirom na originalnost/ponovljivost analiza i podataka

Podaci	Metoda/ Analiza	Metoda/analiza jednaka prvotnoj	Različita metoda/analiza
Prvotni podaci		reproduktibilnost	robusnost
Ponovno prikupljeni podaci		(izravna) replikacija	konceptualna replikacija / test uopćivosti

Umjesto „konceptualne” replikacije danas se češće govori o testu uopćivosti (engl. *generalizability test*, Nosek i sur., 2021, str. 722):

S obzirom na to da je svaka replikacija drugačija od svakoga prijašnjeg istraživanja, svaka je replikacija test uopćivosti, ali obrnuto ne vrijedi. Test uopćivosti je replikacija samo kada svi ishodi testa mijenjaju naše povjerenje u prvotni nalaz [...]. Na primjer, ako pozitivan ishod povećava naše povjerenje i širi uopćivost, ali negativan ishod samo identificira potencijalnu rubnu uvjetovanost i ne mijenja naše povjerenje u prvotni nalaz, tada govorimo o testu uopćivosti, a ne o replikaciji. Primjenom toga okvira, pojam „konceptualna replikacija” često je bio korišten pri opisivanju testova uopćivosti, a ne replikacija jer ih se uvriježno interpretira kao da potvrđuju prvotnu interpretaciju nalaza, a gotovo nikada kao da je odbacuju.

To definicijsko raslojavanje pojmova samo je površina žestokih i detaljnih rasprava koje se vode u tzv. reformskoj literaturi (reformskoj u smislu da okuplja znanstvenike koji su zainteresirani za poboljšanje i reformu praksi istraživanja). Kroz višegodišnje rasprave i istraživanja kriza je u literaturi opisana kao posljedica interakcije uvriježenih istraživačkih praksi (prikupljanja i analize podataka, dizajna istraživanja) i institucionaliziranih normi objavljivanja radova u znanstvenim časopisima, odnosno takozvane kulture objavljivanja (engl. *publication culture*; v. NWO, 2018). U shvaćanju metodologije koje je uvriježno među (posebice kvantitativnim) psiholozima (Milas, 2005) kritičari tvrde da je svaki korak hipotetičko-deduktivnoga modela psihologijske znanosti izložen zabrinjavajućim ugrozama koje dovode u pitanje mogućnost dobrog zaključivanja (Munafò i sur., 2017).

Munafò i suradnici (2017) sve te ugroze prikazuju kroz svoju koncepciju prijetnji replikabilnosti. U prvoj fazi generiranja i specificiranja hipoteza može doći do pristranosti, u postavljanju nacrtu istraživanja uglavnom se ne vodi računa o statističkoj snazi, u provođenju istraživanja i prikupljanju podataka ne provodi se kontrola kvalitete, u analizi podataka dolazi do tzv. analitičke fleksibilnosti (engl. *analytic flexibility*; drugi su povezani pojmovi *p-hacking* (*p-hakiranje*), *significance fishing* (*pecanje značajnosti*) te *researcher degrees of freedom* (*stupnjevi slobode istraživača*); v. Gelman i Loken, 2013; Head i sur., 2015; Simmons i sur., 2011) te, nakraju, u objavljivanju radova do publikacijske pristranosti (*publication bias*; Franco i sur., 2014) koju je još 1979. godine identificirao psiholog Robert Rosenthal i nazvao je problemom dokument – ladica (*file-drawer problem*; Rosenthal, 1979).

Analitička fleksibilnost, odnosno mogućnost da vješti istraživači provlače podatke kroz analize sve dok ne dobiju interpretabilne i, što je najvažnije, objavljuje rezultate okosnica su kritika. Poneki autori pripisuju znanstvenicima malicioznost i nazivaju takvo ponašanje *p-hakiranjem*, ali je mnogo raširenije shvaćanje da je takvo ponašanje u sivoj zoni prihvatljivoga u psihologiji i u drugim kvantitativnim društvenim znanostima, ili čak da je znak metodološke i statističke vještine istraživača (za raspravu o motivu u pozadini *p-hakiranja* v. Nelson i sur., 2018, str. 517–518). Opis takve analitičke fleksibilnosti Simmonsa i suradnika (2011, str. 1359), koju su oni nazvali stupnjevima slobode istraživača, jedan je od onih koji su odigrali važnu ulogu u reformskim argumentima psihologa:

Krivac [za to da je „neprihvatljivo lako objaviti statistički značajne rezultate za *bilo koju* hipotezu”] je konstrukt koji nazivamo *stupnjevi slobode istraživača*. Tijekom prikupljanja i analize podataka istraživači moraju donijeti mnogobrojne odluke: trebamo li prikupiti više podataka? Trebamo li isključiti neke podatke? Koji bi uvjeti trebali biti spojeni, a koji uspoređeni? Koje bi kontrolne varijable trebalo uzeti u obzir? Trebaju li određene mjere biti spojene ili transformirane, ili oboje? Rijetko je, a ponekad i nepraktično, da istraživači donesu sve te odluke unaprijed. Dapače, često se (i prihvaćeno) isprobavaju različite analitičke alternative, traži se baš ona kombinacija koja daje „statističku značajnost” i napiše se samo ono što je „funkcioniralo”.

Do kakvih strukturalnih problema u akumulaciji znanja dovodi takvo ponašanje članova jedne znanstvene zajednice? Argumenti psihologa u krizi najčešće upućuju na razradu strukturalnih utjecaja analitičke fleksibilnosti i sličnih nevidljivih pristranosti koje utječu na „zdravlje” literature.

Publikacijska pristranost, statistička snaga i veličina učinka

Normativno prihvaćena netransparentna analitička fleksibilnost, kako tvrde kritičari, dovodi do dviju zabrinjavajućih posljedica: nepoznate (ali velike, a koliko velike, još je uvijek predmet rasprave; za pregled v. Nosek i sur., 2021) proporcije objavljenih istraživanja koja se ne mogu replicirati s jedne strane te, s druge strane, nepouzdanosti cjelokupne literature pojedinih znanstvenih područja. Teško je

procijeniti koja je literatura koliko zahvaćena, a problematične norme objavljivanja i upitne istraživačke prakse (engl. *questionable research practices*, QRPs) izrazito su raširene u znanstvenoj zajednici (John i sur., 2012).

Zašto je publikacijska pristranost problematična? Zamislimo sljedeći primjer. Više skupina psihologa počinje se istraživački baviti novom terapijskom intervencijom za kliničku depresiju. Tijekom godine dana u cijelome je svijetu provedeno stotinu istraživanja koja ispituju utjecaj te terapijske intervencije na simptome depresije. Od tih stotinu istraživanja samo ih je pet dobilo statistički značajne učinke nove terapijske intervencije. S obzirom na to da časopisi prihvaćaju samo značajne rezultate, ostalih 95 istraživanja nikada nije ugledalo svjetlo dana jer njihovi rezultati nisu objavljeni. Kada bismo kao čitatelji pristupili literaturi o toj novoj kliničkoj intervenciji, našli bismo pet radova koji potvrđuju njezinu efikasnost.

Ovo je ekstreman primjer kojim smo željeli ilustrirati što je publikacijska pristranost. Problem je da, čak i u slučajevima kada ona nije toliko jako izražena, preferiranje značajnih rezultata dovodi do „napuhavanja” veličina učinaka u literaturi, čak i onda kada ti učinci zaista postoje (Franco i sur., 2014), posebice kada smo usto svjesni da nije došlo do porasta u statističkoj snazi istraživanja od 1960-ih godina do danas unatoč upozorenjima statističara i metodologa. O tome svjedoči i nalaz da je u tome razdoblju prosječna statistička snaga istraživanja za detekciju tzv. malih veličina učinka ($d = 0.20$) tek 0.24 (Smaldino i McElreath, 2016). To znači da bi testovi u prosjeku trebali biti neznačajni u 75 % slučajeva u literaturi. Suprotno tomu očekivanju, psihologijska je literatura prepuna značajnih rezultata jer je produkt znanstvenih zajednica koje su pod rastućim pritiskom za objavljivanje pozitivnih rezultata niske i stagnirajuće statističke snage.

Literatura uhvaćena između pritiska za objavljivanje pozitivnih rezultata i institucionalizacije tih pritisaka u uredničke politike časopisa teško da može ispunjavati potrebe znanstvene zajednice (Ferguson i Heene, 2012, str. 559):

[A]verzija prema nultoj hipotezi i ustrajnost publikacijske pristranosti te poricanje njezina postojanja dovode do situacije u kojoj su psihologijske teorije praktički neuništive. Umjesto rigidnoga pridržavanja objektivnoga procesa replikacija i falsifikacija rasprave u psihologiji lako degeneriraju u ideološko grudanje, koje služi tomu da bi nakraju teorije loše kvalitete preživjele zauvijek.

Shodno tomu uvidu u strukturalne pristranosti cijele literature, kritičari upozoravaju da čak ni metode za sintetiziranje velikoga broja istraživanja poput metaanaliza nisu panaceja. Primjerice, Carter i suradnici (2019) napravili su simulacijsko istraživanje korisnosti različitih metoda korekcija pristranosti u metaanalizama s obzirom na razne očekivane strukturalne pristranosti i karakteristike literature / metaanalitičkih uzoraka. U simulacijama su varirali razine istinitosti istraživanoga efekta, heterogenost istraživanja, broj istraživanja u metaanalitičkome uzorku, publikacijsku pristranost i rasprostranjenost upitnih istraživačkih praksi. Njihove simulacije pokazuju da različite metode za korekciju metaanaliza ne konvergiraju prema sličnim rezultatima (ili najboljoj metodi) s obzirom na različite

malformacije literature koje su ugradili u simulacije. Metaanalitičar mora biti dobro upoznat sa stanjem literature koju pokušava sintetizirati, a u slučaju da to nije ili ne može biti potencijalno svojim metaanalizama samo stvara dodatan šum u kanalu umjesto da sintetizira dokaze (Carter i sur., 2019, str. 140):

[N]aglašavamo da je metaanaliza u psihologiji teška. Opažljivi faktori kao mali uzorci – u primarnoj literaturi i na razini metaanalize – međudjeluju s heterogenošću i pristranošću, a heterogenost i pristranost imaju jačinu i funkcionalni oblik koji nije moguće spoznati (npr. raspodjeljuju li se istiniti efekti normalno). Shodno tomu, jednako je teško interpretirati rezultate metaanaliza u psihologiji kao i u bilo kojoj pojedinoj replikaciji [...]. Istraživači u psihologiji ne bi trebali očekivati da će provođenje metaanaliza na postojećoj literaturi proizvesti konkluzivne rezultate za okončavanje rasprava. Umjesto toga mišljenja smo da je najveća korist metaanaliza u usmjeravanju pažnje na postojeće slabosti i snage pojedine literature.

