

**Slobodan Perović**

*KRIZA REPLIKABILNOSTI REZULTATA U SAVREMENOJ  
EKSPERIMENTALNOJ PSIHOLOGIJI U SVETLU KRITERIJUMA  
REPLIKABILNOSTI U SAVREMENOJ FIZICI<sup>1</sup>*

*APSTRAKT: Od 2010. godine nemogućnost replikacije eksperimentalnih rezultata (objavljenih u vodećim časopisima u oblasti) u velikom broju ponovljenih psiholoških eksperimenata počeo je sve više da se percipira kao metodološka kriza. Replikabilnost se generalno govoreći smatra za osnovni standard naučnog eksperimenta, oslanjajući se na uzorne klasične eksperimente u oblasti fizike. No uslovi replikabilnosti variraju zavisno od proučavanog fenomena i u samoj fizici, pa samim tim i kriterijumi adekvatnosti eksperimenta. Tri karakteristična primera eksperimenata koje analiziramo doveli su do tri različita epistemička stava u vezi adekvatnosti eksperimenata na osnovu replikabilnosti rezultata. Ti stavovi i konkretni primeri mogu biti relevantni u razumevanju prirode krize replikabilnosti u psihologiji.*

*KLJUČNE REČI: filozofija nauke, psihologija, fizika, filozofija naučnih eksperimenata, replikabilnost eksperimenata*

## **1. Uvod**

Kriza replikabilnosti rezultata u psihologiji obično se predstavlja kao kriza istraživačke metodologije s obzirom da se replikabilnost eksperimentalnih rezultata tradicionalno smatra za jednu od ključnih karakteristika metoda naučnog eksperimentisanja. (Pashler et al. 2012) Činjenica da veliki broj pokušaja da se reprodukuju rezultati mnogobrojnih eksperimenata u psihologiji od 2010. nije uspeo doveo je do percepcije situacije kao krize i njenih kontroverznih tumačenja. (Maxwell et al. 2015)

Jedna moguća reakcija na krizu replikabilnosti rezultata u psihologiji (kao i slične

---

1 Rad je prezentovan na 23. naučnom skupu "Empirijska istraživanja u psihologiji" održanom 2017. godine na Filozofskom fakultetu u Beogradu. Rad je podržan sredstvima sa projekta #17904 Ministarstva za obrazovanje, nauku i tehnološki razvoj Republike Srbije.

krize u biomedicinskim naukama) bi mogla da bude, međutim, i da percepcija da postoji značajna kriza zapravo nije sasvim opravdana ukoliko se sagleda iz perspektive istorije nauke i sličnih relevantnih slučajeva. Mi ćemo ispitati da li postoje osnove za takav blaži stav prema navedenoj krizi.

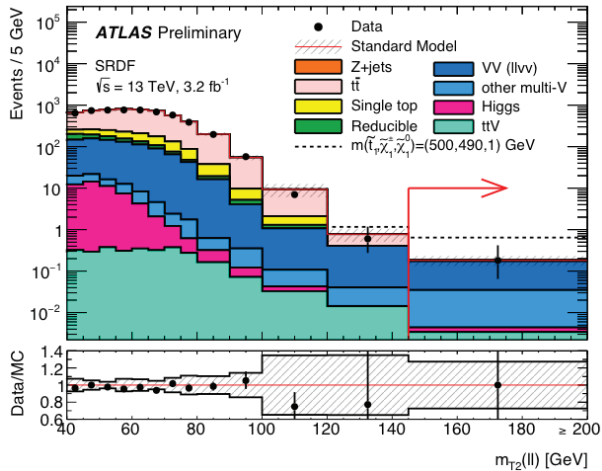
Čak i kada su u pitanju očigledni metodološki prestupi u savremenoj psihologiji možemo navesti poznati primer eksperimenta iz biologije i eksperimenata iz fizike koji su, po nekim tumačenjima, očigledni primeri sličnih kontroverzi. Prvi je primer navodni problem sa načinom na koji je Mendel uradio statističku analizu dobijenih podataka u svojim eksperimentima sa graškom, a koji se percipiraju kao eksperimenti na osnovu kojih je on odredio osnovne zakonitosti rekombinacije gena. (Franklin et al. 2008) Drugi primer su Edingtonovi eksperimenti sa pomračenjem Sunca koji su bili prvi i tada ključni test Opšte teorije relativnosti, pri čemu je Edington navodno eliminisao neke podatke koji nisu potvrđivali testiranu hipotezu. (Stanley 2003)

