

Poznań, 14 września 2022

dr hab. Paweł Łupkowski, prof. UAM
Wydział Psychologii i Kognitywistyki
Uniwersytet im. Adama Mickiewicza w Poznaniu

**Recenzja rozprawy doktorskiej magister Natalii Lipp
„Poznawczo-społeczne uwarunkowania poczucia realizmu
w wirtualnej rzeczywistości”**

(promotor dr. hab. Sławomir Śpiewak, prof. UJ;
promotor pomocniczy dr Paweł Strojny)

Celem recenzowanej rozprawy jest identyfikacja czynników wpływających na wykonanie zadań w wirtualnej rzeczywistości. Autorka dysertacji rozбивa ten cel na dwa cele szczegółowe. Pierwszym z nich jest walidacja empiryczna modelu immersji, poczucia bycia obecnym i wykonywania zadań (IPP). Drugim zaś jest ustalenie tego, jakie czynniki wpływają na powstanie subiektywnego poczucia realizmu w wirtualnej rzeczywistości. Praca stanowi zatem bardzo ważny głos w badaniach nad tym interesującym (i nabierającym aktualnie coraz większego znaczenia) zagadnieniem funkcjonowania w VR. Warto już na wstępie zaznaczyć, że wyniki takich badań posiadają duże znaczenie zarówno od strony poznawczej jak i od strony praktycznej (na co wskazuje również sama Autorka w podsumowaniu pracy).

Wyniki uzyskane w ramach opisanego w dysertacji programu badawczego stanowią istotny wkład w naszą wiedzę dotyczącą zachowań ludzi w wirtualnej rzeczywistości. Przedstawione analizy wyników eksperymentów a w szczególności pogłębione ich dyskusje oceniam bardzo pozytywnie.

Poniżej przedstawiam bardziej szczegółowe uwagi i komentarze dotyczące rozprawy.

Część teoretyczna

Recenzowana dysertacja składa się z dwóch klasycznych części: części teoretycznej oraz części badawczej opisującej trzy przeprowadzone eksperymenty. W części teoretycznej wprowadzono w sposób zwięzły i zrozumiały wszystkie niezbędne elementy ramy teoretycznej wykorzystywanej do zaprojektowania oraz analizy danych z eksperymentów (w szczególności wspomniany już model IPP). Za najmocniejszy element tej części pracy uważam przegląd literatury przedmiotu dotyczącej badań wpływu realizmu na wykonanie zadań oraz poczucie bycia obecnym w wirtualnej rzeczywistości. Przegląd literatury obejmuje imponującą liczbę pozycji i przeprowadzony jest według ścisłych oraz jasno przedstawionych kryteriów, co uważam za jego ogromną zaletę. Już samo to zestawienie stanowi istotne źródło wiedzy dla wszystkich zainteresowanych badaniami z wykorzystaniem VR. Zaprezentowany przegląd literatury doskonale pokazuje potrzebę przeprowadzenia badań przedstawionych w recenzowanej dysertacji.

Warto również podkreślić, iż samo zastosowanie przez Autorkę modelu IPP motywowane jest także przeglądem literatury przedmiotu, której celem było sprawdzenie podstaw empirycznych tego modelu. Jak zauważa Autorka takie podstawy muszą zostać dopiero dostarczone i jest to jeden z celów jej pracy. Tym samym Autorka na stronie 55 deklaruje, że traktuje IPP raczej jako hipotezę bez weryfikacji empirycznej. W świetle tego stwierdzenia brakuje mi nieco dokładniejszego uzasadnienia dla przyjęcia modelu IPP. Dlaczego Autorka uważa, że to właśnie ten model wyznacza dobrą ramę teoretyczną dla proponowanych badań?

Istotnym elementem części teoretycznej jest również podjęta przez Autorkę próba konceptualizacji pojęcia realizmu w wirtualnej rzeczywistości. Dobrych podstaw dostarcza tutaj wspomniany już przegląd literatury przedmiotu wskazujący na komponenty tego pojęcia, takie jak: rozdzielczość, liczba poligonów, oświetlenie, czy złożoność sceny. Ustalenia te znajdują swoje odbicie w projekcie przeprowadzonych eksperymentów. Żałuję jedynie że autorka nie pokusiła się o zaproponowanie swojej własnej definicji regulującej pojęcia realizmu wirtualnej rzeczywistości, którą mogłaby wykorzystać we wszystkich prowadzonych przez nią badaniach (o czym piszę dokładniej w uwagach do przeprowadzonych badań).

