

prof. dr hab. Adam Chuderski
Zakład Kognitywistyki
Instytut Filozofii
Uniwersytet Jagielloński w Krakowie



Recenzja rozprawy doktorskiej p. mgr Angeliki Olszewskiej
pt. „Wgląd to coś więcej niż rozwiązywanie zadań. Wykorzystanie wskaźników subiektywnego
doświadczenia do pomiaru wglądu i opracowania nowego narzędzia”.

Przedstawiona mi do recenzji rozprawa doktorska ma formę maszynopisu (tzw. pracy pisemnej) o objętości 98 stron. Dodatkowe materiały, w tym użyte narzędzia badawcze oraz fragmenty szczegółowych analiz wyników, zostały umieszczone w repozytorium OSF. Rozprawa stanowi element postępowania o nadanie stopnia doktora p. mgr Angelice Olszewskiej, prowadzonego w Uniwersytecie Humanistycznospołecznym SWPS na podstawie Ustawy prawo o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 (z późn. zm.), w szczególności Art. 187 tejże Ustawy.

Rozprawa poświęcona jest badaniu w psychologii już od przynajmniej wieku zjawisku wglądu. Jest ono zazwyczaj rozumiane jako nagłe i niespodziewane zrozumienie pewnej idei, która przed momentem wglądu była całkowicie dla podmiotu niedostępna (czyli odmiennie niż w przypadkach stopniowego odkrywania, nabywania wiedzy itp.), a po uzyskaniu wglądu stała się niemal trywialna, czemu często towarzyszy poczucie pewności i przyjemności. Zatem wgląd zazwyczaj charakteryzowany jest łącznie przez komponent poznawczy (nowa idea w umyśle) i subiektywny (metapoznawczy i/lub emocjonalny). Dwukrotnie piszę tu „zazwyczaj”, gdyż w literaturze przedmiotu brak jest precyzyjnej definicji tego konstruktu, na co słusznie wskazuje i czemu próbuje zaradzić sama Autorka.

Jak deklaruje Autorka pracy, ogólnym celem rozprawy jest „rozwinięcie wiedzy na temat wglądu i rozwiązanie niektórych problemów związanych z tym zjawiskiem”. Cele szczegółowe pracy, ponumerowane oraz dobrze wyeksplikowane w pracy, są relatywnie heterogeniczne, przez co stanowi ona raczej zbiór trzech względnie niezależnych badań aniżeli jeden spójny ciąg zabiegów badawczych. Tym, co łączy owe trzy badania, jest konceptualizacja wglądu dokonana w części 1 oraz pomiar subiektywnego doświadczenia, dokonany w części 2. Ze względu na taki heterogeniczny charakter pracy nie zawiera ona jednej głównej konkluzji, lecz daje kilka odpowiedzi na nieco bardziej szczegółowe i konkretne pytania badawcze, podsumowane w części 3.

Pracę rozpoczyna przytoczenie dziesięciu definicji wglądu obecnych w literaturze. Autorka na ich podstawie proponuje siedem tzw. twierdzeń bazowych opisujących wgląd, które mogą zostać poddane weryfikacji empirycznej. Do pięciu z nich Autorka odnosi się – mniej lub bardziej – w swoich badaniach. Następnie autorka omawia dotychczasowe badania leżące u podłoża każdego z siedmiu twierdzeń bazowych. Pierwszą część pracy kończy przedstawieniem testów typowo angażujących wgląd oraz miar jego subiektywnego doświadczenia, które używane są w tym obszarze badań.

Druga część pracy przedstawia trzy badania własne Autorki. W pierwszym badaniu, przeprowadzonym na dwóch osobnych próbkach osób badanych: studentach przebadanych tradycyjnie oraz ochotnikach z platformy Prolific, podjęto się replikacji badania innych autorów (Laukonnen i Tangen, 2017), w którym grupa, której prezentowano figury bistabilne, poprawnie rozwiązała więcej werbalnych problemów wglądowych (typowych dla tego obszaru badań), w porównaniu do grupy, której prezentowano figury monostabilne. Poszerzenie oryginalnego badania polegało na zastosowaniu po każdym z zadań skal Likerta mierzących cztery wymiary doświadczenia: nagłość, przyjemność, pewność i doświadczenie aha!. W badaniu Autorki, przeciwnie do oryginalnego badania, grupy nie różniły się poprawnością wykonania zadań w żadnej z dwu próbek. Jedyna istotna różnica grupowa dotyczyła wyższej oceny doświadczenia aha! w grupie z figurą bistabilną w drugiej próbce osób badanych.

