

Profesor Stapel na dopingu. O upiększaniu psychologii społecznej

Jarosław Klebaniuk

Instytut Psychologii, Uniwersytet Wrocławski

W artykule pokrótce przedstawiono sprawę skandalu wokół holenderskiego psychologa Diederika A. Stapela, który systematycznie fałszował wyniki badań, a jego oszustwa przez wiele lat nie zostały wykryte. Wskazano sposoby, których używał, a które pozwoliły mu stać się jednym z najpłodniejszych i najbardziej uznanych psychologów społecznych. Przypomniano niektóre ustalenia i ogólne zalecenia raportu trzech komisji uniwersyteckich dla praktyki badawczej. Przedstawiono dane, świadczące o tym, że postępowanie Stapela nie było czymś wyjątkowym, a większość psychologów przyznaje się do manipulowania danymi, aby zwiększyć swoje szanse na publikację wyników badań. W dyskusji zaproponowano, by tak zmodyfikować reguły publikacyjne, aby nie wywierały presji, niekiedy prowadzącej do nieprawidłowości. Zwrócono też uwagę na potrzebę uznawania replikacji badań przeprowadzonych przez innych naukowców za wartościowe z punktu widzenia rozwoju dyscypliny.

Słowa kluczowe: skandal, oszustwo, Stapel, raport, publikacja naukowa, replikacja badań

Diederik Alexander Stapel był uważany za wzorcowy przykład dobrego naukowca i nauczyciela akademickiego: pracowity, twórczy, aktywny, przychylny studentom i współpracownikom, dużo publikujący – budził podziw i uznanie. Od 2010 roku sprawował urząd dziekana Wydziału Nauk Społecznych i Behawioralnych, a rok wcześniej otrzymał od Towarzystwa Eksperymentalnej Psychologii Społecznej (*The Society of Experimental Social Psychology*) prestiżową *Career Trajectory Award*, która przyznawana jest dorocznie tylko jednemu naukowcowi; wcześniej otrzymał ją między innymi Neil Macrae, a później – John Jost. Diederik Stapel znalazł się więc w gronie najślawniejszych, najbardziej rozpoznawalnych, najwybitniejszych psychologów społecznych – był uznawany za członka ścisłej elity uprawianej dyscypliny naukowej.

Pod koniec sierpnia 2011 roku trzech młodzi badacze zgłosili szefowi instytutu psychologii jednej z holenderskich uczelni, że podejrzewają dziekana o fałszowanie danych. O sprawie błyskawicznie został poinformowany rektor i podjęto adekwatne do wagi sprawy działania. 31

października w sieci ukazał się wstępny raport (*Interim...*, 2011) ze śledztwa trzech komitetów, z których jeden został we wrześniu 2011 roku powołany przez rektora Uniwersytetu w Tilburgu (gdzie profesor Stapel pracował w latach 2007–2011), a dwa pozostałe na uniwersytetach: w Groningen (2000–2006) i w Amsterdamie (1994–1999). Z raportu wynika, że ponad wszelką wątpliwość Diederik Stapel dopuszczał się na przestrzeni wielu lat fałszowania danych, a na podstawie tych nieprawdziwych wyników ukazało się co najmniej trzydzieści recenzowanych publikacji.

Metody, jakimi posługiwał się oszust, były na tyle skuteczne, że długo udawało mu się działać bez przeszkód. Postępował na trzy sposoby: albo kompletnie zmyślał dane, albo fabrykował takie powiększenie ich bazy, żeby uzyskać potwierdzenie hipotezy, albo modyfikował dane – kompletne lub niepełne. Jego działania zazwyczaj rozpoczynały się od bardzo starannego i konsultowanego ze współpracownikami przygotowania badań. Dyskutowano więc nad pomysłem, hipotezami, sposobami operacjonalizacji zmiennych i rodzajem manipulacji eksperymentalnej. Przygotowywano wszelkie materiały, w tym kwestionariusze, słodycze dla uczestników i tym podobne namacalne przejawy prowadzenia badań. Wiele z nich było możliwych do sfinansowania dzięki grantom, które „wybitny”

Jarosław Klebaniuk, Instytut Psychologii, Uniwersytet Wrocławski, ul. Dawida 1, 50-527 Wrocław,
e-mail: j.klebaniuk@psychologia.uni.wroc.pl

naukowiec bez trudu otrzymywał. Materiały zazwyczaj znajdowały się w bagażniku samochodu Stapela i tam były widziane po raz ostatni. To, co następowało później, nie miało już nic wspólnego z nauką uczciwością.