Za bardzo interesującą uważam również część końcową wprowadzenia teoretycznego, w której autorka odnosi teorie związane z VR do znanych koncepcji z zakresu psychologii poznawczej. Warto podkreślić, że fragment ten jest bardzo dobrze wykorzystany w dyskusji wyników eksperymentów i doskonale umieszcza uzyskane wyniki w szerszym kontekście badań z zakresu psychologii poznawczej.

Część empiryczna

Hipotezy. W paragrafie: „Główne problemy badawcze i stawiane hipotezy” Autorka wprowadza sześć hipotez badawczych, które postawiła w trzech opisanych eksperymentach. Myślę, że byłoby dużym ułatwieniem dla czytelnika gdyby wyjaśnienie dotyczące kolejności konstruowania hipotez i ich wzajemnej zależności pojawiło się jeszcze przed ich wprowadzeniem (zamiast dopiero na stronie 84).

W sformułowaniach hipotez badawczych zabrakło mi także operacjonalizacji pojęcia realizmu (na taką potrzebę wskazuje analiza literatury przedmiotu przedstawiona w części teoretycznej). Dopiero podczas lektury fragmentów poświęconych poszczególnym eksperymentom dowiadujemy się że pojęcie realizmu było w nich pojmowane inaczej. W szczególności różnica ta widoczna jest pomiędzy eksperymentem pierwszym a eksperymentami drugim i trzecim. Zasadne moim zdaniem byłoby wprowadzenie tych operacjonalizacji już na poziomie hipotez badawczych.

W przypadku hipotezy pierwszej brakuje mi również uzasadnienia dla przyjętego jej kierunku. Dlaczego zakładamy że wskazany wpływ będzie pozytywny? Szczególnie, że w dyskusji eksperymentu pierwszego na stronie 112 Autorka przywołuje szereg badań wskazujących na fakt, że zwiększenie rozdzielczości tekstur czy dodawanie nowych tekstur (co stanowi operacjonalizację pojęcia realizmu na potrzeby eksperymentu pierwszego) nie skutkuje lepszym wykonywaniem zadań. Myślę również, że tutaj warto byłoby także doprecyzować pojęcie wykonywania zadań (czy chodzić będzie o poprawność czy na przykład o czas). Czytelnik dowiaduje się o tym również dopiero na poziomie opisu eksperymentu pierwszego.

Nieco zaskakujące jest dla mnie także uzasadnienie hipotezy czwartej podane na stronie 82. Autorka pisze w nim: „Niewiele wiadomo o wpływie płynności na wykonywanie zadań o charakterze percepcyjnym. Przy założeniu, że ocena realizmu symulacji to wynik płynnego przetwarzania i przy założeniu, że realizm symulacji wpływa na wykonywanie zadań – niezależnie czy poprzez dostarczanie wskazówek czy poprzez inne procesy można się spodziewać, że wzbudzenie płynności przetwarzania wpłynie pozytywnie na wykonywanie zadań”. Warto zauważyć tutaj że wskazane „założenia” to tak naprawdę treść hipotez pierwszej i trzeciej. Z podobnie niefortunnym sformułowaniem mamy do czynienia na stronie 120, gdzie autorka pisze: „Zgodnie z planem badawczym zmieniona została manipulacja realizmem sytuacji. W tym badaniu manipulowano dwoma charakterystykami percepcyjnymi – złożonością i wyrazistością symulacji w układzie czynnikowym 2×2 (wyrazistość $\times$ złożoność), założywszy, że ocena realizmu jest wynikiem płynności przetwarzania”. Ponownie, wskazane tutaj założenie jest tak naprawdę treścią jednej z przyjętych hipotez badawczych.

Symulacja. Przy opisie metod badawczych na stronie 84 Autorka wskazuje powody, dla których wybrano tę a nie inną symulację na potrzeby eksperymentu. Co ciekawe wskazuje również, że jednym z powodów był wysoki stopień realizmu niestety nie definiując co oznacza w tym przypadku – po pierwsze, realizm a po drugie, jak rozumiany jest jego wysoki stopień. Spodziewałbym się że w obliczu trudności interpretacyjnych jasno wskazanych w rozdziale teoretycznym na tym etapie czytelnik otrzyma bardziej precyzyjne wyjaśnienia dotyczące realizmu użytych symulacji.

