

Warszawa, 3 stycznia 2025 r.

Michał Krawczyk
Wydział Nauk Ekonomicznych UW
ul. Długa 44/50, 00-241 Warszawa

Recenzja rozprawy doktorskiej pt. „Decision Making and Numeracy: The Role of Context in Adaptive Strategy Selection” autor: Supratik Mondal

Ocena ogólna

Praca doktorska składa się z trzech powiązanych ze sobą publikacji naukowych. Jedna z nich stanowi autorskie powtórzenie wcześniejszego badania, opublikowane w polskim czasopiśmie, pozostałe dwie są oryginalnymi badaniami, współtworzonymi przez doradcę i opublikowanymi w znanych międzynarodowych czasopismach.

Praca wywołuje u mnie mieszane uczucia. Postawione pytania są interesujące, biorąc pod uwagę wymienioną literaturę. Analiza ekonometryczna jest nowoczesna i jasno przedstawiona. Z drugiej strony, kilka zagadnień pozostaje niejasnych. Poniżej wymieniam kilka kwestii, które uważam za problematyczne (choć moja ogólna ocena jest raczej pozytywna).

Struktura pracy doktorskiej

Dla mnie podstawą pracy doktorskiej są artykuły. Ponieważ są one publikowane, są (lub: powinny być) samowystarczalne. Jednak tutaj są one poprzedzone streszczeniem, „zarysem ogólnym” czy „podsumowaniem”. W efekcie występuje wiele niepotrzebnych powtórzeń. Co więcej, w spisie treści można łatwo przeoczyć właściwe artykuły — są one jedynie podrozdziałami rozdziału 10, wedle bibliografii (?).

Streszczenie

Streszczenie jest dość długie i niezbyt ekscytujące, ponieważ zawiera między innymi definicje (powtórzone później w dwóch innych miejscach).

Wprowadzenie i przegląd teoretyczny

Ta część nieco odbiega od tematu. Moim zdaniem, powiązania między umiejętnością liczenia, szachami i sztuką abstrakcyjną są w najlepszym wypadku słabo przedstawione. Podobnie trudno jest określić, jaki jest związek między akapitami na s. 14. Bardzo zaskakujący jest fakt, że „maximization of expected utility” i „feelings” są przykładami „actors” (s. 11). Nie jest jasne, ile wynosi jednopunktowy przyrost wyniku w zakresie umiejętności liczenia (s. 11).

W dalszej części omówię kluczowe aspekty tych trzech artykułów.

Ramy teoretyczne

Artykuły 1 i 2 wprowadzają kluczowe rozróżnienie między warunkami tzw. „high payoff” oraz „low payoff” oraz sytuacjami wyboru. Intuicyjnie wybór pomiędzy 50% szansą na 50 euro i 51% szansą na 49 euro można intuicyjnie określić jako warunki „high payoff” (ponieważ kwoty są wyższe niż zwykle w krótkich eksperymentach), podczas gdy tutaj byłyby to warunki „(very) low payoff”, ponieważ różnica w oczekiwanych wartościach jest niewielka. Czytając Artykuł 1 mam wrażenie, że lepiej byłoby nazwać „low payoff choices” wyborami „low stakes” lub „small valuation difference choices” lub, po prostu, „hard choices” i postąpić analogicznie dla „high payoff choices”. Czytelnik jest również zaskoczony odkryciem (choć jest to ukryte w aneksie), że uczestnicy średnio poświęcali mniej czasu na dokonanie wyboru w przypadku wyborów w warunkach „high payoff”. Naturalnie, dłużej się zastanawiamy przy podejmowaniu decyzji w sytuacji, gdy stawki są wysokie (np. przy wyborze pracy), niż gdy są niskie (np. przy wyborze przekąski). Nie byłoby jednak zaskakujące, że wyborów dokonujemy szybciej, gdy *różnice w ocenie* są wysokie, tj. jedna opcja jest znacznie lepsza od drugiej, a więc wybór jest łatwy (jedna praca jest znacznie lepsza od drugiej) niż wtedy, gdy jest trudny (obie opcje mają swoje zalety). Jest to logicznym odpowiednikiem wartości samej opłacalności (payoffs) — wybór między przekąskami może być równie trudny, co łatwy. Istnieje obszerna literatura, np. wykorzystująca modele DDM (Drift Diffusion Models), na temat tego, dlaczego wyborów dokonuje się przez dłuższy czas, gdy decydent jest bardziej obojętny. Istnieją artykuły, które stwierdzają, że zazwyczaj poświęcamy zbyt dużo czasu na przypadki, w których obie opcje są prawie równie dobre (np. Oud, B., Krajbich, I., Miller, K., Cheong, J. H., Botvinick, M., oraz Fehr, E. (2016). Irrational time allocation in decision-making. *Proceedings of the Royal Society B: Biological Sciences*, 283(1822), 20151439.); w tym kontekście naturalną hipotezą jest, że osoby o wysokich zdolnościach numerycznych lepiej rozpoznają tę pułapkę i zamiast tego poświęcają swoją uwagę przypadkom, w których ma to znaczenie. Interesujące byłoby wymodelowanie tego w kontekście DDM (a może już to zrobiono?). Ogólnie rzecz biorąc, wydaje się, że nieoptymalny wybór określić nie tylko dezorientuje czytelnika, ale także wprowadza Kandydata w błąd podczas wyszukiwania literatury.

