

Dr hab. Łukasz Okruszek

Pracownia Neuronauki Społecznej

Instytut Psychologii PAN

Ocena rozprawy doktorskiej mgr Katarzyny Zemły (tytuł rozprawy: „Examining the efficacy of Guided Imagery relaxation technique in reducing stress, modulating brain wave activity, and enhancing attention control”). Promotorzy: prof. Grzegorz Sędek, prof. Grzegorz Wójcik).

Przedstawiona praca doktorska stanowi próbę odniesienia do sugestii poprawek i istotnych ograniczeń pierwotnie przedstawionego manuskryptu, które omówione zostały obszernie w poprzedniej wersji recenzji. Z tego też powodu, omówienie nowej wersji pracy przedstawione zostanie w odniesieniu do przedstawionych oryginalnie uwag (przedstawione kursywą), które w różnym stopniu zaadresowane zostały w nowej wersji pracy.

Po pierwsze, jak wskazywałem wcześniej wskazać można szereg istotnych zalet przedstawionej pracy (*Z tego punktu widzenia tematyka przedstawionej pracy doktorskiej mgr Katarzyny Zemły, której celem była analiza wpływu użycia technik kierowanej wizualizacji na wskaźniki neurofizjologiczne i poznawcze jest istotny zarówno z punktu widzenia poszerzenia wiedzy akademickiej, jak i potencjalnych klinicznych aplikacji. Jedną z silnych stron przedstawionej pracy jest to, że obejmuje badanie w schemacie eksperymentalnym, w którym w sposób przypadkowy uczestników przydzielono do grupy poddanej punktowo interwencji guided imagery lub proszono o wykonywanie zadania związanego z odpamiętywaniem faktów semantycznych. Dodatkowo, warte odnotowania jest to, że pod pewnymi względami badanie to zostało starannie zaplanowane i przeprowadzone zgodnie z przyjętymi standardami, np. zadbano o homogeniczność próby osób badanych, tak by wykluczyć możliwość wpływu zmiennych socjodemograficznych na obserwowane w badaniu wskaźniki. Z mojego punktu widzenia, biorąc pod uwagę aktualnie promowane w zakresie neuronauki klinicznej standardy pomiaru zjawisk (np. Research Domain Criteria; <https://www.nimh.nih.gov/research/research-funded-by-nimh/rdoc>), istotną zaletą pracy jest również próba ujęcia wpływu omawianych interwencji na zmienne mierzone na różnych poziomach opisu, uwzględniając subiektywny samoopis przy użyciu zmiennych kwestionariuszowych, pomiar zadaniowy przy użyciu paradygmatów celowanych w procesy uwagowe i funkcje wykonawcze oraz analizę wskaźników fizjologicznych (w tym przypadku pomiar mocy pasm alfa i beta obserwowanych w trakcie zadania/GI u uczestników). Z tego punktu widzenia praca obejmuje trzy różne poziomy pomiaru zgodnie z definicjami RDoC i, co godne podkreślenia, Autorka stara się przetestować*

mechanizmy leżące u źródeł obserwowanych wyników poprzez łączenie poziomów pomiaru przy użyciu hipotez mediacyjnych. Dodatkowo, z uwagi na złożoność uzyskiwanych danych, uzasadnione jest rozszerzenie prezentowanego materiału o analizy opierające się na podejściu data-driven i użycie klasyfikatora do analizy wyników zadania.). Ocenę tą podtrzymuję.

Niestety, jak wskazywałem, wątpliwości budził i wymagał zaadresowania szeregu kwestii związanych z aspektami merytorycznymi, metodologicznymi i formalną prezentacją wyników.

