

Kryzys psychologii, psychologia kryzysu

Mateusz Olechowski

Wydział Psychologii, Uniwersytet Warszawski

W poniższym komentarzu odnoszę się do wybranych części diagnozy i zaleceń zawartych w tekście Jarosława Klebaniuka „Profesor Stapel na doping. O upiększaniu psychologii społecznej”. Rozwijam wątki zasługujące według mnie na dokładniejszą analizę („podejrzane praktyki badawcze”, „stopnie swobody badacza”) oraz przedstawiam elementy diagnozy i zalecenia nieobecne w tekście oryginalnym (moc statystyczna, krytyczne myślenie statystyczne, sposób pisania artykułów, krótkie doniesienia z badań), które wydają się niezbędne dla zrozumienia obecnej kondycji psychologii akademickiej i jej uzdrowienia. Wzywam do aktywnego stawienia czoła trudnemu wyzwaniu, jakim jest obecny kryzys psychologii (społecznej) i podjęcie zdecydowanych kroków w celu jego zażegnania.

Słowa kluczowe: kryzys psychologii, stopnie swobody badacza, podejrzane praktyki badawcze, moc statystyczna, krytyczne myślenie statystyczne, krótkie doniesienia z badań

Z wielką satysfakcją powitałem tekst Jarosława Klebaniuka „Profesor Stapel na doping. O upiększaniu psychologii społecznej”. Ten krytyczny głos spełnia ogromnie ważny postulat Allana Bussa, który we wstępie do swojego studium o „socjologii wiedzy psychologicznej” stwierdził, że „byłoby idealnie, gdyby każdy psycholog zarezerwował część swojej «energii psychicznej» na zadawanie «znaczących pytań» dotyczących przestrzeni swoich kompetencji i zainteresowań naukowych z perspektywy psychologii jako całości, zakotwiczonej w społeczeństwie” (Buss, 1975, s. 988). Zgodnie z tym podejściem, „autorefleksja, autoanaliza i autoocena” działań społeczności naukowców oraz wynikająca z nich samoświadomość są kluczowe dla jej rozwoju (Buss, 1975, s. 988).

Na najwyższe uznanie zasługuje fakt, że mimo iż Jarosław Klebaniuk wychodzi od historii skandalu z udziałem Diederika Stapela, nie wpada w sidła łatwej narracji o „zgniłym jabłku”. W takiej opowieści osoba, która przewiniła, jest przedstawiana jako najohydniejszy bękart danej społeczności, który zdradziecko zatruł jej zdrową tkankę. „Zgniłe jabłko” jest odsądzane od czci i wiary przez innych członków społeczności, za czym stoi m.in. chęć oczyszczenia siebie z jakichkolwiek powiązań

z popełnionym złem i zachowanie nieskazitelnej czystości moralnej – zarówno przed sobą, jak i przed innymi.

Klebaniuk robi coś zupełnie innego. Po pierwsze, stara się wejść w skórę bohatera skandalu, aby zrozumieć pobudki jego działania. Po drugie, analizuje potężne zewnętrzne siły oddziałujące na wszystkich psychologów akademickich, które w ekstremalnych przypadkach katalizują formację takich badaczy, jak Stapel. Po trzecie, zwraca uwagę na powszechność różnego rodzaju naciągania danych pod publikację i na wynikające z tego niebezpieczeństwa. Po czwarte – i najważniejsze – postuluje wprowadzenie ważnych i potrzebnych zmian w sposobach pracy psychologów i publikowania, zachowaniach recenzentów oraz polityce czasopism naukowych. W poniższym komentarzu skupiam się raczej na „upiększaniu psychologii” (Klebaniuk, 2012, s. 213), na sposoby dopuszczalne w ramach obecnej szarej strefy etyczno-metodologicznej niż na te jednoznacznie potępiane, takie jak fałszerstwo danych. W tym kontekście niedawny skandal związany z opublikowaniem przez *Journal of Personality and Social Psychology* artykułu Daryla Bema (2011; Carey, 2011a) z dziewięcioma badaniami wskazującymi na istnienie zjawiska prekognicji jest równie ważną inspiracją dla mojego tekstu, co jeszcze świeższa sprawa Stapela.

