

# Nieco danych świadczących o tym, że Dutton i Aron 40 lat temu wykazali niekoniecznie to, co zamierzali

Katarzyna Szczucka

*Szkoła Wyższa Psychologii Społecznej, Wydział Zamiejscowy we Wrocławiu*

Artykuł Donalda Duttona i Arthura Arona (1974), przedstawiający wyniki trzech badań, można zaliczyć do kategorii klasyków psychologii społecznej. Zgodnie z interpretacją autorów, wystarczy zaindukować u mężczyzny silne pobudzenie autonomiczne (*via* przejście przez wysoki i niestabilny most lub antycypację otrzymania silnych szoków elektrycznych), aby osiągnąć efekt w postaci zwiększenia pociągu seksualnego do obiektu (atrakcyjnej kobiety), który pojawia się tuż po lub przed doznaniem awersywnego bodźca. Tymczasem interpretacje poczynione przez autorów wydają się co najmniej wątpliwe, uzyskane zaś przez nich dane – niekonkluzywne. W niniejszym artykule prezentuję krytyczną interpretację rezultatów uzyskanych przez Duttona i Arona oraz wyniki przeprowadzonych przeze mnie ponownych analiz danych przedstawionych w ich artykule. Ponadto przedstawiam krótki przegląd teorii odnoszących się do związku pobudzenie–atrakcyjność, jak również rezultaty metaanalizy.

Słowa kluczowe: *związek pobudzenie–atrakcyjność, błędne przypisanie źródła pobudzenia, błędne interpretacje, Dutton i Aron*

*Utram bibis? Aquam an undam?*

(John Fowles, *The Magus*)

Jednym z klasycznych artykułów, którego wyniki są znane zapewne większości studentów pierwszego roku psychologii, jest praca Donalda G. Duttona i Arthura P. Arona, która została opublikowana w 1974 roku w najbardziej prestiżowym czasopiśmie naukowym z zakresu psychologii społecznej – *Journal of Personality and Social Psychology*. Jest to artykuł, który można zaliczyć do kategorii *hits* – tzn. jest egzemplifikacją takiej pracy naukowej, która nie tylko tuż po publikacji, ale również w dłuższym czasie przykuwa uwagę uczonych, co bezpośrednio przekłada się na wskaźniki bibliometryczne (por. Lange, 2005). Wzmianka o tym artykule pojawia się w niemal każdym podręczniku psychologii społecznej (por. np. Aronson, Wilson, Akert, 1997; Hogg, Vaughan,

2011; Kenrick, Neuberg, Cialdini, 2002; Rogers, 2003; Wojciszke, 2011) w kontekście błędnego przypisania źródła pobudzenia (*misattribution of arousal*).

W mojej pracy chcę wykazać, iż wyniki badań Duttona i Arona (1974): (1) nie pozwalają na jednoznaczną konstatację; (2) są w gruncie rzeczy przykładem tzw. *negative results* (por. Lehrer, Leschke, Lhachimi, Vasiliiu, Weiffen, 2007). Dostrzeżone błędy chcę przedstawić, dzieląc je na: (1) błędy odnoszące się do irrelevantnych (lub nieadekwatnie interpretowanych) ram teoretycznych; (2) błędy metodologiczne; (3) błędy statystyczne; (4) błędy w interpretacji uzyskanych wyników.

## **BADANIE 1 DUTTONA I ARONA**

Celem autorów badania (Dutton, Aron, 1974) było wykazanie, iż osoby po przejściu przez Capilano Suspension Bridge (chwiejny, niestabilny most linowy o szerokości 1,5 m, długości 140 m, przerzucony nad rzeką na wysokości 70 m), indagowane przez atrakcyjną pomocnicę eksperymentatora, będą odczuwały większe pobudzenie seksualne niż osoby z grupy kontrolnej (przechodzące przez

---

Katarzyna Szczucka, Szkoła Wyższa Psychologii Społecznej Wydział Zamiejscowy we Wrocławiu, ul. Ostrowskiego 30, 52-238 Wrocław, e-mail: katarzyna.szczucka@swps.edu.pl  
Chciałabym złożyć wyrazy wdzięczności mojemu Mistrzowi, profesorowi Dariuszowi Dolińskiemu, za inspirację i zachętę, nie raz wyrażaną, do napisania owego artykułu, wsparcie oraz niezmiennie indukowanie wiary we własne możliwości.

most stabilny, przebiegający na wysokości 3 m nad niewielkim strumieniem). Zmienne zależne, będące wskaźnikiem pobudzenia seksualnego, były następujące: (1) Test TAT – liczba i jakość słów o konotacji seksualnej zawartych w opowiadaniu pisany przez osoby badane po zaprezentowaniu im obrazka 3GF z TAT. Opowiadania były oceniane przez sędziów kompetentnych (porangowane od 1 – *brak treści o konotacji seksualnej* do 5 – *wysokie nasycenie treściami o konotacji seksualnej*) (była to implementacja procedury Barclaya i Habera, 1965); (2) dwójki rodzaju dane behawioralne (*behavioral data*), będące w zamyśle autorów wskaźnikiem wzrostu percypowanej przez osoby badane atrakcyjności pomocnicy eksperymentatora w warunkach uprzednio odczuwanego silnego pobudzenia: ZZ1 – zgoda na wzięcie numeru telefonu od eksperymentatorki; ZZ2 – zadzwonienie do eksperymentatorki. Osobami badanymi byli wyłącznie mężczyźni, natomiast osobami indagującymi – kobieta i mężczyzna (w obu warunkach: most chybliwy vs. stabilny).

### Wątpliwości natury teoretycznej

Zacznijmy od tego, iż Donald Dutton i Arthur Aron (1974), próbując znaleźć ramy teoretyczne dla swoich badań nad związkiem między pobudzeniem a atrakcyjnością (*arousal-attraction link*), powołują się na jedną z teorii nurtu atrybucyjnego dwuczynnikową teorię emocji (*two-factor theory of emotion*) Stanleya Schachtera i Jerome'a Singera (1962). Zgodnie z założeniami owej teorii, warunkami koniecznymi doświadczania konkretnej emocji (tu: pobudzenia o charakterze seksualnym) są: (1) niejednoznaczne (*ambiguous*) faktyczne źródło pobudzenia oraz (2) wskazówka środowiskowa, uzasadniająca i określająca, jaką emocję podmiot przeżywa. Tymczasem w badaniach Duttona i Arona źródło pobudzenia było zawsze jednoznaczne (chwiejny most zawieszony około 70 metrów nad ziemią) i zdecydowanie bardziej wyraziste (*salient*) w porównaniu z inną wskazówką zewnętrzną – atrakcyjną kobietą. Podobne założenia – odwołujące się do efektu błędnej atrybucji (*misattribution effect*) – zawarł w swojej teorii transferu pobudzenia (*excitation transfer theory*) Dolf Zillmann (Zillmann, Katcher, Milavsky, 1972). Zgodnie z tą teorią szczątkowe pobudzenie, będące wynikiem działania wcześniejszej stymulacji, powinno prowadzić do intensywniejszej reakcji na bodziec afektujący podmiot aktualnie. Do transferu (innymi słowy – błędnej atrybucji źródła pobudzenia) nie dojdzie jednak wówczas, gdy podmiot będzie świadomy tego, jaki czynnik de facto pobudzenie wywołał (Cantor, Zillmann, Bryant, 1975). Tymczasem, wbrew powyższym założeniom, na które Dutton i Aron powołują się we wstępie, nadmieniamy w dalszej części artykułu, iż „związek pomiędzy

doświadczanym pobudzeniem seksualnym a silnymi emocjami może pojawić się nawet wówczas, gdy odczuwamy konkretną emocję”<sup>1</sup> (Dutton, Aron, 1974, s. 511).

### Błędy metodologiczne

Pierwsza niejasność – czy Dutton i Aron (i nie tylko oni – zob. też: Foster, Witcher, Campbell, Green, 1998; Hogg, Vaughan, 2011; Wojciszke, 2011) badanie pierwsze (i kolejne) słusznie określają mianem eksperymentu (Dutton, Aron, 1974, s. 511)? Czy faktycznie mamy tutaj do czynienia z manipulacją eksperymentalną (zob. też: Foster i in., 1998)? Trudno wszakże nazwać eksperymentem czy eksperymentem terenowym (*field experiment*; Dutton, Aron 1974, s. 513; Lewandowski, Aron, 2004, s. 361; Reizenzein, 1983, s. 253) proces badawczy, który nie zawiera konstytuujących eksperyment atrybutów, takich jak manipulacja, randomizacji (I i II stopnia) czy kontrola zmiennych ubocznych i zakłócających (Brzeziński, 2004). Autorzy nie przydzielali osób badanych losowo do poszczególnych warunków, co oznacza, iż de facto nie mieli wpływu na to, który mężczyzna będzie przechodzić przez most chybliwy, który zaś przez stabilny. Jedynym czynnikiem, którym manipulowali, była płeć osób przeprowadzających badanie. Możemy zatem w owym przypadku mówić co najwyżej o modelu quasi-eksperymentalnym (por. Aussems, Boomsma, Snijders, 2011; Campbell, Stanley, Gage, 1963; Jackson, 2009; Whitley, 2002). Jego słabością jest ograniczona trafność wewnętrzna (*internal validity*) – wykazanie istnienia związku pomiędzy hipotetyczną przyczyną (ZN: siłą pobudzenia) a skutkiem (ZZ: reinterpretacją owego pobudzenia) nie jest tożsame z wykazaniem istnienia zależności przyczynowo-skutkowej pomiędzy owymi czynnikami (por. Shadish, Cook, Campbell, 2002). Aby móc pójść krok dalej w wnioskowaniu inferencyjnym (tj. nie tylko wykazać istnienie związku pomiędzy ZN a ZZ, ale również zależności przyczynowo-skutkowej) badacz powinien udowodnić, iż: (1) ZN poprzedza w czasie ZZ oraz że (2) nie istnieją inne możliwe wyjaśnienia istnienia związku między ZN a ZZ. Szczególnie ów ostatni warunek trudno sfalsyfikować w modelu quasi-eksperymentalnym (Aussem, Boomsma, Snijders, 2011). Zgodnie z powyższą logiką, efekt końcowy w postaci pomiaru wskaźników zmiennych zależnych mógł być uwarunkowany co najmniej kilkoma zmiennymi ubocznymi: (1) z zakresu różnic indywidualnych (np. motywacją do poszukiwania doznań<sup>2</sup>, percepcją eustresu,

<sup>1</sup> Nie zaś niespecyficzne pobudzenie, co zakłada teoria Schachtera i Singera (1962).

<sup>2</sup> O którym to konstrukcie autorstwa Malvina Zuckermana, znanym od 1964 r., autorzy nie wspominają (por. Zuckerman, Bone, Neary, Mangelsdorff, Brustman, 1972).

dystresu, optymalnego poziomu pobudzenia); (2) odnoszącymi się do szeroko pojętych czynników sytuacyjnych, o czym wspominają sami autorzy (Dutton, Aron, 1974, s. 513). Tymczasem Dutton i Aron w sposób w mojej opinii nieuzasadniony założyli, iż dla grupy mężczyzn, którzy przeszli przez most wysoko zawieszony i niestabilny, aktywność ta była jednoznacznie awersywna i jako taka interpretowana będzie jako lękotwórcza: „[osoby badane] doświadczają silnej emocji (lęku)” (s. 511).