W Artykule 2 Autorzy wyraźnie rozważają dwa wymiary: bezwzględną różnicę EV (AED) i względną różnicę EV (RED), kontrolując każdy z nich. Jest to wyraźna poprawa metodologiczna (która w rzeczywistości pokazuje, że wniosek z Artykułu 1 był przedwczesny). Jednak Kandydat nadal nazywa niski poziom różnicy RED „low payoff”, a wysoki RED „high payoff”. W tym projekcie staje się to jeszcze bardziej mylące, ponieważ aby utrzymać AED na tym samym poziomie, wybory „low payoff” muszą w rzeczywistości wiązać się ze znacznie większymi EV — EV muszą być wysokie, aby dany AED był stosunkowo mały. Jestem również zdezorientowany ostatnim akapitem sekcji 2.2 drugiego artykułu (nawiasem mówiąc, numeracja stron, która byłaby spójna w całej rozprawie, ułatwiłaby życie recenzentowi). Autorzy wydają się sugerować, że AED i trudność to różne pojęcia, podczas gdy w przypisie 1 przypisują, że — przynajmniej w ich próbie — to właśnie niski poziom AED sprawia, że niektóre wybory są trudniejsze.

Kluczowe pojęcie „adaptive choices” zostało przedstawione w nieco zagmatwany i sprzeczny sposób. Informuje się nas, że „a group of decisions can be considered „adaptive” when, as a collective, it outperforms the existing normative prediction or matches it without utilizing a similar amount of resources”. W przypadku decyzji obarczonych ryzykiem, wiążących się z niskimi nagrodami, maksymalizacja EV jest najsilniejszym przewidywaniem normatywnym (i dlatego Kandydat ma rację, używając go w całej rozprawie), a z tego samego powodu określenie „outperform” oznaczałoby większe

oczekiwania. Przekroczenie maksymalizacji EV oznaczałoby zatem zarobienie, zgodnie z oczekiwaniami, więcej niż jest to możliwe — a to bardziej przypomina sztukę magii niż sztukę adaptacji („adaptability”). Dla kontrastu, w sekcji 10.2 (ponownie, brakuje numerów stron) stwierdzono, że „participants’ choices would be considered adaptive if participants, on average, followed the EV consistent strategy in the high-payoff condition and changed their strategy to EV inconsistent choices in the low-payoff condition”. Jest to dość nieprecyzyjne, ale wydaje się sugerować, że adaptacyjny decydent musi spodziewać się zarabiać MNIEJ niż decydent maksymalizujący EV (ponieważ nie są oni konsekwentni podczas niektórych wyborów). Staje się to bardziej jasne, gdy Kandydat omawia ograniczenia czasowe. Punktem odniesienia przy definiowaniu „adaptability” nie jest zatem teoretyczna, natychmiastowa osoba maksymalizująca EV, ale rzeczywista powolna (i niedoskonała?) osoba maksymalizująca EV. Być może powinno to zostać wyraźnie zaznaczone przy definiowaniu „adaptability”. To jednak sprawia, że definicja staje się jeszcze bardziej rozmyta, ponieważ najpierw wymaga definicji operacyjnej (rzeczywistej, powolnej, niedoskonałej) osoby maksymalizującej EV jako punktu odniesienia.