Jak wiadomo, sformułowanie trafnych postulatów musi poprzedzić rzetelna diagnoza. Z tego względu wydaje się rzeczą niezbędną, aby bliżej przyjrzeć się wynikom bezprecedensowego badania John, Loewensteina i Preleca

Mateusz Olechowski, Wydział Psychologii, Uniwersytet Warszawski, ul. Stawki 5/7, 00-183 Warszawa,
e-mail: mateusz.olechowski@psych.uw.edu.pl

(2012) na temat „podejrzanych praktyk badawczych” lub PPB (*questionable research practices*; QRPs) wspomnianego przez Klebaniuka. PPB to takie zachowania badaczy podczas zbierania i analizowania danych, które trudno jednoznacznie zakwalifikować jako nieetyczne, ale które są kontrowersyjne (i kwestionowane) w sensie metodologicznym (np. nieraportowanie zmiennych, które „nie zadziałały”) i często służą do zwiększenia szansy i „jakości” publikacji. Autorzy słusznie zauważają, że „choć to przypadki oszustw w nauce są ostatnio w świetle Jupiterów, to działania w szarej strefie akceptowalnych praktyk są zdecydowanie bardziej rozpowszechnione i mogą bardziej zaszkodzić nauce w perspektywie długoterminowej” (John i in., 2012, s. 524). W pracy o wiele mówiącym tytule „Fałszywie istotna psychologia. Nieujawniana elastyczność podczas zbierania i analizowania danych umożliwia wykazanie istotności czegokolwiek” Simmons, Nelson i Simonsohn (2011) posłużyli się symulacjami, aby empirycznie oszacować, jak szkodliwe są cztery typy takich arbitralnych decyzji, nazwanych przez nich „stopniami swobody badacza” lub SSB (*researcher degrees of freedom*). Z ich analiz wynika, że przy teście różnic między dwoma niezależnymi próbami po 20 obserwacji każda, prawdopodobieństwo otrzymania fałszywie istotnego wyniku przy $p < 0,05$ wynosi 1) 9,5%, jeśli mierzymy dwie zmienne zależne skorelowane na poziomie $r = 0,5$; 2) 7,7%, jeśli zbieramy dodatkowe 10 obserwacji na celkę, bo wynik nie był istotny statystycznie przy 20 obserwacjach; 3) 11,7%, kontrolując płć i interakcję manipulacji z płcią; 4) 12,6%, jeśli wybieramy czy raportujemy wszystkie warunki badawcze, czy tylko niektóre (dwa dowolne warunki lub wszystkie trzy) oraz 5) 61% (!), jeśli zastosujemy wszystkie powyższe SSB. Inna seria symulacji, prowokacyjnie zatytułowana „Dlaczego wyniki większości badań są fałszywe”, zwraca uwagę na podobne niebezpieczeństwa (Ioannidis, 2005). Lektura tych prac uzmysławia, jak łatwo zwiększa się tak hołubiony przez nas (ale być może rzadki w praktyce) pięcioprocentowy margines błędu.

Co dokładnie pokazuje badanie John i współpracowników? Wśród 2155 amerykańskich psychologów akademickich¹ (stopa zwrotu ankiety 36%) 67% przynajmniej raz nie zaraportowało w artykule wszystkich zmiennych zależnych użytych w badaniu; 58% zdecydowało, czy zbierać więcej danych dopiero po sprawdzeniu, czy wyniki są istotne statystycznie; 27% nie zaraportowało w artykule wszystkich warunków eksperymentalnych użytych w badaniu; 23% skończyło zbierać dane wcześniej niż planowało ze względu na uzyskanie poszukiwanego wyniku; 23% „zaokrągliło” w artykule poziom istotności (np. podało $p < 0,05$, gdy w rzeczywistości $p =$

0,054); 50% zaraportowało w artykule tylko badania, które „wyszły”; 43% zdecydowało, czy usunąć obserwacje dopiero po sprawdzeniu wpływu tego kroku na wyniki; 35% zaraportowało w artykule niespodziewany wynik jako uprzednio sformułowaną hipotezę badawczą; 4,5% stwierdziło w artykule, że zmienne demograficzne nie zmieniają wyników, mimo że nie było tego pewnymi lub wiedziało, że zmieniają; 1,7% sfalszowało wyniki badań. Wydaje się, że można z pełną odpowiedzialnością stwierdzić za Klebaniukiem, iż są to wyniki alarmujące. Autorzy sprawdzili także, czy powyższe częstości różnią się między poszczególnymi dziedzinami psychologii. Niestety, wynik nie jest budujący dla psychologii społecznej, która razem z psychologią poznawczą i neuronauką wiedzie niechlubny prym wśród dziewięciu dziedzin. Niejasny status etyczny PPB podkreśla zaś fakt, że aż około 35% respondentów przynajmniej raz miało wątpliwości co do etyczności swojego postępowania (John i in., 2012).