Kolejne pytanie, które pojawia się odnośnie do błędów metodologicznych – dlaczego grupy (indagująca kobieta vs. mężczyzna) są nierównoliczne:  $n = 93$  vs.  $n = 66$  (tabela 1)?

Kończąc ów podrozdział, chciałabym nawiązać do problematyki trafności zewnętrznej (*external validity*) związanej z zasadnością generalizacji uzyskanych wyników – osobami badanymi byli wyłącznie „samotnie idący mężczyźni pomiędzy 18. a 35. rokiem życia” (s. 511). W planie badawczym zabrakło też warunków, w których atrakcyjny mężczyzna podchodziłby do samotnie idących kobiet, jak również warunków, w których średnio/mało atrakcyjna kobieta i średnio/mało atrakcyjny mężczyzna indagują zarówno kobiety, jak i mężczyzn.

### Błędy statystyczne i wyniki powtórnych analiz

**Warunek: indaguje kobieta.** Dane odnośnie do liczebności w poszczególnych celkach tabeli kontyngencji zawiera tabela 1.

**ZZ: Test TAT.** Różnice międzygrupowe okazały się statystycznie istotne – osoby badane z grupy eksperymentalnej uzyskały większy średni wynik testu wyobraźni seksualnej ( $M = 2,47$ ) w porównaniu z grupą kontrolną ( $M = 1,41$ ),

$t(36) = 3,19; p = 0,01$ , dwustronnie. Współczynnik rzetelności ocen sędziów kompetentnych wyniósł 0,87.

**ZZ: Uległość wobec prośby.** Autorzy nie podali wyników dla zmiennej zależnej w postaci uległości wobec prośby o wypełnienie kwestionariusza, co wydaje się zastanawiające – skoro w warunkach zaindukowania silnego pobudzenia pobudzenie owo, wedle hipotezy, powinno zostać błędnie zinterpretowane jako pociąg seksualny do atrakcyjnej kobiety, indagowani mężczyźni powinni znacznie częściej w tych warunkach, w porównaniu z kontrolnymi (brakiem pobudzenia), spełniać prośbę eksperymentatorki. Tymczasem wyniki przeprowadzonych przeze mnie analiz wykazały, iż wskaźniki uległości dla obu grup były podobne: 69,7% vs. 66,7%, *odds ratio* (OR) = 1,15; 95% *confidence interval* Walda (CI) = 0,41 – 3,24;  $\chi^2(1, N = 66) = 0,07; p > 0,7$ .

**ZZ1: Przyjęcie numeru telefonu.** Autorzy nie podają wyników analiz dla pierwszego wskaźnika behawioralnego, tymczasem ponownie okazują się one nieistotne (por. tabela 1): 78,3% vs. 72,7%, OR = 1,35; 95% CI = 0,345 – 5,284;  $\chi^2(1, N = 45) = 0,19; p > 0,6$ . Należało tymczasem oczekiwać, iż wskutek błędnego (*hic et nunc!*) przypisania źródła pobudzenia, indagowani mężczyźni w warunkach eksperymentalnych (78,3%) powinni – w sposób statystycznie istotny, z odpowiednio dużym efektem – częściej wyrażać chęć otrzymania numeru telefonu od atrakcyjnej młodej kobiety, niż w warunkach kontrolnych (72,7%). Tymczasem taka zależność nie wystąpiła.

Za najlepszą miarę wielkości efektu dla zmiennych dychotomicznych  $2 \times 2$  – a z takimi będziemy mieli do czynienia w analizowanych przeze mnie powtórnie

Tabela 1

Dane behawioralne i wskaźniki wyobraźni seksualnej w każdej z grup eksperymentalnych

Osoba indagująca	Liczba wypełnionych kwestionariuszy	Liczba osób, które wzięły numer telefonu	Liczba osób, które zadzwoniły	Liczba użytecznych kwestionariuszy
Kobieta				
Most kontrolny	22/33	16/22	2/16	18
Most eksperymentalny	23/33	18/23	9/18	20
Mężczyzna				
Most kontrolny	22/42	6/22	1/6	20
Most eksperymentalny	23/51	7/23	2/7	20

Za: Dutton, Aron, 1974, s. 513.

danych pochodzących z dwóch pierwszych badań Duttona i Arona (1974) – w literaturze przedmiotu jest uznawany iloraz szans (*odds ratio*, OR), określane również mianem *cross-product ratio* (Agresti, 2002; Field, 2010; Fisher, 1962; Fleiss, Lewin, Paik, 2003; Haddock, Rindskopf, Shadish 1998), zaliczający się do tzw. *d-family of effect size*, nie zaś  $\phi$  (*phi coefficient*), należący do *r-family effect size* (Ellis, 2010; Howell, 2010), ponieważ ta ostatnia miara wielkości efektu jest funkcją  $\chi^2$ , co niesie za sobą wiele deficytów (Fleiss in., 2003; Goodman, Kruskal, 1954, 1979; Haddock i in., 1998). *Odds ratio* ( $\omega$ ) jest miarą szansy wystąpienia określonego stanu rzeczy w grupie poddawanej oddziaływaniu jakiegoś czynnika i szansy wystąpienia tego stanu w grupie kontrolnej, przy czym szansę definiuje się jako stosunek prawdopodobieństwa wystąpienia efektu do prawdopodobieństwa jego niewystąpienia w owej grupie i przybiera wartości od zera do nieskończoności (Davies, 1998).

Wartość OR wyniosła 1,35, co oznacza, że szanse osiągnięcia sukcesu (przyjęcie numeru telefonu) były o 35% większe w grupie eksperymentalnej w porównaniu z grupą kontrolną, innymi słowy – na każdą porażkę przypada 1,35 sukcesu, tj. na każdą 1,35 osoby, która przyjęła numer telefonu od eksperymentatorki, przypada jedna osoba, która numeru telefonu nie przyjęła (Agresti, 2007; Davies, 1998; Fleiss i in., 2003; Howell, 2010). Jeśli OR jest większe od 1, wówczas sukces jest bardziej prawdopodobny od porażki (Agresti, 2007), jednak wartości OR bliskie 1 oznaczają istnienie słabej relacji między zmiennymi (Haddock i in., 1998). Przedział ufności (*confidence interval*, CI) zawiera 1, co należy interpretować jako brak istotnych (tj. nieprzypadkowych, będących wynikiem manipulacji eksperymentalnej) różnic między grupami lub jako wskaźnik tego, że związek między zmiennymi jest słaby lub nieistniejący (Agresti, 2007; Haddock i in., 1998). Innymi słowy, jeśli przedział ufności zawiera wartość OR = 1 (CIs oszacowują granice, w których zawiera się prawdziwe OR populacji<sup>3</sup>) oznacza to, iż obie grupy są niezależne (brak relacji między zmiennymi w tabeli kontyngencji jest określane mianem niezależności (*independence*; Aron, Coups, Aron, 2013; Field, 2010), tj. proporcje badanej charakterystyki są równe, czyli – co potwierdza

wartość  $p^4$  – zależność będąca skutkiem zastosowanej manipulacji eksperymentalnej nie występuje. A zatem fakt przyjęcia numeru telefonu, skoro CI zawiera hipotezę zerową, nie zależy od różnych warunków, w których osoby badane się znalazły (most zawieszony wysoko vs. nisko).

**Z2: Zadzwonienie do eksperymentatorki.** Wyniki okazały się statystycznie istotne (tabela 1): 50% vs 12,5%, OR = 7; 95% CI = 1,22 – 40,125;  $\chi^2$  (1,  $N = 34$ ) = 5,44;  $p < 0,02$ . Wprawdzie rezultaty są statystycznie istotne, lecz należałoby zadać pytanie, o czym właściwie świadczą. OR wynosi 7, co oznacza, że szanse osiągnięcia sukcesu (wykonania telefonu) są siedmiokrotnie większe od szans doznania porażki (por. Agresti, 2007). Jeśli OR > 3 oznacza to silną asocjację między zmiennymi (Agresti, 2007; Haddock i in., 1998), jednak granice CI są szerokie (1,22 – 40,125, ergo: szanse zadzwonienia są od 22% do ponad 40 razy większe), co wskazuje, że jest to mało precyzyjny estymator danych populacyjnych (Sim, Reid, 1999; Sze, Say, 2010). Na szerokość CI wpływają trzy czynniki: heterogeniczność próby, wielkość próby oraz wybrany przez badacza poziom ufności<sup>5</sup> (por. Sim, Reid, 1999). Szeroki zakres owego przedziału może zatem wskazywać na sporą heterogeniczność badanej próby (ten czynnik, tj. niehomogeniczność próby, jest jednak często poza kontrolą badacza – por. Hurlburt, 1994), może być skutkiem tego, że próba jest zbyt mała – szerokość CIs (CIs *width*) zmniejsza się wraz ze zwiększaniem liczebności próby (Fidler, Cumming, 2005), może też wskazywać na istnienie alternatywnych wytłumaczeń uzyskanych wyników – dane są spójne z szeroką rangą możliwych hipotez (źródło internetowe: <http://www.pmean.com/11/WideInterval.html>). Poza tym, jak nadmieniąją Huw Davies i Iain Crombie (2009) szerokie CIs oznaczają uchwycenie różnorodnej rangi wielkości efektu, a co za tym idzie – jakkolwiek estymacja wielkości efektu staje

<sup>3</sup> 95% CI nie oznacza, że jesteśmy pewni na 95% tego, iż przedział ufności zawiera prawdziwą wartość OR w populacji – jesteśmy pewni na 100%, że w 95/100 badaniach owe granice zostaną uchwycione (Field, 2010; Howell, 2010). Innymi słowy – CI opisuje, na ile niepewny jest estymowany przez nas interwał średnich/proporcji jako predyktor poziomu populacyjnego; średnia populacyjna jest stałą i nieznaną wartością, która znajduje się ze 100% pewnością wewnątrz lub na zewnątrz CI (Howell, 2010; Sze, Say, 2010).

<sup>4</sup> Jeśli 95% CI dla OR zawiera wartość 1 – wartość  $p$  będzie większa niż 0,05 i, alternatywnie, jeśli 95% CI dla OR nie zawiera wartości 1, wartość  $p$  będzie mniejsza od 0,05 (Sze, Say, 2010). Oba koncepty – statystyczna istotność i CI – bazują na tym samym procesie inferencyjnym: jeśli CI nie zawiera wartości 1, będącej dla OR wskaźnikiem braku relacji pomiędzy zmiennymi, wówczas otrzymana wartość (różnica w proporcjach) jest statystycznie istotna na 100% minus wartość CI. A zatem jeśli np. granice CI wahałyby się od -0,8 do 0,4, wówczas z pięcioprocentowym poziomem ufności odrzucilibyśmy  $H_0$ . Innymi słowy – CI zawiera dość informacji, aby móc orzekać o poziomie statystycznej istotności, jednak sama wartość  $p$  nie zawiera dość informacji, aby móc obliczyć CI (Howitt, Cramer, 2011).

<sup>5</sup> Na szerokość CIs wpływa to, jaki procent niepewności przyjmujemy – 10% (90% CI), 5% (95% CI) czy 1% (99% CI); 90% CIs będą węższe od 99% CIs (por. Sze, Say, 2010).

się nieprecyzyjna i w konsekwencji tego typu badania są mało informacyjne.