Prije nego što se okrenemo rješenjima za dosad opisane probleme koje je artikulirao reformski pokret u psihologiji, zaključit ćemo ovaj opis nastanka krize kratkim pregledom triju ključnih događaja za nastanak diskursa o krizi u psihologiji – slučajevima Diedrika Stapela i Daryla Bema te objavom velikoga replikacijskog istraživanja *Kolaboratorija za otvorenu znanost*.

Stapel i Bem – kriza u najavi

Dvije ključne epizode za početak replikacijske krize u svijetu psihologije dogodile su se 2011. godine. Prva je slučaj višegodišnje karijerne akademske prijevere nizozemskoga eksperimentalnoga socijalnog psihologa Diedrika Stapela (za detaljan prikaz v. Derksen, 2021). Stapelu, u to vrijeme međunarodno renomiranomu eksperimentalnomu socijalnom psihologu, utvrđeno je potpuno fabriciranje podataka ili manipuliranje njima u 55 objavljenih radova. Jedini razlog zbog kojega je 2011. godine uhvaćen je to što su ga njegovi doktorandi, nakon što su posumnjali u valjanost podataka kojima ih je opskrbljivao iz svojih navodnih istraživanja, prijavili upravi Sveučilišta u Tilburgu koja je reagirala i pokrenula istragu. Iako je Stapel izgubio posao nakon zaključenja procesa utvrđivanja istine o njegovim fabrikacijama, znanstvenoj je zajednici ostalo pitanje: kako je moguće da je jedan Stapel toliko godina objavljivao na najvišoj razini na temelju lažiranih podataka te ima li još takvih „Stapela“?

Drugi slučaj, onaj rada Daryla Bema, nije uključivao prijevaru, ali odjeknuo je još jače, kao najava duboke metodološke krize eksperimentalne socijalne psihologije, ali i cjelokupne kvantitativne psihologije. Iste je 2011. godine Daryl Bem objavio rad *Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect* u prestižnom APA-inu časopisu *Journal of Personality and Social Psychology* (Bem, 2011). U radu je prikazao rezultate devet eksperimenata provedenih na preko 1000 sudionika. Bem je eksperimentima želio ispitati jednu vrstu retroaktivne kauzacije, parapsihološkoga

efekta pri kojemu iskustva koja se osobi dogode u budućnosti utječu na njihovo ponašanje u prošlosti. U provedenih devet eksperimenata Bem je koristio četiri tzv. vremenski obrnute (engl. *time-reversed*) standardne eksperimentalne paradigme u socijalnoj psihologiji. Npr. kod udešavanja Bemovo retroaktivno udešavanje provedeno je na sudionicima nakon što je izmjeren efekt udešavanja. Bemova je logika elegantna: ako postoji bilo kakav suptilan prekognitivni psi-efekt, neki će ljudi doživjeti utjecaj udešavanja koje će im se tek dogoditi u budućnosti.

Komentar Wagenmakersa i suradnika (2011, str. 426) u istome broju časopisa već je upućivao na odjek koji će Bemov rad imati među psiholozima: „Znanstveno polje psihologije trenutno koristi metodološke i statističke strategije koje su preslabe, previše prilagodljive te nude previše prigoda za istraživače da zavaravaju sami sebe i svoje kolege.” Oni su identificirali Bemove „dokaze” kao artefakt koji je posljedica relativno dobro poznatih metodološko-statističkih zamki (miješanje konfirmatornih i eksploratornih istraživanja, kao i loše razumijevanje toga što značajnost smisleno govori o hipotezi i podacima te osjetljivosti testova značajnosti na velike uzorke). Problem je s artefaktom na koji su pažnju skretali istraživači poput Wagenmakersa da Bemovo istraživanje nije idiosinkratično za proučavanje psi-efekata, već je primjer uvriježenoga načina provođenja istraživanja u kvantitativnoj psihologiji. Posebice kritični metodolozi toga su bili svjesni jer je odjednom car bio gol – upozorenja o metodološkim i statističkim manjkavostima više nisu bila samo pedagoška upozorenja o tome što bi bilo dobro raditi kada se provode istraživanja. Bemovi eksperimenti kao da su svima pokazali da nešto nije u redu s kriterijem za dokazivanje fenomena u psihologiji. Jedino što je kod njih bilo artificijelno je to što će prema psi-efektima velika većina psihologa biti inicijalno toliko skeptična da dokaze neće prihvatiti olako – ne tako lako kako se u literaturi prihvaćaju dokazi za nebrojeno mnogo drugih psihologijskih učinaka koji su manje kontroverzni.

Veliko replikacijsko istraživanje OSC-a

Još 2012. godine u posebnome broju časopisa *Perspectives on Psychological Science* izišao je kratak opis nove inicijative pod nazivom *Reproducibility Project*. Cilj je toga projekta bio „[p]ronaći smislenu procjenu reproduktibilnosti [prema današnjoj terminologiji ustvari replikabilnosti]” (Open Science Collaboration, 2012, str. 657) literature psihologijske znanosti. Da bi se takvo što postiglo, više se skupina znanstvenika okupilo u tzv. *Kolaboratorij otvorene znanosti* (engl. *Open Science Collaboration*) te su tijekom više godina odabrali stotinu istraživanja iz triju psihologijskih časopisa i pokušali ih replicirati. Izvješće (Open Science Collaboration, 2015) o tome velikom pothvatu suradnje objavljeno je u časopisu *Science* 2015. godine te predstavlja najsustavniju kritiku replikabilnosti psihologijske znanosti koju je najavio široki reformski pokret u psihologiji. Prema definiranome protokolu odabrano je 100 istraživanja objavljenih tijekom 2008. godine u trima psihologijskim časopisima (*Journal of Experimental Psychology: Learning, Memory, and Cognition*; *Journal of Personality and Social Psychology*;

Psychological Science) i u svakome od radova ponovljeni su posljednji opisani eksperiment ili nacrt. Članovi *Kolaboratorija* odabrali su jedan nacrt te su ga uz superviziju, prateći replikacijski protokol, pripremili i ponovno ga proveli. Replikacijski je protokol „artikulirao postupak odabira istraživanja i glavne efekte iz priloženih članaka, kako kontaktirati originalne autore te zatražiti popis potrebnih materijala, pripremu istraživačkoga protokola i plana analize, osiguravanje recenzije tako napravljenoga protokola prvotnih autora i drugih članova projekta, javnu registraciju protokola, provedbu replikacije, pisanje završnoga izvješća te kontrolu kvalitete cijeloga procesa” (Open Science Collaboration, 2015, str. 944).

Na taj je način 270 znanstvenika provelo i analiziralo stotinu replikacija. Sistematizirajući različite nacрте i analize, svi su rezultati konvertirani u korelacijske koeficijente i intervale pouzdanosti. Potom su uspoređene dobivene korelacije i intervale pouzdanosti u svakoj od stotinu replikacija s onima u prvotnim istraživanjima (Open Science Collaboration, 2015). Ponuđeno je nekoliko usporedbi da bi se prosudila replikabilnost prvotnih istraživanja. Prva je jednostavno prebrojavanje značajnih efekata u replikacijama naspram prvotnim istraživanjima. Od prvotnih je istraživanja 97 bilo značajno, dok ih je u replikaciji bilo tek 35 (36.1 %; 95 % CI = [26.6 %, 46.2 %]) (Open Science Collaboration, 2015). Druga usporedba koja je napravljena procjena je veličina efekta u replikacijama naspram prvotnim istraživanjima. Pri procjeni nalazi li se veličina efekta prvotnih istraživanja u 95 %-tnome intervalu pouzdanosti replikacija ta je usporedba bila ponešto povoljnija i sugerirala je da se u 47.4 % slučajeva prvotna veličina efekta nalazila u rasponu intervala pouzdanosti efekata u replikacijama. Treća je izravna usporedba veličina učinaka; u tome su slučaju 82 prvotna istraživanja pokazala veće veličine učinaka od replikacijskih (prvotna: $M = 0.403$, $SD = 0.188$; naspram replikacijskima: $M = 0.197$, $SD = 0.257$) (Open Science Collaboration, 2015). Subjektivne procjene replikacijskih timova kada se postavilo pitanje jesu li uspjeli replicirati prvotna istraživanja upućivale su na to da su istraživači u samo 39 od 100 slučajeva procjenjivali uspjeh replikacija. Autori sami zaključuju: „[Z]ajednički, ti rezultati nude jasan zaključak: velika proporcija replikacija proizvela je slabije dokaze od prvotnih rezultata unatoč tomu što su korištene materijale ponudili prvotni autori, što su replikacijski nacrti metodološki vjerni prvotnima i što imaju visoku statističku snagu za dokazivanje prvotnih učinaka” (Open Science Collaboration, 2015).

