

dr hab. Marcin Leszczyński, prof. UJ
Uniwersytet Jagielloński
Katedra Kognitywistyki
Laboratorium Badań Mózgu
ul. Ingardena 3, 30-060 Kraków

Kraków, 20 kwietnia 2026



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

RECENZJA ROZPRAWY DOKTORSKIEJ

Autor: mgr Dawid Ratajczyk

Tytuł: *Perceiving Artificial Agents: A Multimethod Investigation into the Shape and Variables of the Uncanny Valley*

Promotor: dr hab. Paweł Łupkowski, prof. UAM

Promotor pomocniczy: dr inż. Marcin Jukiewicz

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków

Przedstawiona do recenzji rozprawa doktorska mgr. Dawida Ratajczyka pt. *“Perceiving Artificial Agents: A Multimethod Investigation into the Shape and Variables of the Uncanny Valley”*, napisana pod kierunkiem dr. hab. Pawła Łupkowskiego, prof. UAM (promotor) oraz dr. inż. Marcina Jukiewicza (promotor pomocniczy), dotyczy hipotezy doliny niesamowitości (ang. *Uncanny Valley Hypothesis*). Hipoteza ta zakłada, że roboty, które są niemal, ale nie w pełni podobne do ludzi, mogą wywoływać uczucie niepokoju i dyskomfortu. Praca składa się z trzech części: wprowadzenia, przeglądu literatury oraz podsumowania własnych badań. Jej podstawę stanowią trzy artykuły empiryczne, które – wykorzystując komplementarne metody (analizę języka naturalnego komentarzy z serwisu YouTube, eksperymenty z wykorzystaniem rzeczywistości wirtualnej (ang. virtual reality; VR) oraz pomiary psychofizjologiczne) – mają na celu lepsze zrozumienie kształtu oraz zmiennych związanych z doliną niesamowitości.



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

Autor formułuje trzy jasno określone cele badawcze:

1. Wyjaśnić kształt i zmienne związane z doliną niesamowitości.
2. Zidentyfikować poznawcze i społeczne moderatory, takie jak przekonania o wyjątkowości natury ludzkiej oraz wiedza na temat doliny niesamowitości.
3. Zbadać interakcję między wyglądem robota oraz atrybucją umysłu w realistycznym kontekście.

Rozdział 1 – Wstęp

Rozdział pierwszy stanowi klarowne i dobrze ugruntowane teoretycznie wprowadzenie do problematyki doliny niesamowitości. Autor przedstawia główne założenia hipotezy doliny niesamowitości oraz przegląd najważniejszych sposobów wyjaśniania zjawiska, obejmujących zarówno mechanizmy percepcyjne, procesy związane z teorią umysłu, jak i ujęcia społeczno-kulturowe. Na szczególną uwagę zasługuje podkreślenie wieloczynnikowego charakteru zjawiska oraz wskazanie braku konsensusu w literaturze. Autor trafnie identyfikuje lukę badawczą dotyczącą relacji między percepcją wyglądu, atrybucją umysłu oraz czynnikami społecznymi, co prowadzi do logicznego sformułowania celów pracy. Rozdział jest dobrze zorganizowany, napisany jasnym i klarownym językiem.

Rozdział 2 – Teoria

Rozdział drugi przedstawia tło teoretyczne dotyczące efektu doliny niesamowitości. Autor porządkuje definicje oraz główne kierunki badań, wskazując zarówno na klasyczne ujęcie relacji między podobieństwem do człowieka a reakcjami emocjonalnymi, jak i trudności związane z operacjonalizacją takich pojęć jak *humanlikeness* czy *uncanny feeling*. Szczególną wartością rozdziału jest systematyczne zestawienie różnych wyjaśnień efektu – od teorii ewolucyjnych, przez mechanizmy percepcyjne, po podejścia poznawcze związane z atrybucją umysłu. Autor przekonująco argumentuje, że zjawisko ma charakter wieloczynnikowy. Rozdział kończy się

