



O NIEKTÓRYCH PRAKTYKACH BADAWCZYCH, KTÓRE PROWADZĄ DO GROMADZENIA I PUBLIKOWANIA NIEWIARYGODNYCH WYNIKÓW BADAŃ NAUKOWYCH

*Piotr Francuz, Katolicki Uniwersytet Lubelski Jana Pawła II,
Katedra Psychologii Eksperymentalnej*

Podstawowym kryterium wiarygodności wiedzy gromadzonej w każdej dziedzinie nauki jest kryterium metodologiczne. Określa ono zestaw dopuszczalnych na jej gruncie procedur i narzędzi badawczych oraz form publikowania informacji dotyczących odkrywanych prawidłowości. Każda metoda ma oczywiste zalety, ale ma również i ograniczenia. Ich świadomość, a ponadto uczciwe przestrzeganie reguł przyjętego sposobu postępowania stanowią niekwestionowaną gwarancję gromadzenia wiedzy ze znakiem najwyższej jakości, czyli wiedzy, która jest godna wiary.

Celem niniejszego artykułu jest krytyczna analiza niektórych praktyk badawczych obecnych w naukach empirycznych. Zostanie ona przeprowadzona na przykładach pochodzących przede wszystkim z dziedziny zachowań konsumenckich. W ciągu minionych ponad 30 lat wielokrotnie zwracano uwagę na różne trudności, a nawet poważne uchybienia metodologiczne, które budziły uzasadniony niepokój teoretyków zachowań konsumenckich (zob. np. Calder, Phillips, Tybout 1982; Ferber 1977; Jacoby 1978; Kollat, Blackwell, Engel 1972; Lynch 1982; Wells 1993). Szczegółowy przegląd literatury w tym zakresie pozwala na określenie kilku kategorii problemów o charakterze metodologicznym, które skutecznie izolują badaczy zachowań konsumentów od rzeczywistości oraz stanowią czynnik znacznie wyhamowujący zarówno rozwój teorii, jak i aplikacji w tej dziedzinie. Należy tu dodać, że omawiane dalej zagadnienia w porównywalnym stopniu odnoszą się również do wielu innych obszarów badawczych.

Co najmniej od połowy XX wieku dziedzina zachowań konsumenckich budzi niesłabnące zainteresowanie zarówno praktyków, bezpośrednio zaangażowanych w produkcję, promocję i sprzedaż różnych produktów, jak również naukowców. Stanowi ona interdyscyplinarne pole działań badawczych angażujących m.in. psychologów, socjologów, ekonomistów i antropologów. Wśród naukowców podejmujących tę problematykę szczególną pozycję zajmują psychologowie. Wyniki ich badań przeniesione na płaszczyznę zachowań ludzi na rynku stanowią istotne źródło inspiracji licznych prac badawczych publikowanych w podręcznikach i czasopismach z zakresu szeroko rozumianego *consumer behavior*.



Psychologia a zachowania konsumenckie

Powiązania między psychologią a dziedziną wiedzy o zachowaniach konsumenckich wydają się dość oczywiste. Podręcznikowa definicja zachowań konsumenckich, rozumianych jako „umysłowa, emocjonalna i fizyczna aktywność ludzi zaangażowanych w wybieranie, nabywanie, używanie i dysponowanie produktami i usługami w celu zaspokojenia swoich potrzeb” (Wilkie 1994), najwyraźniej lokalizuje rdzeń zainteresowań badaczy zachowań konsumenckich na gruncie psychologii. Można by nawet powiedzieć, że nauka o zachowaniach konsumenckich jest psychologią zastosowaną do analizy zachowań ludzi na rynku usług i produktów.

Spośród wielu powodów zainteresowania się psychologa problematyką zachowań konsumenckich kilka wydaje się szczególnie ważnych.

Po pierwsze, do badań zachowań konsumenckich z powodzeniem można aplikować znane na gruncie psychologii teorie percepcji, uczenia się i pamięci, myślenia, emocji i motywacji oraz modele zachowań społecznych, a także szczegółowe strategie badawcze wypracowane na gruncie metodologii psychologii.

Po drugie, szeroki kontekst, w jakim analizowane są zachowania konsumenckie, dostarcza atrakcyjnego materiału badawczego: reklamy, znaki towarowe lub przedmioty codziennego użytku, w których nabycie osobiście może być zaangażowana osoba badana. Czynniki te w istotnym stopniu wpływają na „naturalizację” badania laboratoryjnego, ponieważ zmuszają badaczy do uwzględniania w eksperymentach psychologicznych takich zmiennych, jak subiektywne cele, potrzeby i oczekiwania ludzi podejmujących określone decyzje konsumenckie.