Kada se artikulišu stavovi u vezi pomenutih problematičnih eksperimenata u psihologiji čini se da često istraživači imaju na umu određeni idealni model eksperimenta; nešto poput logičko-pozitivističkog modela eksperimenta kao pukog i epistemološki neproblematičnog sredstva za potvrđivanje hipoteza. Takođe se, očekivano, fizika percipira kao oblast nauke u kojoj se eksperimenti takve vrste realizuju, zapravo sve u skladu sa krilaticom Ernesta Raderforda da je samo fizika nauka a sve ostale oblasti su skupljanje poštanskih markica. No ispostavlja se da je i u samoj fizici situacija znatno kompleksnija nego što se to čini na osnovu ovakvog stava.

## **2. Fizika čestica kao standard: idealna reprikabilnost rezultata**

Prvi primer te kompleksnosti su eksperimenti u savremenoj fizici čestica. Ta oblast eksperimentalne fizike je neobična u poređenju sa drugim oblastima fizike, pre svega po veličini eksperimentalnih aparata i finansijskoj zahtevnosti. Laboratorije u kojima se testiraju hipoteze o fundamentalnim česticama i silama na mikrofizičkom nivou se dizajniraju i grade decenijama, a eksperimenti se u njima izvode godinama, često samo u jednoj jedinstvenoj laboratoriji te vrste.

U takvim uslovima praktično je nemoguće ponoviti eksperiment u smislu u kojem se može ponoviti u oblastima u kojima postoji niz manjih laboratorija slične namene koje izvode slične eksperimente. Ali u nekom smislu to jeste oblast koja zadovoljava taj idealni model replikabilnosti jer nakon što se konstruiše sudarač čestica nove generacije koji funkcioniše na novom graničnom nivou energije sudara, aproksimativno svakih dve do tri decenije, prva stvar koju fizičari izvedu na njemu jeste da se, kao postupak kalibracije, reprodukuju sve one čestice koje su prethodno otkrivene. To nije replikacija u doslovnom smislu, no pitanje je da li je ona moguća a i potrebna ukoliko se pod tim novim uslovima mogu reprodukovati čestice. (Dijagram 1) Ovakav postupak zadovoljava sve uslove izgleda standardne replikabilnosti rezultata.



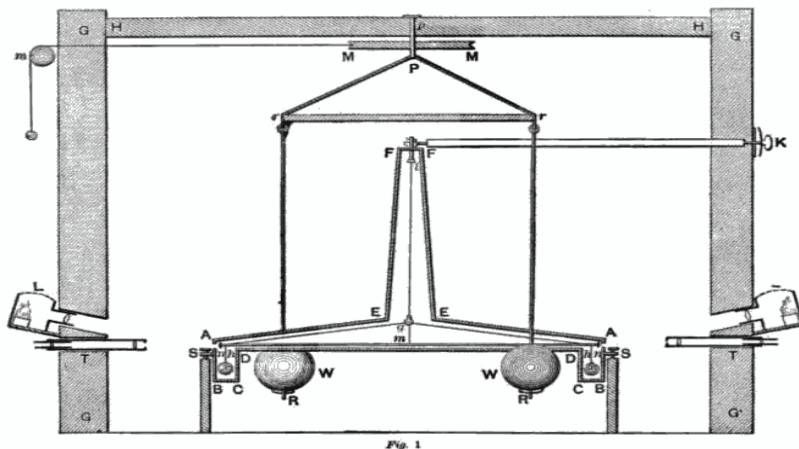
Dijagram 1: Svaki novi sudarač čestica prvo proizvede već otkrivene čestice u procesu kalibracije instrumenta. Dijagram prikazuje različite oblasti energije (svaki obojen drugom bojom – tj. u drugoj nijansi sive) u kojima su pronađene različite elementarne čestice.