Tak jak Jonathan Schooler, cytowany przez Klebaniuka (za: Carey, 2011b), także John i współpracownicy porównują PPB do „sterydów”, zwracając jednak uwagę także na inny, być może istotniejszy aspekt tych praktyk, znaczący zwłaszcza z psychologicznego punktu widzenia – według nich PPB dają „znaczną przestrzeń na racjonalizację i samooszukiwanie się” badaczy (John i in., 2012, s. 524). Z bogatego dorobku psychologii społecznej wiemy dobrze, że ludzkie poznanie i sądy są pod silnym wpływem motywacji (por. Ditto, Pizarro i Tannenbaum, 2009; Jost, Glaser, Kruglanski i Sulloway, 2003; Kunda, 1990) i że wszyscy podlegamy „iluzji obiektywności” (Pyszczynski i Greenberg, 1987). Z badań wiadomo także, że działania niezgodne z naszymi przekonaniem wywołują nieprzyjemny dysonans, który łatwo zlikwidować, dostosowując te ostatnie do zachowania (Festinger, 1957/2007). Z tych powodów PPB są tak niebezpieczne – nie tylko umożliwiają naukowcom naciąganie wyników ze szczerym przeświadczeniem o poprawności merytorycznej swoich decyzji, lecz także utwierdzają ich w tym przekonaniu z każdą kolejną publikacją. Być może zbyt często zapominamy, że my, psychologowie, także podlegamy mechanizmom, które badamy.

W swoim tekście Klebaniuk zwraca uwagę na „problem szuflady” (*file drawer effect/problem*), stwierdzając, że „szuflady pełne nieopublikowanych badań milczą”, co „zafalszowuje obraz nauki” (Klebaniuk, 2012, s. 216). Co do istnienia opisywanego zniekształcenia publikacyjnego (*publication bias*) i powagi jego konsekwencji trudno mieć jakiegokolwiek wątpliwości (por. np. Ferguson i Brannick, 2012). Klebaniuk słusznie wskazuje też na politykę czasopism naukowych oraz presję publikacyjną wynikającą z modelu *publish or perish*

(które napędzają PPB) jako winnych tego stanu rzeczy. Jego postulat spoglądania łaskawszym okiem na wyniki zerowe lub (częściowo) „nieudane” badania przez czasopisma naukowe jest ważny i słuszny. Należy jednak pamiętać także o innym bardzo istotnym czynniku napędzającym „problem szuflady” – powszechnym braku zainteresowania psychologów kwestią mocy statystycznej ($1 - \beta$), a w konsekwencji, bardzo niską skutecznością wykrywania efektów w ich badaniach. W artykule „Czy badania mocy statystycznej wpływają na moc statystyczną badań?” Sedlmeier i Gigerenzer (1989) dali negatywną odpowiedź na to pytanie, pokazując, że sytuacja jest alarmująca. Porównali medianę mocy statystycznej w *Journal of Abnormal and Social Psychology* uzyskaną w klasycznym badaniu Jacoba Cohena do tej z 1984 roku i stwierdzili, że bardzo niska moc z roku 1960 (mediana 0,46 dla średniego efektu) po 24 latach nie zmieniła się $(0,44)^2$. Należy przy tym pokreślić, że Cohen (1988) sugeruje 0,80 jako moc pożądaną. Moc 0,44 oznacza, że jeśli badany efekt istnieje naprawdę i jest średniej wielkości, to w perspektywie długoterminowej uzyskamy istotność statystyczną tylko w 44% badań. Innymi słowy, ponad połowa badań łąduje w szufladzie bez szans na publikację tylko dlatego, że próba jest zbyt mała. Czy teraz jest inaczej? W podręczniku publikacyjnym Amerykańskiego Towarzystwa Psychologicznego (*APA Manual*) kwestia mocy pojawiła się po raz pierwszy, choć tylko zdawkowo, „już” w czwartej edycji, czyli w 1994 roku (Fidler, 2010). Wydaje się jednak, że daleko zbyt optymistyczne byłoby stwierdzenie, iż wyniki analizy Sedlmeiera i Gigerenzera można uznać za relikwiny przeszłości. Wystarczy przejrzeć czołowe czasopisma psychologii społecznej, aby zobaczyć, że kwestie mocy nadal nie są w nich poruszane. Nie dalej jak podczas pisania tego tekstu, chciałem dowiedzieć się od pewnego zagranicznego psychologa ze stopniem doktora i międzynarodowym dorobkiem publikacyjnym, jaka jest średnia wielkość efektu w dziedzinie, w której planowaliśmy badanie, ponieważ chciałem zadbać o należyłą moc statystyczną. Spotkałem się z zupełnym niezrozumieniem i nie jest to odosobnione podejście. Oczywiście intuicyjną i anegdotyczną wiedzę należy poprzeć rzetelną analizą na dużej, najlepiej reprezentatywnej próbie publikacji lub psychologów przed wyciągnięciem ostatecznych wniosków. Nic nie wskazuje jednak na to, że w psychologii społecznej jest dziś lepiej.