Wreszcie, po trzecie, sfera zachowań konsumenckich, która z jednej strony angażuje specyficzne procesy poznawcze i emocjonalno-motywacyjne osób funkcjonujących na rynku, a z drugiej makro- i mikroekonomiczne procesy funkcjonowania tegoż rynku otwierają przed psychologiem niezwykle szeroką przestrzeń badawczą, w której bez trudu może znaleźć miejsce do realizacji swoich teoretycznych i praktycznych zainteresowań.

Warto więc, jak sądzę, nieco bliżej przyjrzeć się teoretycznym i praktycznym osiągnięciom badaczy zachowań konsumenckich z perspektywy metodologicznej oraz ich merytorycznej wartości dla wiedzy o funkcjonowaniu człowieka.

Wartość predykcyjna wyników badań konsumenckich

Inspiracją do podjęcia krytycznej analizy wyników badań z zakresu zachowań konsumenckich stała się dla mnie lektura artykułu Scotta Armstronga (1991), który postawił sobie za cel udzielenie odpowiedzi na pytanie: „Czy eksperci w dziedzinie analiz zachowań konsumenckich (tzn. naukowcy i praktycy marketingowi) trafniej przewidują te zachowania niż nieekspertki (np. studenci)?” Odpowiedź na to pytanie pozornie wydaje się oczywista, stąd też przypuszczenia autora zostały wyrażone w postaci następujących hipotez:



(1) Eksperti trafniej przewidują zachowania konsumenckie niż nieeksperti; (2) Naukowcy trafniej przewidują zachowania konsumenckie niż praktycy.

Do badań Armstrong wykorzystał materiały opublikowane w *Journal of Consumer Research* w latach 1977-1987. Według ściśle określonych reguł wybrał 20 artykułów, na podstawie których przygotował krótkie opisy eksperymentów wraz ze stawianymi w nich hipotezami badawczymi. W sumie materiał przekazany osobom badanym do oceny obejmował 105 hipotez dotyczących różnych aspektów zachowań konsumenckich.

Badania przeprowadzono w trzech losowo dobranych grupach: (1) naukowców, członków *Association for Consumer Research*, (2) praktyków marketingowych, nienaukowców, członków *American Marketing Association* oraz (3) studentów niezwiązanych ani z praktyką, ani z edukacją marketingową. Wszystkie osoby badane otrzymały zestawy hipotez i ich zadaniem było dokonanie oceny tych twierdzeń w kategoriach prawdziwości-falszywości.

Wyniki badań Armstronga nie potwierdziły żadnej z wysuniętych przez niego hipotez. Okazało się, że ani eksperci nie przewidywali trafniej niż nieeksperti (52,6% do 56,6%), ani naukowcy trafniej niż praktycy (51,3% do 58,2%). Co więcej, proporcja prawidłowych odpowiedzi u praktyków marketingowych i studentów nie różniła się istotnie od losowego rozkładu, a u naukowców była nawet istotnie niższa niż wynikałoby to z rozkładu losowego.

Przed opublikowaniem tych danych Armstrong zapytał również 50 recenzentów należących do zespołu wydawców *Journal of Consumer Research* o ich oczekiwania, co do wyniku badania naukowców, praktyków i studentów. Przewidywania 43 recenzentów, którzy odpowiedzieli na te pytania zasadniczo pokrywały się z hipotezami Armstronga, ale ich wariancja była zaskakująco duża. Na przykład w odniesieniu do praktyków marketingowych, którzy dokonają prawidłowej oceny hipotez, szacowany procent wyników wahał się w granicach 15-90%, w odniesieniu do naukowców rozrzut wynosił 40-90%, natomiast w odniesieniu do studentów – 20-80%.

W dyskusji nad wynikami badań Armstrong zaproponował dwa wyjaśnienia. Stwierdził, że wyjątkowo niskie wyniki ekspertów (naukowców i praktyków marketingowych) mogą przede wszystkim wynikać stąd, iż dobre diagnozy stawiają oni tylko w wąskiej specjalności, w której są fachowcami. Z kolei w odniesieniu tylko do naukowców uznał, że być może w toku rozwiązywania zadań testowych mogli oni wytworzyć sobie szczególne nastawienie co do samego badania i zawartych w nim hipotez. Mogli np. podejrzewać, że eksperymentator zastawił na nich jakieś „pułapki” i hipotezy wyrażające zdroworozsądkowo oczywiste relacje między zmiennymi eksperymentalnymi traktowali jako fałszywe.