**Warunek: indaguje mężczyzna.** Dane odnośnie do liczebności w poszczególnych celkach tabeli kontyngencji zawiera tabela 1.

**ZZ: Test TAT.** Różnice międzygrupowe okazały się nieistotne statystycznie – osoby badane z grupy eksperymentalnej uzyskały podobny średni wynik testu wyobraźni seksualnej ( $M=0,8$ ) do osób z grupy kontrolnej ( $M=0,61$ ),  $t(38) = 0,36$ ; nieistotne statystycznie.

**ZZ: Uległość wobec prośby.** Dutton i Aron (1974) ponownie nie podają danych odnoszących się do uległości osób badanych wobec kierowanej do nich prośby – w tych warunkach, tak jak w grupie, którą indagowała kobieta, wyniki okazały się statystycznie nieistotne. W warunkach eksperymentalnych uległo 45,1%, w warunkach kontrolnych zaś – 52,4%,  $OR = 0,75$ ; 95% CI = 0,33 – 1,69;  $\chi^2(1, N = 93) = 0,49$ ;  $p > 0,4$ .

**ZZ1: Przyjęcie numeru telefonu.** Autorzy nie podają wyników analiz dla pierwszego wskaźnika behawioralnego, tymczasem wyniki ponownie okazują się nieistotne (por. tabela 1): 30,4% vs. 27,3%,  $OR = 1,16$ ; 95% CI = 0,321 – 4,247;  $\chi^2(1, N = 45) = 0,05$ ;  $p > 0,8$ .

**ZZ2: Zadzwonienie do eksperymentatora.** Wyniki okazały się nieistotne statystycznie (por. tabela 1): 28,6% vs. 17%,  $OR = 2$ ; 95% CI = 0,134 – 29,81;  $\chi^2(1, N = 13) = 0,26$ ;  $p > 0,6$ .

Podsumowując ów podrozdział – (1) odnośnie do ZZ: test TAT zarówno w warunkach, gdy osobą indagującą była kobieta (statystycznie istotne;  $p = 0,01$ ), jak i mężczyzna (*ni*), wyniki okazały się zgodne z hipotezą badawczą; (2) ZZ: uległość wobec prośby – każdorazowo wyniki okazały się nieistotne statystycznie; indagująca kobieta –  $p > 0,7$ , wyniki niezgodne z hipotezą badawczą; indagujący mężczyzna –  $p > 0,4$ , wyniki zgodne z hipotezą; (3) ZZ1: przyjęcie numeru telefonu – w obu warunkach wyniki nie osiągnęły konwencjonalnego poziomu statystycznej istotności: eksperymentatorka –  $p > 0,6$ , wyniki niezgodne z hipotezą badawczą; eksperymentator –  $p > 0,8$ , wyniki zgodne z hipotezą; (4) ZZ2: zadzwonienie do eksperymentatora/-ki – w przypadku, gdy osobą proszącą o wypełnienie kwestionariusza była kobieta, wyniki okazały się statystycznie istotne ( $p = 0,02$ , zgodne z hipotezą), gdy był nią zaś mężczyzna – wyniki były nieistotne ( $p > 0,6$ , zgodne z hipotezą). A zatem, koncentrując się wyłącznie na warunkach, w których do osób badanych podchodziła atrakcyjna, młoda kobieta, wyniki dwu analiz na cztery potwierdzają hipotezę badawczą. Czy jednak na pewno?

### Błędy interpretacyjne

Dutton i Aron wspominają, opisując rezultaty pierwszego badania, o manipulacji pobudzeniem (*arousal manipulation*), w kontekście danych uzyskanych od 30 mężczyzn w wieku 18–35: „Piętnastu mężczyznom, którzy przeszli przez most eksperymentalny, zadano pytanie: «Jak sądzisz, jak duży lęk odczuwa przeciętna osoba, która przechodzi przez ów most?». Średnia wyniosła 79 na 100-stopniowej skali, na której 100 oznaczało ekstremalny lęk. Piętnastu mężczyzn przechodzących przez most kontrolny uzyskało średnią 18 na tej samej skali ( $t = 9,7$ ;  $df = 28$ ;  $p = 0,001$ , dwustronnie). W odpowiedzi na pytanie: «Jak duży lęk odczuwałeś, przechodząc przez ten most?» mężczyźni przechodzący przez most eksperymentalny uzyskali średnią 65, zaś przez kontrolny – 3 ( $t = 10,6$ ;  $p = 0,001$ ;  $df = 28$ , dwustronnie). Tak więc można skonkludować, że większość ludzi odczuwa lęk na moście eksperymentalnym, ale nie na kontrolnym” (Dutton, Aron, 1974, s. 512). Owa konstatacja jest w moim odczuciu nieuzasadniona: (1) w powyższym przypadku nie może być mowy o jakiegokolwiek „manipulacji”; (2) mamy tu do czynienia z nieuprawnioną generalizacją wyników uzyskanych z niewielkiej próby incydentalnej ( $N = 30$ ). Zadać należałoby pytanie, na jakiej podstawie Dutton i Aron ekstrapolują dane pochodzące z owej próby na o wiele liczniejszą próbę faktycznych uczestników przeprowadzonych dwóch badań terenowych (por. tabela 1). Autorzy nie wspominają o tym, kto indagował mężczyzn: atrakcyjna kobieta? atrakcyjny mężczyzna? nieatrakcyjna, średnio atrakcyjna kobieta? nieatrakcyjny, średnio atrakcyjny mężczyzna? Kolejna nasuwająca się wątpliwość dotyczy rozpiętości skali pomiaru – dlaczego była 100-stopniowa, zamiast 101-stopniowej, z punktem 0, który oznaczałby brak lęku? Być może brak lęku był na owej skali oznaczony jako 1 – to jednak z tekstu nie wynika. Dlaczego z góry założono, iż owa trzydziestka mężczyzn będzie odczuwała lęk, nie zaś – nieewaluowane negatywnie niespecyficzne pobudzenie lub ekscytację, zadowolenie, dumę (np. wynikające z tego, że „dałem radę i przeszedłem przez ów most”)? Tłumaczenie, iż „aby zapobiec jakimkolwiek podejrzeniom, nie sprawdzano wielkości odczuwanego pobudzenia u mężczyzn przechodzących przez most eksperymentalny [w badaniu właściwym – przyp. autorki]” (s. 512) jest nieuzasadnione, z perspektywy zaś przedmiotu badania – kluczowe.

W tekście źródłowym można przeczytać: „Łącznie było indagowanych bądź przez eksperymentatorkę, bądź przez eksperymentatora 85 osób badanych” (s. 511) – tymczasem w badaniu wzięło udział 159 osób, spośród których 90 zgodziło się wypełnić kwestionariusz (por. tabela 1). W warunkach, w których osobą proszącą była kobieta,

spośród 45 wypełnionych kwestionariuszy „7 było beużytecznych, gdyż albo były niekompletne, albo były wypełnione w języku obcym” (s. 512). Nurtuje mnie pytanie, ile spośród owych siedmiu kwestionariuszy było w innym języku i dlaczego w takim razie do dalszych pomiarów (zgoda na wzięcie numeru telefonu i/lub zadzwonienie) zakwalifikowano ową siódmkę ( $n = 45$ , nie zaś 38). Ta sama wątpliwość odnosi się do warunków, w których osobą indagującą był mężczyzna – „pięć z owych kwestionariuszy było beużyteczne, pozostało zatem 20 w obu warunkach – eksperymentalnych i kontrolnych” (s. 512) – wynikiem czego była owa „beużyteczność”?

Powyższe zastrzeżenia czy też wątpliwości być może niektórzy uznają za krytykanctwo. Kolejną już nie tak błahą kwestią jest problematyka odnosząca się do aspektu temporalnego – kluczowego z perspektywy interpretacji uzyskanych wyników. Po pierwsze – autorzy nie podają, jaki czas minął od momentu przejścia osób badanych przez most eksperymentalny do momentu, kiedy indagowali je współpracownicy eksperymentatorów: „po tym, jak osoby badane przeszły albo most eksperymentalny, albo kontrolny, były zagadywane przez osobę przeprowadzającą eksperyment” (s. 511). Z punktu widzenia wyników późniejszych badań (Cantor i in., 1975), to, czy minęła minuta, pięć, czy 10 minut stanowi istotny parametr. Oczywiście jest, iż Dutton i Aron nie mogli znać owych wyników badań, niemniej nie poinformowanie o tych wskaźnikach należy traktować jako uchybienie metodologiczne. Po drugie – Dutton i Aron piszą, iż po wypełnieniu kwestionariusza w obu warunkach badawczych eksperymentatorzy oferowali osobom indagowanym, jeśli wyrażały taką wolę, możliwość wyjaśnienia im wyników eksperymentu – zapisywali numer telefonu oraz „zachęcali każdą z osób badanych do tego, aby zadzwoniła, jeśli chciałaby dodatkowo porozmawiać” (Dutton, Aron, 1974, s. 512). Tymczasem w innej swej pracy Arthur Aron pisze, iż współpracownicy informowali osoby badane o tym, że mogą „zadzwonić tego samego wieczora, jeśli chcą dowiedzieć się więcej o badaniu” (Lewandowski, Aron, 2004, s. 362). Z tekstu pochodzącego z oryginalnego artykułu nie wiadomo: (1) kiedy osoby badane mogły/ powinny zadzwonić – tego samego wieczora? nazajutrz? dowolnego dnia? (2) czy osoby badane dzwoniły mniej więcej w tym samym czasie i o tej samej porze? Autorzy de facto nie mieli kontroli nad owym wskaźnikiem zmiennej zależnej. Po trzecie i najważniejsze – autorzy fakt zadzwonienia po tak długim (jak długim? można przypuszczać, że co najmniej kilkugodzinnym) czasie interpretują (w sposób nieuzasadniony ani teoriami, ani wynikami badań, na które się powołują) jako wskaźnik błędnego określenia źródła pobudzenia i, w efekcie,

większego pociągu seksualnego odczuwanego w stosunku do eksperymentatorki jedynie w warunkach eksperymentalnych. „Zakładając, że zainteresowanie eksperymentem powinno być jednakowe w grupie kontrolnej i eksperymentalnej, założono, że różnice w proporcjach osób, które zadzwoniły, powinny odzwierciedlać różnice w pociągu seksualnym” (Dutton, Aron, 1974, s. 512). Tymczasem osoby badane w momencie wykonywania telefonu nie były już pobudzone – w każdym razie, zgodnie z założeniami dwuczynnikowej teorii emocji Schachtera i Singera (1962) czy teorii transferu pobudzenia Zillmanna i in. (1972), nie z powodu eksperymentatorki. O czym więc świadczą wskaźniki tej zmiennej zależnej?