Mam kilka uwag do prezentacji tych wyników oraz poczynionych analiz. Na przykład, jako że procedura badania była identyczna w obu próbkach za wyjątkiem sposobu dostępu do osób badanych, można było w analizie połączyć obie próbki, a jedynie dodać do modelu zmienną kodującą próbkę i sprawdzić, czy nie będzie istotnych interakcji z tą zmienną. Takie połączenie próbek podniosło by moc testów.

Nazwanie poprawności rozwiązania zadań „trafnością” nie jest dobrym pomysłem, gdyż trafność jest ugruntowanym pojęciem dotyczącym własności narzędzi psychometrycznych i inne jego użycie wprowadza w konfuzję. Nie rozumiem, dlaczego nie można było po prostu pisać o „poprawności”.

Czy nie jest dziwne, że osoby obserwujące figurę jednostabilną średnio ok. 4 razy deklarowały przełączenie się na inny percept, jeżeli, jak zaznacza sama Autorka na str. 36, tylko jeden percept był możliwy? W próbce B aż 34 osoby z tej grupy (na 100) deklarowały więcej niż 10 przełączeń, z czego jedna miała ich aż 210 (naciskała spację częściej niż co 500 ms). Te dane nie wyglądają na miarodajne.

A co z trafnością wskaźników subiektywnych, jeżeli osoba badana nie udzieliła odpowiedzi w danej zagadce? Przy braku odpowiedzi trudno w ogóle zinterpretować te wskaźniki (czego dotyczyły?) oraz zapewne były one niższe, niż po podaniu rozwiązania (rozwiązanie nie mogło być nagłe, skoro w ogóle nie nadeszło, trudno też było mieć wobec niego poczucie pewności). Sugeruje to dodatnia korelacja między indywidualną poprawnością rozwiązywania zadań a średnią wartością każdego ze wskaźników (za wyjątkiem doświadczenia aha! w drugiej z próbek). Wiadomo także, że rozwiązaniom

niepoprawnym towarzyszem niższe wskaźniki doświadczenia wglądu (Chuderski i in., 2021; Webb i in., 2017 – obie prace cytowane w rozprawie). Autorka powinna była przeprowadzić także analizę wariancji oraz analizę korelacji dla wskaźników subiektywnych jedynie dla poprawnych rozwiązań (tylko badając 8 osób nie podało żadnego poprawnego rozwiązania i te osoby trzeba by z konieczności pominąć). Takie analizy byłyby bardziej zasadne, niż wrzucanie do jednego worka wskaźników zarejestrowanych po odpowiedziach poprawnych, odpowiedziach niepoprawnych oraz przy braku odpowiedzi.

Zupełnie nie potrafię zrozumieć następującego zdania w dyskusji wyników (str. 40): „Kolejnym ograniczeniem, które z kolei wynika z interpretacji samych autorów jest wykorzystanie efektu Grattona do potencjalnej stymulacji szybszego wyszukania niedominującej reprezentacji poznawczej problemu.” W obecnej pracy nie manipulowano przecież sekwencyjnym następstwem konfliktów, zwykle prowadzącym do efektu Grattona. Jak ten efekt, typowy dla takich zadań jak test Stroopa czy flankerów, miałby być zaimplementowany czy uzyskany w badaniu wglądu? Nawet po, wspomnianej w tekście, zmianie zagadek na wzrokowe czy przestrzenne wciąż nie będzie miejsca w tym badaniu na wystąpienie efektu Grattona. Laukonnen i Tangen faktycznie wspominają o efekcie Grattona, ale w ich wywodzie pełni on rolę teoretycznego uzasadnienia badania, podczas gdy nigdzie w swojej pracy nie proponują, aby go wywoływać/mierzyć. Manipulację eksperymentalną stanowi tam kostka Neckera.