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków

uzasadnieniem przyjętej strategii metodologicznej. Autor wskazuje na ograniczenia dotychczasowych badań oraz potencjał nowych podejść, takich jak wirtualna rzeczywistość oraz konieczność prowadzenia obserwacji w bardziej naturalnym środowisku. Rozdział ten jest pomocny, choć zagadnienie prowadzenia badań w bardziej naturalistycznych warunkach jest potraktowane stosunkowo powierzchownie. W tym rozdziale, jak i w artykułach pojawia się pojęcie “ecological validity”, którego celem jest uzasadnienie stosowania bardziej naturalnych warunków lub bardziej złożonych miar. Pojęcie to ma uzasadnić użycie narzędzia jako wartości samej w sobie. Nie jest jednak jasne, jakie nowe informacje miałyby wynikać z zastosowania bardziej „naturalistycznych” metod. Praca zyskałaby, gdyby Autor precyzyjniej wskazał, jakie konkretne procesy lub mechanizmy miałyby być uchwycone dzięki wykorzystaniu VR, a które pozostają poza zasięgiem tradycyjnych eksperymentów. W świetle przedstawionych wyników można bowiem odnieść wrażenie, że zastosowanie VR nie wnosi istotnej wartości dodanej (np. brak różnic w podstawowych miarach). Podobna wątpliwość dotyczy miar psychofizjologicznych – nie jest do końca jasne, jakie dodatkowe aspekty reakcji badanych pozwalają one uchwycić w porównaniu z prostszymi metodami pomiaru.

Rozdział 3 – Podsumowanie badań własnych

Rozdział trzeci przedstawia syntetyczne podsumowanie wyników trzech badań empirycznych. Wyniki wskazują, że zależność między podobieństwem do człowieka a reakcjami emocjonalnymi jest bardziej złożona niż sugeruje klasyczna krzywa doliny niesamowitości. Ogólny trend ma raczej charakter liniowy, natomiast nieliniowość ujawnia się w odniesieniu do wybranych cech (np. rysów twarzy). Istotnym wkładem pracy jest identyfikacja moderatorów (np. przekonań o wyjątkowości człowieka) oraz rozróżnienie poszczególnych elementów reakcji emocjonalnej w kontekście doliny niesamowitości (eeriness vs. likability).

Jednocześnie nie jest w pełni jasne, w jaki sposób wyniki trzech artykułów



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

łączą się w spójną całość. Artykuł 1 sugeruje raczej zależność liniową, podczas gdy artykuł 2 wskazuje na efekt bardziej zbliżony do klasycznej doliny niesamowitości, choć najwyższy poziom niepokoju i dyskomfortu nie dotyczy najbardziej podobnych do człowieka robotów, lecz agenta o stosunkowo niskim stopniu podobieństwa (Ybot). Powstaje zatem pytanie dlaczego w artykule 2 poziom niepokoju i dyskomfortu spada dla najbardziej podobnych do człowieka robotów, podczas gdy w artykule 1 wykazano liniowy wzrost niepokoju i dyskomfortu wraz ze wzrostem podobieństwem do człowieka. Jak Autor interpretuje tę rozbieżność? Podobnie w artykule 3 obserwujemy rozbieżności między różnymi miarami: większe poczucie niepokoju wobec agenta o wyglądzie robota współwystępuje z wyższym rytmem serca w warunkach, w których badani są przekonani, że wchodzi w interakcję z człowiekiem. Wskazuje to, że różne miary mogą odzwierciedlać odmienne procesy (np. percepcyjne vs. społeczno-poznawcze). Zarówno artykuły jak i rozdział trzeci zyskałyby, gdyby Autor wyraźniej zaproponował sposób interpretacji tych pozornie niespójnych wyników. Ich pełne zrozumienie wymagałoby bardziej systematycznej próby integracji wszystkich przeprowadzonych analiz oraz użytych wyników w ramach jednej, spójnej interpretacji.

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków

Artykuły empiryczne

Podstawę rozprawy stanowią trzy artykuły empiryczne, które wykorzystują zróżnicowane i komplementarne podejścia metodologiczne do badania efektu doliny niesamowitości. Łączy je próba uchwycenia tego zjawiska z różnych perspektyw: analizy komentarzy społeczności YouTube, kontrolowanych eksperymentów oraz pomiarów psychofizjologicznych. Poniżej omawiam każdy z artykułów, zwracając uwagę zarówno na ich wkład w zrozumienie badanego zjawiska, jak i na ich potencjalne ograniczenia.

