

Prof. dr hab. Tomasz Zaleśkiewicz
Wydział Psychologii we Wrocławiu
SWPS Uniwersytet Humanistycznospołeczny

**Recenzja rozprawy doktorskiej „Analiza efektu odwrócenia preferencji w procesie
dyskontowania społecznego” autorstwa pana Jana Ruska**

Podstawowym celem recenzowanego projektu doktorskiego było zaplanowanie i przeprowadzenie serii trzech badań, których wyniki miały się przyczynić do lepszego zrozumienia efektu dyskontowania społecznego. O ile problematyka dyskontowania nagród w czasie od wielu lat wzbudza olbrzymie zainteresowanie badaczy, to badania nad dyskontowaniem społecznym są zdecydowanie rzadziej spotykane. Zatem już na wstępie moje recenzji chcę pochwalić Autora za wybór tematyki, która, co także ma znaczenie, nie tak łatwo poddaje się empirycznemu testowaniu.

Pracę otwiera ciekawa i spójna część teoretyczna. Autor starannie wyjaśnia istotę efektu dyskontowania, koncentrując się na różnicach między różnymi kontekstami (np. dyskontowanie nagród w czasie vs. dyskontowanie społeczne). Zaimponowało mi to, że we wprowadzeniu przywoływane są nie tylko prace najnowsze, ale również te starsze, które, mimo upływu lat, mogą być rozpatrywane jako solidna baza teoretyczna dla stawiania hipotez w aktualnie prowadzonych badaniach. Podczas lektury napotykamy na odwołania do klasyków psychologii i ekonomii – Kurta Lewina czy Irvinga Fishera. Analizując różnice między wartością a użytecznością, Autor cytuje też Jeremy’ego Benthama – jednego z twórców filozoficznej koncepcji utilitaryzmu – którego dokonania są w ostatnich latach coraz częściej przywoływane w pracach psychologów zajmujących się problematyką oceniania i podejmowania decyzji (np. Kahneman, Wakker i Sarin).

Tym, co pozostawia pewien niedosyt, jeśli chodzi o zawartość wprowadzenia teoretycznego, jest mało precyzyjne zarysowanie własnego modelu teoretycznego. Nie jest dla mnie jasne, jak koncepcja teoretyczna Autora prezentuje się w odniesieniu do teorii, które są już znane i opublikowane; które elementy są tożsame, a które mają charakter odróżniający. Tego rodzaju prezentacja pozwoliłaby na to, by lepiej ocenić oryginalny wkład propozycji teoretycznej Autora w aktualną wiedzę o naturze dyskontowania społecznego i,

szerzej, podejmowania przez ludzi decyzji w kontekście społecznym. Gdybym był na miejscu Autora recenzowanej pracy, spróbowałbym zaproponować jakiś mechanizm psychologiczny, który jest odpowiedzialny za istnienie efektu dyskutowania społecznego i, w szczególności, efekt odwrócenia preferencji. Badania eksperymentalne mogłyby wtedy być zaplanowane w taki sposób, aby nie tylko testować istnienie obu powyższych efektów, ale także siłę mechanizmu psychologicznego (poznawczego, emocjonalnego, i/lub społecznego), który za owe efekty odpowiada. Nowoczesne badania psychologiczne stawiają sobie za cel nie tylko poszukiwanie i weryfikowanie zależności między zmiennymi, ale także odnajdywanie mediatorów i moderatorów tej zależności.

Jeśli chodzi o projekt empiryczny, to można o nim powiedzieć sporo dobrego, ale też rzucają się w oczy pewne wady, które obniżają jakość całego przedsięwzięcia. Zaczę od zalet. Niewątpliwie, Autor włożył sporo wysiłku w to, aby starannie przygotować zarówno materiały badawcze, jak i procedury poszczególnych eksperymentów. Dobrze zaplanowano operacjonalizację zmiennych, co w badaniach tego typu ma duże znaczenie. Dobrym rozwiązaniem było, w moim przekonaniu, zastosowanie procedur komputerowych, choć wydaje się, że nie wykorzystano w pełni ich możliwości. Na przykład, jeśli dobrze zrozumiałem wyjaśnienia przedstawione w opisach metod, losowe przypisywanie osób do warunków eksperymentalnych było wykonywane „ręcznie”, a nie w sposób zautomatyzowany, który byłby dużo korzystniejszy. Niestety, jak już wyżej wspomniałem, nie udało się uniknąć usterek metodologicznych, które w niektórych miejscach są dosyć rażące. Omówię je krótko z podziałem na poszczególne badania.