U trenutku pisanja ovoga teksta veliko replikacijsko istraživanje *Kolaboratorija otvorene znanosti* citirano je 7233 puta. U isto vrijeme *Altmetric*¹ pokazuje izrazito velik odjek toga rada diljem svijeta jer ga spominju i o njemu izvješćuju 227 medijskih priopćenja, 150 blogova, 5 dokumenata javnih politika, 2043 računa na *Twitteru*, 57 stranica na *Facebooku* i 37 *Wikipedijih* stranica. Na stranu kvantitativni indikatori odjeka, taj je rad u 2015. godini predstavio u isto vrijeme dijagnozu širokoga i zabrinjavajućeg problema replikacije svim kvantitativnim psiholozima, ali ujedno ponudio pristup kako taj problem opisati i potencijalno riješiti. Riječima

¹ V. <https://www.altmetric.com/details/4443094?src=bookmarklet>.

samih autora iz zaključka: „Bilo koji pokušaj interpretacije tih rezultata kao poraza za psihologiju ili znanost općenito mora se nositi s činjenicom da ovaj projekt prikazuje znanost kakva bi trebala biti. Mnogobrojne su hipoteze da današnja prevladavajuća kultura u znanosti možda negativno utječe na reproduktivnost [replikabilnost] rezultata. Ideološki odgovor na to bio bi odbaciti argumente, diskreditirati njihove izvore te samo nastaviti po starome. No, znanstveni proces nije ideološki” (Open Science Collaboration, 2015, str. 950). S tim su zaključkom zaista najavili obrise reformskoga pokreta koji će uzeti zamaha u psihologiji u godinama koje dolaze.

Reformski pokret u psihologiji kao odgovor na krizu

Od Bema i Stapela do danas u psihologiji se razvio reformski *grassroots* pokret (Derksen, 2019; Flis, 2019; Wiggins i Christopherson, 2019; za taksativno-kvantitativne indikatore reformskih promjena v. Nosek i sur., 2021, str. 735–738). U ovome ćemo dijelu pregleda reformski pokret prikazati kroz niz glavnih inicijativa i aktivnosti, a ne nužno kako se kronološki razvijao. Bitno je spomenuti da se socijalni život reformskoga pokreta velikim dijelom odvija *na internetu* – u razgovorima glavnih sudionika na *Twitteru* i *Mastodonu*, u većemu broju grupa na *Facebooku* u kojima sudjeluju znanstvenici zainteresirani za replikaciju i metodološku reformu, u nizu *podcastova* reformatora i zainteresiranih studenata te putem *preprint*-servera (za više v. Crüwell i sur., 2019; Derksen i Field, 2021; Field, 2022; Spellman, 2015). Reformski je pokret, s obzirom na njegovu horizontalnu/volontersku/*internetsku* prirodu, teško svesti na brojčane indikatore produkcije poput citata ili broja sudionika. Najbolji se opis te raspršene zajednice može pronaći u doktorskoj disertaciji koja opisuje reformski pokret etnografski i kroz mrežne analize profila i objava na *Twitteru* (Field, 2022). Naš je pregled nužno ograničen, no potrudit ćemo se predstaviti ulazne točke u toj živoj raspravi koja se odvija na internetu i u znanstvenim časopisima.

Predregistracija i registrirana izvješća

Jedno je od glavnih uvjerenja reformatora da su psihologijski znanstvenici predugo imali preniske kriterije za prihvaćanje raznih tvrdnji u literaturi kao znanstvenih dokaza. Drugim riječima, psiholozi su sebe i druge prelako uvjerali u postojanost ili čak istinitost fenomena čije su učinke kvantitativno mjerili. Jedan je od glavnih odgovora reformskoga pokreta na slabost dokazivanja u psihologiji zastupanje predregistracije (engl. *preregistration*) kao intervencije koja će riješiti mnoge probleme koji navodno zahvaćaju literaturu, ponajprije nerazlikovanje eksploratornoga istraživanja nasuprot konfirmatornomu istraživanju, a zatim i publikacijsku pristranost i analitičku fleksibilnost pri obradi podataka.

Kod predregistracije znanstvenici koji provode istraživanje registriraju plan analize kojim se „obvezuju na niz analitičkih koraka bez prethodnih spoznaja o

ishodima samoga istraživanja” (Nosek i sur., 2018, str. 2601). Takvo se obvezivanje često postiže deponiranjem plana analize na neki od nezavisnih mrežnih registara, a u psihologiji se najčešće koristi onaj na platformi *Open Science Foundation*. Registrirano izvješće (engl. *Registered Report*) implementacija je predregistracije u obliku znanstvenoga članka. Najjednostavnije rečeno, registrirano je izvješće ono u kojemu se opća metodološka praksa predregistracije formalizira u jedan od koraka samoga postupka recenzije. U registriranome izvješću autori najprije na recenziju šalju uvod te opis korištenih metodologija i planiranih analiza. Dakle, takva je predregistracija kao dio registriranoga izvješća recenzirana, a u nekim slučajevima čak i objavljena. Ako takav registrirani protokol prođe recenziju, časopis poziva autore da prikupe podatke i napišu cjelovit članak, uključujući dijelove s rezultatima i raspravom. S obzirom na prvi korak, časopis se donekle obvezuje na objavljivanje rukopisa neovisno o tome jesu li rezultati značajni. I u drugome koraku postoji recenzija, ali u principu je rad objavljujiv ako je prošao prvi korak recenzije protokola.²

Predregistracija i registrirana izvješća u biti predstavljaju intervenciju u to kako psiholozi objavljuju i kako provode istraživanja. Takav format osigurava razliku između eksploratornoga i konfirmatornog istraživanja jer istraživači moraju jasno specificirati hipoteze koje pokušavaju potvrditi ili opovrgnuti ili, ako ih nemaju, upozoriti čitatelja da rade eksploratorne analize (više o toj razlici v. u Wagenmakers i sur., 2012). Isto tako, jasno se razlikuje što je u radu predikcija, a što postdikcija, odnosno objašnjenje dobivenih rezultata nakon što su rezultati prikupljeni. Nakraju, upućuje se same urednike da se ne fokusiraju na rezultate pri procjeni je li nešto objavljivo, već na metode, hipoteze i teorijsku razradu. Drugim riječima, takav način objavljivanja rješava ogromne probleme s pristranošću i pritiskom za objavljivanje značajnih rezultata (Nosek i sur., 2018, str. 2602):

U principu, zaključci iz predregistriranih analiza bit će replikabilniji nego analize koje koriste testove značajnosti nultih hipoteza koje nisu predregistrirane zato što međudnos raznih istraživačevih odabira pri analizi i rezultata ne može biti pod utjecajem motivacije, sjećanja ili pristranosti pri rasuđivanju.

Većina reformatora smatra da pravilno primijenjena predregistracija rješava probleme analitičke fleksibilnosti. Primjerice, Scheel i suradnici (2021) ispitivali su razliku u proporciji objavljenih pozitivnih rezultata u registriranim izvješćima naspram klasičnomu formatu objavljivanja istraživanja koja testiraju hipoteze. Njihova je analiza pokazala da u klasičnim istraživanjima 96 % istraživanja prijavljuje pozitivne rezultate, a u registriranim izvješćima samo 44 %. Uzimajući to u obzir, nije čudno da je predregistracija središnja reformaska intervencija u posljednjih deset godina iako još uvijek nema nedvosmislenih „istraživanja čiji rezultati potvrđuju da su rezultati u registriranim izvješćima u prosjeku replikabilniji

² Za primjer registriranoga izvješća v. upute autorima časopisa *Royal Society Open Science*: <https://royalsocietypublishing.org/rsos/registered-reports>.

od drugih” (Nosek i sur., 2021, str. 731). Kako god, predregistracija i registrirana izvješća među najuspješnijim su reformskim strategijama – Chambers i Tzavella (2022) prijavljuju da je od osmišljavanja i uvođenja prvih registriranih izvješća u časopisima *Cortex* i *Perspectives on Psychological Science* u 2012. godini do danas prisutno više od 300 časopisa koji nude takav format objavljivanja u raznim disciplinama (za trend i opis rasta u posljednjih deset godina posebno v. Graf 2. u njihovu radu).

Reformatori pokušavaju promijeniti kulturu objavljivanja u psihologijskim časopisima i razviti različite inicijative za osvještavanje psihologa o metodološkim problemima koji su ovdje opisani. Primjer su jedne takve inicijative *TOP Guidelines* (*The Transparency and Openness Promotion Guidelines*; Nosek i sur., 2015), modularni standardi za povećanje transparentnosti u objavljivanju. Predregistracija je dio toga paketa za povećanje transparentnosti, a sve s ciljem postavljanja jasnih i jačih kriterija za dokazivanje postojanja fenomena i učinaka u psihologijskoj literaturi.

Centar za otvorenu znanost i PsyArxiv

U psihologiji se reformski pokret koji je nastao kao odgovor na replikacijsku krizu često poistovjećuje s pokretom za otvorenu znanost. Otvorena znanost mnogo je širi političko-znanstveni pokret³ koji zahvaća sve znanstvene discipline i institucije te ima dulju povijest od same replikacijske krize jer se povezuje s pokretom za otvoreni pristup znanstvenim časopisima, otvorenim podacima (Fecher i Friesike, 2014) i softverskim pokretima otvorenoga koda. No, miješanje psihologijske metodološke reforme i otvorene znanosti nije ni slučajno niti sasvim pogrešno. Gotovo su svi reformatori u psihologiji istovremeno i promotori raznih aspekata otvorene znanosti te je uvriježeno mišljenje da nove prakse transparentnosti i dijeljenja otvorene znanosti predstavljaju već gotova rješenja za replikacijsku krizu. Nije slučajno da se skupina istraživača koji su objavili veliko replikacijsko istraživanje 2015. godine nazvala *Kolaboratorij za otvorenu znanost*.