Badanie pierwsze. Uważam operacjonalizację hipotezy pierwszej zaproponowaną na potrzeby tego eksperymentu za interesującą. Usunięcie tekstur ze sceny symulacji jest bardzo intuicyjnym krokiem. Szkoda tylko że Autorka na stronie 100 pisze jedynie, że warunki różnią się poziomem teksturowania kluczowych elementów symulatora nie definiuje jednak nigdzie i nie podaje liczby tekstur, które wyznaczają trzy rodzaje użytych scen. Sytuację utrudnia również fakt, że Rysunek 5 na stronie 101 prezentuje jedynie dwa warianty sceny użytej w eksperymencie. Czytelnik nie dowiaduje się jak wygląda scena częściowo nierealistyczna. Na wspomnianym rysunku nie widać również pałek ani pachołków, które były elementem manipulacji eksperymentalnej. Braki te stają się dotkliwe szczególnie podczas lektury raportu wyników na stronie 106. Czytamy tutaj, że „W celu weryfikacji hipotezy o wpływie realizmu symulacji na poziom wykonywania zadań sprawdzono, czy liczba zdobytych punktów różni się w zależności od ilości tekstur w wirtualnej scenie”. Nigdzie jednak nie doszukałem się wartości liczbowych wyznaczających tekstury w różnych wariantach sceny.

W związku z Rysunkiem 5 (panel B) ciekawe wydaje się również pytanie o to, jak podawane były wskazówki dla osoby badanej dotyczące kierunku ruchu samolotu. Czy było to w formie strzałki widocznej nad kadłubem samolotu tak, jak to ilustruje wskazany rysunek? W opisie eksperymentów nie znalazłem takiej informacji. Czy taki rodzaj podawania informacji o kierunku nie koliduje z zamiarem stworzenia realnej sceny wirtualnej rzeczywistości? W świecie realnym nie spodziewamy się wskazówek wizualnych wyświetlanych nad realnymi obiektami.

Opis procedury eksperymentalnej badania pierwszego jest niestety miejscami niejasny. Po pierwsze, dowiadujemy się o liczebności całej grupy badanych natomiast nie wiemy jak zostali oni podzieleni na trzy grupy eksperymentalne. Nie znamy liczebności tych grup ani proporcji kobiet i mężczyzn w nich obecnych. Na stronie 102 czytamy także: „Badany miał trzy razy przeprowadzić samolot po płycie lotniska po wyznaczonej trasie – za każdym razem zadanie różniło się poziomem trudności (łatwe, umiarkowane i trudne)”. Czy kolejność tych zadań była randomizowana czy dla każdego badanego była dokładnie taka jak wskazana w tekście? Jeśli to drugie, to w badaniu mamy do czynienia z efektem trenin-

gu. W kolejnym zdaniu czytamy: „Po zakończeniu zadania uczestnik wypełniał kwestionariusze: oceniał trudność zadania, wysiłek w nie włożony oraz subiektywny poziom wykonania”. Czy ocena ta dotyczyła pojedynczego zadania czy też wszystkich trzech zadań na raz? Jeśli tak, to jaki był tego powód i na ile informatywna jest taka deklaracja osoby badanej?

W kontekście eksperymentu pierwszego interesuje mnie również zasadność manipulacji trudnością zadania. Wskazane trzy warianty zadania różniące się różnymi trasami samolotu i rozstawieniem punktów kontrolnych wydają się tylko komplikować samą procedurę. W obliczu faktu, że poziom wykonania zadania mierzony był zdobytymi przez badanego punktami projekt, w którym jest tylko jedna wersja zadania (z trudnością ustaloną wcześniej na podstawie badania pilotażowego) wydaje się być wystarczający. Rozwiązanie to wydaje się o tyle efektywne, że zostało w zasadzie zaimplementowane w eksperymencie drugim i trzecim. Ciekawy jestem zatem motywacji stojącej za przyjętym rozwiązaniem. Interesujący jest również brak opisu badania pilotażowego. Czy zostało ono przeprowadzone, a jeśli tak, to jak wpłynęło na kształt badania właściwego?

Badanie drugie. Na potrzeby badania drugiego (podobnie zresztą, jak i badania trzeciego) wprowadzono odmienną operacjonalizację pojęcia realizmu, co skutkowało inną manipulacją eksperymentalną. Rodzi to w mojej ocenie problemy interpretacyjne podczas porównania wyników z trzech eksperymentów. Wskazuje też na wspomniany już problem braku jednolitej definicji pojęcia realizmu przyjętej na potrzeby całej rozprawy.

W mojej opinii czynnikiem utrudniającym porównanie wyników eksperymentu 1 i 2 jest także charakterystyka grupy badanej. Warto zwrócić uwagę na to, że grupy te dość mocno różnią się wiekiem badanych. Średnia wieku w grupie drugiej jest wyższa o 6 lat od średniej dla grupy pierwszej (również odchylenie standardowe wieku w grupie drugiej wskazuje na większe zróżnicowanie wiekowe osób badanych). Warto byłoby zwrócić uwagę na wiek jako jedną z analizowanych zmiennych (por. np. badanie Liu, L., Watson, B., & Miyazaki, M. (1999). VR for the elderly: Quantitative and qualitative differences in performance with a driving simulator. *CyberPsychology & Behavior*, 2(6), 567–576).