Eksperymenty inkryminowany naukowiec przeprowadzał zawsze sam lub z pomocą nieokreślonych asystentów, podobnie sam wprowadzał dane i analizował je. Na pytania, w której szkole zrobił badanie, odpowiadał ogólnikowo i wykrętnie. Danymi w postaci surowej dzielił się niechętnie, sam przeprowadzając obliczenia. Jedynym wyjątkiem były te sytuacje, w których wręczał komuś nieprzeanalizowane jeszcze – rzekomo z braku czasu – dane z badań przeprowadzonych przed laty, proponując wspólne napisanie artykułu. Niekiedy nawet pozorował realizowanie bieżących pomysłów badawczych innych naukowców, zresztą ze znakomitym skutkiem. Stapelowi zawsze wychodziły badania – tak to przynajmniej wyglądało na zewnątrz.

Gdyby nie determinacja trzech młodych demaskatorów – najprawdopodobniej powodowanych zdrowym odruchem moralnym, nie zawiścią – to proceder jeszcze by trwał. Wcześniej trzech innych badaczy próbowało zwrócić uwagę przełożonych, lecz bezskutecznie. Także spostrzeżenia, nagłaśniane przez profesorów, że „dane są zbyt dobre, aby były prawdziwe”, były lekceważone. Czyżby jedynie zbieg okoliczności sprawił, że odważna postawa badaczy, którzy ryzykowali swoją rozpoczynającą się dopiero karierą naukową, doprowadziła do ujawnienia oszustwa? Czy psychologia naukowa – przynajmniej ta holenderska, choć nic nie wskazuje na to, aby w innych krajach sytuacja znacząco się różniła – nie posługuje się mechanizmami, które pozwalałyby wcześniej wykrywać, a najlepiej zapobiegać podobnym niemoralnym zachowaniom? Wreszcie – jakie są przyczyny, dla których ktoś może uznać zmyślanie i deformowanie danych za dopuszczalny, a nawet wartościowy instrument pracy naukowej?

Częściowe odpowiedzi na te pytania znajdujemy w „Raporcie dotyczącym złamania uczciwości naukowej popełnionego przez prof. D. A. Stapela”. Mowa w nim o „porażce racjonalnego, systematycznego, rzeczowego, metodologicznego i publicznego krytycyzmu, który stanowi kamień węgielny nauki” (*Interim...*, 2011, s. 11). Przyczynami tej porażki były wyłącznie bilateralne relacje między profesorem a doktorantami, brak pracy zespołowej i drugiego superwizora (nie jest to także w Polsce uznaną praktyką – z uwagi na duże obciążenie dydaktyczne kadry). Doktoranci nie prowadzili samodzielnie eksperymentów, korzystając z „uprzejmości” promotora, dostarczającego im gotowe dane. Pytania o szczegóły proceduralne pojawiały się zbyt rzadko, a to dlatego, że były zbywane i spotykały się z niechętną reakcją.

Prawdopodobnie jednak największą porażką była ta, którą ponieśli recenzenci i redaktorzy, nie prześwietlając danych, choć te ewidentnie były „zbyt dobre, aby były prawdziwe”. Nie dość że prawie zawsze potwierdzały hipotezy, to siła efektów była nieprawdopodobnie duża, natomiast dane brakujące, nietypowe i poza zakresem (*outliery*) prawie nie występowały. Przypisywano to niespotykane wysokim kwalifikacjom badawczym Stapela. Nie wychwycono nawet dziwnych, nieprawdopodobnych zależności, niemożliwych (na przykład powtarzania się identycznych, co wynikało z mechanicznego przeklepania) wzorów danych, dziwnych korelacji czy zaskakującego ich braku, takich samych średnich i odchyień standardowych, niemożliwych rozkładów zmiennych. Demaskatorzy ostatecznie dostrzegli te nieprawidłowości, lecz praktyka nieujawniania danych sprawiała, że trudno było o nieodparte dowody. Nic dziwnego, że wśród ogólnych rekomendacji raportu znalazła się i ta, aby zawsze podawać szczegółowe informacje o sposobie i miejscu zebrania danych, a bazę je zawierającą, jak również kwestionariusze i inne materiały użyte w badaniach przechowywać przez minimum pięć lat i udostępniać na żądanie.