Badanie 1

Zarówno w tym badaniu, jak i w dwóch pozostałych badaniach, poważne wątpliwości budzi sposób doboru próby. Jest to kwestia niebanalna, do której obecnie przywiązuje się olbrzymią wagę, przede wszystkim w kontekście tzw. kryzysu replikacyjnego. Autor podaje w przypadku każdego badania, jaka była liczebność próby, ale nic nie wiadomo o tym, jak ona została określona. Podejrzewam, że niestety nie zastosowano żadnej metody, a jeśli się myślę, to mam nadzieję, że Autor wyjaśni tę kwestię podczas obrony. Podana jest też informacja, że kilka osób zostało usuniętych „z powodu niespełnienia kryterium wieku”. Jakże konkretnie to było kryterium? Na jakiej podstawie je zastosowano? Zwracam na to uwagę, ponieważ kwestia usuwania *ex post* osób badanych z prób również wzbudza obecnie duże kontrowersje. Autor nie prerejestrował swoich badań, więc w tym kontekście usuwanie osób

już po zakończeniu badania jest mocno problematyczne. Jeśli dobrze zrozumiałem, osoby badane otrzymywały rekompensatę w postaci bonu do sieci empik, ale jeśli z jakichś powodów nie ukończyły badania, to tej rekompensaty nie mogły odebrać. Otóż, wg mnie, takie postępowanie nie jest właściwe i budzi wątpliwości z powodów etycznych. Uczestnicy powinni otrzymywać tzw. „show-up fee” niezależnie od tego, czy ukończyli badanie, czy nie. Jeśli przyznanie nagrody za uczestnictwo jest uzależnione od tego, czy ktoś weźmie udział w całej procedurze, czy nie, to może dochodzić do niestaranego wypełniania zadań badawczych, co znacząco obniża jakość danych.

Kolejne uwagi dotyczą instrukcji badawczych i konstrukcji eksperymentu. Obrazek w załączniku 6 zawiera instrukcję dla osób badanych, z której wynika, że dopóki decyzja nie była zatwierdzona, mogła być wielokrotnie zmieniana. Czy ta kwestia była w jakikolwiek sposób rejestrowana i kontrolowana? Nie wiem, czy program umożliwiał zapisywanie tego, w którym podejściu ktoś dokonywał ostatecznego wyboru, ale jeśli tak, to warto do tego wrócić w analizach. Na str. 64 mówi się, że na ekranie prezentowano „instrukcję dotyczącą tworzenia w wyobraźni listy 100 znajomych osób”. Jaka to była instrukcja? Czy jakoś zdefiniowano to, jak rozumieć określenie „znajoma osoba”? A skoro już o tym piszę, to chciałbym zatrzymać się przy tym zadaniu nieco dłużej. Generowanie listy 100 znajomych osób wydaje mi się bardzo trudnym i obciążającym poznawczo zadaniem, o ile w ogóle jest ono wykonalne. A co by się stało, gdybyśmy poprosili osoby badane o wykonanie znacznie prostszego zadania, polegającego na stworzeniu w myślach listy 10 znanych sobie osób, co wydaje się bardziej realne? Czy wtedy wyniki były takie same? Co ciekawe, Autor zwraca uwagę na ten problem w sekcji „Ograniczenia”, co świadczy o tym, że w jakimś stopniu podziela moje wątpliwości, a mimo to konsekwentnie stosuje to zadanie w swoich kolejnych badaniach.

