

Dr hab. Andrzej Rynkiewicz

Uniwersytet Warszawski

Wydział Psychologii

Recenzja pracy doktorskiej mgr Patrycji Chwiłkowskiej pt. „The Influence of Positive Emotions on Cardiovascular Recovery in Esports”

Redukcja stresu nie zawsze jest możliwa. Zatem ważnym celem badań naukowych staje się redukcja konsekwencji stresu lub przyspieszenie procesu powrotu do stanu spoczynkowego po przeżytym stresie (recovery). Ten temat porusza praca doktorska mgr Patrycji Chwiłkowskiej. W obliczu trudności związanych z badaniem stresu w warunkach naturalnych, Autorka zdecydowała się na symulację, która jednak wychodzi nieco poza sztuczne warunki laboratoryjne, bo wykorzystuje środowisko popularnej gry komputerowej Counter Strike. Scenariusz gry osadzony jest w warunkach bitwy militarnej i wymaga od gracza m.in. zręczności, spostrzegawczości, planowania oraz współpracy w zespole. Przez niektórych gra postrzegana jest jako brutalna, bo obejmuje konieczność symulowanego zabijania ludzi. Jest to więc sztuczne odzwierciedlenie silnie stresującego pola walki ale także aranżacja rywalizacji o charakterze sportowym. Uczestnicy badania zaangażowali się w kilkunastominutową rozgrywkę CS, a później prezentowano im kilka krótkich filmów o określonym zabarwieniu afektywnym. W trakcie całej procedury mierzono wskaźniki psychofizjologiczne powiązane z aktywacją współczulną i przywspółczulną. Ponadto na początku i na końcu procedury zmierzono poziom różnych wskaźników afektywnych za pomocą kwestionariuszy. Jest to więc typowa procedura eksperymentu z manipulacją międzygrupową oraz dwukrotnym pomiarem zmiennych zależnych. Autorka oczekiwała, że wskaźniki stresu wzrosną u wszystkich uczestników pod wpływem gry i będą systematycznie opadać po jej zakończeniu, ale spadek ten będzie szybszy i głębszy w grupie, której zaprezentowano filmy o zabarwieniu emocjonalnym przyjemnym.

Chociaż procedura eksperymentu jest dość typowa, to warto zwrócić uwagę na zastosowanie symulowanego grą komputerową środowiska – to zabieg niezbyt często stosowany w badaniach naukowych. Ma on swoje zalety i wady. Jest to sposób na wprowadzenie do laboratorium procedur o trochę podwyższonej trafności ekologicznej. Z drugiej strony, złożony scenariusz gry i różny jej przebieg ma te same wady, co badania terenowe – nie sposób dokładnie kontrolować sytuacji badawczej. Nie mniej, uważam ten zabieg za wartościowy element badania.

Warto również zwrócić uwagę na wysiłek poniesiony w celu poszerzenia puli wskaźników o psychofizjologię. Autorka zastosowała pomiar akcji serca, ale również wcale nietatwą w użyciu i późniejszej analizie kardiografię impedancyjną. W rezultacie otrzymano szereg cennych wskaźników obrazujących tzw. przestrzeń autonomiczną: HR, PEP, CO i SV. Dodatkowo Doktorantka poddała elektrokardiogram przekształceniom wydobywając z niego modny współcześnie wskaźnik zmienności pracy serca w domenie czasowej (RMSSD). Ten zestaw miar pokazuje wprawę Autorki w zakresie pomiaru oraz analizy danych psychofizjologicznych.

Do zalet pracy należy również analiza szybkości powrotu stanu uczestników badania do poziomu spoczynkowego po zakończonej rozgrywce (recovery). Autorka zastosowała tu pomysłowe wskaźniki zaczerpnięte z innych prac. Były to: czas powrotu wskaźników fizjologicznych do poziomu baseline (time-to-baseline) oraz czas powrotu do określonego progu (time-to-threshold), a także wskaźniki nieparametryczne bazujące na analizie częstości. To część analiz zasługująca na uwagę i pokazująca niezły warsztat badawczy Doktorantki. W moim odczuciu można było pójść nawet dalej i przy tak dużej liczbie równomiernie rozłożonych punktów pomiarowych wyznaczyć linię regresji lub inny matematyczny wskaźnik z obszaru analizy szeregów czasowych. W ten sposób dałoby się odczytać coś w rodzaju trendu zmian pomijając chwilowe wahania stanu fizjologicznego. Nie mniej, zastosowane wskaźniki w wystarczającym stopniu zapewniają odfiltrowanie spontanicznych fluktuacji krótkookresowych mogących zaburzać obraz efektów.