Artykuł pt. *“Shape of the Uncanny Valley and Emotional Attitudes Toward Robots Assessed by an Analysis of YouTube Comments”* przedstawia



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków

badanie wykorzystujące analizę języka naturalnego, przeprowadzone w celu lepszego zrozumienia emocjonalnych reakcji na roboty. Autor analizuje ponad 220000 komentarzy z serwisu YouTube. Wyniki sugerują, że wraz ze wzrostem podobieństwa robotów do człowieka rośnie poziom negatywnych reakcji, szczególnie tych związanych z niepokojem i dyskomfortem (eeriness). Autor podkreśla również, że różne miary reakcji emocjonalnej na roboty (np. atrakcyjność vs. niepokój i dyskomfort) nie są ze sobą tożsame. Dodatkowo Autor zaobserwował, że ocena robota zmienia się wraz z jego wielkością – mniejsze roboty są postrzegane bardziej pozytywnie, a efekt ten jest mediowany przez przypisywanie im charakteru zabawowego.

Artykuł jest napisany w sposób zrozumiały. Jednak zyskałby na bardziej narracyjnym ujęciu wyników, które ułatwiłoby czytelnikowi zrozumienie wpływu poszczególnych czynników. W obecnej formie wyniki sprawiają wrażenie zbioru niepowiązanych analiz, co utrudnia ich całościową interpretację.

Artykuł prezentuje wartościowy materiał empiryczny, jednak pewne aspekty metodologiczne mogłyby zostać dodatkowo wzmocnione. Choć liczba analizowanych komentarzy jest imponująca, główną jednostką analizy są roboty ($N = 33$), co znacząco ogranicza liczebność danych na poziomie kluczowych zmiennych. W konsekwencji niektóre zakresy zmiennych wydają się niedostatecznie reprezentowane, w szczególności środkowy zakres skali „surface look” (Rys. 3 oraz Rys. 6). Może to wpływać na stabilność estymacji, zwłaszcza w przypadku bardziej złożonych dopasowań (np. funkcji sześciennych), które mogą być wrażliwe na pojedyncze obserwacje odstające. W przyszłych badaniach warto byłoby rozważyć rozszerzenie próby o większą liczbę robotów, zwłaszcza w tym zakresie skali, co pozwoliłoby na jeszcze bardziej wiarygodne oszacowanie obserwowanych zależności.

Z tym wiąże się również kwestia doboru modeli. Autor stosuje dopasowania liniowe, kwadratowe i sześciennne, co pozwala uchwycić różne możliwe kształty zależności w danych. Warto byłoby jednak bardziej szczegółowo



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

przedstawić kryteria wyboru tych funkcji oraz ich uzasadnienie teoretyczne. Część wyników – zwłaszcza dotyczących zmiennej „surface look” – sugeruje, że inne typy dopasowań (np. funkcje wykładnicze; Rys. 3) mogłyby również stanowić użyteczną alternatywę. Dokładne teoretyczne uzasadnienie testowanych relacji mogłoby dodatkowo wzmocnić artykuł.

Dodatkowo warto byłoby szerzej opisać sposób weryfikacji założeń analiz regresji. Wizualna inspekcja danych może sugerować pewne odstępstwa od założenia homoscedastyczności, dlatego uzupełnienie tej części o analizę reszt pozwoliłoby jeszcze bardziej wzmocnić rzetelność wniosków.

Na poziomie interpretacyjnym interesującym uzupełnieniem byłoby również pogłębienie dyskusji dotyczącej rozbieżności w literaturze, do których Autor się odwołuje (np. liniowy vs. nieliniowy charakter zależności), co pomogłoby lepiej osadzić uzyskane wyniki w szerszym kontekście badań nad doliną niesamowitości.

Wreszcie, choć Autor trafnie wskazuje na zwiększoną ekologiczność podejścia opartego na analizie komentarzy z YouTube, warto byłoby bardziej precyzyjnie wskazać, jakie nowe wnioski wynikają z zastosowania tej metody w porównaniu z bardziej kontrolowanymi eksperymentami.

Artykuł pt. *“The Importance of Beliefs in Human Nature Uniqueness for Uncanny Valley in Virtual Reality and On-Screen”* przedstawia badanie eksperymentalne, którego celem jest zbadanie wpływu prezentacji robota (VR vs. ekran komputera stacjonarnego), stopnia podobieństwa do człowieka (typ agenta: Robot, Ybot, AJ oraz Brian) oraz przekonań o wyjątkowości natury ludzkiej na postrzeganie robotów. Praca ta została przygotowana przez trzyosobowy zespół, w którym Doktorant pełnił rolę lidera. Wskazuje na to zarówno kolejność autorów, jak i opis ich indywidualnego wkładu.