Szczerze mówiąc, nie jestem zdziwiony, że replikacja wyniku Laukonnen i Tangena dot. poprawności rozwiązania zadań wglądowych nie powiodła się. Istotność ich efektu była bliska kryterium $\alpha = 0,05$ (ich $p = 0,034$), a jak podaje rysunek 3 siła tego efektu była niewielka ($d \approx 0,3$) i jej 95% przedział ufności obejmował zero. W badaniu Autorki (1B) użyto także ponad dwukrotnie większej próby (198 vs 80). Może, w świetle kryzysu replikacyjnego w psychologii (i nie tylko), po prostu wynik Laukonnen i Tangena był przypadkowy i efekt kostki Neckera nie istnieje. Niestety, w literaturze nt. wglądu takich wątpliwych efektów publikuje się sporo i ja sam poświęciłem sporo swojego czasu na ich obalanie. Tym bardziej należy wyrazić uznanie dla wysiłku Autorki, aby tenże efekt niezależnie sprawdzić.

Kończąc omówienie badania 1, chciałbym zauważyć, że następujący fragment ze streszczenia pracy: „... w ramach rozszerzenia w swoim badaniu dokonałam również pomiaru doświadczenia wglądu i zaobserwowałam, że wcześniejsze obserwowanie figur dwuznacznych sprzyjało silniejszemu doświadczeniu olśnienia” niestety pomija fakt, że Laukonnen i Tangen także zastosowali pomiar subiektywnego doświadczenia, pytając osoby badane czy uzyskały one rozwiązanie problemu poprzez wgląd, analizę, czy w inny sposób (i też zauważyli wzrost raportowanego wglądu po pokazaniu kostki Neckera). Ich sposób pomiaru subiektywnego doświadczenia jest uznaną metodą pomiaru wglądu, więc rozszerzenie badania przez Autorkę ma charakter jedynie ilościowy (zwiększenie liczby miar oraz wykorzystanie skal zamiast pojedynczego wyboru opcji), ale nie jakościowy.

Badanie drugie także stanowiło rozszerzoną replikację badania innych autorów (Bar-Hillel i in., 2019). Podobnie jak badanie pierwsze, przeprowadzono je tradycyjnie na studentach oraz przy wykorzystaniu platformy online. W oryginalnym badaniu tzw. Test Refleksyjnego Myślenia (CRT) korelował silniej z zadaniami wglądowymi niż z tzw. odległymi skojarzeniami, mimo iż to te dwa ostatnie testy uważa się za angażujące wgląd. Autorka w swoim badaniu zamieniła zadania wglądowe na te użyte w badaniu pierwszym, dodała test sylogizmów, werbalną wersję CRT (oryginalna wersja wymaga obliczeń numerycznych) oraz dwa inne narzędzia (jednakże bez szerszego uzasadnienia ich wprowadzenia do badania). Po każdej z jednostek testów stosowała takie same skale Likerta jak w badaniu 1.

Autorka przedyskutowała także możliwe wyjaśnienie wyników Bar-Hillel i in. odwołując się do detekcji konfliktu pomiędzy myśleniem intuicyjnym a myśleniem analitycznym, zaproponowanych przez Keitha Stanovicha (por. także prace Johnatana Evansa). Po pierwsze, nie jest to jednak całkowicie trafne przedstawienie najnowszej wersji teorii Stanovicha, ponieważ w jego teorii to procesy typu 3 (umysł refleksyjny, racjonalny), a nie procesy typu 2 (ze względu na konotacje metafizyczne, Stanovich stara się nie używać terminów System 1 i System 2 używanych przez Autorkę) odpowiadają za detekcję i rozwiązywanie konfliktu pomiędzy procesami typu 1 (tzw. umysłem autonomicznym, intuicyjnym, heurystycznym) a właśnie procesami typu 2 (umysłem algorytmicznym, analitycznym). Chyba najlepsze umówienie tej teorii można znaleźć w pracy Evansa i Stanovicha (*Perspectives on Psychological Science*, 8, 223-241). Po drugie, postulowanie coraz większej liczby typów procesów oraz słabe wsparcie empiryczne tej teorii stało się powodem silnej jej krytyki (Kruglański, 2016, *PPS*, 8, 242-287; Osman, 2004, *PBR*, 11, 988-1010, 2016, *PPS*, 8, 248-252; Ritchie, 2017, *Intelligence*, 61, 46), którą należało by przytoczyć w pracy, aby prezentacja teorii Stanovicha nie była li tylko jednostronna.