Kluczowym problemem, na który wskazywałem była pierwotna konstrukcja autoreferatu (*W sytuacji, gdy praca opiera się na opublikowanych uprzednio materiałach, sensem przedstawionego autoreferatu jest wskazanie na ogólne mechanizmy, których obraz wynika z cyklu przedstawionych badań i nie są możliwe do uchwycenia z poziomu każdej z przedstawionych publikacji z osobna. W tej sytuacji, z mojej perspektywy, sensem takiego autoreferatu jest z jednej strony klarowne wprowadzenie w założenia pracy badawczej, główne problemy badawcze i hipotezy cyklu badań oraz opis mechanizmów, na które wskazuje całość prac przeprowadzonych w ramach cyklu doktorskiego. Niestety w obecnej formie przedstawiony autoreferat nie osiąga żadnego z wymienionych celów i w dużej mierze dubluje treści zawarte w poszczególnych artykułach co prowadzi do istotnych redundancji – biorąc pod uwagę konieczność zaprezentowania w sposób umożliwiający replikację prezentowanych wyników (do zastrzeżeń w tym zakresie odniosę się w dalszej części recenzji), zarówno metodologia jak i szczegółowe wyniki powinny zostać przedstawione na poziomie każdej z włączonych prac. W tej sytuacji przeznaczanie dużej części autoreferatu na powielenie w sposób minimalnie przetworzony informacji przedstawionych w poszczególnych pracach nie wydaje się realizować ogólny cel przedstawienia sensu całości cyklu. W zasadzie większość kluczowych informacji jest w sposób tożsamy zaprezentowana w autoreferacie i kolejnych artykułach a główną modyfikację stanowi zmiana formatowania cytowań). W tym zakresie odnotowuję istotną poprawę – autoreferat zyskał na objętości (jeżeli dobrze rozumiem, w tym momencie trzykrotnie przekracza sugerowaną w tym zakresie długość?). Jednocześnie, mimo iż nie wszystkie z zaprezentowanych treści wydają mi się esencjonalne, to przedstawiona konstrukcja jest dużo bardziej czytelna a pogłębione wątki lepiej (co nie znaczy, że zawsze w sposób wyczerpujący) adresują tematykę.*

Jednocześnie, jak wskazywałem w pierwotnej wersji recenzji, konstrukcja narracji nie zawsze odzwierciedla istotność wątków (*Dodatkowo, istotnym problemem, który obserwować można zarówno na poziomie autoreferatu, jak i poszczególnych artykułów jest sposób konstrukcji narracji, w którym wątki istotne dla zbudowania problemu badawczego wymieszane*

i niejednokrotnie przesłaniane są przez wątki proksymalne i wprowadzone z nie do końca zrozumiałych dla czytelnika przyczyn – czemu służą na przykład wielokrotne nawiązania do psychoonkologii czy też badań z użyciem wirtualnej rzeczywistości? Na podstawie lektury początkowych fragmentów, bogato osadzonych w literaturze dotyczącej psychoonkologii oraz zastosowania technik wirtualnej rzeczywistości w terapii, jak również koniunkcji obu tematów). W poprawionej wersji recenzji tego typu problemy są również obecne (dalej nie rozumiem stopnia obecności wątku VR), choć z uwagi na proporcje i lepsze ich wprowadzenie (e.g. w przypadku wątków onkologicznych), problem ten nie jest aż tak nasilony.

Kolejny problem, na który wskazywałem to istotne braki definicyjne (e.g. *Z tej perspektywy istotnym staje się to czego autoreferat **nie zawiera**. Dla przykładu, w sposób bardzo zdawkowy zostają w nim zdefiniowane kluczowe mechanizmy poznawcze i neuronalne, na poziomie których obserwować można, zgodnie z narracją Autorski, zmiany zachodzące pod wpływem GI. Z perspektywy neuropsychologicznej uderzający jest brak odwołania do modelu teoretycznego funkcji wykonawczych, który stałby za ich operacjonalizacją w projekcie.*). Aktualna wersja pracy jest w tym zakresie zdecydowanie bardziej przemyślana i koherentna niż wersja pierwotna – rozdział 1.5 („Executive Functions, Cognitive Functions, and Attention Control”) istotnie poszerza pierwotnie przedstawiony zakres informacji, jasno wskazując w jaki sposób przedstawione narzędzia odnoszą się do poszczególnych mechanizmów związanych z kontrolą uwagi (s. 11). Niestety wciąż widoczny jest brak jasnej koncepcji teoretycznej, która w sposób jednoznaczny ujednoliciłaby przedstawione i analizowane mechanizmy wykonawcze.