Praca napisana jest bardzo przejrzystym językiem. Przy tak wielu wskaźnikach z różnych obszarów bardzo pomocna jest konsekwentna terminologia, o co skrupulatnie zadbała Autorka. Praca zawiera ponad 60 stron opisu teoretycznego, niecałe 30 stron opisu metody i wyników oraz 10 stron wniosków. Te proporcje nie są optymalne i wskazują na duży wysiłek włożony w przygotowanie badania, ale mniejszy włożony w interpretację jego wyników. Pewnym usprawiedliwieniem może być fakt, że badanie empiryczne nie potwierdziło stawianych hipotez, ale mimo to od rozprawy doktorskiej oczekuje się znacznie bardziej wnikliwej analizy zebranych obserwacji i wniosków.

Przedstawione badanie koncentruje się w zasadzie na dwóch mechanizmach. Z jednej strony ważne jest to, co działo się podczas rozgrywki, a z drugiej to, co działo się po jej zakończeniu. W obu przypadkach mam pewne uwagi do zaproponowanych modeli teoretycznych oraz wskaźników.

Stan psychiczny odczuwany przez uczestników **podczas rozgrywki** potraktowany został dość chaotycznie. Rozbudowane wprowadzenie teoretyczne prezentuje różne modele stresu, wyzwania, mobilizacji i oddziaływania emocji pozytywnych. Niektóre z nich zawierają komponent fizjologiczny, inne nie. Niektóre bardziej pasują do wyjaśnienia zjawisk psychicznych dziejących się w czasie rozgrywki, a inne w czasie po rozgrywce. Z opisów tych nie wynika jednak postulowany obraz psychicznych zmian podczas inicjowanych zdarzeń. Dalej w treści dominuje podejście, że stan w czasie rozgrywki, to raczej stres. Jednak trudno w pełni zgodzić się z założeniem, że esport jest źródłem stresu, a tym bardziej, że jest źródłem nieprzyjemnych emocji. Fakt, że gra CS2 ma tak wielu sympatyków świadczy, że gracze chętnie do niej powracają. Jest to więc raczej rodzaj wyzwania, które np. Lazarus włącza w definicję stresu, ale pojawia się ono również w koncepcjach motywacji (np. Brehm i Self, 1989). Nie chcę przesądzać, czy stosując grę komputerową Doktorantka badała stres, czy nie badała stresu, ale chcę zwrócić uwagę, że problem ten nie został w pracy należycie omówiony. W rezultacie dla opisu odczuć uczestników gry pojawiają się zamiennie określenia: stres, emocje negatywne, ekscytacja, wyzwanie, mobilizacja, zaangażowanie. Autorka zwraca uwagę na pojawiający się dylemat w rozdziale teoretycznym opisując tzw. Biopsychosocial Model of Challenge and Threat – z jednej strony wyzwanie uruchamia motywację dążenia, a z drugiej strony stan zagrożenia uruchamia motywację unikania. Ale nie przekłada się to na teoretyczne konkluzje, chociaż to właśnie ten model wydaje się najbardziej adekwatnie pasować do kontekstu gry komputerowej. Warto podkreślić, że to nie sama gra wzbudza stres, czy jest traktowana jako wyzwanie, ale działają w ten sposób różne jej komponenty. Doktorantka je wymienia – za stres odpowiedzialne są rywalizacja, zagrożenie problemami technicznymi, nieporozumienia w zespole graczy. W tej konkretnej sytuacji wszystkie one zostały wyeliminowane: a) gra była realizowana w