### 3. Drama: merenje gravitacione konstante (G) torzijom

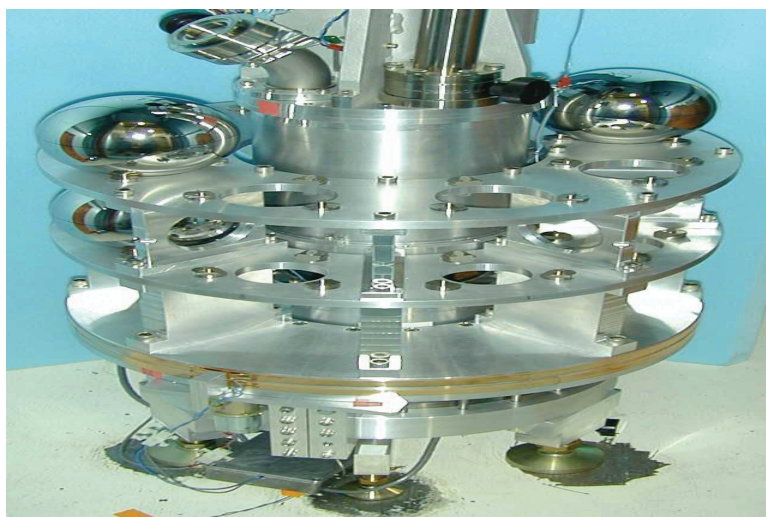
No u drugoj oblasti fizike, eksperimenti merenja gravitacione konstante  $G$  torzijom pružaju sasvim drugačiju sliku. Ovaj klasični eksperiment je osmislio još Kevendiš (Slika 1). Dve masivne olovne kugle se privlače usled dejstva gravitacione sile, što dovodi do uvrtnanja žica spojenih sa kuglama, koje je proporcionalno intenzitetu sile privlačenja. Dvadesetovekovne varijante tog eksperimenta se izvode u nekoliko laboratorija (Slika 2). (Gundlach i Merkowitz 2000, Park i Faller 2010) Eksperiment koristi isti princip ali je nešto drugačiji; usled gravitacionog dejstva između kugla umesto uvrtnanja žice dolazi do proporcionalne rotacije postolja. Ugao te rotacije je proporcionalan gravitacionoj konstanti. U proteklih dve decenije prilikom izvođenja eksperimenta u tri centra dogodila se epizoda koju su sami istraživači percipirali kao dramatičnu.

Tri tima su nezavisno izvodila ovu vrstu eksperimenta. Tim u laboratoriji u Novom Meksiku (Park i Faller 2010) je dobio vrednosti koje su značajno odstupale od vrednosti dobijenih u druge dve laboratorije.<sup>2</sup> Ovo je neslaganje percipirano kao dramatično, i zahtevalo je reviziju zvanično usvojenih rezultata koji se tiču ove osnovne konstantne na osnovu inicijalnog eksperimenta. Istraživači su prirodno bili ubeđeni da posredi mora da bude nekakva greška koja dovodi do neslaganja sa rezultatima u drugim eksperimentima i koja će se nakon detaljnije analize pokazati. Ne samo da su istraživači u laboratoriji u Novom Meksiku pokušali da utvrde koji aspekt merenja je bio neadekvatan, već su to pokušali i gostujući istraživači iz rivalskih timova ali bez uspeha.

<sup>2</sup> Vidi <https://jila.colorado.edu/news-highlights/big-g-isnt-problem-measuring-it>.



Slika 1: Kevendišov eksperimentalni aparat za merenje G putem torzije.

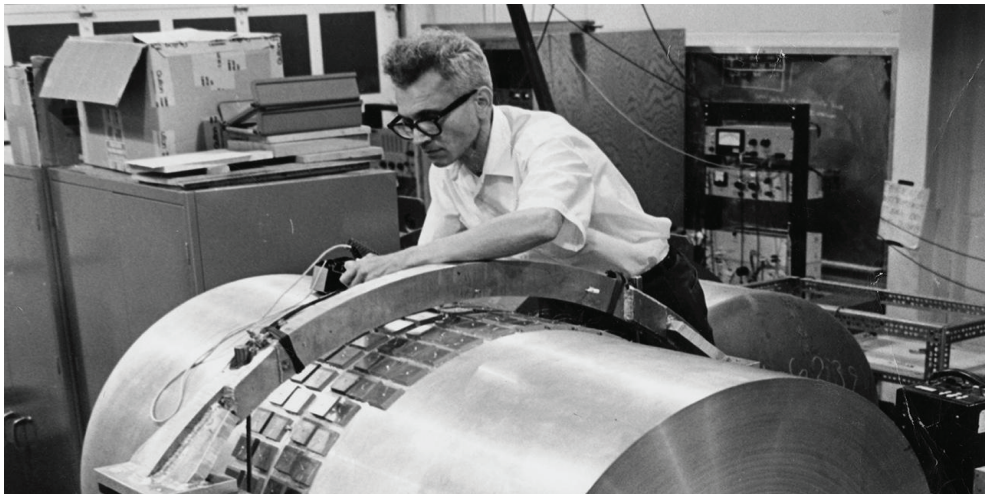


Slika 2: Savremeni eksperimentalni aparat za merenje G torzijom.