W badaniu 2A nie potwierdziła się różnica w sile korelacji CRT-zadania wglądowe w porównaniu do korelacji odległe skojarzenia-zadania wglądowe dla oryginalnego (numerycznego) CRT, ale różnica taka wystąpiła dla werbalnej wersji CRT.

W badaniu 2B już obie wersje CRT korelowały istotnie silniej z zadaniami wglądowymi niż z odległymi skojarzeniami. Porównano także wskaźniki subiektywne dot. wglądu między zadaniami wglądowymi, odległymi skojarzeniami, obiema wersjami CRT oraz sylogizmami zawierającymi niezgodność logicznej poprawności konkluzji z wiedzą potoczną.

Słabszą korelację zagadek wglądowych z odległymi skojarzeniami niż tych pierwszych z CRT można łatwo wyjaśnić na trzy sposoby, bez odwoływania się do mglistych wyjaśnień w kategoriach detekcji konfliktu. Po pierwsze, jak wspomina Autorka, CRT po prostu zawiera zadania wglądowe używane od dziesiętków lat w badaniach nad wglądem. „Autor” CRT (Frederick) jedynie zmienił interpretację tego, co niby mierzą te zadania (refleksyjność). Jego interpretacja nie jest niestety trafna, co pokazała

niedawno meta-analiza badań z wykorzystaniem CRT (Otero i in., 2022, *Intelligence*, 90, 101614): CRT jest najsilniej związany z inteligencją ogólną oraz zdolnościami numerycznymi, podczas gdy nie ma istotnego związku z refleksyjnością. Szkoda, że Autorka przeoczyła tę meta-analizę, bo jest ona kluczowa dla interpretacji wyników badania 2. Podobne wyniki przyniosło zresztą moje niedawne badanie (Jastrzębski i Chuderski, 2022, *Intelligence*, 95, 101705). Zatem, CRT silnie koreluje z zadaniami wglądowymi, bo (i) po prostu ich trafność fasadowa jest zbieżna (jednostki obu tych testów są do siebie semantycznie i konstrukcyjnie podobne; wspomina o tym Autorka), a także (ii) oba silnie korelują z inteligencją. Z kolei nie bardzo wiadomo co dokładnie mierzą odległe skojarzenia, bo faktycznie słabo korelują z innymi miarami myślenia dywergencyjnego i konwergencyjnego (por. Lee i in., 2014, *Psychology of Aesthetics, Creativity, and the Arts*, 8, 446-460). Ponadto, jak pokazaliśmy z Janem Jastrzębskim na dużym zbiorze danych (Chuderski i Jastrzębski, 2018, *Insight: The Origin of New Ideas*, Routledge – rozdział cytowany w rozprawie), korelacja odległych skojarzeń z inteligencją płynną ($r = 0,36$) jest słabsza niż korelacja problemów wglądowych z inteligencją płynną ($r = 0,58$). Odległe skojarzenia słabiej korelują z zadaniami wglądowymi niż CRT, bo po prostu (i) mają odmienną trafność fasadową oraz (ii) są słabiej ładowane przez czynnik inteligencji ogólnej, w porównaniu do zadań wglądowych. Zapewne ma także znaczenie to, że CRT i zadania wglądowe angażują myślenie, a odległe skojarzenia – pamięć (tzw. asymetria Bunswicka; Wittmann & Süß, 1999, w Ackerman, Kyllonen i Roberts (red.), *Learning and Individual Differences: Process, Trait and Content Determinants*, APA).