Središnji je institucionalni okvir za otvorenu znanost u psihologiji neprofitna organizacija *Centar za otvorenu znanost* (engl. *Center for Open Science*, COS) u SAD-u i njihov *internetski* servis OSF (*Open Science Framework*). OSF je „oruđe koje promovira otvorene i centralizirane načine [znanstvenoga] rada tako što obuhvaća različite aspekte i proizvode istraživačkoga ciklusa koji uključuju razvoj istraživačke ideje, postavljanje nacrt, pohranu i analizu prikupljenih podataka te pisanje i objavljivanje izvješća i radova” (Foster i Deardorff, 2017). Jednostavno rečeno, OSF se prometnuo u glavnu infrastrukturu zajednice reformatora putem koje se dijele razni elementi već provedenih ili tekućih istraživačkih projekata.

³ *Politički* u tehničkome smislu, a ne u pejorativnome, jer pretpostavlja određenu razinu kolektivne organizacije, lobiranja institucija, reorganizacije rada i izravnoga utjecaja na javne politike. Neki ga autori kritiziraju kao *ideološki*, ali ta rasprava prelazi okvire ovoga pregleda (Mirowski, 2018; za obranu otvorene znanosti v. Uygun Tunç i sur., 2022).

Većini istraživača u psihologiji OSF je vjerojatno poznat preko jednoga od njihovih najuspješnijih servisa, a to je *PsyArxiv*. *PsyArxiv* je, po uzoru na *ArXiv* u fizici, *preprint*-server za psihologiju. Istraživači iz cijeloga svijeta mogu ga koristiti da bi deponirali verzije svojih rukopisa prije nego što ih pošalju na recenziju. Time se ubrzava širenje informacija unutar znanstvene zajednice i umanjuje moć urednika pojedinih časopisa jer rukopis može biti dostupan znanstvenoj zajednici čak i ako nije objavljen. *PsyArxiv* je jedan od projekata razvijenih u *Društvu za poboljšanje psihologijske znanosti* (*Society for the Improvement of Psychological Science*, SIPS).

Društvo za poboljšanje psihologijske znanosti

SIPS (<https://improvingpsych.org/>) je nastao kao svojevrsna institucionalizacija reformskoga pokreta u znanstveno društvo. Kako kažu u misiji društva, „SIPS okuplja istraživače koji rade na poboljšanju metoda i praksi psihologijske znanosti”. Cilj im je poboljšanje psihologijske znanosti kroz unaprjeđenje obrazovnih i istraživačkih standarda znanstvenika u psihologiji, uvođenje boljih institucionalnih politika na razini časopisa, drugih znanstvenih društava, odjela i sveučilišta, provedba empirijskih provjera trenutnoga stanja i reformi znanstvenih praksi te izgradnja zajednice reformatora. Društvo je izrazito mlado te privlači velik broj studenata, doktoranada, poslijedoktoranada i docenata.

Društvo organizira svoju godišnju konferenciju u neobičnome formatu. Naime, manji je dio te konferencije određen unaprijed jer se većina programa konferencije razvija tijekom samoga susreta. Sudionici malo prije samoga događaja i tijekom njega predlažu različite aktivnosti (tzv. *hack-a-thon* i *unconferences*) koje nemaju klasičan oblik znanstvenih konferencija *ex cathedra*, već više izgledaju kao suradničke radionice. Sudionici predlažu različite inovativne inicijative, pronalaze istraživače koji misle da su takve inicijative vrijedne te se uključuju u zajednički rad koji često nastavlja i nakon same konferencije. Takav zajednički rad početna je suradnja na samoj konferenciji koja često prerasta u višemjesečnu ili ponekad višegodišnju suradnju tako oformljene grupe istraživača koja zajedno objavljuje, vodi neku inicijativu ili osniva novu instituciju otvorene znanosti. Drugim riječima, konferencije su SIPS-a inkubator ideja i inicijativa koje se tijekom godina razvijaju unutar reformskoga pokreta. Neke od njih postaju i uspješni projekti koji nastavljaju postojati neovisno o SIPS-u. Jedan je takav primjer *Psihologijski znanstveni akcelerator* (*Psychological Science Accelerator*, PSA).

Psihologijski znanstveni akcelerator i druge distribuirane istraživačke mreže

PSA (<https://psysciacc.org/>) je nastao kao inicijativa organizirana na jednome od sastanaka u okviru konferencije SIPS-a gdje su okupljeni psiholozi došli na ideju da osnuju „CERN za psihologe” koji žele provoditi velika suradnička istraživanja. Danas je PSA globalno distribuirana mreža psihologijskih laboratorija (u trenutku pisanja ovoga rada u PSA-i sudjeluje preko 500 istraživačkih skupina iz 71 zemlje). Mreža koordinira i provodi velika međukulturalna istraživanja – demokratski odabire

koja će se istraživanja provesti, provodi prikupljanje podataka, analize, objavljivanje te trening/obrazovanje znanstvenika koji sudjeluju u radu mreže. *Akcelerator* trenutno organizira ili je organizirao devet istraživanja te su neka od njih već objavljena (npr. Jones i sur., 2021).

PSA je jedan od izrazito uspješnih primjera temeljne strategije reformatora – koordinacije i provedbe velikih suradničkih istraživanja s desecima ili stotinama istraživačkih skupina. Primjeri za takva istraživanja su mnogi, a ranije opisano istraživanje *Kolaboratorija za otvorenu znanost* (2015) jedan je od njih. Drugi je primjer inicijativa *ManyBabies* (<https://manybabies.github.io/>) u kojoj razvojni psiholozi organiziraju suradnje za istraživanja na djeci i dojenčadi te pokušavaju replicirati i poboljšati istraživačke standarde razvojne psihologije. Treći primjer koji ćemo spomenuti serija je suradničkih projekata *Many Labs* koja je završila 2022. godine i obuhvatila 13 provedenih velikih replikacijskih istraživanja s tisućama sudionika (za pregled v. Williamson, 2022; za pojedina istraživanja v. npr. Ebersole i sur., 2016, 2020; Klein i sur., 2018).

Velika suradnička istraživanja još su jedna ekspresija metodološke kritike reformatora: učinci u psihologiji gotovo su univerzalno maleni te su potrebni veliki resursi, metodološka sofisticiranost i transparentnost da bi se zaista detektiralo ono što postoji. PSA i druge suradničke mreže, često organizirane kroz SIPS, predstavljaju organizacijsku inovaciju kojom se pokušava pronaći rješenje za probleme psihologijske znanosti.

Dosad smo spomenuli veći broj inicijativa i intervencija u istraživački rad psihologa koje predlažu reformatori. Koliko su uspješne te intervencije te kako reformatori identificiraju probleme koje treba rješavati? Nudeći odgovore na ta pitanja, uz reformski se pokret razvijala i njegova istraživačka ruka koja se naziva metaznanost.

Metaznanost

Problemi identificirani u replikacijskoj krizi naizgled su autohtoni za poddiscipline psihologije. To, naravno, nije zaista tako iako je psihologija jedno od područja koje je najviše opisano u diskursu krize te vjerojatno područje s najjačim reformskim pokretom kao odgovorom. Problemi s replikacijom i lošim metodološkim standardima uglavnom zahvaćaju sve discipline koje uvriježeno koriste testiranje značajnosti kao kriterij za ulazak članaka u literaturu (za međuodnos literature, reforme i metodologije v. Flis, 2022) te su prepoznati i u drugim područjima poput biomedicine ili drugih kvantitativnih društvenih znanosti (za perspektivu iz područja biomedicine, koja je posebice kritična prema izuzimanju psihologije kao posebne „žrtve” replikacijske krize, v. Nelson i sur., 2022). Pri tome se identificiranju problema s replikabilnošću i lošim metodološko-statističkim standardima u širem kontekstu od psihologije razvila nova mlada „znanost o znanosti” koju njezini praktikanti nazivaju metaznanost.

Sociolozi Peterson i Panofsky (2021, str. 2–3) ovako opisuju metaznanost:

Društveni pokret znanstvenika koji pokušavaju koristiti oruđa znanosti – posebice kvantifikaciju i eksperimentiranje – da bi dijagnosticirali probleme u istraživačkim praksama te poboljšali efikasnost istraživanja. Okuplja podatkovne znanstvenike, eksperimentalne i statističke metodologe i aktiviste otvorene znanosti u projekt koji se bavi intelektualnim radom i javnim politikama.

Drugim riječima, metaznanstvena istraživanja daju reformatorima u psihologiji podstrek za reformske intervencije. I ne samo to, već reformatori iz psihologije sudjeluju u metaznanosti kao ključni generatori problema i interpretacija (v. Flis, 2019 za primjer kako je psihologijski jezik pristranosti pri rasuđivanju postao ključ za interpretaciju replikacijske krize). Primjeri su metaznanstvenih istraživanja scijentometrijski opisi literature Fanellija (npr. 2011, 2013), metodološko-statistički tutorijali Lakensa (npr. 2017, 2013) i evaluacije reformskih intervencija (npr. Hardwicke i sur., 2018). Metaznanost, kako to točno identificiraju Peterson i Panofsky (2021), predstavlja raskrižje na kojemu se nalaze tri dijela reformskoga pokreta: aktivizam za otvorenu znanost, potom znanost o znanosti te metodološke i statističke kritike.