Najwyższy poziom niepokoju i dyskomfortu wywołuje agent o umiarkowanym, a nie najwyższym stopniu podobieństwa do człowieka (Ybot). Choć można to interpretować jako częściowo zgodne z klasycznym ujęciem doliny niesamowitości, wynik ten pozostaje trudny do pogodzenia z rezultatami

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

artykułu pierwszego, w którym obserwowano raczej liniowy wzrost niepokoju wraz ze wzrostem podobieństwa do człowieka. Warto byłoby wyraźniej odnieść się do tej rozbieżności i zaproponować jej interpretację. Autorzy nie zaobserwowali istotnych różnic między warunkami VR i ekranowym ani związku między poziomem poczucia obecności a ocenami poczucia niepokoju. Istotnym moderatorem okazały się natomiast przekonania esencjalistyczne – osoby silniej wierzące w wyjątkowość człowieka reagowały silniej na najbardziej „niepokojącego” agenta.

Artykuł mógłby dodatkowo zyskać na bardziej rozbudowanym uzasadnieniu teoretycznym głównego kontrastu (VR vs. ekran). Autorzy sugerują, że rzeczywistość wirtualna może wpływać na percepcję robotów poprzez zwiększone poczucie obecności. Warto byłoby szerzej wyjaśnić, na czym dokładnie polega to doświadczenie oraz dlaczego miałyby prowadzić do odmiennych reakcji percepcyjnych lub emocjonalnych.

Ponadto, jeśli przyjąć, że VR oddziałuje poprzez zwiększenie zaangażowania i realizmu doświadczenia, interesującym kierunkiem byłoby rozważenie, czy zastosowany paradygmat pasywnej obserwacji w pełni oddaje potencjał tej manipulacji. W wielu ujęciach teoretycznych kluczową rolę odgrywa aktywna interakcja, co rodzi pytanie, czy wyniki byłyby inne, gdyby uczestnicy bezpośrednio wchodzili w interakcję z robotami.

Podobnie jak w artykule pierwszym, doprecyzowania wymaga również odwołanie do pojęcia ekologiczności. Jego wyraźniejsze osadzenie w ramach teoretycznych – poprzez wskazanie konkretnych procesów poznawczych lub emocjonalnych, które dzięki zastosowaniu VR mogą być lepiej uchwycone – mogłoby dodatkowo wzmocnić interpretację uzyskanych wyników.

Artykuł pt. *“Matter Over Mind: The Joint Impacts of Appearance and Mind Type on the Uncanny Valley Effect in Virtual Reality”* przedstawia badanie dotyczące wpływu wyglądu agenta oraz przypisywanego mu typu umysłu na reakcje emocjonalne, wykorzystując zarówno miary kwestionariuszowe, jak i

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

psychofizjologiczne. Praca ta została przygotowana przez czterosobowy zespół, w którym Doktorant pełnił rolę lidera. Wskazuje na to zarówno kolejność autorów, jak i opis ich indywidualnego wkładu

Wyniki sugerują, że to przede wszystkim wygląd agenta determinuje poziom niepokoju – postaci o bardziej robotycznym wyglądzie wywołują silniejsze reakcje negatywne. Jednocześnie typ przypisywanego umysłu wpływa na inne miary, w tym na aktywność fizjologiczną. Autorzy zaobserwowali interesującą rozbieżność: większe poczucie niepokoju wobec agentów o wyglądzie robota współwystępuje z szybszym rytmem serca w warunkach, w których uczestnicy są przekonani, że wchodzi w interakcję z człowiekiem. Sugeruje to, że każda z tych miar może odzwierciedlać odmienne procesy – percepcyjne (związane z wyglądem) oraz społeczno-poznawcze (związane z atrybucją umysłu). Jest to szczególnie interesujący i obiecujący kierunek interpretacji, który mógłby zostać jeszcze szerzej rozwinięty w artykule.

Praca mogłaby dodatkowo zyskać na bardziej szczegółowym omówieniu i uzasadnieniu zastosowania miar psychofizjologicznych. Choć Autorzy wskazują, że mogą one stanowić uzupełnienie danych samoopisowych, warto byłoby precyzyjnie określić, jakie konkretne procesy mają one odzwierciedlać. Doprecyzowanie hipotez dotyczących tych miar mogłoby jeszcze wzmocnić interpretację uzyskanych wyników i pozwolić w pełniejszy sposób wykorzystać ich potencjał w wyjaśnianiu czynników warunkujących efekt doliny niesamowitości.