W odniesieniu do wskaźników subiektywnych w badaniu 2B, z powyższych powodów (zbieżna trafność fasadowa) nie powinno dziwić, że zadania wglądowe oraz CRT prowadziły do podobnych poziomów doświadczenia aha!, nagłości i przyjemności. Faktycznie, jest nieco zaskakujące, że werbalne CRT oraz odległe skojarzenia przyniosły wyższe wartości tych wskaźników niż zadania wglądowe, ale mimo istotności statystycznej, różnica bezwzględna jest niewielka (3-4 punkty na 25 możliwych). Wydaje mi się, że wyciągnięcie z tego badania konkluzyjnych wniosków na temat różnic w poziomie wskaźników może być trudne ze względu na małą liczbę zadań per test (5) oraz ich idiosynkratyczność.

Trzecie z badań wprowadziło nowe narzędzie do wywoływania subiektywnego poczucia wglądu, nazwane percepcyjnym zadaniem wglądowym. W tym zadaniu obraz pewnego przedmiotu stopniowo przechodził albo z bardzo nieostrego do ostrego, albo z całkowicie zakrytego do odkrytego. Zwykle dopiero po określonym wyostrzeniu/odkryciu obrazu osoby badane były w stanie rozpoznać przedmiot na obrazie – Autorka zamierzała sprawdzić, czy osoby te przeżyją w takim momencie doświadczenie aha! oraz poczucie nagłości porównywalne ze wskaźnikami w zadaniach wglądowych i odległych skojarzeniach, a wyższymi niż wskaźniki w analogiach i matrycach. Badanie 3A służyło walidacji zadania. Pozwoliło ono ustalić, że stopniowe odkrywanie obrazu prowadzi do wyższych wartości wskaźników subiektywnych niż wyostrażanie obrazu (jedynie dla wskaźnika pewności relacja ta była przeciwna).

Badanie na drugiej próbie (3B), która niestety w tekście w ogóle nie została scharakteryzowana, porównywało wskaźniki nagłośności oraz doświadczenia aha! pomiędzy percepcyjnym zadaniem wglądowym a innymi zadaniami potencjalnie wymagającymi albo nie wymagającymi wglądu, wymienionymi powyżej. Wyniki wskazywały na faktycznie względnie wysoki poziom wskaźników subiektywnych dla percepcyjnego zadania wglądowego, które były wyższe niż w analogiach i matrycach, a porównywalne do odległych skojarzeń i problemów zapalczanych (nieistotnie niższe).

W tabelach 18 i 19 występuje błąd dla pozycji zagadki i RAT, gdyż ich różnica względem PIT jest ujemna, podczas gdy powinna być dodatnia (podobnie jak dla pozycji ICAR, VA i PMT). Brakuje rysunku 11. Tabela 18 jest niespójna z rysunkiem 10, bo różnice między PIT a ICAR, PMT i VA zaznaczone jako istotne w tabeli nie są zaznaczone jako istotne na rysunku (jeżeli dobrze rozumiem, kropkami).

Test ekwiwalentności pochodzący z pakietu TOSTER nie jest najlepszym narzędziem do wykazywania braku różnic między dwiema średnimi, za względu na jego arbitralność oraz założenie rozkładu normalnego dla wyliczanych przedziałów ufności. Powszechnie uznaną miarą braku efektu jest uzyskanie znaczącego czynnika Bayesa dla hipotezy zerowej (np. $BF_{01} > 3$ lub jeszcze lepiej > 5), który opiera się na faktycznym rozkładzie zmiennych i precyzyjnie kwantyfikuje wiarygodność hipotezy zerowej. Byłoby to łatwe do zrobienia, bo przecież Autorka do swoich analiz użyła bayesowskiego pakietu JASP. Tak samo można było sprawdzić w badaniu 1 czy efekt prezentacji kostki Neckera jest konkluzywnie zerowy czy niekonkluzywny (to ostatnie zachodzi, gdy BF_{01} jest między 0,5 a 2,0).