Takav je eksperiment teško proglasiti lošim iako je adekvatnost rezultata pod sumnjom usled neslaganja sa rezultatima do kojih su došla druga dva tima. Rezultate je nemoguće reprodukovati, no oni su objavljeni, jer se eksperiment nakon detaljnih analiza nije pokazao neadekvatnim. To je zapravo postao prihvatljiv, pa čak i standardan epistemološki obrazac – stanje drame, takoreći, bez razrešenja - u zajednici eksperimentatora koja meri G.

#### 4. Kontroverza: detekcija gravitacionih talasa

Sledeći primer je kontroverza prilikom pokušaja merenja gravitacionih talasa. Džozef Veber je konstruisao eksperimentalni aparat koji se stajao od velikog cilindra koji je trebalo da, prema postavljenoj hipotezi, interaguje sa gravitacionim talasima na određeni način, ukoliko takvi talasi postoje. (Slika 3) Takva vibracija bi se prenela na kristale na vrhu cilindra, tj. na takozvani piezoelektrični uređaj koji bi ih pretvorio u elektromagnetne signale i koje bi konačno registrovao računar. Kao konačni ishod bi se dobila kriva elektromagnetne aktivnosti koja je predviđena hipotezom, ili pak ne bi bilo takvih oscilacija ukoliko talasi nisu detektovani.



Slika 3: Džozef Veber u svojoj laboratoriji za detekciju gravitacionih talasa.

Šezdesetih godina prošlog veka Veber je izveo eksperimente i objavio rezultate koji su navodno potvrdili hipotezu o postojanju gravitacionih talasa. Druga dva tima su izgradila slične eksperimentalne aparate (Webber 1960) ne bi li proverili rezultate, no bez uspeha. To je dovelo do velike kontroverze. Istraživači iz druga dva tima su smatrali da je Veberov eskeperiment loš, smatrajući da u pronašli i razloge za to, iako se nisu slagali međusobno oko tih razloga tj. oko toga u čemu je eksperiment tačno pogrešio. Jedan tim je tvrdio da je u pitanju loša kalibracija Veberovog aparata, a drugi da je u pitanju neadekvatna analiza podataka i neadekvatan softver koji je korišćen za te svrhe. Otud je ova relativno neobična epizoda u istoriji fizike izučavana temeljno od strane filozofa, istoričara i sociologa nauke. (Collins 2010, Franklin 2012) Veber je izgubio svoju dobru reputaciju i nikada je nije povratio.

**5. Replikabilnost rezultata u psihologiji: drama ili kontroverza?**

U čemu se sastoji ključna razlike ili sličnosti između eksperimenata u fizici čestica, eksperimenata sa merenjem  $G$ , kao i pokušaja detekcije gravitacionih talasa?

U velikim hadronskim sudaračima se sudaraju protoni, a u leptonskim sudaračima elektroni i pozitroni na veoma visokim energijama koje se mere u mega, giga, i tera elektron voltima. U blizini takvih uređaja ne postoje nikakvi slični fizički događaji koji se odvijaju na tim energijama koji će da interferiraju sa eksperimentom, i rezultiraju lažnim signalima koji samo izgledaju kao traženi fenomen a zapravo su šum, iako su detektori osetljivi na jače elektromagnetne uticaje u blizini koji se ipak lako prepoznaju. U slučaju eksperimenata sa gravitacionom konstantom, zbor sveprisutnosti gravitacione sile privlačenja, na eksperiment značajno utiču događaji i objekti u neposrednom okruženju, poput prirodnih uzvišica blizu laboratorije, ili čak hodanja zaposlenih u zgradi zbog čega se merenja vrše noću. Detektori su izloženi konstantnom šumu gravitacionih uticaja koji se ne mogu na rutinski način odvojiti kao šum. To takođe važi i za bilo kakav eksperimentalni aparat konstruisan za potrebe detekcije gravitacionih talasa, uz još druge brojne prepreke u astrofizičkom domenu koje proizvode šum na aparatu.