Analizując gorącą debatę dotyczącą kwestii zawyżonych korelacji mózg–zachowanie w neuronauce społecznej, Klaus Fidler (2011, s. 164) opisuje niską moc jako „wyjątkowo zwodnicze źródło zniekształceń”. Zwraca też uwagę, że przy tak niskich próbach, jakie stosuje się w badaniach fMRI, mediana korelacji istotnych statystycznie

waha się między 0,75 i 0,80 – mniejsze efekty nie uzyskują istotności i zapychają szuflady naukowców. Z tego powodu wykonuje się więcej badań (cały czas o niskiej mocy) w poszukiwaniu $p < 0,05$, co oczywiście zwiększa szansę, że uzyskamy wynik fałszywie istotny (Ioannidis, 2005). Tak więc, aby badania z wynikami zerowymi spełniły swoją ważną, uzdrawiającą rolę na łamach czasopism, ich autorzy muszą zadbać o moc statystyczną. Ten postulat dotyczy oczywiście w takim samym stopniu publikowanych badań „udanych”, które oprócz niskiej mocy mają też często statystyki sztucznie „podkręcone” przez PPB, co tylko potęguje ich zniekształcenie. Akurat w tej konkretnej sprawie winne są nie tyle czasopisma bezwzględnie wymagające $p < 0,05$, co przede wszystkim sami badacze, którzy nie poświęcają wystarczającej uwagi kwestii mocy statystycznej.

Z czego wynika ten brak świadomości znaczenia mocy statystycznej? Prawdopodobnie po części odpowiada za ten stan rzeczy model uprawiania nauki podkreślany przez Klebaniuka, wymagający od naukowców ciągłego publikowania, a w konsekwencji przeprowadzania dużej liczby badań. Wydaje się jednak, że problem leży dużo głębiej, a mianowicie my, psychologowie, mamy wielki problem z myśleniem statystycznym (Gigerenzer, 2004). Po pierwsze, obecnie obowiązujący kanon statystyczny w psychologii (*Null Hypothesis Significance Testing*; NHST) jest przedstawiany studentom jako monolityczny, istniejący od zawsze i niepodlegający żadnym negocjacjom, mimo że powstał w połowie lat 50. jako trochę przypadkowa hybryda wczesnej teorii Ronalda Fischera oraz teorii Jerzego Neymana i Egona Pearsona (Gigerenzer, 2004). Podejście NHST zostało następnie „wmurowane” w psychologię poprzez kolejne wydania bezkrytycznych podręczników (por. Gigerenzer, 2004; Gliner, Leech i Morgan, 2002). Gigerenzer (2004) spytał pewnego autora (niebędącego statystykiem) znanego i wielokrotnie wznawianego podręcznika do statystyki, dlaczego od pewnego wydania zaczął pomijać informacje o teorii Fischera i Neymana–Pearsona (oraz statystyce bayesowskiej). W odpowiedzi usłyszał, że winowajców jest trzech: 1) jego koledzy naukowcy, bo są zainteresowani tylko tym, jak się opublikować; 2) władze uniwersytetu, bo oceniają naukowców tylko na podstawie liczby publikacji, co wzmacnia podejście naukowców i 3) wydawcy, bo wymagali, aby podręcznik był prostą książką kucharską, bez kontrowersji. Takie nauczanie nie zachęca do myślenia statystycznego, wręcz przeciwnie. Po drugie, wśród psychologów powszechne są błędy w rozumowaniu statystycznym nawet w ramach tak prostego systemu, jak NHST. Przykładowo, Haller i Krauss (2002) przeprowadzili badanie na wydziałach psychologii niemieckich