Kritika reforme

Reformski prijedlozi ne nailaze uvijek na suglasnost. Već smo spomenuli otpor određenih eksperimentalnih socijalnih psihologa u raspravi o konceptualnoj i izravnoj replikaciji. Zrelije, zanimljive kritike došle su iz područja kognitivne znanosti, a posebice od znanstvenika koji se bave kognitivnim modeliranjem i njihova otpora predregistraciji. Veći broj kognitivnih znanstvenika iskazao je otpor inicijativama za predregistraciju, argumentirajući zbog čega za njihova istraživanja prednosti predregistracije nisu toliko velike koliko reformatori tvrde, a s druge strane potencijalno predstavljaju problem za kvalitetno istraživanje. Oni smatraju da poboljšanje funkcioniranja inferencijalne statistike neće riješiti problem da psiholozi uglavnom testiraju „naslijepo”, ne vođeni predikcijama izvedenima iz svojih teorija, već raznim drugim ciljevima. Čak i kada bi metodološko-statistička mašinerija u osnovi testiranja hipoteza bila poboljšana, tvrde da to nikako ne bi pomoglo činjenici da su same postulirane hipoteze izrazito loše definirane. I ne samo da su loše definirane, nego istraživačku zajednicu navode na krivi trag. Drugim riječima, prema toj kritici predregistracije u većini psihologijske znanosti problem nije analitička fleksibilnost pri korištenju testiranja značajnosti, već manjak robusne i suradnički razvijene teorije (Szollosi i sur., 2020, str. 94):

Rješavanje problema loše primjene statistike s predregistracijom ne kompenzira slabu teoriju. Zamislimo da radimo slučajnu predikciju ishoda eksperimenta. Kada bi se zaista dogodio ishod koji smo predvidjeli, ne bismo to smatrali „teorijom” koja je korisna za neke buduće nove predikcije. Zašto bismo smatrali da je teorija bolja ako je predregistrirana? Kada nemamo nikakav teorijski

razlog za provedbu statističkoga testa, rješavanja statističkih problema [...] trebali bismo smatrati znanstveno besmislenima.

U reformskim je krugovima taj otpor predikciji artikulacija *nove – stare* potrebe za boljim teorijama (za širu raspravu v. van Rooij i Baggio 2020a, 2020b; van Rooij i Blokpoel, 2020; Fried, 2020a, 2020b; Guest i Martin, 2021; Smaldino, 2019, 2020a, 2020b; za pregled v. Flis, 2022). Tako je početna kritika predregistracije prerasla u cijelu novu reformsku literaturu o formalizaciji i konstrukciji teorije. Središnje je pitanje te trenutne rasprave kako razvijati suradničku, kumulativnu, matematički formaliziranu teoriju u psihologiji.

Tehnički najzahtjevniju kritiku reformatorima uputio je tim Berne Devezer (2021) koji tvrdi da metaznanstvena istraživanja koja motiviraju reformske intervencije nemaju dovoljno dobru logičko-teorijsku podlogu. Drugim riječima, oni tvrde da su metaznanstvenici žrtve sličnih problema kao i „obični” znanstvenici koje kritiziraju. Da bi oprimirali svoju kritiku, autori koriste statističke modele da bi istražili tri uvriježene tvrdnje reformatora: (1) da je reproduktibilnost u osnovi svake znanosti, (2) da se isti podaci ne bi trebali koristiti dva puta u istoj analizi⁴, i (3) da eksploratorna istraživanja impliciraju manjkavu primjenu statistike. Devezer i suradnici izvlače potpuno suprotne zaključke od onih uobičajenih u reformskome pokretu te postavljaju pitanje ima li metaznanost dovoljno jasne teorijske formalizme u pozadini svojih proklamacija za promjenu znanstvenih politika i institucija. Ako ih nema, trebamo li slijediti prijedloge metaznanstvenika pri reformi znanstvenih sustava?

To su tek dvije zanimljive kritike reformskoga pokreta. S obzirom na gotovo normativni ikonoklazam usmjeren na sve što se percipira kao „tradicionalne” znanstvene institucije, poput časopisa, recenzija, metodoloških recepata, ne treba čuditi da se unutar reformskoga pokreta pojavljuju i radikalne kritike nekih od osnovnih postulata same reforme. Budućnost će pokazati hoće li se te kritike „ispuhati” ili će pak proizvesti neke dublje promjene u reformi, a možda i u psihologijskoj znanosti.

Zaključak

Reformske su rasprave u psihologiji izrazito kompliciran set statističkih, metodoloških i normativnih pitanja povezanih sa znanstvenim javnim politikama i funkcioniranjem znanstvenoga sustava na sveučilišnoj, nacionalnoj i iznadnacionalnoj razini. Iako su u slučaju psihologije započele u eksperimentalnoj socijalnoj psihologiji, danas su se proširile na gotovo sve poddiscipline psihologije

⁴ Jedna je od reformskih tema koja se može povezati s reproduktibilnošću koliko se puta koriste podaci u analizi, ali ovdje se misli na to da se statistički modeli koji opisuju/reduciraju podatke prikupljene u psihologiji najčešće postavljaju i testiraju na istim podacima. Mnogi reformatori to smatraju lošom praksom, s čime se Devezer i suradnici nužno ne slažu.

(npr. u komparativnoj psihologiji, v. Brecht i sur., 2021; u kliničkoj psihologiji, v. Tackett i sur., 2019).

Predmet rasprave kompleksan je i opisan na više razina, ali svodi se na jedno: znanstvenici u psihologiji imaju preslabe kriterije za dokazivanje svojih tvrdnji. Neki smatraju da ti problemi imaju statistički uzrok i statističko rješenje. Drugi pak drže da je problem u tome kako psiholozi konstruiraju, objavljuju i testiraju teorije. Kako god, oba su supstantivna znanstvena problema uhvaćena u mrežu znanstvenih institucija koje uključuju širi kontekst izdavačke industrije, uredničkih politika časopisa, pritiska za objavljivanje u znanstvenoj zajednici te uvjeta i pravila za napredovanje u znanstvenim zvanjima.

Koji god naš stav o replikacijskoj krizi i reformi bio, dužnost je svakoga istraživača u psihologiji informirati se o tim raspravama. Posebice zbog toga što će ono što je danas predmet rasprave sutra postati zahtjev nacionalnih ili europskih zaklada za znanost ili međunarodnih časopisa o kojima ovise i hrvatski znanstvenici. Institucionalni okviri našega istraživačkoga i nastavnog rada više ili manje ovise o tim raspravama te bismo s te strane svakako trebali sudjelovati u njima.

Reformski je pokret rasadište inovativnih, radikalnih i informiranih inovacija u statistici, metodologiji, konstrukciji teorija, znanstvenome objavljivanju, prekograničnoj znanstvenoj suradnji i filozofiji znanosti. Samo u ovome pregledu dotakli smo se filozofsko-empirijskih definicija ponovljivosti nacрта, formata članaka u znanstvenim časopisima, velikih međunarodnih istraživačkih suradnji i mnogih drugih naizgled nepovezanih tema moderne znanosti. U hrvatskoj su psihologiji teme otvorene znanosti i reformskoga pokreta uglavnom podzastupljene. Iako za njih postoji velik interes, posebice kod studenata i mlađih istraživača, metodološke inovacije reformskoga pokreta uglavnom nisu implementirane na razini naših studijskih programa ili časopisa. Malen je i broj radova koji govore o tim temama (npr. Vlašiček i Flis, 2021). To je velik propust jer znamo da hrvatski znanstveni sustav ima kapaciteta za postavljanje sustava koji implementiraju najbolje prakse otvorene znanosti (v. npr. opis portala *Hrčak* kao primjera dobre prakse, Leonelli, 2018). *Centar za znanstvene informacije* Instituta Ruđer Bošković predvodi u svojim aktivnostima, kao i drugi knjižničari sveučilišnih knjižnica (v. npr. <https://www.crossda.hr/> na Filozofskome fakultetu u Zagrebu). No, kao što je u ovome pregledu prikazano, otvorena znanost i reformski pokret nude mnogo više prilika od samoga otvorenog pristupa podacima ili pedagoških inovacija. Reformski je pokret ulaz u mnogobrojne međunarodne mreže kontakata i neformalnih skupina u različitim područjima istraživanja koje imaju potencijal za osvježavanje hrvatske psihologije i za njezinu integraciju u suvremene rasprave o budućnosti znanosti.

Kako započeti? S obzirom na opseg i kompleksnost rasprava, očekivano je da istraživačima i studentima koji se s njima nisu prije susreli te kritike i reforme mogu izgledati potpuno nedohvatljivo. Da bi ohrabрили zainteresirane, neki autori pozivaju na to da bi se otvorenoj znanosti trebalo pristupiti kao švedskom stolu s kojega možemo uzeti ono što nam se sviđa (Bergmann, 2018) umjesto da joj pristupamo kao setu principa koji najprije moramo u potpunosti svladati, a tek onda primijeniti

u svojem radu. Kathawalla i suradnici (2021) savjetuju sljedeći slijed koraka za polagan ulazak u svijet otvorene znanosti, svaki primjeren raznim fazama istraživačkoga ciklusa. Tako sa svakim dijelom ciklusa možemo povezati neke aktivnosti ponuđene na „švedskome stolu” otvorene znanosti, a one su različito kompleksne po tome koliko ih je teško primijeniti u vlastitome radu. Autori predlažu sljedeće:

- za fazu konceptualizacije: *osnivanje časopisnoga kluba* za čitanje o reformskim temama; *analizu našega tijeka rada* na svakome projektu;
- za fazu izrade nacрта: izrada *predregistracije* ili *registriranoga izvješća*; *planiranje dijeljenja podataka*;
- za fazu analize: *korištenje reproduktibilnoga koda* za sve analize koje koristimo;
- za fazu izvješćivanja: primjena tzv. *transparentnoga pisanja* pri izradi istraživačkih izvješća;
- za fazu diseminacije: korištenje *preprinta* i *dijeljenja podataka*.

Takav hodogram nudi jednostavne upute za uvođenje reformskih tema u svakodnevni rad. Stvar je mnogo kompliciranija kada govorimo o promjeni rada cijelih znanstvenih institucija ili o uvođenju novih kriterija u javne politike povezane sa zapošljavanjem i slično. Smatramo da je u Hrvatskoj još prerano govoriti o takvim inovacijama povezanim s otvorenom znanošću. No, svakako je vrijeme da se i hrvatska psihologijska javnost uključi u te rasprave i doprinese im svojim razmišljanjima i stavovima te postavi sljedeća pitanja studentima i znanstvenicima na svojim katedrama: ima li psihologija kao znanost problem s replikacijom i slabim inferencijalnim strategijama? Ako ima, što ćemo sad?