## BADANIE 2 DUTTONA I ARONA

W dyskusji wyników pierwszego badania Dutton i Aron stwierdzają, iż uzyskane rezultaty mogą wynikać z różnic międzygrupowych na wejściu, będących skutkiem: (1) tego, że mężczyźni z grupy eksperymentalnej mogli być turystami, którzy przyjechali jedynie na krótką wycieczkę; (2) różnic indywidualnych (Dutton, Aron, 1974, s. 513). W celu wyeliminowania alternatywnych wyjaśnień skonstruowano nieco odmienny plan badawczy. Osobami badanymi byli, tak jak poprzednio, wyłącznie mężczyźni ( $N = 34$ , lecz tak naprawdę  $N = 69$ : w grupie kontrolnej liczebność wyniosła  $n = 35$ , w eksperymentalnej zaś  $n = 34$ ) idący samotnie, w wieku 18–35 lat, jednak w przypadku badania drugiego byli to wyłącznie ci mężczyźni, którzy przeszli uprzednio przez Capilano Suspension Bridge (zamysł autorów był następujący: badaną próbę powinni stanowić „członkowie tej samej populacji”), przy czym grupę kontrolną stanowili ci mężczyźni, którzy przeszli ów most co najmniej 10 minut wcześniej, eksperymentalną zaś ci, którzy „przeszli przez most” (s. 511). Kolejna różnica polegała na tym, że w przypadku drugiego badania, w świetle poprzednio uzyskanych wyników, autorzy zrezygnowali z udziału eksperymentatora płci męskiej – każdorazowo badanych mężczyzn indagowała atrakcyjna kobieta. Zmienne zależne były analogiczne: (1) liczba i jakość słów o konotacji seksualnej zawartych w opowiadaniu pisanim przez osoby badane po zaprezentowaniu obrazka 3GF z TAT, które następnie oceniali sędziowie kompetentni; (2) dwojakiego rodzaju dane behawioralne, stanowiące wskaźniki percepcji atrakcyjności kobiety przeprowadzającej badanie przez badanych mężczyzn: ZZ1 – zgoda na wzięcie numeru telefonu od eksperymentatorki; ZZ2 – zadzwonienie do eksperymentatorki.

### Błędy metodologiczne

Jak piszą autorzy, eksperymentatorka podchodziła do uczestników badania albo po upływie co najmniej

10 minut od przejścia przez most (warunki kontrolne, brak odczuwanego rezydualnego pobudzenia fizjologicznego) albo... nie wiadomo, po jakim czasie, gdyż Dutton i Aron ponownie, tak jak to miało miejsce w opisie procedury pierwszego badania, nie podają, jaki czas minął od momentu przejścia osób badanych przez most do momentu, kiedy indagowała ich przeprowadzająca badanie kobieta. Tymczasem ów interwał czasowy, który upłynął (na pewno mniej niż 10 minut, ale nie wiadomo dokładnie, ile), jest parametrem z owej perspektywy kluczowym (por. Zillmann i in., 1972; Cantor i in., 1975). Wszelkie pozostałe zastrzeżenia natury metodologicznej są replikacją poprzedniej sekcji *Błędy metodologiczne*.

### Błędy statystyczne i wyniki powtórnych analiz

**ZZ:** **Test TAT.** Różnice międzygrupowe okazały się statystycznie istotne – osoby badane z grupy eksperymentalnej (na bazie wypełnionych 23 z 25 kwestionariuszy) uzyskały większy średni wynik testu wyobraźni seksualnej ( $M = 2,99$ ) w porównaniu z grupą kontrolną (wypełnione 22 z 25 kwestionariuszy);  $M = 1,92$ ;  $t(43^6) = 3,07$ ;  $p = 0,01$ , dwustronnie. Współczynnik rzetelności ocen sędziów kompetentnych wyniósł 0,79.

**ZZ:** **Uległość wobec prośby.** Autorzy ponownie nie przeprowadzili analizy danych odnoszących się do tego, czy indagowani mężczyźni ulegli kierowanej do nich prośbie. Tymczasem moje analizy wykazały, że wskaźniki uległości dla obu grup były podobne: 25/34 w grupie eksperymentalnej i 25/35 w grupie kontrolnej, (73,53% vs. 71,43%),  $OR = 1,11$ ; 95% CI = 0,39 – 3,20;  $\chi^2(1, N = 69) = 0,04$ ;  $p > 0,8$ .

**ZZ1:** **Przyjęcie numeru telefonu.** Dutton i Aron odnośnie do odpowiedzi w teście TAT piszą: „W grupie eksperymentalnej 25 spośród 34 indagowanych mężczyzn zgodziło się wypełnić kwestionariusz. W grupie kontrolnej zgodę wyraziło 25 z 35” (s. 514). W następnym akapicie zaś, w którym prezentują wyniki danych behawioralnych: „W grupie eksperymentalnej 20 spośród 25 osób badanych, które zgodziły się na rozmowę, przyjęło numer telefonu eksperymentatorki. W grupie kontrolnej numer telefonu wzięło 19 osób z 23” (s. 514). Tymczasem, jak wynika z dokładnej analizy tekstu

<sup>6</sup> Skoro  $N = 60$  ( $n = 23$  w grupie eksperymentalnej i  $n = 22$  w grupie kontrolnej), to  $df$  powinny wynosić 43, nie zaś 36 (Dutton, Aron, 1974, s. 514). Osobną kwestię stanowią pomyłki (tu raczej – chochliki drukarskie), które wkradają się do tłumaczonych – w tym wypadku na język polski – tekstów oryginalnych; zamiast „ $df = 19$ ” (Dutton, Aron, 1974, s. 517), w polskiej wersji owego artykułu pojawia się „ $df = 10$ ” (por. Dutton, Aron, 2002, s. 607).

źródłowego, w warunkach kontrolnych numer telefonu przyjęło nie 19/23, lecz 19/25 indagowanych mężczyzn – takie obliczenia poczynili wszakże autorzy w badaniu pierwszym, tj. do puli osób, spośród których obliczano potem odsetek mężczyzn biorących numer telefonu, wchodziły również te osoby badane, których kwestionariusze z różnych powodów okazywały się bezużyteczne (por. tabela 1). Sami autorzy piszą zaś o tym w nawiązaniu do badania drugiego, iż pulę ową tworzą „osoby badane, które zgodziły się na rozmowę” (s. 514), *ergo*: wypełniły kwestionariusz, abstrahując od jego (bez)użyteczności.

Autorzy ponownie nie przedstawiają wyników analiz odnośnie do zmiennej zależnej ZZ1, a prezentują się one w sposób następujący: numer telefonu przyjęło w grupie eksperymentalnej 20 z 25 osób, natomiast w grupie kontrolnej – 19 z 25 (80% vs. 76%),  $OR = 1,26$ ; 95% CI = 0,33 – 4,84;  $\chi^2(1, N = 50) = 0,12$ ;  $p > 0,7$ . Zatem ponownie różnica w częstościach nie zbliżyła się nawet do granicy trendu, wielkość efektu jest niewielka, przedział ufności natomiast zawiera hipotezę zerową. Na podstawie owych danych – zarówno z pierwszego, jak i drugiego badania – można stwierdzić, iż mężczyźni przeżywający silne pobudzenie nie brali statystycznie istotnie częściej numeru telefonu od eksperymentatorki, w porównaniu z warunkami, w których mężczyźni owego pobudzenia nie przeżywali.

**ZZ2:** **Zadzwonienie do eksperymentatorki.** Wspomniany wyżej błąd ma konsekwencje dotyczące faktycznych wyników analiz drugich danych behawioralnych, a mianowicie późniejszego zadzwonienia do eksperymentatorki. Autorzy piszą: „W grupie eksperymentalnej 13 spośród 20 osób badanych zadzwoniło, podczas gdy w grupie kontrolnej zadzwoniło 7 osób z 23 ( $\chi^2 = 5,89$ ,  $p < 0,02$ )” (s. 514). Tymczasem w warunkach kontrolnych do eksperymentatorki zadzwoniło 7 z 19, nie zaś 7 z 23, jako że „w grupie kontrolnej 19 osób z 23 zaakceptowało numer telefonu” (s. 514). Tak więc wyniki zaprezentowane przez Duttona i Arona są błędne – właściwe prezentują się następująco: w grupie eksperymentalnej zadzwoniło 13 z 20 badanych, w kontrolnej zaś – 7 z 19 (65% vs. 37%),  $OR = 3,18$ ; 95% CI = 0,86 – 11,78;  $\chi^2(1, N = 39) = 3,09$ ;  $p = 0,08$ . Można więc mówić w tym wypadku jedynie o trendzie. Przedział ufności zawiera 1, co oznacza, iż związek między zmiennymi jest słaby lub nieistniejący, przypadkowy, niebędący rezultatem zastosowanej manipulacji (por. Agresti, 2007; Haddock i in., 1998), co dodatkowo potwierdza wartość  $p$ .

### Błędy interpretacyjne

Po pierwsze nieuzasadniona jest generalizacja wyników pochodzących z dwu niereprezentatywnych dla populacji

prób. Po wtóre, biorąc pod uwagę zarówno wyniki powyższych reanaliz, jak i zaprezentowane wcześniej wątpliwości natury metodologicznej oraz interpretacyjnej, nie sposób zgodzić się z konstatacją autorów, jakoby „dane behawioralne z eksperymentu 1 również zostały zreplikowane” (Dutton, Aron, 1974, s. 514), jak również ze stwierdzeniem, zgodnie z którym autorzy rzekomo wykazali, iż „eksperyment 2 umożliwia odrzucenie założenia, iż osoby badane pochodzą z różnych populacji, jako wyjaśnienia zaistniałych różnic pomiędzy grupami przechodzącymi przez most eksperymentalny i kontrolny, w odniesieniu do warunków z eksperymentu 1, w których badanie przeprowadzała eksperymentatorka płci żeńskiej” (s. 514). Owszem – wyniki badania pierwszego zostały zreplikowane, osiągając poziom statystycznie istotnego, ale wyłącznie w stosunku do zmiennej zależnej w postaci testu TAT (każdorazowo  $p = 0,01$ ). Natomiast oba wyniki ZZ w postaci przyjęcia numeru telefonu nawet nie zbliżyły się do granicy trendu ( $p > 0,6$  w badaniu 1 oraz  $p > 0,7$  w badaniu 2), natomiast odnośnie do zmiennej zależnej w postaci wykonania telefonu – nie zostały zreplikowane ( $p = 0,02$  w badaniu 1 i  $p = 0,08$  w badaniu 2). Zgodnie z dzisiejszymi standardami, replikacje stanowią fundament nauki, w tym naukowo uprawianej psychologii (Cumming, 2008; Wojciszke, 2006). Jak pisze Geoff Cumming (2008), wartość  $p$  nie jest rzetelnym fundamentem jakiegokolwiek wniosku inferencyjnego<sup>7</sup> – uzyskanie w jednym badaniu wyników statystycznie istotnych niewiele mówi o tym, czy efekt zostanie zreplikowany w kolejnych badaniach. Autor ów wykazał w przeprowadzonej symulacji 25-krotnej replikacji typowego eksperymentu, że wartość  $p$  wahała się od 0,01 do 0,76. Miarami, na której naukowcy współcześnie uprawiający psychologię powinni bazować, są – niezależne od wielkości próby – przedziały ufności i wielkości efektu (por. szóste wydanie *APA Publication Manual*, 2010). Nie mniej istotna jest implementacja myślenia metaanalizy (Cumming, 2008).