Co ciekawe w eksperymencie drugim (podobnie jak w eksperymencie trzecim) Autorka kontroluje wady wzroku osób badanych. Sugeruje to, że wady wzroku nie miały znaczenia dla eksperymentu pierwszego co wydaje się być problematyczne.

W opisie procedury badawczej eksperymentu drugiego na stronie 125 brakuje jasnego sprecyzowania jak liczony był czas reakcji, co jest informacją kluczową dla zrozumienia interpretacji otrzymanych wyników – porównaj chociażby strona 131.

Badanie trzecie. Badanie trzecie uważam za najbardziej interesujące oraz najlepiej dopracowane ze wszystkich przedstawionych w rozprawie. Bardzo pozytywnie oceniam pomysł wykorzystania pomiarów fizjologicznych do zmierzenia pobudzenia u osób badanych. Istotne jest również to że zachowano operacjonalizację pojęcia realizmu opracowaną na potrzeby badania drugiego.

W kontekście pomiarów EDA prezentowanych na stronach 151 i 152 ciekawe wydaje się pytanie o sposób ekstrahowania komponentu tonicznego EDA. Jest to bardzo istotny etap wstępnego przetwarzania sygnału, natomiast czytelnik nie otrzymuje precyzyjnych informacji na ten temat.

Warto w tym miejscu również zauważyć, że pobrany sygnał fizjologiczny pozwala również na analizę obciążenia poznawczego osób badanych, co wydaje się być interesujące w kontekście sprawdzenia modelu IPP (por. Buchwald, M., Kupiński, S., Bykowski, A., Marcinkowska, J., Ratajczyk, D., & Jukiewicz, M. (2019). *Electrodermal activity as a measure of cognitive load: a methodological approach. Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)* (pp. 175–179). IEEE). Naturalnie jest to tylko sugestia do potencjalnych dalszych badań i analiz. Mogą być one szczególnie ciekawe w kontekście uwag Autorki na stronie 157.

Uwagi ogólne

Osoby badane. We wszystkich trzech eksperymentach zauważyłem jeden spójny problem dotyczący osób badanych. Przy każdym eksperymencie autorka wskazuje na liczebność całej grupy badanej nie wskazując na liczebności grup dla poszczególnych warunków. W szczególności autorka nie analizuje tego, czy podgrupy różnią się w sposób istotny pomiędzy sobą w wynikach kwestionariusza *wyrazistości wyobrażonych obrazów*. Kwestionariusz ten dostarcza nam informacji na temat predyspozycji osób badanych i jego wynik stanowi istotną zmienną analizowaną w pracy. Tym bardziej ważne wydaje się upewnienie, że podgrupy w każdym kolejnym eksperymencie nie różnią się między sobą pod względem tej predyspozycji.

Interesujące wydaje się też dla mnie to, jakie dane demograficzne zbierano o uczestnikach badania. Autorka wskazuje jedynie na wiek i płeć badanych. Tymczasem zasadne wydaje się zbieranie informacji o wcześniejszych doświadczeniach badanych z wirtualną rzeczywistością, ponieważ mogą mieć one potencjalny wpływ na zachowania badanego podczas eksperymentu. Na ten fakt zwraca uwagę sama Autorka w dyskusji eksperymentu pierwszego na stronie 113 (gdzie czytamy: „Jak zauważa Slater (2009) szczególnie nowi użytkownicy VR oceniają realizm wyżej i nie zwracają uwagi na niedostatki w teksturach”). Z jakiego powodu zdecydowano się o zbieraniu tak ograniczonych informacji o badanych?