wariancie z komputerowymi botami, a więc bez rywalizacji z drugim człowiekiem; b) gra indywidualna, a więc pozbawiona specyfiki współpracy w zespole; c) gra na cudzym komputerze, a więc brak odpowiedzialności za ewentualne problemy techniczne. Symulacyjna gra pozbawiona źródeł stresu w dalszym ciągu jednak obejmowała różne źródła motywacji – realizacja zadania, potwierdzenie swoich kompetencji, zdobycie punktów itp. Efekt ten widać w tabeli 4, która przedstawia poziom różnych wymiarów afektywnych przed rozpoczęciem gry – dominuje motywacja, pozytywny afekt i ekscytacja, a stres i lęk mają raczej umiarkowany poziom. Autorka oczywiście nie przeoczyła tych danych i odnosi się do nich w dyskusji, sugerując, że mogą być odpowiedzialne za porażkę w weryfikacji postawionych hipotez. Uważam jednak, że efekt ten dało się przewidzieć wcześniej i tak zaplanować badanie, by nie było ono tylko replikacją w zmienionych okolicznościach, ale żeby szło naprzód proponując bardziej odważne podejście do modelu - np. czy *undoing effect* odnosi się w równej mierze do *recovery after stress*, co do *recovery after challenge*. Chcę zakończyć ten wątek konkluzją, że zaprezentowane badanie stanowi w mojej opinii dobry pilotaż i mogłoby być źródłem ciekawych hipotez dla dalszego poszukiwania, ale tak się nie stało, bo cały program badawczy na tym jednym eksperymencie się zakończył.

Stan psychiczny uczestników **po zakończeniu rozgrywki** doczekał się bardziej uporządkowanej analizy, chociaż też mam do opisu pewne zastrzeżenia. Problem wzbudza używane powszechnie w rozprawie określenie *physiological recovery*. W przypadku zmian fizjologicznych wywołanych stresem lub emocjami negatywnymi chodzi o powrót wskaźników do poziomu spoczynkowego. Jednak w przypadku zmian wywołanych emocjami pozytywnymi charakteryzującymi się dużym pobudzeniem już niekoniecznie. W obu przypadkach zmiany fizjologiczne mogą mieć podobny kierunek, np. przyspieszenie akcji serca i akcji oddechowej, spadek HRV itp. Jednak w jednej sytuacji są one równoległe z negatywnymi odczuciami, a w drugiej z pozytywnymi. Czy w tym drugim przypadku też mówimy o *physiological recovery*? W zaprezentowanych badaniach zastosowano grę komputerową wywołującą pobudzenie fizjologiczne oraz stan emocjonalny, który mógł być albo negatywny (stres) albo pozytywny (ekscytacja, mobilizacja). Nie ma pewności, który z tych stanów dominował, ale dane przedstawione w tabeli 4 świadczą, że był to raczej stan przyjemnej mobilizacji. Dlatego zjawiska, które nastąpiło po zakończeniu gry nie nazwałbym określeniem *recovery*. A jeśli nie jest to *recovery*, to dlaczego miałyby ten proces wspierać prezentowanie filmów o zabarwieniu pozytywnym?

Tak jak zaznaczyłem wcześniej, mimo pewnych zastrzeżeń, faza oddziaływań eksperymentalnych po zakończeniu rozgrywki wydaje się opisana dość porządnie. Przytoczone teorie emocji pozytywnych sensownie wpisują się w model redukcji stresu (stres, to oczywiście założenie Doktorantki). Bardziej jednak odnoszą się do płaszczyzny psychicznej (dla mnie to ważne), niż do fizjologicznej (dla Doktorantki to ważne, co sugeruje tytuł rozprawy). Pewnym łącznikiem efektu afektywnej redukcji stresu z poziomem fizjologicznym może być koncepcja niejednorodności nerwu błędnego (polyvagal theory, PVT) sformułowanej przez Stephena Porges. Zgodnie z nią nerw błędny ma dwie gałęzie – brzuszną i grzbietową. Gałąź brzuszna odpowiedzialna jest za regulowanie aktywności wegetatywnej wspierającej zachowania społeczne (np. zmiana intonacji głosu, zmiana napięcia mięśni twarzy itp.). Reguluje ona również oddechową modulację zmian akcji serca nazywaną arytmia oddechową (respiratory sinus arrhythmia, RSA). Zdaniem Doktorantki aktywizacja tej właśnie gałęzi nerwu błędnego powinna być głównym mechanizmem prowadzącym do redukcji niekorzystnych konsekwencji stresu. Warto dodać, że koncepcja ta ma w środowisku psychofizjologów również wielu zwolenników, co i przeciwników. Do tej drugiej grupy należy między innymi jeden z bardziej uznanych badaczy zmienności akcji serca - Paul Grossman. Wypunktował on skrupulatnie wady koncepcji