Jeżeli chodzi o interpretację wyników badania 3B, to mam wątpliwość, czy można łatwo przejść od nich do wniosku, że wgląd w zadaniu percepcyjnym nie wymaga restrukturyzacji poznawczej (str. 84). Autorka wydaje się zakładać, że percept stopniowo wyłania się w umyśle i jest od samego początku mniej lub (potem) bardziej kompletnym perceptem tego samego przedmiotu. Byłoby to założenie sprzeczne z podstawową wiedzą o mechanizmach percepcji, w szczególności z faktem, że percepcja polega na testowaniu hipotez, a nie bezpośrednim odbieraniu informacji (może za wyjątkiem percepcji w mocno ekologicznych kontekstach w sensie Gibsona), oraz z faktem, że ze względu na ograniczenia szerokości kątowej ostrego widzenia oraz pojemności wzrokowej pamięci krótkotrwałej większa część pola wzrokowego to symulacja umysłowa, a nie bezpośrednia informacja (por. prace takich badaczy spostrzegania jak Rensink, Lamme, Wolfe, Biederman, itd.). Dlaczego więc osoba badana na początku próby rozpoznania zakrytego przedmiotu nie mogła by przyjąć błędnej hipotezy nt. obiektu (auto, rura itp.), a dopiero wraz z dodatkowymi danymi odkryć, że to jest inny obiekt (aha! to szpula nici!)?

Uwagi techniczne i językowe:

Tekst jest w zasadzie napisany poprawnie językowo, choć liczba błędów interpunkcyjnych znacznie przekracza liczbę, której oczekuje się od tekstu naukowego. Zdarzają się także fragmenty zdań,

w których zapewne czegoś brakuje, bo fragmenty te nie mają sensu, np. na str. 64 „brak istotnych [?] z zadaniami niewymagającymi wglądu”, a na str. 8 mamy „asocjację skojarzeń” (tzw. masło maślane).

Nie chcę czepiać się wprowadzanych pojęć, ale przynajmniej jedno z nich, dość kluczowe dla rozprawy, budzi moją dużą wątpliwość. W kilku jej miejscach, w nawiązaniu do psychologii postaci, pisze się o „pełnej strukturze” i jej wyłanianiu się. W literaturze anglojęzycznej odpowiednim pojęciem wydaje się być ‘structured whole’ (niem. ‘strukturiertes Ganzes’), które oddaje się jako „ustrukturyzowana całość” (lub „zorganizowana całość”, np. w tłumaczeniu książki ‘Cognitive Psychology’ R. Sternberga), a nie „pełna struktura” (a jeżeli już, to lepiej „kompletna” albo „spójna”). W podobnym tonie wprowadza się także w rozprawie angielski termin ‘ill-defined problems’ jako „problemy z niepełną strukturą”, a przecież mamy od dawna używany termin „problemy źle określone”. Nie do końca zrozumiałem zatem czy „pełna struktura” to jakieś nowe pojęcie proponowane przez Autorkę, czy też kolejne tłumaczenie ‘structured whole’. Jeżeli to pierwsze, to nie zostało ono należycie zdefiniowane w rozprawie, jeżeli to drugie, to wydaje mi się niepoprawne, bo struktura to jedynie sposób organizacji całości, a nie sama całość. Całość, czyli układ, to dopiero struktura plus wypełniające ją elementy.

Zamiast „udowodnienie” proponowałbym używać „pokazanie”, „wsparcie”, czy „dostarczenie danych na rzecz”, bo w naukach empirycznych mało możliwe jest udowodnienie tezy w sposób niepodważalny.

Używanie jako zmiennych średnich zamiast sum wyników w danej skali/teście ułatwiłoby czytelnikowi ich interpretację zgodnie i opisem skal Likerta (tzn. używanie zakresu np. 1-5 zamiast 1-60).

Prezentacja wyników dla wskaźników doświadczenia jest niespójna pomiędzy badaniami 1 a 2 i 3, bo w badaniu 1 podano średnie, a w badaniu 2 i 3 – mediany. Jakie jest uzasadnienie takiej zmiany? Wydaje się, że średnie są bardziej właściwym wskaźnikiem, więc pozostałbym przy nich w całej pracy.

Niektóre wnioski i sugestie są niewystarczająco uargumentowane czy wyjaśnione. Na przykład, w jakim sensie we fragmencie „Badania pokazują, że subiektywne doświadczenie jest nawet ważniejsze” doświadczenie jest ważniejsze? Ważniejsze z jakiego/czyjzego punktu widzenia, wg jakiego kryterium?