Pretekstem do przeprowadzenia eksperymentu sensu stricto laboratoryjnego (w tekście źródłowym – „eksperyment 3”) były, ponownie, kwestie związane z interpretacją

uzyskanych wyników. Jedną z postulowanych alternatywnych przyczyn, mogącą prowadzić do uzyskania artefaktualnych wyników, mogła być, w mniemaniu Duttona i Arona, (nie)świadomość hipotezy badawczej i jej konsekwencje. Po raz kolejny jednak badacze sami sobie przeczą, pisząc w jednym miejscu, iż kobieta przeprowadzająca badania każdorazowo była nieświadoma hipotezy badawczej (Dutton, Aron, 1974, s. 511 i s. 514), w innym zaś, że „gdyby nie dostarczono kobiecie [przeprowadzającej eksperyment – przyp. autorki] wskazówek odnośnie do hipotezy badawczej, nie można byłoby kontrolować bardziej stałych form komunikacji niewerbalnej – takich jak kontakt wzrokowy” (s. 514). Drugą natomiast autorzy wiążą ściśle z postacią (*gestalt*) wykreowaną przez sytuację eksperymentalną, która „mogła sprawić, iż eksperymentatorka wydawała się bardziej bezradna lub przestraszona, niczym «kobieta w tragicznej sytuacji» [z obrazka 3GF TAT – przyp. autorki]. Nie miałyby to miejsca w warunkach kontrolnych” (s. 514). Powyższe wątpliwości autorów w moim odczuciu są irrelevantne i niespójne logicznie – skoro eksperymentatorka w warunkach przeżywanego przez osoby badane silnego pobudzenia miała wydawać się bezradna lub przestraszona, owa hipoteza powinna uzyskać potwierdzenie w pomiarach kwestionariuszowych (zwiększona, w porównaniu z warunkami kontrolnymi, liczba takich słów, jak strach, lęk, bezradność – tymczasem takich danych autorzy nie podali, czy zaś je uzyskali – tego nie wiemy).

### EKSPERYMENT 3 DUTTONA I ARONA

Uczestnikami eksperymentu było 80 ochotników, studentów płci męskiej. Eksperyment przeprowadzano w diadach tej samej bądź odmiennej płci (partnerką była wówczas atrakcyjna kobieta), przy czym druga osoba z diady była zawsze pomocnicą/-kiem eksperymentatora. Współpracownica „wiedziała ona o tym, że badania dotyczą atrakcyjności seksualnej, ale nie znała hipotez” (Dutton, Aron, 1974, s. 514), oprócz tego „była poinstruowana, aby unikać jakiegokolwiek «ponadwymiarowego» kontaktu wzrokowego” (s. 514; mógł on mieć miejsce tylko w dwóch sytuacjach: przywitania i zajęcia miejsca po uprzednim odwieszeniu płaszcza). Zgodnie z instrukcją maskującą – eksperyment miał dotyczyć wpływu szoków elektrycznych na uczenie. Zmiennymi niezależnymi były: osoba w diadzie (tej samej płci – warunki kontrolne lub odmiennej płci – warunki eksperymentalne) oraz siła natężenia szoku elektrycznego (silny vs. słaby). Zmienne zależne były trojakiemu rodzajowi: ZZ1 – pomiar odczuwanego przez osoby badane lęku (oszacowywany na skali pięciostopniowej); ZZ2 – dwójakiemu rodzajowi pomiaru natężenia pociągu seksualnego względem pomocnicy

<sup>7</sup> Bazowanie we wnioskowaniu inferencyjnym wyłącznie na owo magiczne  $p = 0,05$  jest wciąż jeszcze postawą spotykaną we współcześnie przeprowadzanych przez psychologów analizach danych, co jest od niemal 20 lat niezgodne z dyrektywami APA (por. *APA Publication Manual*, 1994; Wilkinson & The *APA Task force on statistical inference*, 1999; *APA Publication Manual*, 2010). Od wielu lat nie milknie też wśród psychologów i statystyków debata nad zasadnością podtrzymywania rytuału „ $p = 0,05$ ” (por. Carver, 1978; Cohen, 1994; Gigerenzer i in., 1989; Oakes, 1986; Rozeboom, 1960).



eksperymentatora<sup>8</sup>: chęć umówienia się z kobietą na randkę i chęć pocałowania jej (określana na skali pięciostopniowej; mierzono średnią z obu odpowiedzi); ZZ3 – wskaźniki wyobraźni seksualnej mierzone za pomocą obrazka 3FG TAT wykorzystywanego w dwu poprzednich badaniach. Im większy wynik uczestnicy eksperymentu uzyskiwali na ww. skalach, tym większy odczuwali lęk lub tym większy pociąg seksualny do pomocnicy eksperymentatora.

### Niejasności statystyczne

**ZZ1: Pomiar odczuwanego lęku.** Wyniki pomiaru tej zmiennej zależnej są następujące (por. tabela 2): „W warunkach, w których osoby badane oczekiwały na otrzymanie silnego szoku elektrycznego, wskazywały na odczuwanie silniejszego lęku w porównaniu z warunkami, w których oczekiwały na słaby szok ( $t = 4,03$ ;  $p < 0,001$ ;  $df = 39$ , jednostronnie). W warunkach, w których osoby badane spodziewały się otrzymać silny szok, ale kiedy towarzyszyła im współpracownica eksperymentatora, osoby badane wykazywały mniejszy lęk w porównaniu z warunkami kontrolnymi ( $n = 20$ ), w których brało udział dwóch mężczyzn ( $t = 2,17$ ;  $p = 0,025$ ;  $df = 19^9$ , jednostronnie). Nie odnotowano statystycznie istotnych różnic w poziomie lęku będącego funkcją antycypacji przez pomocnicę eksperymentatora mocnego vs. słabego szoku” (Dutton, Aron, 1974, s. 515).

Powyższe wyniki wydają się niejasne – uprzednio autorzy pisali bowiem, iż uczestników badania było 80, jednak w tabeli 2 pojawia się celka „brak współpracownika eksperymentatora płci żeńskiej” (*no female confederate*) oraz informacja, iż każdorazowo celka liczy  $n = 20$ ; celka, o której tu mowa jest piąta, ergo: należy wnioskować, iż osób badanych było 100, tak więc mamy tutaj do czynienia z planem eksperymentalnym  $2 \times 2$  z izolowaną grupą kontrolną.

Autorzy dokonują dwóch porównań międzygrupowych testem  $t$  Studenta (por. tabela 2): (1) osoby badane oczekują szoków silnych (celki: 1 i 2) vs. słabych (celki: 3 i 4); (2) osoby badane oczekują silnych szoków, gdy towarzyszy im pomocnica eksperymentatora (celki: 1 i 2) vs. warunki kontrolne (tj. diada mężczyzna–mężczyzna, celka 5). Jeśli tak, to dlaczego, odpowiednio, w porównaniu pierwszym  $df$  wynosi 39, zamiast 78, zaś w drugim – 19, zamiast 58? Ponadto – dlaczego autorzy stosują porównania jednostronne? Uprzednio, w badaniach 1 i 2, stosowali dwustronne. Niestety – nie sposób dokonać powtórnych analiz, nie znając wartości odchyłeń standardowych w poszczególnych grupach.

### Błędy metodologiczne

Sama idea owego planu eksperymentalnego, o czym uprzednio wspominałam, a mianowicie odwoływanie się do psychologii postaci, jest dyskusyjna – brakuje tu przede wszystkim uzasadnienia teoretycznego (relevan- te byłoby na przykład odwołanie się do teorii wzmocnień; takie wyjaśnienie związku pobudzenie–atrakcyjność zaproponowali w 1977 r. Kenrick i Cialdini). Oprócz tego ów eksperyment nie jest, w moim odczuciu, logiczną ciągłością poprzednich badań terenowych, aczkolwiek sam pomysł tego, aby badania kontynuować w warunkach laboratoryjnych jest ze wszech miar słuszny i zgodny z obecnymi standardami (por. np. Wojciszke, 2004).

Do uchybień mniejszej wagi owego planu eksperymentalnego zaliczyłabym to, iż – w warunkach, gdy obecna jest pomocnica eksperymentatora płci żeńskiej jest to zawsze kobieta atrakcyjna, brakuje zaś takich sytuacji, w których kobieta owa byłaby średnio lub mało atrakcyjna. Po wtóre – nic nie wiadomo o tym, jak mało/bardzo atrakcyjny był pomocnik eksperymentatora z warunków określanych przez autorów mianem kontrolnych. Niewątpliwym ograniczeniem jest również niemożność

Tabela 2

*Deklarowany poziom lęku w poszczególnych warunkach eksperymentalnych*

Osoba badana oczekuje:	Współpracownica ma otrzymać silny szok	Współpracownica ma otrzymać słaby szok	Brak współpracownika płci żeńskiej
Silnego szoku	(1) 3,17	(2) 3,05	(5) 3,8
Słabego szoku	(3) 2,42	(4) 2,28	(6) ?

W każdej celce  $n = 20$  (za: Dutton, Aron, 1974, s. 515).

<sup>8</sup> Przy czym odnosi się to wyłącznie do tych warunków, kiedy pomocnica była płci żeńskiej; w diadzie mężczyzna–mężczyzna eksperymentator prosił uczestników eksperymentu, aby pominęli owo pytanie.

<sup>9</sup> Zastanawiające jest, dlaczego  $df = 36$ , zamiast 43, skoro liczebność wynosiła  $N = 45$  (por. Dutton, Aron, 1974, s. 514).

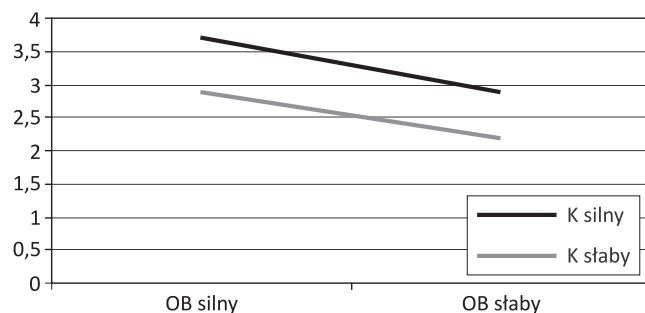
uogólniania uzyskanych wyników – tak jak w poprzednich dwóch badaniach próbę badawczą stanowili wyłącznie mężczyźni.

Kolejne pytania, jakie należałoby zadać w kontekście uchybień natury metodologicznej, odnoszą się do kwestii braku analogii warunków: współpracownica płci żeńskiej vs. męskiej. Otóż dlaczego badacze nie dokooptowali do planu eksperymentalnego warunków, w których pomocnik eksperymentatora płci męskiej oczekuje szoków silnych vs. słabych? Dlaczego zabrakło warunków stricte kontrolnych, w których zarówno współpracownik eksperymentatora, niezależnie od płci, jak i osoba badana nie oczekują żadnych szoków? Oprócz tego, skoro w badaniach 1 i 2 założono, iż nie należy pytać osób badanych o poziom odczuwanego lęku, aby „zapobiec jakimkolwiek podejrzeniom” (Dutton, Aron, 1974, s. 512), to dlaczego w eksperymencie właściwym owego pomiaru dokonano, w dodatku bezpośrednio przed pomiarem ZZ2 i ZZ3? W mojej opinii uchybienia te są funkcją nietrafnego założenia, aby w eksperymencie laboratoryjnym skoncentrować się na eliminacji czynnika „kobieta w sytuacji tragicznej”.