Autorzy twierdzą, że nie udało im się „find any statistical difference in the prediction between the [models]”, odsyłając czytelnika do rysunku 1 w aneksie, który pokazuje udział problemów wyboru, w których dana para modeli daje identyczne przewidywania, z pewnego rodzaju wąsami (nie powiedziano nam, co one reprezentują, ale można by podejrzewać, że jakiś rodzaj przedziałów ufności). Nie mam również pewności, co ma w tym przypadku oznaczać „[no] statistical difference”. Podejrzewam, że jeśli zdefiniujemy stałość_{XY} zmiennych binarnych przyjmując wartość 1, jeśli model X i model Y dają takie same przewidywania dla danego problemu wyboru oraz wartość 0 w przeciwnym razie, to średnia stałość_{XY} nie różni się statystycznie od 1. Jest z tym więcej niż jeden problem. Po pierwsze, innym sposobem na powiedzenie tego jest to, że podczas próby, przewidywania modeli X i Y zawsze się pokrywają, co dla niektórych par jest najwyraźniej *nieprawdą*. Po drugie, i być może ważniejsze, uważam, że jakiegokolwiek pojęcie istotności statystycznej i przedziałów ufności jest tutaj bez znaczenia. O ile mi wiadomo, 72 problemy wyboru nie są losową próbą z żadnej dobrze zdefiniowanej populacji. Istotność statystyczna i przedziały ufności opisują naszą wiedzę na temat populacji macierzystej lub parametrów pewnego rozkładu prawdopodobieństwa w oparciu o wylosowaną próbę. Nie mają one zatem znaczenia dla omawianej sprawy.

Czasami Kandydat wydaje się przedstawiać wyniki tak, jakby były one ogólne, podczas gdy w rzeczywistości (moim zdaniem) zależą one dramatycznie od bodźców użytych w danym badaniu. Na przykład w trzecim artykule, s. 3, Kandydat pisze: „For instance, after drawing just one sample, Hertwig and Pleskac (2008) demonstrated that the probability of choosing an option with a higher EV was approximately 60%. Drawing an additional five samples increased accuracy by 18%, while further increments in the number of draws — from five to 10 samples and from 10 to 20 samples — improved accuracy by 6% and 4%, respectively.”. — To odkrycie wydaje się być bardzo specyficzne dla (prawdopodobnie: arbitralnego) wyboru parametrów „hazardowych” użytych w tym badaniu.

Metody empiryczne

Kandydat intensywnie korzysta z Berlin Numeracy Test, zapewniając „reprezentatywny” przykład: „Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws, how many times would this five-sided die show an odd number?” — to pytanie mówi mi więcej o umiejętnościach matematycznych twórcy testu niż o jego uczestnikach. Nie sądzę nawet, by możliwe było

skonstruowanie prawdziwej pięciościennej kostki do gry, ponieważ, jak mówi nam matematyka w szkole średniej, nie istnieje pięciościenna bryła platońska (wypukły wielościan foremny). Dlaczego gotowość do przyjmowania matematycznie bezsensownych założeń określa się mianem „numeracy”? Podczas gdy tysiące cytatów, które otrzymała ta metoda, stanowią pewne uzasadnienie dla kolejnego uczonego (tutaj: Kandydata), który ją stosuje, czy nie można znaleźć mniej mylącego testu?