Analizując informacje przedstawione w raporcie trzech komisji, łatwo powziąć przypuszczenie, że w wykryciu oszustwa we własnym środowisku paradoksalnie przeszkadza mogłaby uczciwość naukowców. Ktoś, komu do głowy by nie przyszło, że można po prostu sfałszować dane, zapewne będzie miał kłopoty ze spontanicznym wyobrażeniem sobie, iż ktoś inny się tego dopuszcza. (Inna sprawa, że ta uczciwość, jak się przekonamy, też bywa ograniczona). Dopiero bliskie kontakty zawodowe, nietypowość pewnych sposobów postępowania i rezultatów badań mogły wzbudzić podejrzenia. Zapewne też budziły znacznie wcześniej, skoro niektóre zostały nawet wypowiedziane (lecz je zlekceważono). Jednak niełatwo jest podjąć zdecydowane działania przeciw komuś, kto cieszy się autorytetem wynikającym nie tylko z relacji nauczyciel–student czy promotor–doktorant, lecz również z międzynarodowego uznania i liczby publikacji w prestiżowych czasopismach. Moim zdaniem nie bez znaczenia było również miejsce Stapela w strukturze uniwersyteckiej – samodzielne stanowisko i funkcja dziekana to wehikuły sprawowania władzy, także w nieformalny sposób. Wprawdzie na mniejszą skalę niż w polityce czy biznesie, jednak i w środowisku uczelnianym władza może korumpować. Notabene sam bohater skandalu publikował artykuły dotyczące władzy jako przyczyny hipokryzji (np. Lammers, Stapel i Galinsky, 2010).

Konsekwencje zbrodni Stapela przeciwko prawdzie są niezmiernie, lecz nawet najprostszy sposób ich ilościowego ogarnięcia budzi zgrogę. Na platformie EBSCO,

mieszczącej między innymi publikacje i abstrakty psychologiczne z baz PsychArticles, PsychInfo i PsychBooks, wpisanie frazy „Diederik Stapel” 12 listopada 2011 dało 152 rezultaty, a ograniczenie wyszukiwania do pełnych tekstów recenzowanych naukowo wciąż pozostawiło 70 pozycji, w większości w prestiżowych czasopismach. To imponujący, niemal niespotykany w psychologii dorobek. W większości z nich Stapel jest współautorem. Teksty te, a w każdym razie większość z nich, raczej nie zostaną usunięte z bazy, choć Kim I. Mills – rzeczniczka Amerykańskiego Towarzystwa Psychologicznego (*The American Psychological Association*), największego wydawcy naukowych tekstów psychologicznych – zapowiedziała: „Posuniemy się do usunięcia nieważnych badań, gdy takie artykuły zostaną w jasny sposób wykryte” (za: Carey, 2011).

Najprawdopodobniej nie wszystkie artykuły zostały napisane na podstawie sfałszowanych danych, a i skala deformacji nie zawsze była taka sama. Zaufanie do wartości naukowej tych tekstów na zawsze jednak zostało podważone. Ofiarą oszustwa padli więc nie tylko współpracownicy i doktoranci (np. obrona dysertacji doktorantki M. H. C. Meijers planowana na 7 października została przełożona; konieczne będzie uzupełnienie pracy w tej części, w której korzystała z zafałszowanych przez Stapela danych). Również my, psychologowie społeczni, musimy liczyć się z tym, że za każdym razem, gdy będziemy cytować artykuł z renomowanego czasopisma, przypomnimy sobie tę niesławną sprawę i zadamy niepokojące pytanie: czy to aby prawda? I nie będzie to niestety pytanie paranoiczne.