### Błędy interpretacyjne

**ZZ2: Oszacowania pociągu seksualnego.** Odnośnie do tej zmiennej zależnej autorzy wykazali, iż niezależnie od tego, czy współpracownica eksperymentatora oczekiwała wstrząsów silnych, czy też słabych – badani mężczyźni odczuwali większy pociąg seksualny w warunkach, w których oni sami oczekiwali wstrząsów silnych (por. rysunek 1).

Jak ów uzyskany efekt główny siły wstrząsów antycypowanych przez osoby badane ( $F = 22, 8; p = 0,001$ ) oraz nieobecność statystycznie istotnej interakcji interpretują



OB – osoba badana (mężczyzna)

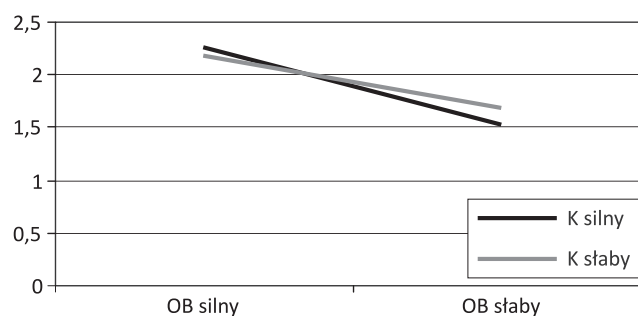
K – kobieta (współpracownica eksperymentatora)

*Rysunek 1.* Wskaźniki pociągu seksualnego w warunkach eksperymentalnych. Wskaźnikiem najsilniejszego pociągu seksualnego jest 5.

autorzy? Otóż jako dowód potwierdzający hipotezę badawczą: „Tak więc w tym badaniu nie pojawił się efekt wpływu «kobiety w sytuacji tragicznej» na pociąg seksualny” (Dutton, Aron, 1974, s. 516). Powyższa interpretacja wydaje się zastanawiająca, biorąc pod uwagę, iż: (1) brak racjonalnego uzasadnienia tego, aby właśnie taką hipotezę poddać empirycznej weryfikacji; (2) alternatywne wyjaśnienia teoretyczne lepiej uzasadniałyby uzyskany efekt (*vide* dalej); (3) mamy tu do czynienia z wieloma niedociągnięciami natury metodologicznej.

**ZZ3: Wskaźniki wyobraźni seksualnej.** Pomiar tej zmiennej zależnej nie wykazał istnienia statystycznie istotnego efektu głównego siły wstrząsów oczekiwanych przez osoby badane (por. rysunek 2), osiągnął on jedynie trend ( $F = 4, 22; p = 0,07$ ) – statystycznie istotna okazała się jedynie interakcja ( $F = 4, 73; p = 0,05$ ).

Trudno owe wyniki, bazując na wyjaśnieniu teoretycznym zaproponowanym przez Duttona i Arona a odwołującym się do efektu błędnej atrybucji, w racjonalny sposób zinterpretować. Jak uzyskane rezultaty mają się do głównej idei przyświecającej eksperymentowi laboratoryjnemu, a mianowicie postaci (*gestalt*) „kobieta w sytuacji tragicznej”? Jak można je zinterpretować, skoro stoją w opozycji do wyników uzyskanych z pomiaru ZZ2? Autorzy wykazali wszakże uprzednio istnienie efektu głównego antycypowanej przez osoby badane siły szoku – bez względu na to, czy pomocnica oczekuje szoku słabego, czy silnego, wydawała się im atrakcyjniejsza. Tutaj ów efekt nie wystąpił – niemniej mamy do czynienia z trendem ( $p = 0,07$ ). Jednakże, bazując na dzisiejszych standardach metodologicznych, uzyskane rezultaty powinny raczej statystycznie niepokoić niż uspokajać (Cumming, 2008). Rzetelną miarą, obecnie obowiązującą, byłaby chociażby wielkość efektu (Cumming, 2008), którą dla



OB – osoba badana (mężczyzna)

K – kobieta (współpracownica eksperymentatora)

*Rysunek 2.* Wskaźniki wyobraźni seksualnej w warunkach eksperymentalnych. Pięć oznacza największy wynik.

analizy wariancji jest omega kwadrat (Howell, 2010) – trudno jednak wymagać od Duttona i Arona, aby znali i stosowali owe standardy *ante factum*.

### PODSUMOWANIE

Cóż zatem tak naprawdę wykazali Dutton i Aron? Jak interpretują uzyskane rezultaty? Czy owe wnioski są uprawnione? Autorzy w „Ogólnej dyskusji wyników” konstatają, iż:

(1) „Wyniki 3. eksperymentu wspierają stwierdzenie, że silna emocja *per se* zwiększa pociąg seksualny osób badanych do współpracownicy eksperymentatora” (Dutton, Aron, 1974, s. 516). Powyższa interpretacja jest nadużyciem, zważywszy na to, iż nie został zreplikowany efekt główny ZZ2 odnośnie do ZZ3, sam zaś plan eksperymentalny zawierał wiele uchybień natury metodologicznej.

(2) „Rezultaty tych badań wydają się potwierdzać istnienie związku pomiędzy emocją a atrakcyjnością seksualną” (s. 516). W świetle moich powtórnych analiz powyższy wniosek jest nieuprawniony – ów związek istnieje nawet wtedy, jak twierdzą autorzy, gdy osoby badane dzwonią po kilkunastu (?) godzinach do eksperymentatorki, nie będąc już pod wpływem uprzednio doświadczonego stanu afektywnego.

(3) „Teoretyczne implikacje tych rezultatów dostarczają dodatkowego wsparcia na korzyść założeń teoretycznych, z których oryginalne hipotezy się wywodzą: Schachtera i Singera (1962) koncepcja poznawczego reetykietowania emocji” (s. 516). I dalej: „(...) dokonując poznawczego reetykietowania lęku jako pobudzenia seksualnego, osoby badane odczuwały mniejszy lęk (...) w warunkach, w których antycypowały doznanie silnego szoku i współpracownica eksperymentatora była obecna, w porównaniu z warunkami, gdy potencjalny obiekt seksualny nie był obecny” (s. 517). Jest to jawna nadinterpretacja. Autorzy wypaczyli założenia teorii Schachtera i Singera – dlaczego osoby badane miałyby pobudzenie, będące funkcją tak wyrazistych (*salient; unambiguous*) czynników, jak antycypacja otrzymania silnych wstrząsów czy przejście przez wąski, niestabilny i zawieszony 70 metrów nad rzeką most, reinterpretować jako pobudzenie o charakterze seksualnym? Wszakże oryginalna teoria Schachtera i Singera (1962) zakłada, iż aby doszło do owego reetykietowania, czynnik indukujący owo pobudzenie musi się charakteryzować wieloznacznością (*ambiguity*).

(4) „Tak więc, bez względu na to, jak zinterpretujemy mechanizm owego związku [pobudzenie – atrakcyjność; przyp. autorki], obecne badania wykazały jasno istnienie owego związku” (s. 517). Autorzy owego związku bezsprzecznie nie wykazali, sama zaś konstatacja: „bez względu na interpretację” budzi we mnie opór.

Ponadto autorzy nie wykazali jednoznacznie, że mężczyźni biorący udział w badaniach 1 i 2 odczuwali lęk (który jest przecież jednoznacznie awersywny) – w trakcie przeprowadzania owych badań eksperymentatorzy przecież nie pytali osób badanych o poziom odczuwanego lęku, aby „nie wzbudzać podejrzeń” (s. 512). Podążając tym tokiem myślenia – w eksperymencie laboratoryjnym zadanie pytania o poziom odczuwanego lęku przed dokonaniem pomiaru ZZ1 i ZZ2 powinno wzbudzać podejrzanie. Dlaczego w takim razie badacze owo pytanie zadali? Źródło pobudzenia może być przecież ewaluowane jednoznacznie pozytywnie (np. oglądanie komedii), neutralnie (np. ćwiczenia fizyczne) lub negatywnie (np. antycypacja otrzymania silnych szoków elektrycznych), przy czym ewaluacje te mogą być subiektywne, jak również obiektywne, i owe dwa rodzaje ewaluacji mogą się na siebie nakładać bądź nie. Interesujące z tej perspektywy wydaje się to, czy oszacowania poziomu odczuwanego lęku byłyby takie same, gdyby osoby badane dokonywały ich na końcu, zamiast na początku badania.

Poza tym warunki z badań 1 i 2 oraz eksperymentu laboratoryjnego były nieekwiwalentne – podczas badań terenowych mężczyźni byli indagowani po tym, jak źródło pobudzenia (nie można stwierdzić, czy na pewno przez wszystkich uczestników badania było ewaluowane jednoznacznie pejoratywnie) zostało usunięte (tj. po przejściu przez most chybliwy), w laboratorium natomiast osoby badane dopiero spodziewały się – jednoznacznie awersywnego, w warunkach antycypacji silnego szoku – doświadczenia owego bodźca. Owe dwie sytuacje badawcze są nieekwiwalentne, również jeśli chodzi o wzbudzenie motywów afiliacyjnych. Dodatkowo w badaniach 1 i 2 mieliśmy do czynienia, co podkreślają sami autorzy, z autoselekcją osób badanych (*self-selecting-subject*), w eksperymencie właściwym zaś – z losowym przydziałem osób badanych do poszczególnych warunków badawczych. Zadać należałoby również pytanie, dlaczego autorzy do planu eksperymentalnego (badanie 3) nie włączyli pomiaru – *per analogiam* – zmiennych zależnych w postaci przyjęcia numeru telefonu i późniejszego zadzwonienia do kobiety poznanej w laboratorium? Wreszcie, nawiązując do tego, jak doniosłą rolę w naukowo uprawianej psychologii odgrywają replikacje, warto podkreślić, iż efekt uzyskany przez Duttona i Arona w eksperymencie laboratoryjnym, a mianowicie to, iż mężczyźni uznali za atrakcyjniejszą pomocnicę eksperymentatora w grupach, które oczekiwały na silne szoki elektryczne, niż w grupach, które antycypowały otrzymanie słabego, „przez niektóre osoby badane określanego jako przyjemny” (s. 515), szoku, nie został zreplikowany ani przez Kenricka, Cialdinię i Lindera (1979), którzy podjęli aż cztery próby wykazania istnienia

owego efektu, ani przez Catherine Riordan i Jamesa Tedeschiego (1983). W jednym z artykułów Douglas Kenrick i Robert Cialdini (1977) skonstatowali, iż ów bazujący na wywołaniu lęku via antycypacja otrzymania silnych szoków (*threat-of-shock*) plan eksperymentalny nie może prowadzić do pojawienia się efektu błędnego przypisania źródła pobudzenia (*misattribution effect*). Takie same wyniki, tj. niewystąpienie związku pobudzenie–atrakcyjność (*arousal–attraction link*) w warunkach, w których bodziec lękotwórczy jest wyrazisty (*salient*), uzyskali, przed opublikowaniem przez Donalda Duttona i Arthura Arona artykułu w 1974 r., Brehm i Anderman (1972; za: Kenrick, Cialdini, Linder, 1979). Wyniki te nie zaskakują i są zgodne z faktycznymi, nie zaś zreinterpretowanymi założeniami dwuczynnikowej teorii emocji (Schachter, Singer, 1962). Czy zatem jest możliwe, aby wyniki uzyskane przez Duttona i Arona były artefaktem? Być może atrakcyjna współpracownica eksperymentatora niezależnie od warunków (bodziec awersywny vs. nieawersywny) wzbudzała zainteresowanie osób badanych per se, nie zaś na zasadzie błędnego określenia źródła pobudzenia. Biorąc pod uwagę fakt, iż powyższe wyniki nie zostały zreplikowane – jak wykazał Douglas Kenrick wraz ze współpracownikami (1979), oszacowania atrakcyjności pomocnicy eksperymentatora nie zależały od wzbudzenia vs. niewzbudzenia lęku, ale od niewielkich różnic w atrakcyjności fizycznej współpracownic eksperymentatorów (*small differences in the physical attractiveness of the confederates*; Kenrick i in., 1979, s. 332). Za bardziej racjonalny należałoby uznać wniosek, iż powyższe dane stanowią kilka dowodów na to, że wyjaśnienie teoretyczne, którym posłużyli się Dutton i Aron, jest nieadekwatne, natomiast związek pobudzenie–atrakcyjność istnieje.