W rzeczywistości nie jest to jedyne ważne założenie, które zostało tutaj pominięte: Kandydat wydaje się zakładać, że boki są ponumerowane od 1 do 5. W rzeczywistości nie wszystkie (faktycznie istniejące) kości n-ścienne mają boki ponumerowane od 1 do n. Na przykład, 10-ścienne kości są często ponumerowane od 00 do 90. Tutaj dokument źródłowy (Cokely i in.) faktycznie pomocnie dodaje, że chodzi o liczby „1, 3 lub 5”, jednak Kandydat to zignorował, nie wiem, dlaczego.

Mogłoby się wydawać, że są to błahe kwestie, ponieważ większość uczestników z łatwością przyjmie domyślne założenie, że pięć boków jest jednakowo ponumerowanych od 1 do 5. Jednak dla niektórych może to brzmieć jak moneta o trzech stronach — dobry powód, by stać się rozproszonym, zdezorientowanym, nieufnym lub zirytowanym. W eksperymentach z udziałem ludzi (zwłaszcza online, gdy nie mogą oni zadawać pytań i są nieustannie kuszeni, by jednocześnie grać w Fortnite lub sprawdzać swoje konto na Facebooku) szczegóły mają znaczenie, więc ostrożność jest wskazana.

W Artykułach 1 i 2 naturalne wydaje się przypuszczenie, że czas reakcji pośredniczy we wpływie interakcji BNT-payoff na wybór: w warunkach high-payoff uczestnicy z wysokim BNT decydują się na stosunkowo intensywnie myślenie i to właśnie sprawia, że wybierają opcję z wyższym EV. Tak więc, gdy RT jest uwzględniony w modelu, bezpośredni efekt zostaje zmniejszony. Dlaczego nie zostało to wyraźnie przedstawione i przetestowane?

„In future studies, I intend to conduct further experiments with between-participant design to establish the causal relationship between adaptive behavior and numeracy” — Nie mam pojęcia, w jaki sposób zastosowanie projektu między uczestnikami samo w sobie pozwoliłoby zidentyfikować taki efekt przyczynowo-skutkowy.

Nie wiem, jak przetestować normalność między grupami (tabela 8 z 10.1.1). Być może myli mnie tylko podpis tabeli.

Rozumiem, że główna różnica między eksperymentami 1 i 2 w Artykule 2 polega na tym, że w pierwszym (późniejszym) zastosowano projekt międzyprzedmiotowy. Plusy i minusy każdego z tych rozwiązań są dobrze znane oraz wydaje się naturalne, aby wyjaśnić, w jaki sposób mają one zastosowanie w tej konkretnej dziedzinie. Natomiast motywacją Autorów wydaje się być „non-inferiority test” (badanie równoważności). Nie znam tego podejścia, być może dlatego, że zdecydowana większość jego zastosowań wydaje się dotyczyć nauk medycznych. Wydaje się to dość naturalne, aby upewnić się, że nowy lek lub procedura nie jest, w najgorszym przypadku, znacząco (lub: niedopuszczalnie) gorsza od już istniejącej metody leczenia. Podejrzewam, że przynajmniej w niektórych zastosowaniach medycznych istnieje rozsądny konsensus co do tego, co jest niedopuszczalnie gorsze, co stanowi pewne wsparcie dla tego podejścia. Dlaczego mielibyśmy tutaj myśleć w kategoriach równoważności jednych warunków eksperymentalnych z innymi? I czy poziomy „acceptable difference” nie są tutaj całkowicie losowe? Ostatecznie takie podejście wydaje się niewiele wносить do interpretacji wyników. W przeciwieństwie do tego, zwiększa to dezorientację czytelnika, gdy napotyka on zwroty takie jak „mean EV consistency in the low-payoff condition is not non-inferior to the mean EV consistency in the high payoff condition”. Wydaje się, że jest to bardzo skomplikowany sposób powiedzenia, że, co nie jest zaskakujące, odsetek osób wybierających lepszą z dwóch opcji jest wyższy, gdy różnica między opcjami jest duża.

Bardzo problematycznym aspektem analizy empirycznej w niniejszej rozprawie jest wykorzystanie rejestracji wstępnych. Oczywiście świetnie, że Kandydat w ogóle zdecydował się z nich skorzystać. Niestety, sposób ich wykorzystania pozostawia wiele do życzenia.

