

Wrocław 12.11.2024

Dr hab. Michał Białek, prof. UWr

Kierownik Zakład Psychologii Zarządzania

Instytut Psychologii, Uniwersytet Wrocławski

Recenzja drugiej rewizji rozprawy doktorskiej

Wpływ obecności rywala na awersję do straty w decyzjach ekonomicznych

Autorstwa mgra Witolda Siekierzyńskiego

Przedstawiona mi do recenzji rozprawa doktorska „Wpływ obecności rywala na awersję do straty w decyzjach ekonomicznych” autorstwa magistra Witolda Siekierzyńskiego bada wpływ obecności rywala na podejmowanie decyzji ekonomicznych. Autor opisuje w niej cykl trzech badań eksperymentalnych, w których analizował, jak obecność rywala wpływa na skłonność do angażowania własnych zasobów poznawczych oraz skłonność do podejmowania ryzyka w sytuacjach decyzji ekonomicznych. W badaniach uczestniczyło łącznie ponad 1800 osób, które podejmowały decyzje dotyczące różnych zasobów ekonomicznych (pieniędzy własnych, cudzych zasobów finansowych oraz decyzji dotyczących ludzi). Decyzje te były prezentowane z perspektywy potencjalnego zysku lub straty. Autor analizował, czy sama obecność rywala oraz jego cechy (takie jak płeć czy narodowość) wpływają na skłonność do podejmowania ryzyka. Na podstawie uzyskanych wyników autor zaproponował model teoretyczny, według którego obecność rywala tworzy dodatkową perspektywę społeczną, w ramach której analizowane są skutki decyzji. Oznacza to, że decydenci mogą rozważać swoje wybory nie tylko w kontekście bezpośrednich skutków finansowych, ale także w kontekście konsekwencji społecznych, takich jak wpływ na status czy relacje z innymi.

W porównaniu z wcześniejszymi wersjami rozprawa została poprawiona w zakresie prezentacji teorii (choć nadal jest tu wiele do zrobienia). Jednak pomimo licznych uwag moich oraz pozostałych recenzentów, plan badań oraz interpretacja wyników nadal są niewłaściwe, same badania noszą znamiona wątpliwych praktyk badawczych (QRP), a interpretacja uzyskanych wyników jest powierzchowna. Liczyłem, że w kontekście licznych krytycznych uwag moich oraz pozostałych recenzentów autor zdecyduje się na przeprowadzenie kolejnego badania, które niwelowałoby deficyty dotychczasowych badań. Niestety, zauważyłem, że naniesiono jedynie kosmetyczne zmiany, które nie odnoszą się do meritum moich zarzutów. W konsekwencji podtrzymuję swoją negatywną opinię na temat rozprawy i uznaję, że nie spełnia ona warunków określonych w Ustawie z dnia 20 lipca 2018 r., – Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018, poz. 1668), i wnioskuję o odmowę nadania stopnia doktora Panu mgr. Witoldowi Siekierzyńskiemu. Poniżej przedstawiam syntetyczne zarzuty w stosunku do rozprawy.

Część teoretyczna nie uzasadnia hipotez, szczególnie moderacyjnych. Nie wiemy, jak literatura naukowa kontekstualizuje badane w tym projekcie zmienne mające moderować zależność między

obecnością rywala a skłonnością do ryzyka. Dotyczy to szczególnie elementów analizowanych w Badaniu 3, takich jak narodowość czy podejmowanie decyzji dla siebie lub jako doradca innej osoby. W efekcie słabego ułożenia projektu w literaturze, uzyskane przez Autora wyniki nie są interpretowane w kontekście teorii, a jedynie opisane słownie.