uniwersytetów. Wzięło w nim udział 30 wykładowców oraz doktorantów uczących statystyki, 39 wykładowców nieuczących statystyki i 44 studentów po zaliczeniu przynajmniej jednego kursu statystyki. Przedstawiono im sześć zdań i poproszono o ocenę, czy na podstawie podanych wyników uprawnione są dane stwierdzenia. Najwięcej trudności sprawiły respondentom następujące pozycje: „Jeśli odrzucasz hipotezę zerową, to znasz prawdopodobieństwo poprawności swojej decyzji” oraz „(...) Gdyby eksperyment został powtórzony dużą liczbą razy, to wynik byłby istotny w 99% przypadków” (przedstawiony wynik był $p = 0,01$). Żadne ze stwierdzeń nie było uprawnione, ale tak odpowiedziało tylko 20% uczących statystykę, 10% nieuczących statystyki i żaden student po kursie ze statystyki! Próba badawcza jest tu mała i niereprezentatywna, ale podobne wyniki zaobserwowano w badaniu brytyjskim (Oakes, 1986; za: Gigerenzer, 2004) i izraelskim, w którym dodano zbiorczą opcję odpowiedzi *żadne stwierdzenie nie jest poprawne* (Falk i Greenbaum, 1995). Skoro nawet nauczyciele akademicy mają trudności z podstawowymi założeniami systemu statystycznego, którego uczą, to trudno spodziewać się, by rozumieli je studenci, a są to przecież przyszli badacze, autorzy artykułów i nauczyciele, także statystyki. W tym kontekście nie dziwi brak zrozumienia psychologów – młodszych i starszych – dla wagi mocy statystycznej, tym bardziej że w obecnym systemie statystycznym jest to kwestia poboczna względem faworyzowanej *alfy* i poziomu istotności, które dają mechaniczną przepustkę do „udanej” pracy magisterskiej czy „udanej” publikacji (por. Cohen, 1994). Co więcej, przy braku myślenia statystycznego rośnie także normatywność stosowania PPB. Zmiana podejścia do nauczania i praktykowania statystyki z bezmyślnego i skrajnie instrumentalnego na myślące i bardziej podmiotowe jest więc jednym z kluczowych wyzwań i sposobów poprawy kondycji psychologii akademickiej.

Wspominany postulat Klebaniuka, dotyczący otwarcia łam czasopism na badania zerowe i mniej „udane” (Klebaniuk, 2012, s. 216), wydaje się niestety nieskuteczny, jeśli nie pójdą za nim dodatkowe zmiany (oprócz zadbania o moc statystyczną). Przyjmując bowiem, że liczba artykułów, które mogą ukazać się w czołowych czasopismach z danej dziedziny w danym momencie jest stała, badanie „ładniejsze” zawsze wyprze badanie „brzydsze” w kolejce do publikacji. Dlaczego? Recenzenci i redaktorzy czasopism nie mają zazwyczaj wystarczających informacji, aby stwierdzić, czy dane badanie jest naprawdę wysokiej jakości, czy raczej stanowi „makijaż upiększający twarz naukowej prawdy”. Bez tej wiedzy recenzenci nie mają podstaw, aby odrzucić badanie „ładniejsze” na

korzyść „brzydszego”. Oczywiście przy odpowiedniej dozie uporu i poświęconego czasu recenzent ma szansę uzyskać informacje obnażające słabości badania, ale biorąc pod uwagę ilość pracy, jaką jest obciążonych wielu naukowców, jest to mało praktyczne rozwiązanie (nie wspominając już o tym, że wszyscy recenzenci są także autorami, a więc wielu z nich stosuje PPB). Aby zmienić ten stan rzeczy potrzeba więc dużo więcej niż udostępnianie danych surowych na życzenie oraz raportowanie szczegółów proceduralnych – to może zabezpieczyć głównie przed fałszowaniem wyników. Obecny system recenzowania i publikacji opiera się bowiem w znacznej mierze na zaufaniu między naukowcami, dając im wielką dowolność w tym, co, jak i czy w ogóle opisują w manuskryptach wysyłanych do recenzji. Badanie John i współpracowników pokazuje, że w obecnej sytuacji system oparty na zaufaniu i dowolności nie działa i trzeba poszukać nowego, systemowego rozwiązania.