Po pierwsze, kluczowym aspektem rejestracji wstępnej jest ograniczenie swobody autorów w wyborze modeli i hipotez. Chodzi o to, aby rygorystycznie trzymali się oni wcześniej zarejestrowanego planu analizy danych, a jeśli kiedykolwiek od niego odstąpią lub go rozszerzą, wyraźnie ostrzegali czytelnika i wyjaśniali, dlaczego to robią. W tym przypadku Kandydat nie wskazuje, które analizy zostały wstępnie zarejestrowane, a które nie. Czytelnik powinien wtedy założyć, że Kandydat skrupulatnie trzyma się planu analizy, bez zmian i dodatków. Jednak, o ile mi wiadomo, plan analizy danych dla Artykułu 2 nie określa „acceptable differences” w badaniu równoważności i w ogóle nie wspomina o technice („interwałach”) Johnsa-Neymana. Wydaje się więc, że ważne decyzje dotyczące analizy danych zostały podjęte po ich zebraniu. Mimo to czytelnik jest przekonany, że jest inaczej.

Po drugie, w Artykule 3 powiedziano nam, że „The analysis of qualitative data was preregistered (Mondal & Traczyk, 2023b).” W bibliografii wpis ten zawiera link <https://doi.org/10.17605/OSF.IO/U59FC>, który wydaje się prowadzić donikąd (doi foundation error 404). W jaki sposób link do rejestracji wstępnej może wygasnąć w ciągu kilku miesięcy od publikacji? Wydaje się również, że badanie 2 z artykułu 3 nie zostało wstępnie zarejestrowane, chociaż Autorzy nie wspominają o tym w artykule.

Ogólnie rzecz biorąc, wstępnie zarejestrowany plan analizy, który jest a) częściowy b) pozornie tylko częściowo dostępny dla czytelnika c) zmodyfikowany w rzeczywistym badaniu bez ostrzeżenia jest gorszy niż brak wstępnej rejestracji w ogóle — daje mniej dociekliwemu czytelnikowi całkowicie fałszywy obraz.

W eksperymencie 2 w Artykule 2 punkty zdobyte w dwóch warunkach mają różne wartości. Jednak autorzy nie biorą tego pod uwagę, tak jakby zakładali, że ich badanym zależy na punktach per se, a nie na pieniądzu.

Jak wyjaśniają Autorzy: „As a bonus, in the Low-payoff condition, participants were told that for every 1000 points, they would receive an additional 0.80 GBP on top of the flat fee. However, in the High-payoff condition, participants were told that for every 500 points, they would receive an additional 0.65 GBP on top of the flat fee”. Z literatury jasno wynika, że aby ułatwić porównywanie wskaźników (np. prawdopodobieństwa), należy zachować wspólne mianowniki. Na przykład, znacznie łatwiej jest rozpoznać, że 16/100 jest większe niż 15/100, niż zdać sobie sprawę, że 8/50 jest większe niż 15/100. Rozumiem, że wartości pieniężne punktów nie zostały tutaj wyraźnie porównane w różnych warunkach, ale nadal nie jest jasne, dlaczego zostały one przedstawione inaczej.

Co gorsza, porównując 11.2 i 11.3 w 10.2.1, widzimy, że uczestnikom powiedziano wprost (...za każde 500 punktów...) i za pomocą przykładu (...więc, jeśli zbierzesz 1500 punktów...), ile wart jest punkt w warunkach „high-payoff”. Dla kontrastu, w warunkach low-payoff — podany jest tylko jeden przykład. Wygląda to na poważny błąd, sprawiający, że oba warunki różnią się w dodatkowy, nieplanowany sposób.