Literatura

- Ali, M. J. (2021). Understanding the altmetrics. *Seminars in Ophthalmology*, 1–3. <https://doi.org/10.1080/08820538.2021.1930806>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604), 452. <https://doi.org/10.1038/533452a>
- 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>

- Bergmann, C. (2018). *How to integrate open science into language acquisition research*. Student workshop at the 43rd Boston University Conference on Language Development (BUCLD), Boston, USA.
<https://docs.google.com/presentation/d/1bdICPzPOFs7V5aOZA2OdQgSAvgOB6WQwel21kVpk9Gg/edit#slide=id.p>
- Brecht, K., Legg, E. W., Nawroth, C., Fraser, H. i Ostojić, Lj. (2021). The status and value of replications in animal behavior science. *Animal Behavior and Cognition*, 8(2), 97–106.
<https://doi.org/10.26451/abc.08.02.01.2021>
- Carp, J. (2012). On the plurality of (methodological) worlds: Estimating the analytic flexibility of fMRI experiments. *Frontiers in Neuroscience*, 6.
<https://doi.org/10.3389/fnins.2012.00149>
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M. i Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2), 115–144.
<https://doi.org/10.1177/2515245919847196>
- Chambers, C. (2017). *The seven deadly sins of psychology: A manifesto for reforming the culture of scientific practice*. Princeton University Press.
- Chambers, C. D. i Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour*, 6(1), 29–42. <https://doi.org/10.1038/s41562-021-01193-7>
- Clairbout, J. F. i Karrenbach, M. (1992). Electronic documents give reproducible research a new meaning. *SEG Technical Program Expanded Abstracts 1992*, 601–604.
<https://doi.org/10.1190/1.1822162>
- Crüwell, S., van Doorn, J., Etz, A., Makel, M. C., Moshontz, H., Niebaum, J. C., Orben, A., Parsons, S. i Schulte-Mecklenbeck, M. (2019). Seven easy steps to open science. *Zeitschrift Für Psychologie*, 227(4), 237–248. <https://doi.org/10.1027/2151-2604/a000387>
- Derksen, M. (2019). Putting Popper to work. *Theory & Psychology*, 29(4), 449–465.
<https://doi.org/10.1177/0959354319838343>
- Derksen, M. (2021). A menagerie of imposters and truth-tellers: Diederik Stapel and the crisis in psychology. U: S. Woolgar, E. Vogel, D. Moats i C.-F. Helgesson (Ur.), *The imposter as social theory: Thinking with gatecrashers, cheats and charlatans* (str. 53–76). Bristol University Press.
- Derksen, M. i Field, S. (2022). The tone debate: Knowledge, self, and social order. *Review of General Psychology*, 26(2), 172–183. <https://doi.org/10.1177/10892680211015636>
- Devezer, B., Navarro, D. J., Vandekerckhove, J. i Ozge Buzbas, E. (2021). The case for formal methodology in scientific reform. *Royal Society Open Science*, 8(3), 200805.
<https://doi.org/10.1098/rsos.200805>
- Dickersin, K. i Min, Y.-I. (1993). Publication bias: The problem that won't go away. *Annals of the New York Academy of Sciences*, 703(1), 135–148. <https://doi.org/10.1111/j.1749-6632.1993.tb26343.x>

- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., Davis, W. E., ... Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68–82. <https://doi.org/10.1016/j.jesp.2015.10.012>
- Ebersole, C. R., Mathur, M. B., Baranski, E., Bart-Plange, D. J., Buttrick, N. R., Chartier, C. R., Corker, K. S., Corley, M., Hartshorne, J. K., IJzerman, H., Lazarević, Lj. B., Rabagliati, H., Ropovik, I., Aczel, B., Aeschbach, L. F., Andrighetto, L., Arnal, J. D., Arrow, H., Babincak, P., ... Szecsi, P. (2020). Many Labs 5: Testing pre-data-collection peer review as an intervention to increase replicability. *Advances in Methods and Practices in Psychological Science*, 3(3), 309–331. <https://doi.org/10.1177/2515245920958687>
- Fanelli, D. (2011). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891–904. <https://doi.org/10.1007/s11192-011-0494-7>
- Fanelli, D. (2013). Why growing retractions are (mostly) a good sign. *Plos Medicine*, 10(12), e1001563. <https://doi.org/10.1371/journal.pmed.1001563>
- Faye, C. (2012). American social psychology: Examining the contours of the 1970s crisis. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(2), 514–521. <https://doi.org/10.1016/j.shpsc.2011.11.010>
- Feest, U. (2019). Why replication is overrated. *Philosophy of Science*, 86(5), 895–905. <https://doi.org/10.1086/705451>
- Fecher, B. i Friesike, S. (2014). Open science: One term, five schools of thought. U: S. Bartling i S. Friesike (Ur.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (str. 17–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_2
- Ferguson, C. J. i 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. <https://doi.org/10.1177/1745691612459059>
- Field, S. M. (2022). *Charting the constellation of science reform* [Neobjavljena doktorska disertacija]. Sveučilište u Groningenu. https://pure.rug.nl/ws/portalfiles/portal/229114777/Complete_thesis.pdf
- Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, 29(2), 158–181. <https://doi.org/10.1177/0959354319835322>
- Flis, I. (2022). The function of literature in psychological science. *Review of General Psychology*, 26(2), 146–156. <https://doi.org/10.1177/10892680211066466>
- Foster, E. D. i Deardorff, A. (2017). Open Science Framework (OSF). *Journal of the Medical Library Association: JMLA*, 105(2), 203–206. <https://doi.org/10.5195/jmla.2017.88>

- Franco, A., Malhotra, N. i Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345(6203), 1502–1505.
<https://doi.org/10.1126/science.1255484>
- Frey, B. B. (2018). Generalizability. *The SAGE Encyclopedia of educational research, measurement, and evaluation*. SAGE Publications, Inc.
<https://doi.org/10.4135/9781506326139.n284>
- Fried, E. I. (2020a). Lack of theory building and testing impedes progress in the factor and network literature. *Psychological Inquiry*, 31(4), 271–288.
<https://doi.org/10.1080/1047840X.2020.1853461>
- Fried, E. I. (2020b). Theories and models: What they are, what they are for, and what they are about. *Psychological Inquiry*, 31(4), 336–344.
<https://doi.org/10.1080/1047840X.2020.1854011>
- Gelman, A. i Loken, E. (2013). *The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time* [Neobjavljeni rukopis].
<http://stat.columbia.edu/~gelman/research/unpublished/forking.pdf>
- Guest, O. i Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789–802.
<https://doi.org/10.1177/1745691620970585>
- Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B. i Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal *Cognition*. *Royal Society Open Science*, 5(8), 180448.
<https://doi.org/10.1098/rsos.180448>
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T. i Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *Plos Biology*, 13(3), e1002106.
<https://doi.org/10.1371/journal.pbio.1002106>
- John, L. K., Loewenstein, G. i 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>
- Jones, B. C., DeBruine, L. M., Flake, J. K., Liuzza, M. T., Antfolk, J., Arinze, N. C., Ndukaihe, I. L. G., Bloxson, N. G., Lewis, S. C., Feroni, F., Willis, M. L., Cubillas, C. P., Vadio, M. A., Turiegano, E., Gilead, M., Simchon, A., Saribay, S. A., Owsley, N. C., Jang, C., ... Coles, N. A. (2021). To which world regions does the valence-dominance model of social perception apply? *Nature Human Behaviour*, 5(1), 159–169.
<https://doi.org/10.1038/s41562-020-01007-2>
- Kathawalla, U. K., Silverstein, P. i Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra: Psychology*, 7(1).
<https://doi.org/10.1525/collabra.18684>

- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams Jr., R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahnik, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzaska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., ... Sowden, W. (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>
- Kukull, W. A. i Ganguli, M. (2012). Generalizability: The trees, the forest, and the low-hanging fruit. *Neurology*, 78(23), 1886–1891. <https://doi.org/10.1212/WNL.0b013e318258f812>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00863>
- Lakens, D. (2017). Equivalence tests: A Practical primer for t-tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- LeBel, E. P., McCarthy, R. J., Earp, B. D., Elson, M. i Vanpaemel, W. (2018). A unified framework to quantify the credibility of scientific findings. *Advances in Methods and Practices in Psychological Science*, 1(3), 389–402. <https://doi.org/10.1177/2515245918787489>
- Leonelli, S. (2018). *Implementing open science: Strategies, experiences and models*. European Commission, Directorate-General for Research and Innovation, Thematic Report No. 4. https://ore.exeter.ac.uk/repository/bitstream/handle/10871/31337/MLE%20OS_4th%20thematic%20report.pdf?sequence=1
- Lynott, D., Corker, K. S., Wortman, J., Connell, L., Donnellan, M. B., Lucas, R. E. i O'Brien, K. (2014). Replication of “Experiencing physical warmth promotes interpersonal warmth”. *Social Psychology*, 45(3), 216–222. <https://doi.org/10.1027/1864-9335/a000187>
- Maatman, F. O. (2021). Psychology’s theory crisis, and why formal modelling cannot solve it. *PsyArXiv*. <https://doi.org/10.31234/osf.io/puqvs>
- Mertens, G. i Kryptos, A.-M. (2019). Preregistration of analyses of preexisting data. *Psychologica Belgica*, 59(1), 338–352. <https://doi.org/10.5334/pb.493>
- Milas, G. (2005). *Istraživačke metode u psihologiji i drugim društvenim znanostima*. Naklada Slap.
- Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, 48(2), 171–203. <https://doi.org/10.1177/0306312718772086>
- Morawski, J. (2019). The replication crisis: How might philosophy and theory of psychology be of use? *Journal of Theoretical and Philosophical Psychology*, 39(4), 218–238. <https://doi.org/10.1037/teo0000129>

- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., du Sert, N. P., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J. i Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021. <https://doi.org/10.1038/s41562-016-0021>
- Mülberger, A. (2022). Early experimental psychology: How did replication work before p-hacking? *Review of General Psychology*, 26(2), 131–145. <https://doi.org/10.1177/10892680211066468>
- National Academies of Sciences. (2019). Understanding reproducibility and replicability. U: National Academies of Sciences, Engineering, and Medicine (Ur.), *Reproducibility and replicability in science* (str. 39-54). National Academies Press (US). <https://www.ncbi.nlm.nih.gov/books/NBK547546/>
- Navarro, D. J. (2021). If mathematical psychology did not exist we might need to invent it: A comment on theory building in psychology. *Perspectives on Psychological Science*, 16(4), 707–716. <https://doi.org/10.1177/1745691620974769>
- Nelson, N. C., Chung, J., Ichikawa, K. i Malik, M. M. (2022). Psychology exceptionalism and the multiple discovery of the replication crisis. *Review of General Psychology*, 26(2), 184–198. <https://doi.org/10.1177/10892680211046508>
- Nelson, L. D., Simmons, J. i Simonsohn, U. (2018). Psychology's renaissance. *Annual Review of Psychology*, 69, 511–534. <https://doi.org/10.1146/annurev-psych-122216-011836>
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, 348(6242), 1422–1425. <https://doi.org/10.1126/science.aab2374>
- Nosek, B. A. i Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, 23(3), 217–243. <https://doi.org/10.1080/1047840X.2012.692215>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C. i Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Nosek, B. A. i Errington, T. M. (2020). What is replication? *PloS Biology*, 18(3), e3000691. <https://doi.org/10.1371/journal.pbio.3000691>
- Nosek, B., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M., Rohrer, J., Romero, F., Scheel, A., Scherer, L. D., Schönbrodt, F. i Vazire, S. (2021). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- NWO. (2018). *Study into the publication cultures: Social science san humanities* [Izvrješće]. https://www.nwo.nl/sites/nwo/files/documents/Study%20into%20Publication%20Cultures_0kt2018_UK.pdf

- Oberauer, K. i Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5), 1596–1618. <https://doi.org/10.3758/s13423-019-01645-2>
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science*, 7(6), 657–660. <https://doi.org/10.1177/1745691612462588>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Pashler, H. i Wagenmakers, E.-J. (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>
- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227. <https://doi.org/10.1126/science.1213847>
- Peterson, D. i Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science*, 51(4), 583–605.
- Peterson, D. i Panofsky, A. (2020). Metascience as a scientific social movement. *Minerva*, 1–28. <https://doi.org/10.1007/s11024-023-09490-3>
- Peterson, D. i Panofsky, A. (2021). Self-correction in science: The diagnostic and integrative motives for replication. *Social Studies of Science*, 51(4), 583–605. <https://doi.org/10.1177/030631272111005551>
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(11), e12633. <https://doi.org/10.1111/phc3.12633>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Scheel, A. M., Schijen, M. R. i Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science*, 4(2), 25152459211007467. <https://doi.org/10.1177/25152459211007467>
- Schooler, J. W. (2014). Metascience could rescue the 'replication crisis.' *Nature*, 515, 7525. <https://doi.org/10.1038/515009a>
- Shrout, P. E. i Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, 69(1), 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Simmons, J. P., Nelson, L. D. i Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Smaldino, P. (2019). Better methods can't make up for mediocre theory. *Nature*, 575(7781), 9. <https://doi.org/10.1038/d41586-019-03350-5>

- Smaldino, P. (2020a). How to translate a verbal theory into a formal model. *Social Psychology*, 51(4), 207–218. <https://doi.org/10.1027/1864-9335/a000425>
- Smaldino, P. (2020b). How to build a strong theoretical foundation. *Psychological Inquiry*, 31(4), 297–301. <https://doi.org/10.1080/1047840X.2020.1853463>
- Smaldino, P. E. i McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3(9), 160384. <https://doi.org/10.1098/rsos.160384>
- Spellman, B. A. (2015). A short (personal) future history of revolution 2.0. *Perspectives on Psychological Science*, 10(6), 886–899. <https://doi.org/10.1177/1745691615609918>
- Stroebe, W. i 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>
- Szollosi, A. i Donkin, C. (2021). Arrested theory development: The misguided distinction between exploratory and confirmatory research. *Perspectives on Psychological Science*, 174569162096679. <https://doi.org/10.1177/1745691620966796>
- Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T. i Donkin, C. (2020). Is preregistration worthwhile? *Trends in Cognitive Sciences*, 24(2), 94–95. <https://doi.org/10.1016/j.tics.2019.11.009>
- Tackett, J. L., Brandes, C. M., King, K. M. i Markon, K. E. (2019). Psychology's replication crisis and clinical psychological science. *Annual Review of Clinical Psychology*, 15(1), 579–604. <https://doi.org/10.1146/annurev-clinpsy-050718-095710>
- Uygun Tunç, D., Tunç, M. N. i Eper, Z. B. (2022). Is open science neoliberal? *Perspectives on Psychological Science*. <https://doi.org/10.1177/17456916221114835>
- van Rooij, I. i Baggio, G. (2020a). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *PsyArXiv*. <https://doi.org/10.31234/osf.io/7qbpr>
- van Rooij, I. i Baggio, G. (2020b). Theory development requires an epistemological sea change. *Psychological Inquiry*, 31(4), 321–325. <https://doi.org/10.1080/1047840X.2020.1853477>
- van Rooij, I. i Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. *PsyArXiv*. <https://doi.org/10.31234/osf.io/r2zqy>
- Vazire, S. i Holcombe, A. O. (2022). Where are the self-correcting mechanisms in science? *Review of General Psychology*, 26(2), 212–223. <https://doi.org/10.1177/10892680211033912>
- Vlašiček, D. i Flis, I. (2021). Otvorena znanost. Kratak pregled pokreta i metodološkog značaja. *Revija za sociologiju*, 51(3), 507–516.
- Wagenmakers, E.-J., Wetzels, R., Borsboom, D. i 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>

- Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L. J. i Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7(6), 632–638. <https://doi.org/10.1177/1745691612463078>
- Wiggins, B. J. i Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39(4), 202–217. <https://doi.org/10.1037/teo0000137>
- Williamson, E. (2022, svibanj). *After 10 years, 'Many Labs' comes to an end – but its success is replicable*. UVAToday. <https://news.virginia.edu/content/after-10-years-many-labs-comes-end-its-success-replicable>

Scientific Reform and Replication Crisis in Psychology

Abstract

In the 2010s, psychological science was engulfed in a discussion about the replicability of already published studies. These discussions soon morphed into a multi-layered criticism of quantitative methods and methodologies in psychology, especially when it comes to inferential statistics. The replicability problems were described as a symptom of a deep crisis of commonly applied research designs and statistical analyses, in combination with the dysfunction of scientific publishing and scholarly communication. In this paper, an overview of the replication crisis and reform movement in psychology in Croatia will be provided for the first time.

Keywords: replication crisis, reform in psychology, metascience

Primljeno: 17. 8. 2021.

Tablica 2.
 Prijevodi FORTT-ovih definicija relevantnih pojmova

Prijevod <i>Altmetrics</i>	Engleski termin <i>Altmetrics</i>	Definicija	Izvor
analitička fleksibilnost	<i>analytic flexibility</i>	Udaljavajući se od tradicionalnih mjera citata, <i>Altmetrics</i> (kratica za <i>alternativne metrike</i>) nude procjenu širega odjeka istraživačkoga rada na temelju raznih izvora kao što su društveni mediji (npr. <i>Twitter</i>), digitalni novinski izvori, broj preuzimanja <i>preprinta</i> itd. <i>Altmetrics</i> su kritizirani zbog toga što senzacionalne tvrdnje uobičajeno privlače mnogo više pažnje od ozbiljnoga istraživačkoga rada (Ali, 2021).	https://fortt.org/glossary/altmetrics/
izravna replikacija	<i>direct replication</i>	Analiitička fleksibilnost tip je stupnjeva slobode istraživača (Simmons i sur., 2011) koji se specifično odnosi na velik broj odabira koji se donose tijekom čišćenja podataka i statističkih analiza. „Raspon ishoda analize s obzirom na različite prihvatljive metode analize” (Carp, 2012, str. 1). Analitička fleksibilnost može biti problematična s obzirom na to da varijabilnost u analitičkim strategijama može značiti varijabilnost u istraživačkim ishodima, posebice kada se primjenjuje više strategija koje nisu transparentno opisane.	https://fortt.org/glossary/analytic-flexibility/
eksploratome analize	<i>exploratory analyses</i>	S obzirom na to da <i>izravna replikacija</i> nema široko prihvaćeno tehničko značenje, ne postoji jasna razlika između izravne i konceptualne replikacije, a niže navodimo više prijedloga za konsenzus [v. na mrežnim stranicama FORTT-aj. Umjesto rasprave o točnosti replikacije bilo bi bolje raspravljati o relevantnim razlikama između replikacije i prvotnoga istraživanja te njihovim implikacijama za pouzdanost i uopćivost prvotnih rezultata.	https://fortt.org/glossary/direct-replication/
konceptualna replikacija	<i>conceptual replication</i>	<i>V. konfirmatorne analize.</i> Pokušaj replikacije u kojemu je učinak koji se proučava jednak onomu u prvotnoj studiji, ali testiran je na drugačijemu uzorku i drugačije zahvaćen nego što je prvotno opisan (npr. korištenjem drugačijih operacionalizacija, obrade podataka i statističke analize i/ili drugačijih konstrukata; LeBel i sur., 2018). Uobičajeno je cilj konceptualne replikacije istraživanje uvjeta koji ograničavaju generalizaciju i opažljivost nekoga učinka (npr. samo u nekim kontekstima, samo na određenim uzorcima, samo kada koristimo određene pristupe mjerenju) u svrhu evaluacije i unapređivanja teorije.	https://fortt.org/glossary/conceptual-replication/