Postavlja se pitanje da li je opravdano očekivanje da postoji opšti standard u vezi replikabilnosti rezultata koji važi u svim oblastima nauke, ili čak samo u različitim oblastima fizike. U kojoj meri možemo da očekujemo dati nivo replikabilnosti rezultata, i čime je takvo očekivanje opravdano? Takođe se nameće pitanje da li su eksperimenti u psihologiji i u bio-medicinskim naukama koji imaju problem sa reproducibilnošću<sup>3</sup> bliže slučaju drame – kada rezultati ne mogu da se repliciraju ali se uzrok toga ne može pronaći pa eksperimenti izgledaju metodološki adekvatno i nakon podrobne analize; ili su bliže slučaju kontroverze gde bi smo mogli bar u principu da identifikujemo metodološke probleme. Takođe se postavlja pitanje koji procenat eksperimenata spada u prvu a koji u drugu kategoriju, ili je pak neka treća vrsta problema po sredi.

Na ovakva pitanja se može naravno samo odgovoriti detaljnijim ispitivanjem konkretnih slučajeva. Međutim, pojava označena kao kriza replikabilnosti nije nužno neočekivana situacija u oblastima eksperimentalne nauke u kojima se radi sa sistemima velike kompleksnosti i osetljivosti, kada je po sredi veći broj promenljivih faktora koji se ipak ne mogu lako kontrolisati pa čak ni čija se kauzalna relevantnost za posmatrane fenomene ne može lako proceniti. To se odnosi i na same instrumente kao i na subjekte istraživanja. Kao provizorni zaključak možemo da kažemo da ne bi bilo previše veliko iznenađenje, uzevši sve ovo u obzir, da kada se u ovakvim oblastima pojavi eksperiment koji se može sa lakoćom ponoviti u drugim laboratorijama to jeste

3 Postavlja se pitanje metodološke ispravnosti i ponovljenih eksperimenata na osnovu kojih se tvrdi da postoji kriza reproducibilnosti.

zapravo značajno otkriće samo po sebi, a da su eksperimenti koji se ne mogu lako ponoviti standardan slučaj pre nego li kontroverza ili primer metodološki sumnjive eksperimentalne nauke.

Slobodan Perović  
Filozofski fakultet, Univerzitet u Beogradu

## Literatura

- Collins, H. (2010). *Gravity's shadow: the search for gravitational waves*. University of Chicago Press.
- Franklin, A., Edwards, A. W. F., Fairbanks, D. J., & Hartl, D. L. (Eds.). (2008). *Ending the Mendel-Fisher Controversy*. University of Pittsburgh Press.
- Franklin, A. (2012). *Can that be right?: Essays on experiment, evidence, and science* (Vol. 199). Springer Science & Business Media.
- Gundlach, J. H., & Merkowitz, S. M. (2000). Measurement of Newton's constant using a torsion balance with angular acceleration feedback. *Physical Review Letters*, 85(14), 2869.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean?. *American Psychologist*, 70(6), 487.
- Parks, H. V., & Faller, J. E. (2010). Simple pendulum determination of the gravitational constant. *Physical Review Letters*, 105(11), 110801.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence?. *Perspectives on Psychological Science*, 7(6), 528-530.
- Stanley, M. (2003). "An Expedition to Heal the Wounds of War" The 1919 Eclipse and Eddington as Quaker Adventurer. *Isis*, 94(1), 57-89.
- Weber, J. (1960). Detection and generation of gravitational waves. *Physical Review*, 117(1), 306.

Slobodan Perović

## **Crisis of Replicability of Results in Contemporary Experimental Psychology in the Light of the Criteria of Replicability in Contemporary Physics** (Summary)

Since 2010, the impossibility of replicating experimental results (published in leading journals in the field) in a number of repeated psychological experiments began to be increasingly perceived as a methodological crisis. Replicability is generally considered to be the basic standard of scientific experiments, modeled on exemplary classical experiments in the field of physics. But the conditions of replicability vary

depending on the studied phenomenon in physics itself, and therefore the criteria for the adequacy of the experiment vary as well. Three characteristic examples of experiments we analyze led to three different epistemic attitudes of their adequacy based on the replicability of the results. We argue that these attitudes can be relevant in understanding the nature of the crisis of replicability in psychology.

KEYWORDS: philosophy of science, psychology, physics, philosophy of scientific experiments, replicability of experiments