Sposób analizy danych mógłby zostać wzmocniony poprzez pełniejsze uwzględnienie ich struktury. W szczególności analiza reakcji przewodnictwa skórniego opiera się na wielu obserwacjach przypadających na jednego uczestnika, co sugeruje zagnieżdżony charakter danych. Ilość stopni swobody sugeruje jednak, że potraktowano pojedyncze pomiary jako niezależną jednostkę obserwacji. Prowadzi to do pomieszenia efektu stałego oraz losowego, zawyżenia liczby stopni swobody (np. $df = 606$ lub 1108 przy 116 obserwacjach) i zwiększenia ryzyka błędu pierwszego rodzaju. Uwzględnienie zagnieżdżonej struktury danych – na przykład poprzez

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków



UNIWERSYTET
JAGIELLOŃSKI
W KRAKOWIE

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków

zastosowanie modeli mieszanych – pozwoliłoby na bardziej precyzyjne rozróżnienie efektów stałych i losowych oraz zwiększyłoby rzetelność wnioskowania statystycznego.

Dobór przedziału czasowego (0.9–6.0 s) oraz oparcie analizy na wartościach maksymalnych stanowią interesujący element metodologii, który mógłby jeszcze zostać doprecyzowany. Autorzy opierają wybór okna analizy na wcześniejszych badaniach, jednak uwzględnienie charakterystyki uzyskanych w tym badaniu danych mogłoby dodatkowo wzmocnić zasadność tego wyboru. Wizualna inspekcja danych (Rys. 6) sugeruje, że część efektów może wykraczać poza przyjęty zakres. Na przykład średnie przebiegi (human user, human AI, robot AI) osiągają wartości maksymalne dopiero około 6 s, czyli dokładnie tam, gdzie kończy się okno analizy. Dodatkowo warunek „robot user” wydaje się osiągać maksimum nieco później (około 7 s). Rozważenie alternatywnych zakresów czasowych lub analiz uzupełniających mogłoby pomóc ocenić stabilność wyników i jeszcze pełniej wykorzystać potencjał zgromadzonych danych.

Podsumowanie

Przedstawiona praca doktorska Dawida Ratajczyka stanowi wartościowe i oryginalne opracowanie jednego z kluczowych problemów współczesnych badań nad interakcją człowiek–technologia. Praca ma przejrzystą i logiczną strukturę, opartą na trzech artykułach empirycznych wykorzystujących komplementarne metody badawcze, co należy ocenić bardzo wysoko. W dwóch projektach Dawid Ratajczyk pełnił rolę lidera a w trzecim jest jedynym autorem, co dodatkowo podkreśla jego samodzielność badawczą. Doktorant wykazuje się bardzo dobrą znajomością literatury przedmiotu oraz umiejętnością formułowania ambitnych i istotnych celów badawczych, które zostały w dużej mierze zrealizowane. Na szczególne uznanie zasługuje zastosowanie zróżnicowanych podejść metodologicznych, pozwalających uchwycić badane zjawisko z różnych perspektyw, a także identyfikacja poznawczych i społecznych moderatorów efektu doliny niesamowitości.

Pewne elementy pracy – zwłaszcza te dotyczące wykorzystania wirtualnej rzeczywistości oraz miar psychofizjologicznych – mogłyby zostać jeszcze silniej osadzone teoretycznie, a niektóre aspekty analiz doprecyzowane. Ponadto, wyniki poszczególnych badań mogłyby zostać w jeszcze większym stopniu zintegrowane w spójną całość interpretacyjną. Uwagi te jednak nie podważają ogólnej wartości pracy. W mojej ocenie praca spełnia wymagania stawiane rozprawom doktorskim, świadczy o szerokiej wiedzy teoretycznej kandydata oraz jego zdolności do samodzielnego prowadzenia badań naukowych, a także stanowi oryginalny wkład w rozwój dyscypliny.



UNIwersytet
JAGIELLOŃSKI
W KRAKOWIE

Katedra Kognitywistyki
Wydział Filozoficzny
ul. Ingardena 3
30-060 Kraków

dr hab. Marcin Leszczyński, prof. UJ