Obie interpretacje budzą jednak pewne wątpliwości. Z wyjątkiem jednej, wszystkie hipotezy stosowane w materiale eksperymentalnym miały charakter na tyle ogólny, że ich weryfikacja nie tylko nie wymagała jakiegось specjalistycznej wiedzy, ale z powodzeniem mogła opierać się na wiedzy potocznej, zdroworozsądkowej. Z kolei drugi argument dotyczy w zasadzie wszystkich grup i tym samym nie wyjaśnia zachowania się tylko grupy naukowców.



Nie wchodząc w dalszą dyskusję nad interpretacją osiągniętych wyników badań, warto, jak sądzę, zastanowić się nad tym, gdzie tkwią źródła nietrafnych przewidywań formułowanych zwłaszcza przez naukowców?

Odpowiedzi na to pytanie nie trzeba szukać daleko. Najpierw wystarczy nieco uważniej przeanalizować strategię badawczą stosowaną przez Armstronga. Warto zwrócić uwagę np. na to, że ekspertów i nieekspertów badał on w zupełnie innych warunkach. Naukowcy i praktycy marketingowi indywidualnie otrzymali zestawy testów drogą pocztową i tą samą drogą przesyłali swoje odpowiedzi, natomiast studenci byli badani grupowo i specjalnie motywowani 25-dolarową nagrodą dla tego, który sformułuje najwięcej poprawnych predykcji. Co więcej, do analizy porównawczej wzięto pod uwagę wyniki 43 studentów oraz wyniki tylko tych ekspertów, którzy je odesłali z powrotem. Nie byłoby w tym nic nadzwyczajnego, gdyby nie fakt, że ostatecznie analizie poddano odpowiedzi tylko 16 na 100 naukowców i 12 na 100 praktyków, którzy otrzymali zestawy testowe.

W kontekście tak bezkrytycznie zastosowanej przez Armstronga strategii badawczej pod znakiem zapytania staje więc wartość osiągniętych przez niego wyników. I bynajmniej nie zmienia tego fakt, że w analogicznych badaniach przeprowadzonych nieco wcześniej przez Stephen Hocha (1988) ich autor doszedł do podobnych wniosków, co Armstrong. Niewłaściwa standaryzacja warunków badania oraz formułowanie ogólnych wniosków na podstawie danych pochodzących od niereprezentatywnych grup osób badanych musi budzić poważne wątpliwości, co do trafności i rzetelności osiągniętych wyników.

O kilku wątpliwych praktykach metodologicznych

Naukowa oprawa zdrowego rozsądku

Przegląd problemów o charakterze metodologicznym, na które chciałbym zwrócić uwagę, rozpocznę od zagadnienia, które najprościej ujmując można sprowadzić do pytania o merytoryczną wartość weryfikowanych hipotez. Krótko mówiąc, wyniki wielu badań konsumenckich, obudowane skomplikowaną aparaturą pojęciową, w gruncie rzeczy odzwierciedlają zdroworozsądkowo oczywiste fakty.

Donald Lichtenstein, Scot Burton i Eric Karson (1991) zajęli się problemem wartości perswazyjnej sloganów reklamowych w rodzaju: „Kosztowało \$49.95. Teraz tylko \$29.95”, które to struktury semantyczne ustalają relację między tzw. ceną odniesienia i aktualnie oferowaną, sugerując jakąś specjalną obniżkę na pewną klasę produktów. W wyniku przeprowadzonych eksperymentów nastawionych na testowanie licznych hipotez dotyczących porównywanych cen, autorzy badań ostatecznie doszli do wniosku, że bez względu na wysokość ceny odniesienia, aktualnie oferowana jest tym bardziej atrakcyjna dla klientów, im jest niższa (!). W tym kontekście nie pozostaje już nic innego, jak tylko postawić retoryczne pytanie: czy rzeczywiście do tego typu wniosków koniecznie trzeba dochodzić na drodze badań naukowych?