Co więcej, nie jest jasne, czy, powiedzmy, w warunkach high-payoff, zdobycie 400 punktów oznaczało a) zarobienie 0 (ponieważ $400 < 500$), czy b) zarobienie $400/500$ razy 0,65. Podobnie, nie jest jasne, czy punkt 1500, zmieniający 1499 na 1500, reprezentuje przyrost a) 0,65 lub b) tylko $0,65/500$ itd. Przykłady nie są w tym zakresie pomocne. W zakresie, w jakim niektórzy uczestnicy faworyzowali

interpretację a), mamy dwa potencjalne problemy: 1) w zależności od liczby punktów, które (jak sądzą) zgromadzili, uczestnicy mogą mieć motywację do ryzyka lub jego unikania; na przykład, jeśli uważam, że mam do tej pory około 300 punktów i pozostało mi tylko kilka gier, mogę szukać ryzyka, aby mieć szansę na osiągnięcie 500 punktów 2) różne mianowniki (/500 punktów w warunkach high-payoff i /1000 punktów w warunkach low-payoff) implikują kolejną nieplanowaną strategiczną różnicę między warunkami (ponownie, z dodatkowym zastrzeżeniem, że uczestnikom w warunkach low-payoff nie powiedziano nawet wyraźnie, jaki był ich licznik, więc kto wie, co myśleli).

Ogólnie rzecz biorąc, można odnieść wrażenie, że nie tylko Autorzy przyjęli zaskakujące założenie, że uczestnicy nie dbają o wartość pieniężną punktów, ale także sami Autorzy nie przejmowali się tym zbyt, co skutkowało pewnym zaniedbaniem w tym zakresie.

Można również zauważyć, że w instrukcjach użytych w Artykule 2, uczestnikom powiedziano, że będą musieli stawiać czoła „gambles” (ryzykowne zdarzenia) z dwiema opcjami. Według mojej najlepszej oceny, nazywanie wyboru między dwoma ryzykownymi opcjami (two gambles) „ryzykownym zdarzeniem” (a „gamble”) jest potencjalnie bardzo mylące. (Podobnie w Artykule 3, uczestnikom mówi się, że będą musieli zmierzyć się z „decision problems in the form of lotteries”; i rzeczywiście w trakcie instrukcji zachęca się ich do „wybierania” z puli jednej [sic] loterii). Ponownie, są to drobne szczegóły, ale podczas przygotowywania bodźców eksperymentalnych szczegóły mają znaczenie i nie można ich naprawić w ramach powtórzenia (chyba że zostanie przeprowadzone nowe badanie).

Nie wiem, jaka jest wartość dodana badania symulacyjnego 10.3. Autorzy piszą: „we found that the “blue” decision maker would face 30 choice problems, given that they are following a more energy-intensive EV maximization strategy 75% of the time. However, the “orange” decision maker would earn more reward, compared to the “blue” decision maker, by facing 10 (i.e., 40 choice problems in total) problems more by randomly choosing between options (i.e., EV consistency of 50%)”. — Jeśli się nie mylę, pierwsze „odkrycie” było założeniem, a drugie można łatwo obliczyć na podstawie założeń (parametrów zdarzeń), bez symulowania 60 000 obserwacji.

Nie jest dla mnie jasne, dlaczego Autorzy są zainteresowani „współczynnikiem zmiany” — dlaczego jest to interesujące i dlaczego spodziewamy się, że bardziej numeryczne osoby będą rzadziej się zmieniać?

Przejrzystość

W rozprawie jest sporo stwierdzeń, które trudno mi było rozszyfrować. Oto niektóre z nich.

„More numerate individuals were better at maximizing expected reward following the frame of reference of absolute difference (as captured by the absolute expected value difference distribution) and not the relative difference in reward (as captured by both payoff conditions)”. (s. 2) jest dla mnie niejasne.

„elaborate and in-depth processing of information, utilizing heuristics (Cokely & Kelley, 2009)” (s. 12) brzmi dość zaskakująco. Określenia „in-depth” oraz „heuristics” rzadko idą ze sobą w parze.

„While heuristics can lead to biases and errors, more often than not, they often approximate optimal choices sufficiently well and sometimes even outperform them”. (s. 13) Z definicji niemożliwe jest osiągnięcie lepszych wyników niż optymalny wybór...