Przechodząc do oceny rozprawy mgr Angeliki Olszewskiej, praca ta dotyczy niewątpliwie intrygującego i potencjalnie ważnego problemu badawczego, czyli tego, jakim zjawiskiem umysłowym jest subiektywnie doświadczany wgląd oraz jakie przełożenie ma to doświadczenie na rezultaty aktywności poznawczej (np. skuteczność w odnajdywaniu prawidłowych rozwiązań czy odpowiedzi). W mojej ocenie rozprawa ma mocniejsze i słabsze strony. Zacznę od omówienia tych drugich.

Przed wszystkim, przeprowadzenie badań replikujących wcześniej zastosowane procedury badawcze, z nieznacznym rozszerzeniem zakresu badania Laukonnena i Tangena w badaniu 1 oraz nieco większym

uzupełnieniem badania Bar-Hillel i in. w badaniu 2, chociaż samo w sobie jest oczywiście cenne (dobre replikacje są w nauce tak samo ważne jak nowe procedury), to jednakże nie do końca spełnia kryterium stawiane rozprawom doktorskim przez ustawodawcę, określone w art. 187 ust. 2 jako „oryginalne rozwiązanie problemu naukowego”. Oryginalność badania 1 wydaje mi się niewielka, szczególnie że nie podjęto wysiłku ustalenia, czy efekt kostki Neckera jest konkluzywnie zerowy czy niekonkluzywny, pomimo posiadania dużego zbioru danych (szczególnie po możliwym połączeniu wyników 1A z 1B) oraz istnieniu metody do tego służącej (czynnika Bayesa). Problematyczne jest także „przemilczenie” przez Autorkę istotnego efektu w zakresie raportów subiektywnych uzyskanego przez Laukonnena i Tangena. Badanie 2 charakteryzuje nieco wyższy poziom oryginalności, szczególnie związany z wskaźnikami doświadczenia, bo sama replikacja różnic w korelacji była względnie przewidywalna w świetle znanych wad testu odległych skojarzeń jako miary wglądu i do pewnego stopnia stanowiła wyważenie otwartych drzwi (np. wystarczyło zajrzeć do macierzy korelacji w moim, cytowanym w rozprawie, artykule, Chuderski i Jastrzębski, 2018, gdzie średnia korelacja problemów wglądowych z problemami numerycznymi pochodzącymi z CRT to $r = 0,42$, a z odległymi skojarzeniami to $r = 0,24$). Wyniki dotyczące wskaźników subiektywnych jednak uzupełniają z powodzeniem istniejącą literaturę.

Doceniając wysiłek włożony w części 1. pracy w zdefiniowanie konstruktów wglądu, mam jednak niedosyt teoretycznego wprowadzenia w to obszerne i wielowątkowe zagadnienie. Autorka pominęła wiele z kluczowych badań, stanowiących kamienie milowe w rozwoju teorii wglądu, oraz jedynie bardzo lakonicznie omówiła najpopularniejsze z tych teorii (Weisberg, Ohlsson) i to w taki sposób, że czytelnik nie będący ekspertem w tej dziedzinie zapewne i tak niewiele skorzysta. Po przeczytaniu tej części rozprawy odniosłem wrażenie, że najważniejsi badacze wglądu to Wiley, Webb, Danek, DeCaro i ... Chuderski (sic!), choć są to osoby badające wgląd od zaledwie dekady (no, Wiley może nieco dłużej), podczas gdy giganci tego obszaru badań pozostają poza rozprawą (np. Mednick, Davidson, Sternberg, Schooler, Sun, Ormerod, MacGregor, Beaty, Silvia itd.). Nawet jeżeli we wprowadzeniu pojawiają się jedna czy dwie prace poszczególnych autorów, to niekoniecznie są to ich kluczowe prace. Prezentację teorii, metod i wyników badań nad wglądem i restrukturyzacją uważam zatem za pobieżną.

