

Wrocław 05.06.2022

dr hab. Michał Białek, prof. UWr

Kierownik Zakład Psychologii Zarządzania

Instytut Psychologii, Uniwersytet Wrocławski

Recenzja rozprawy doktorskiej

Pani mgr Karoliny Szutkiewicz-Szekalskiej

Pod tytułem „Wybrane moderatory zniekształceń w sądach porównawczych: kiedy typowe efekty
zawyżania ulegają odwróceniu?”

Promotor: prof. dr hab. Jerzy Karyłowski

Promotor pomocniczy: dr Jakub Niewiarowski

Przedstawiona rozprawa dotyczy porównań losowo wybranego elementu z grupy z resztą tej grupy. Autorka wskazuje, że wśród atrakcyjnych obiektów losowo wybrany obiekt oceniany jest wyżej lub niżej niż grupa referencyjna w zależności od jej wielkości: przy małej grupie referencyjnej obiekt oceniany jest podobnie lub nieznacznie lepiej niż grupa referencyjna, przy dużej i najlepiej heterogenicznej grupie referencyjnej losowy obiekt oceniany jest gorzej niż grupa referencyjna. Odkrycie to jest nowatorskie, poszerza aktualną wiedzę na dany temat, i w moim odczuciu spełnia wymogi stawiane rozprawie doktorskiej przez odpowiednie ustawy i rozporządzenia.

Praca na 183 stronach raportuje sześć eksperymentów, uporządkowanych w ciąg logiczny, tak, że kolejne badanie poszerza wiedzę zdobytą w poprzednim eksperymencie. Na pochwałę zasługuje umiejętność zwięzłego i klarownego przekazywania kluczowych informacji na temat badań: ich układu, hipotez, zmiennych oraz planu badawczego.

Na szczególną uwagę i pochwałę zasługuje sposób pisania pracy – prosty, zrozumiały język znakomicie wspomaga przekazywanie treści. Przegląd literatury jest wyczerpujący; wszystkie poza



modelem dwusystemowym teorii i modele opisane są przystępnie, zrozumiale i wyczerpująco. Ta część pracy zasługuje na szczególne wyróżnienie.

Jest kilka powodów, dla których niestety nie będę wnioskował o wyróżnienie tej pracy. Praca, szczególnie w jej końcowej części, zawiera sporo błędów edytorskich (błędne wartości w tabelach, wyłuszczone cały podrozdział zamiast jedynie jego tytułu). Mam także kilka uwag co do teorii oraz sposobu analiza danych, które jednak w żaden sposób nie wpływają na moja ogólnie pozytywna ocenę całej pracy. Omówię te uwagi poniżej, dając możliwość autorce do podjęcia dyskusji nad wybranymi przez nią sposobami analizy danych oraz wnioskowania na ich podstawie.

1. Teoria dwusystemowa omówiona została w oparciu o jej dość archaiczny model. Od co najmniej dekady mówi się nie o systemach, a o procesach typu 1 i typu 2, które ze sobą współpracują lub konkurują. Co prawda w pracy pojawiają się nawiązania do nowszego modelu, omówionego w książce przez Kahnemana, ale sama część teoretyczna omawia Epsteina oraz wczesną wersję teorii dwusystemowej za Sokołowską. Warto sięgnąć do kanonicznego już artykułu Stanovicha i Evansa z *Perspective on Psychological Science* (Evans & Stanovich, 2013) by zobaczyć na ile model omawiany w pracy jest nieaktualny. Polecam też prace De Neysa i Pennycooka z (2019) roku na ten sam temat.
2. Gdy jesteśmy już przy teorii dwusystemowej, w eksperymentach 4-6 pojawiają się odwołania do trybów przetwarzania oraz ich głębi. Autorka wraz z odpowiedziami badanych zbierała też czasy reakcji. Nie szkodziło zatem przeanalizować to czy procesy kategoryzacyjne przypisywane przez autorkę procesom Typu 2 faktycznie zajmowały badanym więcej czasu niż procesy klasyfikowane jako Typu 1. Oczywiście szybkość przetwarzania jest tylko korelatem procesów Typu 1 i 2, ale jednak pozwoliłyby podeprzeć wnioski stawiane przez autorkę. (oczywiście problemem może być to, że opisy filmów czyta się dłużej niż ogląda plakaty filmowe).
3. Moc do wykrycia interakcji. Moc testu to prawdopodobieństwo zaobserwowania danego prawdziwego efektu. O ile próbki zebrane przez autorkę wydają się imponujące (w sumie w 6 eksperymentach zbadano około 1500 osób), to starczą one jednak tylko do wykrycia efektów głównych, oraz być może interakcji pierwszego rzędu. Interakcje takie, w którym jeden efekt jedynie słabnie wraz ze zmianą wartości moderatora (np., odwrócony efekt zawiżania roślinie wraz ze wzrostem liczby obiektów w grupie referencyjnej) wymagają nawet 16-krotnie większej próby niż obserwacja efektu głównego (Lakens & Caldwell, 2021). Uważam zatem interpretację