Porgesa opierając się na analizie zjawiska arytmii oddechowej. Doktorantka w rozprawie przedstawiła koncepcję i zestawiła ją ze swoimi oczekiwaniami badawczymi. Jednak z racji bardzo żywego sporu toczącego się w środowisku wokół tej popularnej teorii (pojawiają się nawet publiczne zarzuty o pseudonaukowość) oczekiwałbym od Doktorantki również dyskusji z uwagami formułowanymi wobec tej koncepcji. Ponadto, w kontekście koncepcji Porgesa nieadekwatne wydaje się zastosowanie wskaźnika RMSSD do oceny zmienności akcji serca, jeśli wiadomo o zastrzeżeniach wobec niego. Uważa się bowiem, że nie jest on zbyt dobrą miarą RSA i jego wartość potrafi się zmniejszyć, gdy RSA się zwiększa np. przy spowolnionym oddychaniu. Nieco lepsze wskaźniki RSA dałoby się przecież znaleźć wśród miar HRV analizowanych w domenie częstotliwościowej. Ich użyteczność znacznie by jeszcze wzrosła, gdyby uwzględnić częstotliwość akcji oddechowej obliczonej choćby na podstawie surowego sygnału EKG. W pracy dostępny byłby wtedy wskaźnik pozwalający połączyć zjawiska obserwowane w fazie powrotu do stanu spoczynkowego (*recovery*) z mechanizmami postulowanymi w koncepcji Porgesa.

Inny problem wiąże się z faktem, że w eksperymencie indukowano stan afektywny za pomocą filmów. To uznany sposób manipulowania emocjami. Jednak w tej konkretnej sytuacji zmiany fizjologiczne mogą być konsekwencją dwóch czynników: afektywnego, czyli powrotu do stanu spoczynkowego po rozgrywce oraz poznawczego, czyli ukierunkowania uwagi na interesujące bodźce zewnętrzne. Tego drugiego zjawiska Badaczka nie analizowała, a wiadomo, że pojawienie się ważnego bodźca w polu percepcji powoduje zwykle spowolnienie akcji serca i wzrost RMSSD, czyli dokładnie taki sam wzorzec reakcji, jak przy oczekiwanej *recovery*. W moim odczuciu oba efekty nałożyły się na siebie, a biorąc pod uwagę, że oba mają odmienne mechanizmy ale podobne konsekwencje fizjologiczne, mogło to spowodować zniwelowanie różnic międzygrupowych – grupy różniły się rodzajem filmów, ale obie oglądały filmy. Być może w ten sposób można wytłumaczyć widoczne na wykresie 7 nieoczekiwane zmiany RMSSD w fazie *recovery* – wskaźnik ten gwałtownie rośnie po rozpoczęciu prezentacji filmów, a potem się obniża. Na tym samym wykresie widać też niewielkie, ale systematyczne wahania HR zsynchronizowane z momentami prezentacji kolejnych filmów. Szkoda, że ten efekt nie został w rozprawie przedyskutowany, bo wnioski mogłyby stanowić wskazówkę dla przyszłych badań w podobnym paradygmacie.

Oczekiwane efekty eksperymentalne są zadziwiająco słabe – kluczowe testy F są poniżej 1. Nie uważam, że negatywny wynik eksperymentu jest porażką badacza – w psychofizjologii to niestety bardzo częsty rezultat badawczy. Dla psychofizjologa nie jest też zaskakującym, że w odróżnieniu od efektów dla wskaźników fizjologicznych, efekty dla wskaźników kwestionariuszowych i behawioralnych są bardzo silne. Mikroskopijne wielkości efektów fizjologicznych mogą mieć szereg różnych przyczyn. Same pomiary kwestionariuszowe i fizjologiczne są raczej bez zarzutu, bo widać wyraźny efekt fazy badania – wskaźniki mocno się różnią w fazie gry i w fazie *recovery*. Przyczyny tak słabych efektów należy więc szukać w przesłankach hipotez. W rozprawie niestety nie widać takich poszukiwań, bo Autorka w dyskusji koncentruje się głównie na mankamentach technicznych. Naturalną konsekwencją eksperymentu z wynikiem niepotwierdzającym hipotez są kolejne hipotezy i kolejne eksperymenty aż do momentu, gdy uznamy, że temat jest wyczerpany. Trudno jednak to uznać po jednym eksperymencie. Uważam więc, że Doktorantka nie wykazała wystarczającej dociekliwości badawczej, nawet jeśli uwzględnimy tutaj ograniczenia techniczne skomplikowanego eksperymentu i organizacyjne przewodu doktorskiego. Być może podczas obrony Doktorantka mogłaby się pokusić o sformułowanie hipotez *post factum*, które stanowiłyby konkluzję z eksperymentu niepotwierdzającego oczekiwań.