Przydałoby się także choć krótko omówić relacje pomiędzy konstruktami wglądu a innymi ważnymi konstruktami w psychologii, takimi jak twórczość, kreatywność, metapoznanie, myślenie dywergencyjne i konwergencyjne, myślenie analityczne, a także bardziej podstawowymi konstruktami poznawczymi jak uwaga i pamięć, bo teraz ma się wrażenie jakby badania nad wglądem były samotną wyspą bez kontaktu z innymi obszarami badań psychologicznych, a to przecież nie ma miejsca.

Nie całkiem zadowalająca jest także dyskusja uzyskanych wyników, która w rozprawie sprowadza się do ich podsumowania w kontekście zakładanych hipotez, omówienia ograniczeń metodologicznych

badan własnych oraz zaproponowania kierunków przyszłych badan. Brak bardziej ogólnych wniosków na temat możliwych wyjaśnień teoretycznych uzyskanych wyników (jedyne takie odniesienie do Stanovicha uważam, jak wcześniej pisałem, akurat za mało szczęśliwe), a także wskazania tego, w jaki sposób wyniki te mogłyby się przyczynić do rozwoju teorii wglądu. W szczególności, w ostatnim zdaniu rozprawy Autorka pisze, że „uzyskane wyniki pozwalają lepiej zrozumieć zjawisko wglądu”, ale nie wyjaśnia, na czym dokładnie polega wkład pracy w owo lepsze zrozumienie tego zjawiska.

Przejdę teraz do mocniejszych stron pracy, które po krótkiej refleksji wydają się przeważać nad tymi słabszymi. Niewątpliwie najmocniejszą częścią pracy jest badanie 3. Pokazanie, że rozpoznaniu z początku nieoczywistego przedmiotu na obrazie towarzyszą wskaźniki subiektywne (nagłość i olśnienie) typowe dla bardziej złożonych problemów zwykle wymagających restrukturyzacji stanowi niewątpliwie oryginalny wkład Autorki w literaturę nt. wglądu. Tym samym, Autorka rozwiązała problem udoskonalenia pomiaru wglądu, pokazując, że można uzyskać jego subiektywne korelaty prostym i powtarzalnym zadaniem percepcyjnym (jeszcze prostszym, niż metoda sztuczek magicznych zaproponowana przez Danek). Wynik ten dokłada się do coraz liczniejszych obserwacji, że wgląd może stanowić względnie podstawowe poczucie metapoznawcze, odpowiadające nieoczekiwanej łatwości (płynności) przetwarzania informacji, a nie musi być – jak dotąd sądzono – związany jedynie z twórczymi, ważnymi, czy złożonymi ideami (wgląd traci swą magię).

Mocną stroną pracy jest rzetelne zaprojektowanie serii badań empirycznych (szczególnie serii badań 2 i 3), a także metodologicznie poprawna analiza i prezentacja wyników tych badań (z jedynie małymi niedociągnięciami technicznymi, które nie wpływają na ogólną konkluzję pracy). Rozprawa wskazuje zatem, że Autorka posiada odpowiednie kompetencje w projektowaniu, prowadzeniu, analizie i przedstawieniu badań własnych, czyli umiejętność samodzielnego prowadzenia pracy naukowej.

Docenić należy wysiłek Autorki włożony w przeprowadzone badania, tak duże próby osób badanych (nie bez znaczenia w obecnym kryzysie replikacyjnym), jak i użycie wielu różnych metod pomiaru.

Reasumując, przedstawiona mi do recenzji rozprawa doktorska prezentuje w dostatecznym zakresie wiedzę teoretyczną p. mgr Angeliki Olszewskiej w dyscyplinie psychologia, a także dowodzi umiejętności prowadzenia samodzielnej pracy naukowej. Przedmiotem rozprawy jest oryginalne rozwiązanie problemu naukowego, polegające na pogłębieniu rozumienia zjawiska wglądu poprzez analizę towarzyszących mu subiektywnych doświadczeń, a także rozwinięcie nowego – prostego i użytecznego – narzędzia służącego do ich pomiaru. Konkludując, uważam, że powyższa rozprawa **spełnia** warunki określone w art. 187 ustawy Prawo o szkolnictwie wyższym i nauce.

Kraków, 4 kwietnia 2023