Heurystyka nie wydaje się być precyzyjnie zdefiniowana, gdy jest używana po raz pierwszy. Skąd mogę wiedzieć, czy dana różnica „jest znacząca”?

Co to znaczy, że „participants ... on average maximized EV” (s. 5 z 10.1; reszta tego zdania też nie jest dla mnie jasna).

CCA jest naszkicowane tylko na stronie 15 w 10.1. Dlaczego w Tabeli 1 czytelnik musi zgadywać, co oznacza payoff=1 lub choice=1? Tabela 2 jest dla mnie bardzo myląca. Czy rzeczywiście test dla wysokiego BNT różnił się od testu dla niskiego BNT, jak sugeruje tytuł tabeli? Dlaczego? Dlaczego notatka sugeruje coś innego?

Jak definiuje się „long/medium/short” RT? (s. 16 z 10.1)

Dlaczego w 10.1 wyniki modeli 1 i 2 nie są przedstawione razem?

Rysunek 4 to kompletna pomyłka. Zły rendering pliku .pdf przez program OCR?

Materiały uzupełniające 10.1.1 często wprowadzały mnie w błąd. Co dokładnie oznacza, że „CPT/EV choices...” równa się, powiedzmy, .2?

10.2

„the presence of the two conditions [low-payoff and high-payoff I suppose] together” — jak mogą one występować „together”?

W 3.1.2. Autorzy stwierdzają, że „followed the same principle” i „control[led] as many factors as possible (i.e. variance, AED)” bezpośrednio po akapicie krytykującym (ich własne) poprzednie artykuły, które nie kontrolowały AED.

Rozumiem, że „restricting them to 100 and 50, respectively” oznacza, że AED nie może dać więcej niż 50 punktów. W jaki sposób to ograniczenie powoduje (jak sugerowałoby określenie „as a consequence”), że „participants will earn ... less if they [make] EV inconsistent choices”?

Jeśli autorzy zrównoważyli rozmieszczenie opcji tak, że „option A is not option A for each participant” (sic), jaki jest sens sprawdzania, czy opcja A była w jakimkolwiek wymiarze znacząco różna od opcji B?

Panel B Rysunku 1: jak to możliwe, że „possible earning” związane z wyborami zgodnymi z EV w warunkach low-payoff są niższe niż w przypadku wyborów niezgodnych z EV w tych samych warunkach?

Nie rozumiem Rysunku 3. Te proste linie nie mogą bezpośrednio reprezentować danych empirycznych. Przypuszczam, że są to najlepsze dopasowania liniowe, ale nie zostało to wyjaśnione.

Mam również duże trudności ze zrozumieniem dyskusji na temat hipotezy malejącej wrażliwości na wartość. Powiedziano mi, że „as the psychological interpretations of the law of diminishing returns dictate, preferences become more inconsistent when the difference in value is relatively similar”. Podobna do czego? Co ważniejsze, co prosty fakt, że prawdopodobieństwo wyboru lepszej opcji spada, gdy opcje stają się bardziej do siebie podobne (zakładam, że o to tutaj chodzi), ma wspólnego z prawem malejących przychodów (znanym również jako prawo malejącej produktywności krańcowej) lub jego, nieujawnioną, „interpretacją psychologiczną”? Założenie niemal dowolnego procesu stochastycznego w modelu sprawia, że błędy są bardziej prawdopodobne, gdy różnica w użyteczności jest mniejsza. I nawet jeśli przez „malejące przychody” Kandydat rozumie malejącą użyteczność krańcową (pieniądza?), to nadal nie jest dla mnie jasne, w jaki sposób jest to konieczne w tym przypadku. Kandydat twierdzi następnie, że „when the absolute difference between two offers is 50\$ and the relative difference is small (i.e., option A is 400\$ and option B is 450\$), preference for the higher value item is relatively less consistent and in accordance with the hypothesis of diminishing value sensitivity”. Czy „diminishing value sensitivity” oznacza to samo, co „diminishing returns”? Czy preferowanie

wyższej wartości jest „in accordance”? A może fakt, że jest „relatively less consistent” jest „in accordance”? Czy mówimy tutaj o wartościach oczekiwanych? Ponieważ mam pewność, że wybór 450 \$ ponad 400 \$ (bez innych różnic między opcjami) będzie wyjątkowo „consistent” (konsekwentny). Ogólnie rzecz biorąc, uważam ten akapit za bardzo mylący.