#### ALTERNATYWNE WYJAŚNIENIA TEORETYCZNE ZWIĄZKU POBUDZENIE–ATRAKCYJNOŚĆ

Alternatywa teoretyczna wobec wyjaśnienia odwołującego się do efektu błędnego określenia źródła pobudzenia (nurt atrybucyjny) pojawiła się dość szybko – w artykule opublikowanym przez Douglasa Kenricka i Roberta Cialdiniego w 1977 r. Autorzy zaproponowali wyjaśnienie sugerujące, iż teoria wzmocnień negatywnych (*negative-reinforcement theory*) lepiej wyjaśnia rezultaty uzyskane przez Duttona i Arona. Otóż po zaindukowaniu negatywnie odbieranego przez osoby badane pobudzenia, obecność innej osoby może stanowić środek do usunięcia owego, bazującego na lęku, pobudzenia (osoba ta pełni funkcję wzmocnienia negatywnego), co w efekcie prowadzi przez samą jej obecność do wzrostu jej spostrzeganej atrakcyjności (niezależnie od atrakcyjności faktycznej). A to per se jest nagradzające i prowadzi w konsekwencji do redukcji

odczuwanego przez uczestników badania napięcia (por. także Kenrick, Johnson, 1979).

Kolejną propozycją teoretyczną, według której mediatorem związku pobudzenie–atrakcyjność jest odwołanie się do przejawiania reakcji dominującej (*response-facilitation theory*) zaproponował James Allen ze współpracownikami (Allen, Kenrick, Linder, McCall, 1989). Zgodnie z ich teorią mężczyźni, którzy przeszli przez niestabilny most zawieszony nad rzeką Capilano, odczuwając silne pobudzenie, przejawiali reakcję dominującą (tj. nawykową, automatyczną) w odpowiedzi na bodziec w postaci atrakcyjnej kobiety. Rodzaj tej reakcji był przy tym uzależniony od tego, czy osoba bodźcowa była tej samej płci, czy też nie, oraz od stopnia jej atrakcyjności – „pociąg seksualny jest reakcją dominującą w odpowiedzi na pojawienie się atrakcyjnej osoby, z którą potencjalnie można umówić się na randkę, zaś awersja jest reakcją dominującą na pojawienie się nieatrakcyjnej osoby, z którą można by umówić się potencjalnie na randkę” (Allen i in., 1989, s. 261). Zgodnie z założeniami tego modelu wyjaśniającego efekt pobudzenie–atrakcyjność, do wzrostu percypowanej atrakcyjności obiektu bodźcowego dojdzie tylko wtedy, gdy obiekt ów jest atrakcyjny i odmiennej płci, niezależnie od walencji źródła pobudzenia oraz niezależnie od tego, czy to źródło jest wyraziste, czy też nie.

Clou owych dwu powyższych wyjaśnień teoretycznych stanowił poziom pobudzenia. Trzecią z najbardziej znanych koncepcji wyjaśniających omawianą zależność jest wyjaśnienie autorstwa, co ciekawe, m.in. Donalda Duttona i Arthura Arona (Aron, Dutton, Aron, Iverson, 1989). Koncepcja owa abstrahuje od samego pobudzenia i podkreśla rolę nowości sytuacji (*novelty of the situation*), jej nietypowości – ów czynnik sam z siebie wywołuje wzrost atrakcyjności osoby bodźcowej, niezależnie od intensywności doświadczanego przez podmiot (osobę badaną) pobudzenia.

#### WYNIKI METAANALIZY

Metaanaliza jest uważana za najlepszą metodę pozyskiwania rzetelnych danych, jak również integracji owych danych (por. APA *Publication Manual*, 2010; Cumming, 2008; Wojciszke, 2006). W 1998 r. wyniki analizy integracyjnej, przeprowadzonej przez zespół Craiga Fostera, której przedmiotem zainteresowania była eksploracja związku pobudzenie – atrakcyjność, zostały opublikowane w *Journal of Personality and Social Psychology*. Jak wykazano, wyjaśnienie związku pobudzenie–atrakcyjność zaproponowane przez Kenricka i Cialdiniego (1977), odwołujące się do teorii wzmocnień, nie jest spójne z wynikami owej analizy. Potwierdzenie uzyskały natomiast teorie: błędnego przypisania źródła pobudzenia (przy czym wyłącznie

wtedy, kiedy ów proces atrybucji jest procesem automatycznym, nie zaś wydeliberowanym, tak jak zakładali to Schachter i Singer w swojej teorii) oraz teoria odwołująca się do przejawiania reakcji dominującej (Allen i in., 1989). Moderatorami owego związku są: atrakcyjność i płeć współpracownika/-cy eksperymentatora (tzn. w stanie doświadczania pobudzenia za atrakcyjniejsze uważamy piękne osoby płci przeciwnej, za najmniej atrakcyjne zaś – brzydkie osoby płci przeciwnej), natomiast walencja odczuwanego pobudzenia nie ma wpływu na oszacowanie atrakcyjności (co jest spójne z przedstawioną przeze mnie interpretacją wyników uzyskanych przez Duttona i Arona – trudno zgodzić się z hipotezą, iż w badaniach 1 i 2 wysoko zawieszony, niestabilny most indukował u wszystkich osób badanych wyłącznie lęk). Co interesujące – wyniki owych analiz wykazały, że zwiększona percepcja atrakcyjności współpracownika/-cy występuje również wówczas, gdy źródło pobudzenia jest jednoznaczne. Wielkość efektu pobudzenie–atrakcyjność (szacowana na trzech wymiarach: atrakcyjność fizyczna, lubienie i atrakcyjność romantyczna) waha się od małej do średniej (*small to moderate*; Foster i in., 1989, s. 99).

### WNIOSKI

Biorąc pod uwagę powyższe dane, jak również wyniki przeprowadzonych przeze mnie reanaliz należałoby zadać następujące pytania: (1) co Dutton i Aron ponad wszelką wątpliwość wykazali? (2) czy rezultaty tego klasycznego już artykułu powinny zniknąć z annałów naukowo uprawianej psychologii?

Na jedenaście analiz (wyłącznie warunki kobieta–mężczyzna) trzy okazały się statystycznie istotne<sup>10</sup> i bezsprzecznie zgodne z hipotezą: ZZ: test TAT z badania 1 i 2 (każdorazowo  $p = 0,01$ ); ZZ2: oszacowania pociągu seksualnego z badania 3 (efekt główny szoku antycypowanego przez badanych mężczyzn;  $p = 0,01$ ). A zatem autorzy wykazali bez wątpienia, iż osoby bardziej pobudzone (abstrahując od tego, czy było to pobudzenie interpretowane jako awersywne, czy nieawersywne) używały więcej słów o konotacji seksualnej w teście TAT<sup>11</sup>, w porównaniu z osobami niepobudzonymi. Działo się tak w warunkach, w których: (1) na obu mostach indagowała kobieta (badanie 1) – mężczyźni przechodzący przez most

eksperymentalny zawierali więcej treści o takiej konotacji w historyjce niż mężczyźni, którzy przeszli przez most kontrolny; (2) indagowała kobieta wyłącznie na moście eksperymentalnym (badanie 2) – statystycznie istotnie więcej takich treści wystąpiło w grupie eksperymentalnej. Różnic statystycznie istotnych badacze nie odnotowali w warunkach, gdy na obu mostach indagował mężczyzna (badanie 1).

Odnosnie do pozostałych ośmiu analiz, cztery okazały się niezgodne z hipotezą – ZZ: uległość wobec prośby z badania 1 i 2 (każdorazowo  $p > 0,7$ ) oraz ZZ1: przyjęcie numeru telefonu, również z badania 1 i 2 (w obu przypadkach  $p > 0,6$ ). Odnosnie zaś do ZZ2: zadzwonienie do eksperymentatorki – z badania 1 i 2 – chociaż uzyskany wynik przekroczył konwencjonalny poziom statystycznej istotności (każdorazowo  $p = 0,02$ ), nie jest on zgodny z hipotezą w świetle założeń teoretycznych, na które powołują się autorzy. Jeśli chodzi o pozostałe dwie zmienne zależne z eksperymentu 3: ZZ1: pomiar odczuwanego lęku oraz ZZ3: wskaźniki wyobraźni seksualnej – wyniki są niejasne, bo nie wiemy, które grupy z którymi autorzy porównywali (ZZ1) oraz są niespójne z poprzednio uzyskanymi rezultatami analiz (ZZ3 vs. ZZ2).

Konkluzja, w ogólniejszym sensie, odnosząca się do tego, czy należałoby podważyć istnienie (dobrze ugruntowanego w psychologii eksperymentalnej) efektu pobudzenie–atrakcyjność, rzecz jasna brzmieć powinna jednoznacznie: nie. W węższym sensie natomiast, tj. odnośnie do treści analizowanego tu przeze mnie artykułu Duttona i Arona, wnioski, jakie należałoby wyciągnąć, są następujące: (1) powinno się dokonać replikacji wszystkich trzech badań, które ww. autorzy przeprowadzili; (2) w kolejnym kroku należałoby ów plan badawczy, wykonany przez Duttona i Arona, udoskonalić, niwelując wszelkie nieścisłości metodologiczne; ponadto (3) należałoby uwzględnić alternatywne wyjaśnienia teoretyczne – autorzy w dość nonszalancki sposób interpretowali i dopasowywali teorie do hipotez, nie zaś odwrotnie; konieczne, w moim odczuciu, jest także (4) bardziej refleksyjne podejście do cytowań (w artykułach, monografiach, podręcznikach akademickich etc.) – w tym przypadku należałoby dokonać w przyszłości korekt i uściślenia rezultatów uzyskanych przez Duttona i Arona.

### LITERATURA CYTOWANA

- Agresti, A. (2002). *Categorical data analysis* (wyd. 2). Hoboken, NJ: John Wiley & Sons.
- Agresti, A. (2007). *An introduction to categorical data analysis* (wyd. 2). Hoboken, NJ: John Wiley & Sons.
- Allen, J. B., Kenrick, D. T., Linder, D. E., McCall, M. A. (1989). Arousal and attraction: A response-facilitation

<sup>10</sup> To, że różnice między grupami są statystycznie istotne nie oznacza, iż – z punktu widzenia praktycznego – mają one istotne znaczenie (por. Howell, 2010).