- interakcji 3-rzędu w układzie 5x2x2 za wątpliwą, ze względu na małą moce do wykrycia tego efektu w sposób rzetelny (tzw. efekt małego teleskopu; (Simonsohn, 2015).
4. Ku mojemu zdziwieniu, liczbę obiektów w grupie referencyjnej potraktowano jako zmienną kategoryjną. A przecież liczba obiektów pozwala się transformować w każdy sposób – można np. powiedzieć przecież że 4 obiekty to dwa razy więcej niż 2. Zatem można spokojnie potraktować liczbę obiektów jako zmienną ciągłą, i znacząco uprościć schemat badania (np. w eksperymentach 1 oraz 2, zamiast 5 kategorii mieć jedną zmienną, która nie wymaga testów post-hoc i korekty na wielokrotne testowanie) Byłoby to tym bardziej uzasadnione, że autorka twierdzi że związek liczby obiektów w grupie referencyjnej ma związek liniowy z siłą odwróconego efektu zawyżania, ale wykazuje to serią kontrastów między kategoriami.
 5. Tabela 11 zawiera błąd – odchylenia standardowe i siła efektu są identyczne dla każdego testu i nie pasują do wartości p. Na przykład średnia 3.09, przy odchyleniu standardowym 0.39 i próbie $N = 151$ sugeruje różnicę $t(150) = 2.84$, $p = 0.052$. (użyłem kalkulatora online <https://www.graphpad.com/quickcalcs/oneSampleT2/>). W tabeli istotność podano jako $p = .009$. Zakładam że odchylenia standardowe są poprawnie zaraportowane w tabeli, bo te same wartości przywoływane są przy opisie próby w tekście poniżej.
 6. W eksperymencie 6 brakło jakiegokolwiek potwierdzenia że efekt torowania przełożył się na decyzje badanych, analiza od razu odbywa się w podgrupach. W efekcie brakuje dowodu, że efekty obserwowane w różnych grupach, nawet jeśli różnią się istotnością, to różnią się od siebie istotnie statystycznie. Na różnice między siłami efektu wskazywałaby istotna interakcja torowanie x cecha bodźca.
 7. Interpretacja wielkości efektów opiera się na przestarzałej typologii Cohena. Od czasu kryzysu replikacyjnego wiemy, że zdecydowana większość efektów raportowanych we wcześniejszej literaturze jest przeszacowana, a średni efekt w psychologii to $d = 0.02$ (Funder & Ozer, 2019). Aby to zilustrować, różnica między grupą kontrolną a eksperymentalną $d=0.5$ to prawie 8 punktów IQ różnicy, i zdecydowanie nie jest to „średni” efekt prostej manipulacji psychologicznej. Warto zrewidować swoje oczekiwania co do spodziewanych efektów eksperymentów w psychologii, choćby po to by adekwatniej dobierać wielkość prób eksperymentalnych.
 8. Szkoda, że żadne z badań nie było prerejestrowane, ani że dane z eksperymentów nie zostały udostępnione w publicznym repozytorium (np. Open Science Framework, osf.io). Zachęcam do

stosowania praktyk otwartej nauki, to naprawdę pomaga autorom bardziej niż komukolwiek innemu.

Podsumowując, pomimo pewnych zastrzeżeń uważam rozprawę za bardzo dobrą, oraz stwierdzam, że spełnia ona wymogi stawiane rozprawom doktorskim, określone w art. 13 ust. 1 Ustawy z dnia 4 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. Ustaw nr 65 poz. 595 z późn. zm.) i wnioskuję o dopuszczenie Doktorantki do dalszych etapów procedur zmierzających do przyznania jej stopnia doktora.

Bibliografia

- De Neys, W., & Pennycook, G. (2019). Logic, fast and slow: Advances in dual-process theorizing. *Current Directions in Psychological Science*, 28(5), 503–509.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241.
<https://doi.org/10.1177/1745691612460685>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, 2(2), 156–168.
<https://doi.org/10.1177/2515245919847202>
- Lakens, D., & Caldwell, A. R. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*, 4(1), 2515245920951503.
- Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological Science*, 26(5), 559–569.

Zakład Psychologii Zarządzania
KIEROWNIK

dr hab. Michał Bialek, prof. UWr