Prijevod	Engleski termin	Definicija	Izvor
konfirmatome analize	<i>confirmatory analyses</i>	Jedna strana suprotstavljenih pojmova <i>konfirmatorno – eksploratorno</i> (Wagenmakers i sur., 2012), gdje su konfirmatorne analize one analize koje su određene unaprijed i testiraju već postojeće hipoteze. Nedostatak opisa te razlike u objavljenim istraživanjima jedno je od objašnjenja za probleme s replikabilnošću i predlaže se rješavanje toga problema pomoću predregistracije koja jasno razlikuje konfirmatorne od eksploratornih analiza. Drugi istraživači dovode u pitanje te termine i preporučuju njihovu zamjenu terminima <i>istraživanje usmjereno na otkriće i istraživanje usmjereno na testiranje teorija</i> (Oberauer i Lewandowsky, 2019; v. i Szollosi i Donkin, 2019).	https://forrt.org/glossary/confirmatory-analyses/
metaznanost	<i>meta-science</i> ili <i>meta-research</i>	Znanstveno proučavanje same znanosti s ciljem opisivanja, objašnjavanja, evaluacije i/ili poboljšanja znanstvenih praksi. Metaznanost uobičajeno istražuje znanstvene metode, analize, objavljivanje i evaluaciju podataka, reproduktibilnost i replikabilnost rezultata istraživanja te institucionalne poticaje [<i>incentives</i>] istraživanja.	https://forrt.org/glossary/meta-science-or-meta-research/
pecanje značajnosti	<i>significance fishing</i>	V. <i>p-hakiranje</i> .	
<i>p-hakiranje</i>	<i>p-hacking</i>	Korištenje tehnika koje umjetno povećavaju vjerojatnost dobivanja statistički značajnoga rezultata tako da se ispuni standardni kriterij statističke značajnosti (uobičajeno $\alpha = .05$). Npr. provođenje više analiza i prijavljivanje samo onih kod kojih je $p < .05$, selektivno odstranjivanje podataka do $p < .05$, odabir varijabli za analize na temelju statistički značajnih parametara.	https://forrt.org/glossary/p-hacking/
predregistracija	<i>preregistration</i>	Praksa objavljivanja plana istraživanja koji uključuje istraživačke probleme/hipoteze, naert te plan analize podataka prije nego što su podaci prikupljeni i proučeni. Isto tako, moguće je predregistrirati sekundarne analize podataka (Mertens i Kryotos, 2019). Na predregistracijskome je dokumentu naznačeno kada je napravljen te se uobičajeno registrira kod neke neovisne stranke (npr. repozitorija), što omogućuje njegovo javno dijeljenje s drugima (ponekad nakon razdoblja embarga). Predregistracija omogućuje transparentnu dokumentaciju toga što je bilo planirano u određenome trenutku i omogućuje raznim trećim stranama da procijene kakve su promjene napravljene nakon toga trenutka. Sto je predregistracija detaljnija, treće strane mogu bolje procijeniti promjene, a shodno tomu i valjanost napravljenih analiza. Cilj je predregistracije jasno razlikovanje konfirmatornoga naspram eksploratornomu istraživanja.	https://forrt.org/glossary/preregistration/

Prijevod	Engleski termin	Definicija	Izvor
<i>preprint</i>	<i>preprint</i>	Javno dostupna verzija bilo koje vrste znanstvenoga rukopisa / proizvoda istraživanja prije formalnoga objavljivanja koja se smatra tipom zelenoga otvorenog pristupa. <i>Preprinti</i> su uobičajeno smješteni u repozitorije (npr. <i>arXiv</i>) koji pomažu u diseminaciji time što dijele rezultate istraživanja brže nego tradicionalno objavljivanje. Repozitoriji <i>preprinta</i> uobičajeno <i>preprintima</i> pridaju stalne identifikatore (npr. brojeve DOI). <i>Preprinti</i> se mogu diseminirati u bilo kojernu dijelu istraživačkoga ciklusa, ali najčešće se pohranjuju pri prijavi u časopis (prije recenzije). Prihvaćeni radovi i recenzirane verzije članaka isto su tako često pohranjeni na <i>preprint</i> -serverima i nazivaju se <i>postprintovi</i> .	https://forrt.org/glossary/preprint/
problem dokument – ladica	<i>file-drawer problem</i>	V. publikacijska pristranost.	
publikacijska pristranost	<i>publication bias (file-drawer problem)</i>	Neobjavljivanje rezultata na temelju „smjera ili snage rezultata istraživanja” (Dickersin i Min, 1993, str. 135). Pristranost nastaje kada procjena objektivnosti istraživanja ovisi o ishodu istraživanja, često s preferencijom novih i značajnih rezultata kao vjernih za objavljivanje od replikacija i radova koji prihvaćaju nulte hipoteze. Ta se pristranost pojavljuje kao disproporcionalno velik broj značajnih nalaza i „napuhane” veličine učinka. Taj proces vodi do nereprezentativnosti objavljene znanstvene literature za cjelokupni raspon istraživačkoga rada te specifično podcjenjuje nulte rezultate. Takvi rezultati završavaju u tzv. „ladici za dokumente”, otkud ne bivaju objavljeni niti imaju dokumentacije koja ih bilježi.	https://forrt.org/glossary/publication-bias-file-drawer-proble/
registrirano izvješće	<i>Registered Report</i>	Format znanstvenoga objavljivanja koji uključuje inicijalnu recenziju pozadinske pripreme istraživanja i metoda (nacrt, mjerenja i plana analize); dovoljno kvalitetni rukopisi u principu se prihvaćaju (<i>in principle acceptance</i> , IPA) u toj fazi. Uobičajeno je da se ta recenzija prve razine provede prije prikupljanja podataka, no i sekundarne su analize moguće u tome formatu objavljivanja. Nakon analize podataka te pisanja rezultata i rasprave rukopisa recenzija druge razine procjenjuje jesu li se autori dovoljno pridržavali početnoga plana istraživanja i pravilno prijavili sva odstupanja od njega (te jesu li ta odstupanja utjecala na rezultate). Takav postupak preusmjerava pažnju s percipirane zanimljivosti rezultata istraživanja na njegove predložene probleme i metodologije.	https://forrt.org/glossary/registered-report/
replikabilnost	<i>replicability</i>	Krovni termin koji se različito koristi u raznim poljima, a obuhvaća koncepte kao što su izravna i konceptualna replikacija, komputacijska reproduktivnost/replikabilnost, uopćivost analiza i robusnost analiza. Neke ranije korištene definicije uključuju drugačiji tim istraživača koji dolazi do jednakih rezultata koristeći artefakte prvotnih autora (Barba, 2018); istraživanje koje dolazi do jednakih zaključaka nakon prikupljanja novih podataka	https://forrt.org/glossary/replicability/

Prijevod	Engleski termin	Definicija	Izvor
reproduktivnost	<i>reproducibility</i>	(Claerhout i Karrenbach, 1992); istraživanja čiji rezultat ima vrijednost dijagnostičkoga dokaza za tvrdnje iz prijašnjih istraživanja (Nosek i Errington, 2020). Minimalni standard u rasponu aktivnosti („raspon reproduktivnosti”) kojim procjenjujemo vrijednost ili preciznost znanstvenih tvrdnji na temelju prvotnih metoda, podataka i koda. Npr. slučaj kada se podaci i kompjuterski kod prvotnih autora koriste da bi se ponovno dobili rezultati (Barba, 2018), a često se naziva komputacijska reproduktivnost. Reprodutivnost ne podrazumijeva kvalitetu, točnost ili valjanost objavljenih rezultata (Peng, 2011). U nekim se poljima to značenje povezuje s pojmom <i>replikabilnost</i> ili <i>ponovljivost</i> [<i>repeatability</i>].	https://fortt.org/glossary/reproducibility/
robustnost (analiza)	<i>robustness (analyses)</i>	Opstojnost podrške hipotezi pod uvjetima promjena u metodološkim/analitičkim postupcima. Drugim riječima, primjena različitih metoda/analiza da bi se ispitalo mogu li se iz podataka izvesti jednaki zaključci.	https://fortt.org/glossary/robustness-analyses/
stupnjevi slobode istraživača	<i>researcher degrees of freedom</i>	Odnosi se na fleksibilnost koja je često inherentna znanstvenomu procesu, od smišljanja hipoteza, izrade nacrtu i provedbe istraživanja do obrade i analize podataka, kao i interpretacije i izvješćivanja o rezultatima. Zbog manjka precizno definiranih teorija i/ili empirijskih dokaza često je više odluka jednako opravdano. Pojam se nekada koristi za opisivanje oportunističke zlonamjerne upotrebe te fleksibilnosti da bi se postigli željeni rezultati, npr. u odlukama o uključivanju ili isključivanju podataka, iako termin sam po sebi nije vrijednosno opterećen.	https://fortt.org/glossary/researcher-degrees-of-freedom/
test uopćivosti	<i>generalizability</i>	Uopćivost se odnosi na to koliko su rezultati istraživanja primjenjivi na drugim grupama ljudi, kontekstima i situacijama i kako se nalazi istraživanja odnose prema širem kontekstu (Frey, 2018; Kukul i Ganguli, 2012). [Testovima uopćivosti kako su opisani u radu bliža je alternativna FORTT-ova definicija koja u prijevodu glasi:] Primjena doradenih materijala i/ili analitičkih pristupa na novim podacima ili uzorcima da bi se odgovorilo na iste hipoteze (različiti materijali, različiti podaci) u svrhu testiranja uopćivosti učinka koji se proučava.	https://fortt.org/glossary/generalizability/
upitne istraživačke prakse	<i>questionable research practices ili questionable reporting practices (QRPs)</i>	Raspon aktivnosti koje namjerno ili nenamjerno iskrivljavaju podatke u skladu s istraživačevim hipotezama ili propust u izvješćivanju o takvim praksama, a obuhvaća selektivno uključivanje podataka, postavljanje hipoteza nakon što su rezultati poznati (<i>HARKing</i>) i <i>p-hakiranje</i> . Popularizirano u radu Johna i suradnika (2012).	https://fortt.org/glossary/questionable-research-practices-or-/