Jak donosi Benedict Carey (2011), Leslie John z Harvard Business School wraz z dwoma współpracownikami przeprowadziła wśród ponad dwóch tysięcy amerykańskich psychologów anonimowy sondaż (ma zostać opublikowany jeszcze w tym roku), w którym 70% respondentów przyznało się do wygładzania wyników, około jedna trzecia – do publikowania nieoczekiwanych rezultatów jako zgodnych z hipotezami (poprzez modyfikację hipotez po przeprowadzeniu badań), a 1% – do fałszowania danych. Są to alarmujące dane, a jeśli dodamy do tego wyniki analiz Jelte Wichertsa i Marjana Bakker z Uniwersytetu w Amsterdamie, to skaza na obrazie najbardziej zaawansowanego metodologicznie działu psychologii wydaje się jeszcze większa. Otóż znaleźli oni w około połowie z 281 artykułów losowo wybranych z prestiżowych czasopism jakiś błąd statystyczny, 15% zaś zawierało tak poważne uchybienie, że wnioski powinny być inne, zazwyczaj w kierunku odwrotnym do hipotez (za: Carey, 2011).

Ponadto dwaj wspomniani autorzy wraz z Dylanem Molenaarem zwrócili się do autorów 141 artykułów,

opublikowanych w drugiej połowie 2004 roku w czterech wysoko punktowanych naukowych czasopismach psychologicznych, z prośbą o przesłanie danych surowych, które były podstawą analiz. Ponieważ większość autorów nie odpowiedziała na tę prośbę, dalsze badanie oparli na 1148 statystycznie istotnych wynikach w 49 tekstach z dwóch czasopism: *Journal of Personality and Social Psychology* i *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Podstawowy wniosek brzmiał następująco: niechęć do dzielenia się wynikami wiąże się ze słabością dowodów i częstszym występowaniem błędów w analizie statystycznej. Brak skłonności do poddawania danych ponownej analizie przez autorów artykułu był szczególnie nasilony, gdy groziło to zakwestionowaniem statystycznej istotności wyników (Wicherts, Bakker i Molenaar, 2009). Intersubiektywna komunikowalność i intersubiektywna sprawdzalność rezultatu badawczego, a więc to, co nazywane jest „słabą zasadą racjonalności”, w praktyce napotykają więc opór części psychologów społecznych. Także to, co Kazimierz Ajdukiewicz nazwał mocną zasadą racjonalności (por. Brzeziński, 2000) – racjonalne uznawanie przekonań, niewyciąganie wniosków wykraczających poza niezawodność samego sposobu wnioskowania, warto by przypominać autorom artykułów, w których siły efektu są marginalne, ich istotność zależy od wybranego testu statystycznego, a wielkość wyjaśnianej wariancji nie przekracza paru procent.

Profesor Jonathan Schooler z Uniwersytetu Kalifornijskiego w Santa Barbara tak skomentował rewelacje dotyczące Stapela: „Duży problem stanowi kultura, w której naukowcy podkreślają swoją pracę w taki sposób, że opowiada historię ładniejszą, niż to, co rzeczywiście odkryli. To prawie tak, jakby wszyscy byli na steroidach, a żeby rywalizować trzeba również brać sterydy”. Nic dziwnego, że w związku z tym słusznie postuluje: „Mamy technologię [niezbędną do tego], by dzielić się danymi i publikować nasze pierwotne hipotezy i teraz jest na to czas. To oczyściłoby postępowanie w dyscyplinie w zdecydowany sposób” (za: Carey, 2011).

Pozostaje rzecz najważniejsza – motywy postępowania Diederika Stapela. On sam nie stawiał się wprawdzie przed żadną komisją, zasłaniając się złą kondycją, wcześniej zaś, poproszony o listę publikacji rozdziałów w książkach i materiałów pokonferencyjnych, stwierdził, że „fizycznie i emocjonalnie nie jest w stanie odpowiedzieć”. (Ten zaledwie 45-latek mógł się zresztą autentycznie rozchorować, gdy cały jego misternie budowany świat sukcesu i chwały legł w gruzach praktycznie z dnia na dzień). W oświadczeniu napisał: „Nie potrafiłem się oprzeć presji ciągłego publikowania i osiągnięcia coraz lepszych wyników.