Bardzo wątpliwym elementem badania 1 jest zastosowana manipulacja poziomem reprezentacji. Osoby „w warunkach wysokiego konkretności” (str. 67) miały za zadanie pomyśleć o tym, kim konkretnie jest przywołana z pamięci osoba. A jaką instrukcją otrzymywały osoby w drugim warunku? Jeśli dobrze zrozumiałem, to tam żadnej instrukcji nie było. Skąd zatem mamy wiedzieć, w jaki sposób w tym drugim warunku osoby badane myślały o przypominanych sobie znajomych? Jaką mamy pewność, że również nie myślały o nich w sposób konkretny – szczegółowy? Sprawę dodatkowo komplikuje to, że nie zastosowano kontroli manipulacji. To jest, wg mnie, poważna usterka metodologiczna, która niezwykle

utrudnia interpretowanie wyników. Zatem wyniki, które w pracy przypisuje się działaniu manipulacji, są wysoce niepewne i bezwzględnie wymagają one powtórzenia.

Badanie 2

Tutaj można by powtórzyć wszystkie wątpliwości dotyczące sposobu doboru próby, które już dotychczas opisałem. Jak sądzę, planując badanie 2, Autor znał już wyniki badania 1, więc mógł je bez problemu wykorzystać do określenia wielkości próby w badaniu 2. Szkoda, że tego nie zrobiono, bo ponownie pojawiają się wątpliwości dotyczące jakości próby badawczej.

Ale w tym kontekście pojawia się problem jeszcze poważniejszy. Na str. 84 Autor pisze, że część osób brała udział w badaniu w warunkach laboratoryjnych (osobisty kontakt z eksperymentatorem), a druga część uczestniczyła w nim zdalnie. To jest niestety dość poważny błąd metodologiczny i Autor zdaje sobie z tego sprawę, bo analizuje tę kwestię w rozdziale o ograniczeniach projektu. Jeśli nawet dzieje się tak, że z ważnych powodów warunki badania nie są jednakowe dla wszystkich, to należy bezwzględnie uwzględnić to w analizach, kontrolując, czy struktura wyników utrzymuje się na tym samym poziomie. To jednak nie jest jedyny problem związany z organizacją badania 2. Jedną z hipotez odnosi się do różnic w odpowiedziach między osobami studiującymi i niestudującymi. Otóż sytuacja jest taka, że w grupie badania stacjonarnego mamy 36 osób studiujących i 3 osoby niestudujące, podczas gdy w pomiarze online proporcje było odwrotne – 38 osób studiujących i 80 osób niestudujących. Jeżeli w analizie zostaną odkryte różnice między tymi grupami, to skąd mamy wiedzieć, czy odpowiada za nie fakt studiowania lub nie studiowania, czy też sposób, w jaki ktoś brał udział w badaniu? Dwa opisane tu przeze mnie problemy powodują, że wiarygodność wyników badania 2 stoi pod dużym znakiem zapytania.

Badanie 3

Kolejny raz powtarza się problem z doбором próby badawczej – jej wielkość nie jest w żaden sposób uzasadniona. Co więcej, liczebność próby jest wyraźnie mniejsza niż w dwóch poprzednich badaniach. Nie jest też jasne, dlaczego Autor ponownie wraca do organizowania badania z udziałem tylko studentów, skoro opisując poprzednie badanie przedstawiał argumenty, że próba powinna obejmować także te osoby, które nie studiują.

To badanie jest też najslabiej i najmniej precyzyjnie opisane. Wątpliwości budzą postawione hipotezy. Na przykład, w Hipotezie 1a są postawione dwie tezy jednocześnie, co powoduje trudności w jej weryfikowaniu. Nie jest dla mnie jasna hipoteza 1b, przede

wszystkim w jej punktach (ii) oraz (iii); mam wrażenie, że czegoś brakuje w sformułowaniu tych elementów hipotezy

W opisie narzędzi badawczych uwagę zwraca to, że badanie prowadziły co najmniej dwie osoby, a w dodatku te osoby różniły się płcią. W pracy nie ma informacji na temat tego, czy sprawdzono ewentualny efekt płaci eksperymentatora. Tutaj też w sposób jawny mówi się o tym, że to eksperymentator(ka) decydował(a) o przydziale osób badanych i że ten proces nie był zautomatyzowany. Raz jeszcze powtórzę, że to nie jest zbyt fortunne rozwiązanie, tym bardziej, że w procedurze komputerowej taki automatyczny przydział mógł być łatwo zaimplementowany.