<sup>11</sup> Zadać należałoby pytanie, dlaczego w badaniach 1 i 2 do tej zmiennej zależnej stosowano testy dwustronne, wygenerowawszy uprzednio hipotezę kierunkową (por. Howell, 2010), podczas gdy w eksperymencie 3 badacze stosują testy jednostronne (ZZ1: Pomiar odczuwanego lęku).

- alternative to misattribution and negative-reinforcement models. *Journal of Personality and Social Psychology*, 57, 261–270, doi:10.1037/0022-3514.57.2.261.
- American Psychological Association (1994). *Publication manual of the American Psychological Association* (wyd. 4). Washington, DC: APA.
- American Psychological Association (2010). *Publication manual of the American Psychological Association* (wyd. 6). Washington, DC: APA.
- Aron, A., Coups, E. J., Aron, E. N. (2013). *Statistics for psychology* (wyd. 6). New Boston: Pearson.
- Aron, A., Dutton, D. G., Aron, E. N., Iverson, A. (1989). Experiences of falling in love. *Journal of Social and Personal Relationships*, 6, 243–257, doi: 10.1177/0265407589063001.
- Aronson, E., Wilson, T. D., Akert, R. M. (1997). *Psychologia społeczna. Serce i umysł*. Poznań: Wydawnictwo Zysk i S-ka.
- Aussems, M-C. E., Boomsma, A., Snijders, T. A. B. (2011). The use of quasi-experiments in the social sciences: A content analysis. *Qual Quant*, 45, 21–42, doi: 10.1007/s11135-009-9281-4.
- Barclay, A. M., Haber, R. N. (1965). The relation of aggressive to sexual motivation. *Journal of Personality*, 33, 462–475, doi: 10.1111/j.1467-6494.1965.tb01398.x.
- Brzeziński, J. (2004). *Metodologia badań psychologicznych*. Warszawa: Wydawnictwo Naukowe PWN.
- Campbell, D. T., Stanley, J. C., Gage, N. L. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton, Mifflin and Company.
- Cantor, J. R., Zillmann, D., Bryant, J. (1975). Enhancement of experienced sexual arousal in response to erotic stimuli through misattribution of unrelated residual excitation. *Journal of Personality and Social Psychology*, 32, 69–75, doi: 10.1037/h0076784
- Carver, R. P. (1978). The case against significance testing. *Harvard Educational Review*, 48, 378–399.
- Cohen, J. (1994). The earth is round ( $p = 0,05$ ). *American Psychologist*, 49, 997–1003, doi: 10.1037/0003-066X.49.12.997.
- Cumming, G. (2008). Replication and  $p$  intervals –  $p$ -values predict the future only vaguely, but confidence intervals do much better. *Perspectives on Psychological Science*, 3, 286–300.
- Davies, H. T. O. (1998). When can odds ratios mislead? *British Medical Journal*, 316, 989–991, doi: http://dx.doi.org/10.1136/bmj.316.7136.989.
- Davies, H. T. O., Crombie, I. K. (2009). What are confidence intervals and  $p$ -values? Pobrano z: [http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what\\_are\\_conf\\_inter.pdf](http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/what_are_conf_inter.pdf).
- Dutton, D. G., Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, 30, 510–517, doi: 10.1037/h0037031.
- Dutton, D. G., Aron, A. P. (2002). Nieco danych świadczących o wzroście pociągu seksualnego w warunkach silnego lęku. W: E. Aronson (red.), *Człowiek istota społeczna. Wybór tekstów* (s. 592–608). Warszawa: Wydawnictwo Naukowe PWN.
- Ellis, P. D. (2010). *The essential guide to effect sizes: An introduction to statistical power, meta-Analysis and the interpretation of research results*. Cambridge: Cambridge University Press.
- Fidler, F., Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *International Statistical Institute, 55<sup>th</sup> Session*. Pobrano z: <http://www.stat.auckland.ac.nz/~iase/publications/13/Fidler-Cumming.pdf>.
- Field, A. (2010). *Discovering statistics using SPSS (and sex, drugs and rock n'roll)*. London: Sage Publications.
- Fisher, R. A. (1962). Confidence limits for a cross-product ratio. *Australian Journal of Statistics*, 4, 41, doi: 10.1111/j.1467-842X.1962.tb00285.x.
- Fleiss, J. L., Lewin, B., Paik, M. C. (2003). *Statistical methods for rates and proportions* (wyd. 3). Hoboken, NJ: John Wiley & Sons.
- Foster, C. A., Witcher, B. S., Campbell, W. K., Green, J. D. (1998). Arousal and attraction: Evidence for automatic and controlled processes. *Journal of Personality and Social Psychology*, 74, 86–101, doi: 10.1037/0022-3514.74.1.86.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., Krüger, L. (1989). *The empire of chance*. Cambridge: Cambridge University Press.
- Goodman, L. A., Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49, 732–764, doi: 10.2307/2281536
- Goodman, L. A., Kruskal, W. H. (1979). *Measures of association for cross classifications*. New York: Springer Verlag.
- Haddock, C. K., Rindskopf, D., Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3, 339–353, doi: 10.1037/1082-989X.3.3.339.
- Hogg, M. A., Vaughan, G. E. (2011). *Social psychology* (wyd. 6). Harlow: Pearson Education Limited.
- Howell, D. C. (2010). *Statistical methods for psychology* (wyd. 7). Belmont, CA: Wadsworth.
- Howitt, D., Cramer, D. (2011). *Introduction to statistics in psychology* (wyd. 5). Harlow, Essex: Pearson Education.
- Hurlburt, R. T. (1994). *Comprehending behavioral statistics*. Pacific Grove, CA: Brooks, Cole Publishing.
- Jackson, S. J. (2008). *Research methods and statistics. A critical thinking approach* (wyd. 3). Belmont, CA: Wadsworth.
- Kenrick, D. T., Cialdini, R. B. (1977). Romantic attraction: Misattribution versus reinforcement explanations. *Journal of Personality and Social Psychology*, 35, 381–391, doi: 10.1037/0022-3514.35.6.381
- Kenrick, D. T., Cialdini, R. B., Linder, D. E. (1979). Misattribution under fear-producing circumstances: Four failures to replicate. *Personality and Social Psychology Bulletin*, 5, 329–334, doi: 10.1177/014616727900500312
- Kenrick, D., Johnson, G. (1979). Interpersonal attraction in aversive environments: A problem for the classical conditioning paradigm? *Journal of Personality and Social Psychology*, 37, 572–579, doi: 10.1037/0022-3514.37.4.572.

- Kenrick, D. T., Neuberg, S. L., Cialdini, R. (2002). *Psychologia społeczna. Rozwiązane tajemnice*, przeł. A. Nowak. Gdańsk: Gdańskie Wydawnictwo Psychologiczne.
- Lange, L. L. (2005). Sleeping beauties in psychology: Comparisons of 'hits' and 'missed signals' in psychological journals. *History of Psychology*, 8, 194–217, doi: 10.1037/1093-4510.8.2.194
- Lehrer, D., Leschke, J., Lhachimi, S., Vasiliu, A., Weiffen, B. (2007). Negative results in social science. *European Political Science*, 6, 51–68, doi:10.1057/palgrave.eps.2210114.
- Lewandowski, G. W., Aron, A. P. (2004). Distinguishing arousal from novelty and challenge in initial romantic attraction between strangers. *Social Behavior and Personality*, 32, 361–372, doi: 10.2224/sbp.2004.32.4.361.
- Oakes, M. L. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Reisenzein, R. (1983). The Schachter theory of emotion: Two decades later. *Psychological Bulletin*, 94, 239–264, doi: 10.1037/0033-2909.94.2.239
- Riordan, C., Tedeschi, J. (1983). Attraction in aversive environments: Some evidence for classical conditioning and negative reinforcement. *Journal of Personality and Social Psychology*, 44, 683–692, doi: 10.1037/0022-3514.44.4.683
- Rogers, W. S. (2003). *Social psychology. Experimental and critical approaches*. Maidenhead: Open University Press.
- Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, 57, 416–428, doi: 10.1037/h0042040
- Schachter, S., Singer, J. E. (1962). Cognitive, social and physiological components of the emotional state. *Psychological Review*, 69, 379–399, doi: 10.1037/h0046234
- Shadish, W. R., Cook, T. D., Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Sim, J., Reid, N. (1999). Statistical inference by confidence intervals: Issues of interpretation and utilization. *Physical Therapy*, 79, 186–195.
- Sze, H. T., Say, B. T. (2010). The correct interpretation of confidence intervals. *Proceedings of Singapore Healthcare*, 19, 276–278. Pobrano z: [http://www.singhealthacademy.edu.sg/Documents/Publications/Vol19No32010/16\\_Statistics.pdf](http://www.singhealthacademy.edu.sg/Documents/Publications/Vol19No32010/16_Statistics.pdf).
- Whitley, B. E., Jr. (2002). Group comparison research. W: M. W. Wiederman, B. E. Jr. Whitley (red.), *Handbook for conducting research on human sexuality* (s. 223–254). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Wilkinson, L., The APA Task Force on Statistical Inference (1999). Statistical methods in psychology journals. Guidelines and explanations. *American Psychologist*, 54, 594–604, doi: 10.1037/0003-066X.54.8.594.
- Wojciszke B. (2006). Systematycznie modyfikowane autoreplikacje: logika programu badań empirycznych w psychologii. W: J. Brzeziński (red.), *Metodologia badań psychologicznych. Wybór tekstów* (s. 44–68). Warszawa: Wydawnictwo Naukowe PWN.
- Wojciszke, B. (2011). *Psychologia społeczna*. Warszawa: Wydawnictwo Naukowe Scholar.
- Zillmann, D., Katcher, A. H., Milavsky, B. (1972). Excitation transfer from physical exercise to subsequent aggressive behavior. *Journal of Experimental Social Psychology*, 8, 247–259, doi: 10.1016/S0022-1031(72)80005-2.
- Zuckerman, M., Bone, R., Neary, R., Mangelsdorff, D., Brustman, B. (1972). What is the sensation seeker? Personality trait and experience correlates of the sensation-seeking scales. *Journal of Consulting and Clinical Psychology*, 39, 308–321, doi: 10.1037/h0033398.

## Some evidence that Dutton and Aron 40 years ago had shown not necessarily what they intended to show

Katarzyna Szczucka

*University of Social Sciences and Humanities, Faculty of Wrocław*

### ABSTRACT

The article by Dutton and Aron (1974), in which the authors present results of three studies, is one of the classics in the history of social psychology. According to the original interpretation, the sufficient condition of an increase in sexual attraction toward an object (attractive woman), presented shortly after or during waiting for an aversive stimulus, is an induction in male participants of a strong autonomic arousal (*via* crossing by the high suspended bridge or anticipating receipt of strong electric shocks). However, this interpretation seems unconvincing, and the results – inconclusive. This article presents a critical interpretation of the results obtained by Dutton & Aron (1974) through reanalysis of the original data. In addition the article presents a short review of theories concerning the arousal-attraction link.

Keywords: *arousal-attraction link; effect of arousal on attraction; misattribution effect; misinterpretations; Dutton & Aron*

Złożono tekst: 3.05.2012

Złożono poprawiony tekst: 20.12.2012/15.06.2013

Zaakceptowano do druku: 1.07.2013