Chciałem zbyt dużo i zbyt szybko. Poszedłem złą drogą w systemie, w którym jest mało kontroli, a wielu badaczy pracuje indywidualnie” (*Skandal...*, 2011). Wcześniej, ustosunkowując się do raportu, napisał: „Przeczytałem raport z poczuciem konsternacji i wstydu (...). Muszę podkreślić, że błędy, które popełniłem, nie były motywowane własnym interesem. Nie utożsamiam się z zarysowanym obrazem człowieka, który próbował wykorzystać młodych badaczy dla własnej korzyści. Popełniłem błędy, ale byłem i pozostaję szczerze oddany dziedzinie psychologii społecznej, młodym badaczom i innym kolegom. Z tego powodu żałuję cierpienia, którego przysporzyłem innym” (*Interim...*, 2011, s. 21). Słowa te, jeśli są szczerze, a pewności co do tego nie mamy, wskazują, że holenderski psycholog nie czuje się aż tak bardzo winny, skoro uważa się za oddanego dyscyplinie, którą tak opacznie uprawiał. Sądzi być może, nie do końca zresztą bezpodstawnie, że robił dla niej to, czego od niego oczekiwano.

Sprawa Stapela pokazuje, że praktyka uprawiania psychologii społecznej przez sporą część naukowców wymaga zmian. Większa przejrzystość badań, czyli bezwarunkowe udostępnianie danych surowych, a także informacji o szczegółach proceduralnych (te o miejscu prowadzonych badań, o grupie osób badanych, muszą być znane co najmniej współpracownikom i recenzentom, bo z uwagi na rodzaj upublicznianych informacji nie zawsze mogą być powszechnie ogłaszane), powinny być standardem. Wskazane byłoby jednak także łaskawsze traktowanie przez redakcje czasopism naukowych i przez recenzentów tych wyników badań, które zupełnie nie potwierdziły hipotez, czy nawet tych, w których manipulacja eksperymentalna nie w pełni zadziałała. Tymczasem od naukowców wymaga się publikacji w czasopismach, i to najlepiej prestiżowych, o wysokich indeksach cytowań. Obowiązuje zasada *publish or perish* (publikuj albo znikaj). W nauce mają zatem szansę utrzymać się (i słusznie!) ci, którzy badają bardzo dużo i część (najlepiej lwia) z tych badań im wychodzi, lecz także ci, którzy po prostu dostarczają publikowane w czasopismach rezultaty, jak to przez wiele lat robił okryty niesławą główny bohater tego tekstu. Wyścig szczurów w psychologii społecznej wywołuje pokusy, aby – niczym niedozwolonego dopingu – zażywać sfabrykowanej prawdy. Łatwiej jest przecież samodzielnie „wydrukować” jeden wynik zgodny z hipotezami (Stapel był w tym mistrzem), niż uczciwie przeprowadzić liczne badania, z których jakieś od czasu do czasu spełnią oczekiwania bardzo wymagających redaktorów i recenzentów. Wygrywa szczur skuteczniejszy, niekoniecznie uczciwszy.

Jako czynny psycholog społeczny jestem przekonany, że co roku tysiące tak zwanych nieudanych badań

łądzą w szufladach bez żadnej szansy na publikację (sam oczywiście mam takie we własnym biurku – mnóstwo pracy, żadnych wymiernych efektów). To nie tylko zafałszowuje obraz nauki – eksponowane są głównie czyste, eleganckie wyniki, wskazujące na klarowne zależności – lecz także motywuje różnych Stapelów do pomagania wynikom w taki sposób, aby nadawały się do publikacji. W psychologii można dzięki temu uogólniać wyniki, a nawet układać je w jakieś prawa, bo publikowane są głównie te potwierdzające lub częściowo potwierdzające hipotezy. Pozostałe badania są uznawane za „nieudane”. W pewnym sensie nauka zamienia się w ten sposób w grę pozorów. Szuflady pełne nieopublikowanych badań milczą. Zależności („efekty”) z wysoko punktowanych czasopism stanowią makijaż upiększający twarz naukowej prawdy. A może to już nie makijaż, lecz maska?