Obecność rywala w zdecydowanej większości scenariuszy była bez kluczowego znaczenia dla problemu decyzyjnego. W treści zadania nie było powiedziane, że badany z w jakikolwiek sposób rywalizuje ze wspomnianymi w scenariuszach osobami czy firmami (poza warunkiem bycia doradcą klienta w Badaniu 3). Informacja, że istnieje inny gracz albo firma, niekoniecznie stanowi właściwą, wystarczająco silną manipulację. Aby ktoś stanowił rywala, musi zachodzić porównywanie się do tej osoby. Na przykład, czy sama informacja zawarta w Badaniu 3, że istnieje w twojej okolicy inna firma podobnej wielkości, powoduje, że traktuje się właśnie tę firmę jako rywala? Istnieje duże prawdopodobieństwo, że manipulacja nie była dostatecznie wyrazista, by wywołać realne i silne efekty psychologiczne.

Kolejny problem to scenariusze opisujące straty finansowe i niefinansowe. Po pierwsze, konieczność zwolnienia pracowników niekoniecznie jest stratą (może być neutralna albo nawet oceniana pozytywnie oszczędnością), a jeśli już, to jest stratą zarówno dla zwalnianego, jak i zwalniającego (większą dla tego pierwszego). Badany może zatem mieć poczucie, że decyduje zasobami własnymi i cudzymi lub/ oraz że ponosi koszty moralne krzywdzenia pracowników. Nie ma zatem pełnej symetrii między warunkiem niefinansowym (zwalniania większej lub mniejszej liczby pracowników) a finansowym (zarobieniem większej czy mniejszej kwoty). Kolejny problematyczny aspekt, to fakt że warunek straty obiektywnie nadal jest zyskiem (po prostu zarabiamy 1500 zamiast potencjalnych 2000, przedstawiony jako „tracimy 500”), podczas gdy warunek zwalniania pracowników w najlepszym przypadku pozwala nam nie zwolnić nikogo. Zatem porównuje się obiektywne zyski z obiektywnymi stratami jako przypadki straty. Mam zatem duże wątpliwości co do poprawności porównań pomiędzy kategoriami finansowymi i niefinansowymi.

Kolejnym problemem jest zmiana prawdopodobieństwa wygranej lub straty z 50% (zasoby własne) na 33% (zasoby cudze) pomiędzy warunkami Badania 3. Porównanie między decyzjami dotyczącymi własnych i cudzych zasobów dotyczy zatem dwóch elementów jednocześnie (prawdopodobieństwa wygranej i właściciela zasobów), a nie, jak autor chciałby - tylko jednego.

Operacjonalizacja zmiennych pozostawia wiele do życzenia – czasy reakcji traktowane są jako miara zaangażowania poznawczego w podejmowanie decyzji. Jednak pomiar czasów reakcji zawiera czytanie oraz namysł nad decyzją, jest zatem zakłócony tempem czytania. Scenariusze z rywalem są nieco dłuższe, co może wpływać na różnice w czasie wykonania. Fundamentalnym problemem jest to, że zmienna zależna "czas reakcji" była arbitralnie ograniczona w sposób niespójny w badaniu 1 (poniżej 15 sekund i górne 5%) i badaniu 2 (ponad 2SD od średniej) oraz określona nieznanym mi przedziałem w badaniu 3 (opartym o przewidywania doświadczonych pracowników firmy badawczej). Innymi słowy, autor usunął dane po ich zebraniu w oparciu o kryteria ustalone post-hoc (nie przed zebraniem danych jak twierdzi pisząc o kryteriach *a priori*, a po ich zebraniu i wcześniejszej analizie), stosując zupełnie różne kryteria. Działanie takie nosi znamiona jednej z wątpliwych praktyk badawczych (QRP), mianowicie P-hackingu, który polega na wielokrotnym testowaniu danych różnymi metodami statystycznymi do momentu uzyskania pożądanego wyniku. W praktyce p-hacking objawia się dodawaniem lub usuwaniem zmiennych kontrolnych oraz dzieleniem danych na podgrupy. Często stosuje się też wykluczanie wyników odstających bez metodologicznego uzasadnienia oraz zmianę punktów końcowych w trakcie prowadzenia analizy. Obejmuje arbitralne usuwanie "niepasujących" obserwacji, modyfikowanie wartości granicznych