Istotnie rozbudowana została również część wstępu dotycząca mocy pasma alfa i jej zmiany pod wpływem medytacji, co częściowo adresuje podnoszone uprzednio w tym zakresie wątpliwości (*Niestety wprowadzenie w tematykę założeń przyjętych w zakresie analizy wskaźników związanych z mocą poszczególnych oscylacji (alfa/ beta). W jaki sposób są one powiązane z aktywnością mózgu? Jakie są ich neuronalne generatory? W jaki sposób ta aktywność przekłada się na obserwowane na poziomie zachowania czy sprawności systemu poznawczego zmienne? W jaki sposób interwencja miałaby przekładać się na aktywność sieci je generujących, manifestującą się zmianą mocy poszczególnych pasm obserwowaną w EEG?*).

W poprzedniej wersji pracy sygnalizowałem istotne problemy w zakresie konstrukcji postawionych hipotez (*Niestety, praca nie zawiera również jasnego i czytelnego wprowadzenia hipotez. Jeżeli dobrze zrozumiałem intencje Autorki, to główna hipoteza pracy brzmi literalnie „brief GI session can lower stress levels in healthy male individuals who lack prior exposure to such sessions or have a history of chronic health issues.”. Jeżeli tak to oczekiwałbym*

porównania wskaźników związanych w sposób czytelny ze stresem, od miar deklaracyjnych po wskaźniki fizjologiczne jednoznacznie z nim konotowane, tymczasem bezpośrednio przechodzi do opisu eksperymentu (...) Nawet pominięty fakt cyrkularności przedstawionej hipotezy (...) tego typu dyskurs, w ramach którego w jednocześnie omawia hipotezę, jej operacjonalizację, aspekty metodologiczne i stricte techniczne związane z jej weryfikacją (a momentami również dostępne w literaturze wyniki), stawia pod znakiem zapytania sformułowania problemu badawczego i powiązanych z nim pytań/hipotez.). W porównaniu z pierwotną wersją pracy, w aktualnie przedstawionym autoreferacie istotnie bardziej czytelnie nakreślone są hipotezy badania i ich związek z wcześniejszą literaturą. Jednocześnie hipotezy nie są wciąż sformułowane jednoznacznie i explicite, co utrudnia jednoznaczne odniesienie się do nich i, w niektórych przypadkach określenie kierunkowości. Szczególnie widoczny jest ten problem w przypadku hipotez mediacyjnych.

Zaadresowane zostały wskazane usterki edycyjne pracy, związane m.in. z użyciem anglicizmów czy konstrukcją niektórych zdań. Poprawione zostały również niektóre problemy z formatowaniem cytowań, choć przegląd bibliografii wciąż pokazuje, że nie wszystkie referencje sformatowane zostały w pełni zgodnie z APA. Niestety nie zostały również zaadresowane niektóre problemy edycyjne wskazane w poprzedniej recenzji (*'Interpolatiion' jako faze preprocessingu w jednej z wizualizacji*) czy dublowanie dużych fragmentów tekstu, (wielokrotne redundancje w zakresie omówienia kryteriów włączenia i wyłączenia osób baanych (...))Niestety dublujące się sekcja, w których te same czynniki omawiane są w sekcjach *'inclusion criteria'* i *'exclusion criteria'* obecne są zarówno w autoreferacie, jak i poszczególnych artykułach). Siłą rzeczy dalsze użycie tych samych artykułów prowadzi do zgłaszanych już poprzednio problemów z fragmentami, których styl i forma rodzi duże wątpliwości co do jakości procesu recenzyjnego w czasopismach, do których zgłoszono wyniki (e.g. otwierający abstrakt artykułu numer 2: *Coronavirus disease 2019 (COVID-19) has caused everything from daily hassles, relationship issues, and work pressures to health concerns and debilitating phobias.*).

Istotnie pełniejszy opis procedury eksperymentalnej przedstawiony w nowej wersji pracy w dużym stopniu adresuje wzmiankowane poprzednio niedostatki opisu metodologii (*Kolejnym, problematycznym aspektem pracy jest opis użytej metodologii – złotym na standardem opisu tego typu sekcji byłoby zaprezentowanie odpowiedniego materiału wraz z pełnym protokołem zapewniającym możliwość replikacji zarówno warunków eksperymentu, jak i wszelkich analiz, zaś przy obecnych standardach dyskursu naukowego tego typu protokół powinien umożliwić pełną replikowalność wyników przez niezależne grupy badawcze (por.*