Procedura badawcza bez wątpienia została przygotowana skrupulatnie. To samo można powiedzieć o doborze próby, selekcji danych oraz samych analizach. Powiedziałbym nawet, że te elementy rozprawy są wzorcowe. Wyłapałem jednak kilka mankamentów metodologiczno-technicznych, które krótko omówię.

Pomiary ECG i ICG były realizowane w odmienny sposób na bazie założenia, że dane ICG należy uśredniać dla odcinków 1-minutowych. Nie są dla mnie jasne przyczyny takiej decyzji. Zarówno wskaźniki wydobywane z ECG (np. HR, IBI), jak i wydobywane z sygnałów dostarczanych przez ICG (np. PEP) można obliczać dla każdego cyklu pracy serca. W tym kontekście próbkowanie z różną częstością uważam za nierozsądne.

W analizach zmian poziomu HR i PEP w kolejnych fazach eksperymentu Doktorantka użyła testu Fishera (anova). Jednak dalej wykonując porównania parami średnich w sąsiadujących fazach (np. *baseline i prematch*) nie dokonała typowej korekty przez zastosowanie testów post-hoc lub zmniejszenia alfy. Nie są jasne przesłanki dla tego nietypowego postępowania, które może prowadzić do zwiększania się ryzyka błędu I rodzaju.

Stan afektywny uczestników badania określono za pomocą dwóch narzędzi – Differential Emotions Scale (ocena stanów afektywnych) i skali SAM (ocena wymiarów afektywnych). W drugim podejściu zdecydowano się na pojedyncze pytanie do każdej z trzech skal – znaku, pobudzenia i motywacji. Mam wątpliwości, czy takie rozwiązanie pozwala utrzymać rzetelność pomiaru w akceptowalnych granicach. Jeśli wykorzystuje się narzędzia zawierające wiele pozycji, to błąd losowy pojawiający się w każdej odpowiedzi uśrednia się i niweluje przy obliczaniu sumy wyników – im dłuższe narzędzie, tym całkowity błąd pomiaru mniejszy. Nie jest zatem jasne, dlaczego Doktorantka zdecydowała się na zastosowanie pomiaru za pomocą pojedynczej pozycji, jeśli łatwo dostępne są alternatywy o większej rzetelności, np. UMACL.

Zmienne time-to-baseline i time-to-threshold ujawniły rozkłady skośne utrudniające analizy metodami parametrycznymi. W tej sytuacji Doktorantka zdecydowała, by poddać je logarytmowaniu, co doprowadziło do normalizacji rozkładów. To dość kontrowersyjny zabieg, bo skutkiem ubocznym jest nierównomierna czułość takich miar – po logarytmowaniu różnice w niskim zakresie są przeszacowane, a różnice w wysokim zakresie niedoszacowane. Pojawia się pytanie, dlaczego w tej sytuacji nie zastosowano rangowania i testu nieparametrycznego.

Podsumowując, chcę stwierdzić, że brak satysfakcjonującego wyniku badań eksperymentalnych nie może być podstawą negatywnej oceny badacza, bo eksperymenty realizuje się po to by weryfikować hipotezy, a nie je potwierdzać. Jednak niepomysłny eksperyment powinien być dla badacza impulsem do przemyśleń i do dalszych badań. Jednego i drugiego tu zabrakło, co uważam za poważnym mankament. Ale ustawa o stopniach naukowych z 14 marca 2003 r. w art. 13 wymaga, by rozprawa doktorska stanowiła oryginalne rozwiązanie problemu (przynajmniej próbę), pokazywała wiedzę i umiejętność prowadzenia badań. Przetawiona mi rozprawa mgr Patrycji Chwiłkowskiej spełnia te kryteria, bo pokazuje potencjał badawczy Doktorantki, sprawność metodologiczną i operacyjną oraz wiedzę z opisywanego zakresu. Zatem wnioskuję o wszczęcie dalszych etapów przewodu doktorskiego.

Andrzej Ryzulian