dla outlierów, stosowanie różnych metod imputacji brakujących danych w zależności od ich wpływu na wynik czy też wybiórcze korygowanie błędów tylko w przypadkach niekorzystnych dla hipotezy. Często towarzyszy temu brak przejrzystej dokumentacji procesu czyszczenia lub przedstawianie post-hoc uzasadnień dla podjętych decyzji. Być może wzorując się na starszej literaturze stosującej tego typu metody, autor popełnia podobne błędy. Na szczęście dla nauki, od kilkunastu lat trwa rewolucja metodologiczna w psychologii (i nie tylko), która uzmysłowiła nam, że tak prowadzone badania generują niereplikowalne wyniki i tylko utrudniają zrozumienie studiowanego zjawiska. Warto zatem zadbać o lepsze planowanie i analizy danych, najlepiej w formie prerejestracji.

Innym przykładem działań podobnych do opisanych powyżej QRP jest podzielenie decyzji ryzykownych na te w domenie zysków i strat jako dwie odrębne, jakościowo różne kategorie i przeprowadzenie wszystkich analiz oddzielnie na każdej z tych decyzji używając MANOVA. Działanie takie jest nieuzasadnione teoretycznie (w końcu są to decyzje analogiczne matematycznie). Przecież to właśnie porównania skłonności do ryzyka w domenie zysków i strat były podwaliną stworzenia teorii perspektywy, która była jedną z podstaw uzyskania Nagrody Nobla przez Kahnemana.

W pracy mamy także do czynienia z arbitralnym dzieleniem badania na grupy i podgrupy i dokonywaniem analiz w ich obrębie. Czasem autor testuje znaczenie płci decydenta (łącznie grupy z rywalem kobietą i rywalem mężczyzną w jedną grupę), a innym razem testuje znaczenie płci rywala (łącznie decydentów kobiety i mężczyzn w jedną grupę). Analizowanie poszczególnych zmiennych w osobnych modelach, zamiast włączenia wszystkich predyktorów do jednego modelu, jest problematyczne z kilku powodów metodologicznych. Po pierwsze, może prowadzić do przeszacowania wielkości efektów poprzez nieuwzględnienie współdzielonej wariancji między predyktorami (Cohen i in., 2003). Po drugie, wielokrotne testowanie tej samej hipotezy na tym samym zbiorze danych zwiększa ryzyko błędu I rodzaju ($\alpha = 5\%$) (Tabachnick & Fidell, 2013). Po trzecie, takie podejście uniemożliwia wykrycie potencjalnych interakcji między zmiennymi, które mogą być kluczowe dla pełnego zrozumienia badanego zjawiska (Maxwell & Delaney, 2004). Konsekwencją przyjętego podejścia jest niemożność oszacowania względnego wkładu poszczególnych predyktorów przy kontroli pozostałych zmiennych. W rezultacie raportowane efekty mogą być zawyżone, a istotne zależności między zmiennymi - przeoczone. Rekomendowanym podejściem byłoby włączenie wszystkich teoretycznie istotnych predyktorów do jednego modelu, co pozwoliłoby na bardziej rygorystyczną kontrolę błędu I rodzaju oraz umożliwiłoby testowanie interakcji między zmiennymi. Zatem na przykład analiza płci decydenta i rywala jest wykonywana zawsze oddzielnie, a przecież, jak sam autor wspomina, kluczowa jest tu interakcja między tymi czynnikami – ludzie inaczej rywalizują z osobami swojej płci, a inaczej z osobami płci przeciwnej.