Pavlov i in., 2021). Jednocześnie w dalszym ciągu brak w niej istotnych informacji, podobnie jak poprzednio, autoreferat skupia się na opisie globalnych zalet klasyfikatora GLM niż szczegółach użycia go w ramach prowadzonej pracy (Niestety, opis przedstawiony przez Autorkę jest na tyle szeroki i nieprecyzyjny, że wielokrotnie w trakcie analizy zastanawiałem się nad szczegółami metodologii czy analizy nie znajdując odpowiednich informacji w pracy, lub rekonstruując je z okruszków informacji zawartych w tekście – przykładowo, Autorka wielokrotnie wskazuje, że do klasyfikacji użyty został ‘GLM classifier’, który zaimplementowany został w bibliotece h2o. Intuicyjnie założyłem, że chodzi o zwykłą regresję logistyczną zaimplementowaną w tym pakiecie, choć informację, którą z dziesięciu różnych opcji ‘klasyfikacji GLM’, choć informację ujednoznaczniającą w tym zakresie znalazłem dopiero w sekcji ‘The 14th min Choice Justification’ ostatniej z przedstawionych prac.). W mocy pozostają też uwagi dotyczące wątpliwości dotyczące materiału użytego do analizy (w pracy nr 2 mowa jest o tym, że „All EEG recordings were obtained using a 256-channel dense-array EEG amplifier” oraz „The signal was classified using generalized linear models (GLMs) using the implementation included in the h2o library available for Python.”. Koniunkcja obu tych informacji sugerowałaby, że do analizy użyto sygnału z 256 źródeł. Tymczasem praca nr 3 informuje, że „However, considering our prior expertise [80–82] in analyzing cognitive processing EEG signals, we anticipated detecting variations specifically on the designated cognitive electrodes.” w związku z czym sygnał zbierano na 27 elektrodach. Pomijając fakt, iż pierwszy raz w swojej karierze naukowej spotykam się z pojęciem ‘cognitive electrodes’ i wnioskując po referencjach cytowanych w tym zakresie, nie jest to konstrukt powszechnie używany, to po lekturze całości pracy dalej nie mam dalej pewności co stanowiło sygnał użyty do klasyfikacji.). Doceniam jednocześnie próbę wyjścia z sytuacji, w której opis aparatury w poszczególnych artykułach zawiera fragmenty nie dotyczące w żaden bezpośredni sposób opisywanych badań (praca nr 2: *The event-related potential (ERP) experiments were designed in PST e-Prime 2.0.8.90*), poprzez poinformowanie w autoreferacie, iż „PST e-Prime 2.0.8.90 is utilized for designing the ERP experiments.”. Roszerzenie opisu interwencji w dużym stopniu wyjaśniło wzmiankowane poprzednio wątpliwości, choć dalej stoję na stanowisku iż optymalne byłoby załączenie scenariusza tejże do pracy w formie załącznika. Niestety w dalszym ciągu nie znajduję w pracy wzmiankowanych poprzednio:

- informacji o sposobie przygotowania danych behawioralnych (e.g. w zakresie cenzorowania odstających czasów reakcji w zadaniach behawioralnych), przygotowania i analizy danych okulograficznych.
- rekompensaty/nagrody dla osób badanych

- pozyskiwania zgody opiekunów od osób niepełnoletnich biorących udział w badaniu
- zastosowania poprawki na wielokrotne porównania przy analizach korelacyjnych

Niestety w przedstawionej wersji pracy nie zostały zaadresowane ani w formie uzupełniających nowych analizy ani też skomentowane w formie potencjalnych ograniczeń wskazywane przeze mnie poprzednio istotne problemy związane z podejściem analitycznym przyjętym w pracy: *Opisane w pracy podejście bazujące na użyciu data-driven klasyfikatora w celu znalezienia najbardziej istotnej różnicy między grupami w zakresie wskaźników neuronalnych, po czym użycie korespondujących z nią wskaźników do sprawdzenia post-hoc różnic między grupami (co przedstawione jest jako zaplanowana analiza przewidzianego efektu: We found predicted, significant effect of group ($F(1, 53) = 4.01, p = 0.05, p_2 = 0.070$), jak również dalsze korelacje ze zmiennymi behawioralnymi mierzonymi w ramach zadania to dość mocny przykład krytykowanego szeroko na gruncie neuronauki poznawczej podejścia opartego na double dipping oraz circular correlations (niestety w języku nie funkcjonuje odpowiednia nomenklatura; por. Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S., & Baker, C. I. (2009). Circular analysis in systems neuroscience: the dangers of double dipping. *Nature neuroscience*, 12(5), 535-540; Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on psychological science*, 4(3), 274-290.). (...)* Dodatkowo, z uwagi na brak niezależnego nowego zbioru danych, na którym przeprowadzono by sprawdzian krzyżowy (cross validation) klasyfikatora istotnie ogranicza jego moc predykcyjną.