Nie wiem również, dlaczego wspomniane są poziomy aspiracji (w tej samej sekcji). Nie wydaje się, abyśmy dowiedzieli się o nich cokolwiek z tego eksperymentu.

10.2.1

Przypis 1: dlaczego była potrzebna symulacja? Dlaczego po prostu nie zsumować EV lepszej opcji we wszystkich problemach wyboru?

W podrozdziale 5: Co oznacza „we divided 1.6 GBP based on reward earned following EV consistent choices” — podzielone przez co? Skąd w tym samym akapicie wzięły się liczby 20 i 12?

Na rysunku 1 w aneksie, co ma oznaczać „Similarly between different models in being plotted.”?

10.3

Nie rozumiem, co przedstawia Rysunek 1.

Mogę tylko zgadywać, co oznacza „binned relative block/trial”.

Dlaczego mielibyśmy „speculate that skilled decision makers in the loss or mixed domains in a tie-constrained environments should make slower decisions, ultimately preventing them from making any choice at all”? To bez wątpienia jedynie spekulacje.

Język

Ogólnie rzecz biorąc, tekst jest dość dobrze napisany. Od czasu do czasu Kandydat ulega słowotokowi. Na przykład:

na s. 1, określenie „In the real world” jest zbędne. Czy czytelnik byłby zainteresowany fałszywym światem?

na s. 8. „essential” wystarczy, „most essential” to zdecydowanie za dużo.

Obejmuje to przypadki tautologii:

„Also, we counterbalanced... as well”; „continue the study further” (oba w 10.2).

Oto kilka innych, pomniejszych kwestii:

na s. 9, nie wiem, co oznacza określenie „efficient mathematics”

na s. 12, określenie „variance... between” jest niejasne

Na s. 5 z 10.1: „data collection was done online by appointing 73 participants” (??)

Na s. 7 z 10.1: „Every moment of our life is bombarded with information condensed in a statistical shell (??)

„a statistician... would continuously buy lottery ticket year after year...” — coś tu naprawdę poszło nie tak.

„financial decision-making in the last decade” — prawdopodobnie chodziło o to, że dowody zostały zgromadzone w ciągu ostatniej dekady.

„CV1 is a synthetic predictor variable consist of...” (s. 15 z 10.1).

„lastly it is assumed that each participant to have varied...” (s. 18 z 10.1).

Podsumowanie i wnioski

Podsumowując, istnieje wiele kwestii, które wymagają wyjaśnienia. Być może trzy z nich są najważniejsze:

- 1) sposób przeliczania punktów na pieniądze (i wiedza uczestników na ten temat)
- 2) które analizy były dokładnie zgodne z rejestracją wstępną, a które nie i dlaczego
- 3) w jaki sposób bardziej uważne spojrzenie na literaturę modeli DDM wpływa na interpretację wyników i wkład

Chociaż mogę mieć tendencję do skupiania się na słabościach, dostrzegam również, jak wspomniano wcześniej, ambitne plany, interesujące pytania badawcze i wysokiej jakości analizę danych. Oczywiście fakt, że wszystkie trzy badania zostały opublikowane, również świadczy o ich jakości. Dlatego uważam, że Kandydat powinien mieć możliwość wyjaśnienia pozostałych kwestii podczas obrony. Dla porządku podaję wniosek formalny w języku polskim:

Recenzowana rozprawa spełnia warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2018 r. poz. 1668). Rozprawa stanowi oryginalne rozwiązanie problemu naukowego, prezentuje ogólną wiedzę teoretyczną Kandydata w dyscyplinie Psychologia. Wnioskuje o dopuszczenie mgr. Supratika Mondala do dalszych etapów przewodu doktorskiego.