Inną sprawą domagającą się zmian są replikacje opublikowanych badań. Wprawdzie systematyczne replikacje i autoreplikacje są dobrze znanym postulatem metodologicznym (Wojciszke, 2002), praktyka publikacyjna jednak zdecydowanie im nie sprzyja, zwłaszcza jeśli odnoszą się do czyichś badań. Nie są one cenione przez redakcje czasopism, gdyż jednym z kryteriów kwalifikacji do druku jest to, na ile prezentowane wyniki są nowe, oryginalne i przyczyniają się do rozwoju dyscypliny. Siłą rzeczy replikacje wobec tak postawionych kryteriów są mniej wartościowe, niż badania realizowane po raz pierwszy. Utrudnia to weryfikowanie wcześniejszych (m.in. Stapelowskich) rezultatów, choć jest to przecież jeden z podstawowych postulatów metody naukowej. Nie widzę prostego rozwiązania tego problemu, jednak postęp rozumiany jako weryfikacja nowych hipotez nie może być jedynym celem nauki. Także dotychczasowe efekty wymagają dalszego drążenia, zwłaszcza gdy, co jest raczej normą niż wyjątkiem, mają słaby – w sensie siły związku lub proporcji wariacji wyjaśnionej – charakter. Droga do ustaleń w psychologii nie jest prosta i choć trudno się przebić z nadmiernie zagmatwanym, niezbyt wyraźnym ich obrazem, to może warto czasem poprzestać na małym. Diederik Stapel „odkrywał” zależności klarowne i liczne. Teraz wiele z nich trzeba będzie podać w wątpliwość lub wręcz odwołać. Może warto dać szansę mniej udanym badaniom. Prawda w psychologii społecznej nie musi być piękna.

LITERATURA CYTOWANA

- Brzeziński, J. (2000). Poznanie naukowe – poznanie psychologiczne. W: J. Strelau (red.), *Psychologia. Podręcznik akademicki*, (t. 1, s. 333–354). Gdańsk: Gdańskie Wydawnictwo Psychologiczne.

- Carey, B. (2011). *Fraud case seen as a red flag for psychology research*, http://www.nytimes.com/2011/11/03/health/research/noted-dutch-psychologist-stapel-accused-of-research-fraud.html?_r=1&scp=1&sq=Leslie%20John&st=cse, dostęp 12.11.2011.
- Interim report regarding the breach of scientific integrity committed by prof. D. A. Stapel*, Tilbur, 31 October 2011, <http://www.tilburguniversity.edu/nl/nieuws-en-agenda/commissie-levelt/interim-report.pdf>, dostęp 12.11.2011.
- Lammers, J., Stapel, D. A., Galinsky, A. D. (2010). Power increases hypocrisy: Moralizing in reasoning, immorality in behavior. *Psychological Science*, 21(5), 737–744.
- Skandal w psychologii: znany holenderski uczony fałszował badania*, serwis „Nauka w Polsce”, 6 listopada 2011, http://www.naukawpolsce.pap.pl/palio/html.run?_Instance=cms_naukapl.pap.pl&_PageID=1&s=szablon.depesza&dep=385861, dostęp 12.11.2011.
- Wicherts J. M., Bakker, M., Molenaar, D. (2009). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE Journal*, <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0026828>, dostęp 12.11.2011.
- Wojciszke, B. (2002). *Człowiek wśród ludzi. Zarys psychologii społecznej*. Warszawa: Wydawnictwo Naukowe Scholar.

Professor Stapel on dope. On the beautifying of social psychology

Jarosław Klebaniuk

Institute of Psychology, University of Wrocław

Abstract

In the article discussed is the scandal concerning a Dutch psychologist Diederik A. Stapel, who systematically committed fraud with research data, and it remained uncovered for many years. Described are ways that he used and which let him become one of the most prolific and recognized social psychologists. Described were some findings and general recommendations for research practice, offered in report prepared by three university committees. The article presents data that suggest that Stapel's conduct was not exceptional, and that most psychologists admit fiddling with their research data in order to improve their publication chances. In the discussion, proposed is relaxation of the existing publication rules so that they do not put the excess pressure that may lead to irregularities. It is emphasized that replications of findings collected by other researchers is the valuable means of control and development of the discipline.

Keywords: scandal, fraud, Stapel, report, publication, replication