Również krytyka dotycząca doboru interwencji kontrolnej, została zaadresowana tylko częściowo – mimo zapewnień iż obie interwencje były ekwiwalentne pod kątem przebiegu, nowa wersja pracy wciąż nie odnosi się do kwestii przedstawionych poprzednio w tym zakresie poprzednio: *Obciążenie poznawcze w zadaniu kontrolnym jest istotnie większe niż w przypadku interwencji GI, co rodzi pytanie czy jakkolwiek zaskakującym wynikiem jest to że wzorec neuronalny odpamiętywania informacji osadzonych w pamięci semantycznej różni się od GI? Czy obecność różnic pomiędzy wzorcem aktywności neuronalnej obserwowanym przy użyciu technik neuroobrazowania pomiędzy stanem relaksacji a wykonywaniem zadania pamięciowego byłby jakkolwiek zaskakujący?*

Ponownie odnosząc się do strategii publikacyjnej, zaznaczę iż w kontekście wszystkich mankamentów pierwszej z prac cyklu (*Pierwsza z publikacji zaprezentowanych w cyklu stanowi bardzo luźny dyskurs, pełen kolokwializmów, informacji opierających się na popularno-naukowych źródłach i informacji których nie sposób osadzić w procesie naukowym („Healing is located within the patient. The therapist has no secret powers to control or*

heal.”). Nie próbuję nawet punktować wszystkich budzących moje wątpliwości co do jakości procesu recenzji (zarówno merytorycznych jak i stricte edycyjnych, o czym świadczyć może choćby nieprawidłowe zatytułowanie kluczowego rozdziału) fragmentów tejże publikacji. Przy ogromnej dozie dobrej woli nie znajduję uzasadnienia dla włączenia w cykl publikacji opartej na zapisie pojedynczej sesji relaksacji, którą doktorantka (KZ) przeprowadziła promotorowi (GW) analizując przy tym wizualnie wykresy dotyczące zmian mocy EEG czy wskaźników zebranych przy użyciu smartbanda Xiaomi.), nie rozumiem podtrzymania decyzji o włączeniu jej do cyklu. Analogicznie – w tej sytuacji dołączenie do pracy preprintu pracy będącej w recenzji w Scientific Reports i dotyczącego zagadnień w małym stopniu związanych z dyscypliną psychologii, w którym autorka cyklu jest piątą autorką z sześciu powinno zostać w dużo większym stopniu zaadresowane niż ma to aktualnie miejsce w autoreferacie.

Biorąc pod uwagę całokształt zmian w przedstawionej wersji pracy daleki jestem od jednoznacznej oceny. Z jednej strony Autorka, szczególnie w części początkowej w dużym stopniu zaadresowała przedstawione uwagi, wprowadzając przy tym do pracy w sposób dodatkową literaturę i konceptualizując uzyskane wyniki w sposób zdecydowanie bardziej kompleksowy niż uprzednio. Z drugiej strony, szczególnie w warstwie analitycznej, wiele z poprzednio sygnalizowanych wątpliwości pozostało niezaadresowanych, nawet na poziomie dyskusji o potencjalnych ograniczeniach z nich wynikających, nawet jeżeli kompleksowa re-analiza nie byłaby możliwa z uwagi na ramy wytyczone przez format pracy. Mimo ogromnych wątpliwości w tym zakresie z uwagi na niemożność oczekiwania dalszych działań w cyklu poprawek, uznaję że w przedstawionej formie praca w minimalny sposób spełnia kryteria stawiane pracy doktorskiej w zakresie wyznaczonym przez Ustawodawcę.