Postulat Klebaniuka sam w sobie może mieć małą moc sprawczą także dlatego, że badacze nie mają prywatnego interesu w zmywaniu makijażu ze swoich wyników. Przy obecnym natężeniu rywalizacji o często bardzo ograniczone zasoby (takie jak stanowiska uniwersyteckie i granty), a także przy nadprodukcji wiedzy naukowej, na miejsce wyników pozbawionych makijażu natychmiast wpełzną się te upiększone.

Jakie rozwiązanie może być więc skuteczne? Simmons, Nelson i Simonsohn (2011) zaproponowali szeroki zestaw wymagań dla autorów publikacji naukowych i ich recenzentów, który wydaje się dawać dobrą odpowiedź na to pytanie. Wedle ich zamysłu autorzy wysyłający tekst do recenzji powinni: 1) zdecydować o docelowej wielkości próby przed badaniem i raportować ją w artykule; 2) dysponować minimum 20 obserwacjami w celce; 3) raportować wszystkie zmienne użyte w badaniu; 4) raportować wszystkie warunki eksperymentalne oraz nieudane manipulacje; 5) podawać statystyki także dla pełnej próby, jeśli decydują się wykluczyć jakieś obserwacje oraz 6) podawać statystyki także bez współzmiennych/zmiennych kontrolowanych, jeśli decydują się włączyć jakieś do analizy. Recenzenci tych prac powinni natomiast: 1) egzekwować przestrzeganie powyższych zaleceń przez autorów; 2) być bardziej tolerancyjnymi dla nieperfekcyjnych wyników; 3) wymagać, aby autorzy udowodnili, że wyniki nie zależą od arbitralnych decyzji podczas analizy oraz 4) wymagać dokładnych replikacji, jeśli sposób zbierania i analizy danych nie broni się.

Jak widać, pomysł Simmonsa i współpracowników zawiera kwestie postulowane przez Klebaniuka, ale uzupełnia je o kluczowy element, czyli o zmianę w sposobie pisania artykułów empirycznych, proponując przejście od

pełnego zaufania i pełnej dowolności do ograniczonego zaufania i pełnej transparentności. Taka zmiana zapewniłaby „intersubiektywną komunikowalność i intersubiektywną sprawdzalność rezultatu badawczego” (Klebaniuk, 2012, s. 215) przez recenzentów oraz wszystkich czytelników (a nie tylko tych, którzy zadali sobie trud zdobycia i powtórnego przeanalizowania czyjejs bazy danych). Dopiero gdy czasopisma naukowe przyjmą taki lub podobny (por. postulat Schoolera o raportowaniu pierwotnych hipotez; za: Carey, 2011b) zestaw zaleceń jako obowiązujący autorów, ta mniej piękna twarz psychologii będzie miała szansę ujrzeć światło dzienne, z wielką korzyścią dla niej samej.

Innym istotnym czynnikiem zniekształcającym i upiększającym psychologię, a niewspomnianym przez Klebaniuka, jest współczesny trend w czasopismach psychologicznych promujący publikowanie krótkich doniesień z badań (*short/flash reports*), zawierający często tylko jedno badanie. W 1990 roku wprowadziło je na swoje łamy *Psychological Science* i od tego czasu bardzo wiele czasopism poszło tą drogą, lansując wśród naukowców publikacyjną „strategię zająca” w kontrze do tradycyjnej „strategii żółwia” (Ledgerwood i Sherman, 2012). Ta moda dotarła także do Polski, np. *Roczniki Psychologiczne* wprowadziły niedawno taką formę publikacji (do 9 tys. znaków). Jest przynajmniej kilka czynników czyniących tę strategię tak zachęcającą, np. 1) zwiększona interakcja między badaczami; 2) potencjalnie większa siła oddziaływania psychologii na inne dziedziny i osoby spoza akademii i 3) szybsze wymierne efekty pracy naukowej wymagane przez obecny model kariery (Ledgerwood i Sherman, 2012). Niestety, słabości „strategii zająca” w formie stosowanej obecnie znacznie przewyższają jej zalety (przy założeniu, że w ogóle istnieją). Przede wszystkim krótkie doniesienia zwiększają szansę na „efekt zaniku” (*decline effect*; Lehrer, 2010), polegający na tym, że z upływem czasu (nawet lat) wielkość badanego efektu znacząco się zmniejsza, a nawet całkowicie zanika, ku rozpaczy autorów danej teorii. Wynika to z tego, że artykuły z jednym badaniem: 1) mają większe szanse na uzyskanie fałszywie istotnego wyniku; 2) po publikacji prowadzą do zniekształcenia konfirmacyjnego (*confirmation bias*), bo naukowcy częściej idą tropem wyników fałszywie istotnych niż fałszywie nieistotnych i 3) wraz ze zwiększeniem liczby badaczy próbujących zreplikować dany fałszywie pozytywny wynik szansa na kolejny błąd I rodzaju rośnie i w ten sposób marnowane są zasoby wielu psychologów, a dziedzina kompromituje się sama przed sobą, przed naukami pokrewnymi oraz przed szerszą publicznością (Bertamini i Munafò, 2012; Ioannidis, 2005; Ledgerwood i Sherman, 2012; Lehrer,