Inny przykład badań nastawionych tym razem na ocenę czynników wpływających na skuteczność reklamy pochodzi z pracy Thomasa Olneya, Morrisa Holbrooka i Rajeeva



Batry (1991). Zwrócili oni uwagę na to, że współczesne techniki wyboru programów telewizyjnych za pomocą pilota spowodowały, że telewidzowie znacznie częściej niż kiedyś przełączają aktualnie oglądany program na inny (*zapping*) lub szybko przesuwają do przodu taśmę wideo z uprzednio nagrany programem wtedy, gdy jest on przerywany przez blok reklamowy (*zipping*). Zdaniem autorów eksperymentu pociąga to za sobą konieczność rewizji niektórych poglądów na temat skuteczności reklamy w telewizji, np. z punktu widzenia częstości jej prezentacji, która może być znacznie obniżona przez zasygnalizowane zachowania się telewidzów.

W swoich badaniach wykorzystali oni 146 oryginalnych reklam telewizyjnych. Każda reklama w trakcie wielogodzinnych sesji została oceniona przez kilkunastu niezależnych sędziów pod względem 63 wymiarów skalujących różne aspekty ich ładunku emocjonalnego i informacyjnego. W ciągu ponad pół roku indywidualnie przebadano ponad 100 osób (studentów), angażując w te badania tak duże środki finansowe, że przekroczony został budżet i nie wszystkie zaplanowane przez nich aspekty badań zostały zrealizowane. Do analizy danych eksperymentalnych zastosowano skomplikowane procedury wielozmiennej analizy regresji, po to by ostatecznie dojść do wniosku, że zmienna długości czasu oglądania reklam w istotny sposób jest wyjaśniana zarówno przez ich zawartość informacyjną, jak również przez pozytywny ładunek emocjonalny towarzyszący formie prezentacji ich treści. Pomijając fakt, że osiągnięte wyniki badań noszą wszelkie znamiona oczywistości, można by oczekiwać, że na podstawie tak bogatego materiału badawczego autorzy sformułują jakieś bardziej szczegółowe wnioski praktyczne. Nic z tego. Do końca pozostają wierni idei gromadzenia analitycznych danych, z których niewiele wynika zarówno dla teorii, jak i praktyki dotyczącej zachowań konsumenckich.

W kontekście wyników cytowanych badań na szczególną uwagę zasługuje opinia Jacoba Jacoby'ego (1978), prezesa *Association for Consumer Research* w 1975 roku, który stwierdził, że w literaturze poświęconej zachowaniom konsumenckim jest zbyt wiele publikacji o bardzo małej wartości zarówno praktycznej, jak i teoretycznej. Dwa lata wcześniej Jacoby (1976) wyraził to nieco dosadniej, pisząc, że „zbyt duża część współcześnie publikowanych wyników badań [konsumenckich] nie jest warta papieru, na którym została wydrukowana, ani czasu poświęcanego na ich czytanie”.

Reprezentacja populacji konsumentów w akademickich badaniach marketingowych

Jednym z najważniejszych postulatów metodologicznych funkcjonujących w ramach paradygmatu badań grupowych jest trafne określenie zakresu generalizacji osiągniętych wyników. Zakres ów jest funkcją reprezentatywnych dla interesującej populacji, prawidłowo dobranych do badań grup osób (lub obiektów) testowanych. Teza ta nie jest obca badaczom zachowań konsumenckich i *explicite* jest wyrażona niemal w każdym podręczniku akademickim podejmującym problematykę marketingu (Wentz, 1979), zarządzania rynkiem (DeLozier, Woodside, 1978), zachowań konsumenckich (Wilkie, 1994) czy profesjonalnej sprzedaży (Kurtz, Dodge, Klompmaker, 1988). W szczególności wyraża się ona w stwierdzeniu, że skuteczna strategia marketingowa, czyli taka, która pozwala na trafne



przewidywanie zachowań konsumenckich, musi uwzględniać m.in. precyzyjną segmentację rynku uwzględniającą strukturę potrzeb potencjalnych odbiorców i celów konsumenckich, specyfikę oferowanych do sprzedaży produktów i usług, a także formę ich dystrybucji i promocji.

Zastanawiające jest jednak to, że w praktyce badawczej niemal powszechnie ignoruje się te postulaty. W badaniach zachowań konsumenckich osobami testowanymi są najczęściej studenci wydziałów marketingu oraz szkół biznesu i jakkolwiek ta grupa konsumentów niewątpliwie charakteryzuje się określoną strukturą potrzeb i zainteresowań motywujących np. do oglądania takich czy innych reklam telewizyjnych, to wydaje się zupełnie oczywiste, że ich reakcje na reklamy proszków do prania będą zupełnie inne niż reakcje gospodyń domowych lub malarzy pokojowych.