Efekt metody. W świetle wyników trzech eksperymentów najbardziej interesująca wydaje się dyskusja dotycząca zastosowanych narzędzi samoopisowych. Bardzo podoba mi się pomysł odróżnienia realizmu i poczucia realności oraz zastosowania dla ich pomiaru dwóch odrębnych narzędzi (podskali realizmu scenicznego z polskiej skali realizmu symulacji VR oraz podskali realności z kwestionariusza Presence Ingroup). To czego mi stanowczo zabrakło to pogłębionej analizy tego, jak manipulacje eksperymentalne dotyczące realizmu sceny i elementy operacjonalizacji tego pojęcia powiązane są z pozycjami wykorzystanych narzędzi samoopisowych. Autorka zauważa ten problem dopiero w dyskusji badania drugiego (porównaj strona 132). Co ciekawe, pomimo zauważenia tego problemu nadal stosuje te same narzędzia badawcze przy eksperymencie trzecim. Na świadomość tego problemu wskazuje dyskusja obecna na stronach 155 i 156: „W świetle uzyskanych wyników wydaje się, że w przypadku poczucia realizmu pojawił się efekt metody, czyli ograniczenie pomiarów wynikające z zastosowanego narzędzia, co sygnalizowano już w dyskusji badania drugiego. Polska skala realizmu symulacji VR (Lipp i in. 2021) mierzy ocenę między innymi kolorystyki, światłocieni, proporcji w symulacji, a żaden z tych komponentów nie został zmieniony podczas wprowadzania manipulacji eksperymentalnej”. Dlaczego pomimo świadomości tego problemu pozostaną przy wybranych narzędziach samoopisowych?

Operacjonalizacja pojęcia realizmu. Jak już wspomniałem powyżej, eksperyment pierwszy oraz eksperymenty drugi i trzeci posługują się innego rodzaju operacjonalizacją realizmu na potrzeby manipulacji eksperymentalnej. Stawia to otwarte pytanie o to, czy wyniki tych badań możemy porównywać wzajemnie stosując jeden pojemnik w postaci słowa „realizm”. Robi to Autorka na przykład na stronie 156 interpretując wyniki badania trzeciego i wskazując, że wyniki te pozwalają na przyjęcie hipotezy drugiej mówiącej o tym że zgodność pomiędzy oceną realizmu a możliwościami wyobrażeniowymi ułatwia wykonywanie zadań VR.

Uwarunkowania społeczne. Autorka przy projektowaniu eksperymentów celowo unikała obecności wirtualnych agentów w symulowanej scenie. Uważam to za bardzo dobry zabieg zwłaszcza w świetle potencjalnych zniekształceń uzyskanych wyników z uwagi na interakcje społeczne w VR. Taka decyzja jednak odbiega nieco od tego co sugeruje temat dysertacji. W oczywisty sposób mamy tutaj analizę uwarunkowań poznawczych wykonywania zadań w VR, ale uwarunkowania społeczne analizowane nie są.

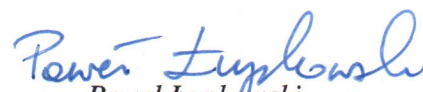
Ocena aspektów formalnych. Pod względem formalnym praca przygotowana jest dość starannie. Obecne drobne niedociągnięcia redakcyjne oraz potknięcia językowe (por. m.in. str. 115, w. 1–2 i 15 dół; 158 w. 8–10 dół; 159, w. 1 dół; 160 w. 10–11 góra, 9 dół) nie zakłócają znacząco odbioru pracy.

Dużym ułatwieniem dla czytelnika byłoby niewątpliwie ponumerowanie rozdziałów i podrozdziałów. Dodanie wykorzystanych narzędzi samoopisowych jako załączników do pracy również traktowałbym jako duże ułatwienie w jej lekturze a w szczególności w analizie wyników eksperymentów.

Zastanawia mnie także dlaczego podczas raportowania wyników eksperymentów autorka nie podaje szczegółowych wyników dla zastosowanych narzędzi samoopisowych. Spodziewałbym się przynajmniej tabelarycznego zestawienia wyników kwestionariuszy wskazujących wyniki maksymalne, minimalne, średnią, odchylenie standardowe czy mediany (z podziałem na podgrupy badanych) – por. raporty na stronie 104 dla zmiennych realizm i poczucie realizmu, strona 127 dla tych samych kwestionariuszy, czy choćby strona 125 dla oceny estetycznej symulacji. Jeśli motywacją było oszczędzanie miejsca, tabele takie można było załączyć w dodatkach do rozprawy.

Na zakończenie chciałbym skomplementować pewien interesujący zabieg polegający na połączeniu wstępu oraz zakończenia pracy cytatem z filmu „Matrix” – uważam to za bardzo udany literacki ukłon w stronę czytelnika.

Konkluzja: Stwierdzam, że przedłożona mi do oceny rozprawa doktorska mgr Natalii Lipp spełnia wszystkie wymogi **spełnia wymogi** stawiane rozprawom doktorskim w odpowiednich przepisach. Wnoszę tym samym do Rady Dyscypliny Psychologia Uniwersytetu Jagiellońskiego o **dopuszczenie Kandydatki do dalszych etapów przewodu doktorskiego.**


Paweł Łupkowski