2010; Pfeiffer i Hoffman, 2009; Sterne i Smith, 2001). Inną słabością krótkich doniesień jest mniejsza spójność wiedzy naukowej (m.in. przez „amnezję cytowań”; por. Garfield, 1982, i mniej miejsca na teoretyczne wpisanie wyników w strukturę wiedzy; por. Abelson, 1997) oraz koncentracja na („błyskotliwych”) efektach zamiast na rozwijaniu teorii i badaniu mechanizmów (Ledgerwood i Sherman, 2012). Oczywiście można próbować kompensować słabości „strategii zająca”, ale na razie nic takiego się nie zapowiada, a kolejne czasopisma z radością otwierają swoje łamy na krótkie doniesienia, wspomagając w ten sposób proces, w którym „nauka zamienia się (...) w grę pozorów” (Klebaniuk, 2012, s. 216).

W świetle dostępnych danych trudno zaprzeczyć, że psychologia, zwłaszcza społeczna, jest dziś w stanie poważnego kryzysu i inercyjnie zmierza ku przepaści. Potrzebne są zdecydowane działania chroniące ją przed zapaścią i dyskredytacją w oczach nauk pokrewnych, odbiorców spoza akademii, naszych rygorów metodologicznych i, może przede wszystkim, w oczach młodych adeptów nauk psychologicznych, do których grona sam należą. Dla społeczności naukowców jest to duże wyzwanie, wymagające intensywnej, ale rzeczowej i odpowiedzialnej dyskusji. Najgorsze, co może wydarzyć się w takim momencie, to tabuizacja tematu, jego zbagatelizowanie lub erupcja defensywnych negacji broniących błędnie rozumianego dobrego imienia psychologii. Mocny reformatorski głos Jarosława Klebaniuka, zapoczątkowujący dyskusję na łamach *Psychologii Społecznej*, daje nadzieję, że tak się nie stanie oraz że czeka nas ciekawa i owocna dyskusja wśród polskich psychologów społecznych, ukoronowana wypracowaniem polityki zmian i jej efektywnym wdrożeniem.

LITERATURA CYTOWANA

- Abelson, R. P. (1997). On the surprising longevity of flogged horses: Why there is a case for the significance test. *Psychological Science*, 8, 12–15.
- Bem, D. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407–425.
- Bertamini, M., Munafò, M. R. (2012). Bite-size science and its undesired side-effects. *Perspectives on Psychological Science*, 7, 67–71.
- Buss, A. (1975). The emerging field of the sociology of psychological knowledge. *American Psychologist*, 30, 988–1002.
- Carey, B. (2011a). *Journal's paper on ESP expected to prompt outrage*, <http://www.nytimes.com/2011/01/06/science/06esp.html?pagewanted=all>, dostęp 12.03.2012.
- Carey, B. (2011b). Fraud case seen as a red flag for psychology research, <http://www.nytimes.com/2011/11/03/health/research/>