Podjmując to zagadnienie w odniesieniu do badań z zakresu psychologii społecznej, David Sears (1986) wskazał na kilkanaście specyficznych cech populacji studentów. Między innymi zwrócił uwagę na wąski przedział wieku reprezentowany przez tę grupę osób oraz na wyższy poziom wykształcenia i nasilenie potrzeb poznawczych. Studenci mają znacznie słabiej skrytalizowane postawy wobec zagadnień społecznych i politycznych niż osoby starsze. Są bardziej egocentryczni i konformistyczni oraz domagają się większej aprobaty społecznej niż osoby w późniejszym wieku. Ponadto charakteryzują się wyraźnie mniejszym zainteresowaniem telewizją, która stanowi podstawowe medium kształtujące opinie, przekonania i postawy ludzi zarówno od nich młodszych, jak i starszych.

Już tylko tych kilka charakterystyk wyraźnie ujawnia, że studenci zdecydowanie nie są typowymi konsumentami, a na pewno nie można ich traktować jako reprezentantów populacji konsumentów w ogóle. Wyprowadzanie zbyt daleko idących wniosków na podstawie danych uzyskiwanych w eksperymentach podejmujących problematykę zachowań konsumenckich na tej tylko grupie osób testowanych jest zatem wyraźnym nieporozumieniem. Na przykład w badaniach skuteczności reklamy już samo rozróżnienie płci osób testowanych jest w pełni uzasadnione. Niestety, tego typu rozróżnienia są dokonywane w nielicznych tylko publikacjach poświęconych analizie zachowań konsumenckich, a co więcej zdarzają się nawet takie prace, w których autorzy poprzestają na stwierdzeniu, że „w badaniach wzięło udział 109 osób z tej samej populacji”, nie podając żadnej dodatkowej informacji, kim były te osoby (Hawkins, Hoch, 1992).

Podsumowując, problem związany z reprezentatywnością prób osób badanych polega na tym, że badacze zachowań konsumenckich traktują jako reprezentatywne dla całej populacji konsumentów wyniki pochodzące od grupy studentów, która obejmuje tylko dość specyficzny segment rynku. Co gorsza publikują oni te zupełnie niewiarygodne informacje w postaci ogólnych prawidłowości dotyczących zachowań na rynku całej populacji konsumentów. I jak tu się dziwić, że nawet recenzenci zespołu wydawców *Journal of Consumer Research* mają tak rozbieżne oceny stanu wiedzy tzw. ekspertów w dziedzinie zachowań konsumentów.



Badania laboratoryjne i sondaże opinii publicznej

Badacze zachowań konsumenckich dość często łudzą się, sądząc, że wyniki badań prowadzonych w warunkach laboratoryjnych oraz wyniki ankietowych badań opinii publicznej trafnie odzwierciedlają zjawiska zachodzące w tzw. świecie realnym. Byłoby oczywiście zbyt daleko idącym uproszczeniem stwierdzenie, że zjawiska wywoływane w eksperymentach laboratoryjnych czy pewne trendy ujawniane na podstawie ankietowych badań opinii publicznej w ogóle nie występują w naturalnych warunkach. Niemniej jednak w odniesieniu do wielu pytań zawartych w ankietach nie można oczekiwać od respondentów odpowiedzi wyrażających coś więcej niż tylko ich wiedzę deklaracyjną (w przeciwieństwie do proceduralnej). Fakt, że w odniesieniu do pewnych spraw respondenci twierdzą, iż zachowują się w określony sposób, wcale nie musi oznaczać, że rzeczywiście tak się zachowują. Podobnie nie można oczekiwać, żeby symulowana w laboratorium za pomocą instrukcji sytuacja, np. kupowania drogiego samochodu przez osobę badaną (studenta), mogła uwzględniać wszystkie poznawcze, społeczno-ekonomiczne oraz emocjonalno-motywacyjne czynniki, które w naturalnych warunkach towarzyszyłyby temu przedsięwzięciu. Przeanalizujemy dwa przykłady.