Konsekwencja przyjętego podejścia analitycznego jest fakt, że trzy opisywane badania analizowane są w 75 tabelach, z których wiele przedstawia analizy kilkunastu, a nawet kilkudziesięciu testów statystycznych. Na przykład sama tabela 31 zawiera analizy 7 podskal testu DOSPERT x 2 domeny (zysk/strata) x 3 obecności rywala (brak/kobieta/mężczyzna) = 42 analizy. To podnosi ekstremalnie poziom błędu I rodzaju. Autor raportuje, że 3 z tych 42 testów są istotne (7% analiz, mniej więcej tyle, co przewiduje przyjęty błąd I rodzaju).

Autor, co ciekawe, analizuje kolejność wykonywania zadań w modelach dotyczących czasów reakcji, ale nie podejmowania decyzji. Są one istotne statystycznie. Dlaczego zatem nie są używane do wyjaśniania decyzji?

I kolejne pytanie – dlaczego autor mierzy zaangażowanie zasobów oraz decyzje, nie decydując się jednak na umieszczenie ich w jednym modelu? Byłoby ciekawe, czy osoby, które dłużej decydują, podejmują zarazem inne decyzje. Bo jeśli zaangażowanie nie prowadzi do innych decyzji, to oznaczałoby, że jest redundantne (mało prawdopodobne) albo źle zmierzone (bardziej prawdopodobne).


Co ciekawe, według Westfala & Yarkoniego (2016), kontrola statystyczna zmiennych zakłócających poprzez ich włączenie jako kowariancji może prowadzić do zafałszowania prawdziwych relacji między zmiennymi, zamiast je "oczyszczać", szczególnie gdy te kowariancje nie były teoretycznie uzasadnione przed analizą danych.

Autor dokonuje dekompozycji efektów moderacji bez podania ogólnej istotności modelu ani istotności dekomponowanej interakcji. Robiąc to i np. obserwując istotny efekt w jednej z 5 podgrup moderatora w Badaniu 3 twierdzi, że zaobserwował istotną interakcję. Otóż nie, sama różnica w istotności (jak pisałem już wielokrotnie w poprzednich recenzjach) nie stanowi dowodu na różnice między efektami. Odrębny test porównuje efekt w podgrupach i wskazuje, czy efekty się różnią (tj. zachodzi moderacja).

Ostatnim elementem są statystyki bayesowskie, które autor stosuje bez niezbędnej analizy rozkładu prior. Rozkład prior to założenie o wielkości efektu, jaki można zaobserwować przy braku prawdziwego testowanego efektu. Domyślny prior (0.707) jest często krytykowany jako uprzedzony wobec częstych w psychologii małych efektów. Przy okazji należy zaznaczyć, że wbrew temu, co pisze autor pod tabelą 12, BayesFactor nie określa szans na prawdziwość hipotezy, a relatywną szansę zaobserwowania danych takich jak zebrane lub bardziej ekstremalnych przy prawdziwości dwóch porównywanych hipotez.

Podsumowując, przedstawiona rozprawa doktorska zawiera szereg fundamentalnych błędów metodologicznych i statystycznych, które podważają wiarygodność prezentowanych wyników i wniosków. Główne problemy obejmują: niewłaściwą operacjonalizację zmiennych, niespójne kryteria analizy danych, nieuzasadnione teoretycznie podziały analiz oraz błędne podejście do interpretacji efektów moderacyjnych. Szczególnie niepokojące jest ignorowanie współczesnych standardów metodologicznych w psychologii, w tym brak prerejestracji badań i stosowanie praktyk zwiększających ryzyko fałszywych odkryć. Autor nie odniósł się konstruktywnie do wcześniejszych uwag recenzentów, a wprowadzone zmiany poza częścią teoretyczną rozprawy mają charakter jedynie kosmetyczny. W związku z powyższym podtrzymuję swoją negatywną ocenę rozprawy i rekomendację o odmowie nadania stopnia doktora.

Dr hab. Michał Białek, prof. UW.

Zakład Psychologii Zarządzania
KIEROWNIK

dr hab. Michał Białek, prof. UW