W końcu, nie jest dla mnie jasne, czemu służyła ostatnia część badania, w której osoba badana dokonywała wyboru między natychmiastowymi lub odroczonymi w czasie nagrodami dla siebie. W planie teoretycznym całego projektu nie były formułowane oczekiwania na temat tego, jakiego rodzaju zależności mogą zachodzić między dyskontowaniem społecznym a dyskontowaniem nagród w czasie. Dlatego można mieć wrażenie (a przynajmniej ja mam takie wrażenie), że ta część badania 3 pojawiła się nieco przypadkowo.

Abstrahując od szczegółowych uwag metodologicznych dotyczących poszczególnych badań, chciałbym jeszcze zwrócić uwagę na dwie kwestie ogólniejsze, które dotyczą całego projektu empirycznego.

Po pierwsze, we wszystkich badaniach wykorzystano wypłaty hipotetyczne (ta kwestia jest też omówiona przez Autora w ograniczeniach badań). Szkoda, że w choćby jednym badaniu Autor nie pokusił się o to, by wprowadzić wypłaty rzeczywiste. Nawet jeśli badacz dysponuje ograniczonym budżetem i wypłaty są stosunkowo wysokie, to można zastosować rozwiązanie, w którym jest losowana jakaś część uczestników, otrzymujących rzeczywiste wynagrodzenie. W nowoczesnych badaniach z pogranicza psychologii eksperymentalnej i ekonomii behawioralnej zdecydowanie odchodzi się od rozwiązania, w którym wszystkie wypłaty są hipotetyczne.

Po drugie, znaczna część badań była prowadzona w warunkach laboratoryjnych, w których osoby badane miały bezpośredni kontakt z eksperymentatorami (lub eksperymentatorkami). Jak Autor wyjaśnia w opisach poszczególnych badań, to eksperymentatorzy wyjaśniali osobom badanym zasady obowiązujące w badaniu i

odpowiadali na ewentualne pytania osób badanych. Brakuje informacji o tym, kim byli eksperymetatorzy. Czy był spełniony warunek tzw. „ślepych eksperymetatorów”. To kolejna kwestia metodologiczna, do której przywiązuje się bardzo dużą wagę w kontekście kryzysu replikacyjnego.

W końcu, chciałbym też zwrócić uwagę na pewne niedociągnięcia ogólnej dyskusji projektu. Ma ona raczej charakter rozszerzonego podsumowania wyników niż solidnej debaty teoretycznej. Byłbym na przykład bardzo ciekaw tego, jak Autor spostrzega zastosowany paradygmat badawczy w kontekście innych paradygmatów, w których ludzie dokonują podziału zasobów między siebie a inne osoby (np. gra Dyktator). Jak zachowania w zadaniach testujących dyskutowanie społeczne korelują z zachowaniami w innych zadaniach mierzących preferencje społeczne?

Podsumowując, nie ukrywam, że miałem pewien dylemat z końcową oceną tego projektu, gdyż nagromadzenie różnych mniejszych lub większych niedociągnięć metodologicznych i teoretycznych jest spore. Natomiast w obronie Autora trzeba powiedzieć, że wielu tych usterek jest On świadomy i otwarcie pisze o nich albo w podsumowaniach poszczególnych badań, albo w sekcji zbierającej ograniczenia całego projektu. To odważna postawa Autora i z całą pewnością zasługuje ona na odnotowanie. Nawet jeśli mam wątpliwości związane z jakością opisywanych efektów, to nie zmieniam swojej opinii, że jest to ciekawy i oryginalny projekt. Warto by był on kontynuowany, przy czym niezbędne wydaje się dużo staranniejsze przygotowanie metodologiczne prowadzonych pomiarów.

Ważąc wszystkie wady i zalety pracy, stwierdzam, że zasługuje ona na ocenę pozytywną. Wnioskuje zatem do Rady Dyscypliny o dopuszczenie pana mgr. Jana Ruska do dalszych etapów postępowania o nadanie stopnia doktora.

14.07.2022
Tomasz Zalewski