- noted-dutch-psychologist-stapelaccused-of-research-fraud.html?_r=1&scp=1&sq=Leslie%20John&st=cse, dostęp 12.03.2012.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (wyd. 2). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997–1003.
- Ditto, P. H., Pizarro, D. A., Tannenbaum, D. (2009). Motivated moral reasoning. W: D. M. Bartels, C. W. Bauman, L. J. Skitka, D. L. Medin (red.), *The psychology of learning and motivation* (t. 50, s. 307–338). Burlington: Academic Press.
- Falk, R., Greenbaum, C. W. (1995). Significance tests die hard: The amazing persistence of a probabilistic misconception. *Theory and Psychology*, 5, 75–98.
- Ferguson, C. J., Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods*, 17, 120–128.
- Festinger, L. (1957/2007). *Teoria dysonansu poznawczego*. Warszawa: Wydawnictwo Naukowe PWN.
- Fidler, F. (2010). The American Psychological Association Publication manual sixth edition: Implications for statistics education. W: C. Reading (red.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute.
- Fiedler, K. (2011). Voodoo correlations are everywhere – not just in neuroscience. *Perspectives on Psychological Science*, 6, 163–171.
- Garfield, E. (1982). More on the ethics of scientific publication: Abuses of authorship attribution and citation amnesia undermine the reward system of science. *Current Contents*, 30, 5–10.
- Gigerenzer, G. (2004). Mindless statistics. *The Journal of Socio-Economics*, 33, 587–606.
- Gliner, J. A., Leech, N. L., Morgan, G. A. (2002). Problems with Null Hypothesis Significance Testing (NHST): What do the textbooks say? *The Journal of Experimental Education*, 71, 83–92.
- Haller, H., Krauss, S. (2002). Misinterpretations of significance: A problem students share with their teachers? *Methods of Psychological Research Online*, 7, 1–20.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Med*, 2, e124.
- John, L., Loewenstein, G., Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524–532.
- Jost, J. T., Glaser, J., Kruglanski, A. W., Sulloway, F. J. (2003). Political conservatism as motivated social cognition. *Psychological Bulletin*, 129, 339–375.
- Klebaniuk, J. (2012). Profesor Stapel na dopingu. O upiększaniu psychologii społecznej. *Psychologia Społeczna*, 3 (22), 213–217.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480–498.
- Ledgerwood, A., Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science*, 7, 60–66.
- Lehrer, J. (2010, 13 grudnia). Annals of science: The truth wears off. *The New Yorker*, s. 52.
- Pfeiffer, T., Hoffmann, R. (2009). Large-scale assessment of the effect of popularity on the reliability of research. *PLoS ONE*, 4, e5996.
- Pyszczynski, T., Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. W: L. Berkowitz (red.), *Advances in experimental social psychology* (s. 297–340). New York: Academic Press.
- Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Clinical and Consulting Psychology*, 58, 646–656.
- Sedlmeier, P., Gigerenzer, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, 105, 309–316.
- Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359–1366.
- Sterne, J. A. C., Smith, G. D. (2001). Sifting the evidence: What's wrong with significance tests? *British Medical Journal*, 322, 226–231.

PRZYPISY

1. Badani zostali losowo przydzieleni do dwóch grup. Przedstawione frekwencje opierają się na odpowiedziach osób z grupy eksperymentalnej, które były specjalnie motywowane do udzielania szczerych odpowiedzi (John, Loewenstein i Prelec, 2012).
2. Przed 1984 r. *Journal of Abnormal and Social Psychology* został podzielony na *Journal of Abnormal Psychology* i *Journal of Personality and Social Psychology*. Autorzy oparli swoje analizy tylko na *Journal of Abnormal Psychology*, ale podobne badanie z tego okresu oparte na *Journal of Abnormal Psychology*, *Journal of Consulting and Clinical Psychology* i *Journal of Personality and Social Psychology* także wykazało niską moc (mediana 0,53 dla średniego efektu; Rossi, 1990).

Crisis in psychology, psychology of crisis

Mateusz Olechowski

Faculty of Psychology, University of Warsaw

Abstract

In this commentary I refer to selected elements of diagnosis and recommendations presented in Jarosław Klebaniuk's 'Professor Stapel on dope. On the beautifying of social psychology'. I elaborate on issues that in my opinion require a more detailed analysis ('questionable research practices', 'researcher degrees of freedom') and I introduce new claims and guidance (statistical power, critical statistical thinking, article writing manner, short/flash reports) that seem essential to understand today's academic psychology's state and to heal it. I urge to actively affront this difficult challenge and take strong actions to end the mounting crisis in (social) psychology.

Keywords: crisis in psychology, researcher degrees of freedom, questionable research practices, statistical power, critical statistical thinking, short/flash reports