Pierwszy przykład pochodzi z badań nad odbiorem programów telewizyjnych przez dzieci (Anderson, 1985; Anderson, Field, 1991). Jeszcze do niedawna najczęstszą strategią analizy preferencji programowych było prowadzenie ankietowych sondaży opinii publicznej. O ile jednak w odniesieniu do widzów dorosłych można było oczekiwać bardziej lub mniej dokładnych odpowiedzi na pytania dotyczące np. dziennego czasu pracy telewizora, czasu spędzanego w tzw. obszarze oglądania czy szacunkowego czasu koncentracji wzrokowej na ekranie telewizora, o tyle np. dzieciom w wieku przedszkolnym raczej nie ma sensu zadawać takich pytań. O udzielanie na nie odpowiedzi prosi się zatem odpowiednio przeszkolonych rodziców.

Celem badań prowadzonych przez Daniela Andersona i jego współpracowników była m.in. ocena wartości danych ankietowych w porównaniu z danymi z bezpośredniej obserwacji zachowań dzieci oglądających telewizję w warunkach domowych. Niezwykle trudne technicznie, ale i zarazem unikatowe badania polegały na zainstalowaniu w 106 domach rodzinnych na terenie Springfield (Massachusetts) kamer wideo, za pomocą których rejestrowano zachowania odbiorców znajdujących się w tzw. obszarze oglądania, czyli w tej części pokoju, w której z każdego jego punktu widać ekran telewizora. W wyniku zestawienia danych dostarczonych przez rodziców z danymi pochodzącymi z analizy zapisów wideo ustalono stosunkowo wysoką korelację w odniesieniu do czasu spędzanego przez dzieci w obszarze oglądania. Stwierdzono jednak również wysokie statystycznie istotne różnice między protokołowanym przez rodziców czasem koncentracji wzrokowej (uwagi) ich dzieci na ekranie telewizora a danymi z zapisu wideo, oraz znacznie wyższą wariancję wyników badań ankietowych pochodzących od rodziców, którzy nie byli specjalnie przygotowywani do obserwacji zachowań swoich dzieci przed telewizorem i nie zgodzili się na zainstalowanie kamer wideo w swoich domach w stosunku do tych, które pochodziły od rodziców z grupy eksperymentalnej. W kontekście wyników tego badania uzasadnione wydaje się pytanie, co w takim razie mierzyła ankieta?



Drugi przykład pochodzi z badań nad zmianą przekonań konsumenckich, prowadzonych przez Anthony'ego Pratkanisa i Anthony'ego Greenwalda (1985). Działalność wielu agencji reklamowych sprowadza się do kreowania takich komunikatów, które względnie trwale zmieniałyby nastawienia (opinie) odbiorców w kierunku zgodnym z ich treścią. W zdecydowanej większości przypadków efekty tych oddziaływań są jednak krótkotrwałe. Po chwilowym, następującym bezpośrednio po odbiorze komunikatu uzgodnieniu treści własnych przekonań z tezą wyrażoną w komunikacie, odbiorcy najczęściej powracają do swoich uprzednich przekonań. Niemniej jednak w latach 30. Ruth Peterson i Louis Leon Thurstone (1933) odkryli zjawisko polegające na tym, że niektóre komunikaty kształtują opinie odbiorców w kierunku zgodnym z ich treścią nie bezpośrednio po ich odbiorze, ale dopiero po dłuższym czasie (np. kilku lub kilkunastu miesiącach). Carl Hovland, Arthur Lumsdaine i Fred Sheffield (1949) nazwali je efektem śpiocha, lub jak wolą inni – uśpienia (*sleeping effect*).

Wysiłki wielu badaczy podejmujących próby wyjaśnienia tego fenomenu najczęściej kończyły się niepowodzeniem: zjawisko ujawniało się w badaniach w sposób niesystematyczny i trudno było znaleźć jakąś wspólną dla wyników różnych badań płaszczyznę teoretyczną. Na jego temat sformułowano wiele, nierzadko sprzecznych ze sobą hipotez. Pratkanis i Greenwald (1985) dokładnie przeanalizowali ten materiał i opracowali precyzyjny program telewizyjnej ekspozycji komunikatów perswazyjnych, których obejrzenie gwarantowałoby uzyskanie efektu uśpienia. Badania rzeczywiście skończyły się sukcesem. I chociaż okazało się, że zjawisko można wywołać w laboratorium, to w warunkach naturalnych nadal nie można nim w racjonalny sposób sterować.

Obydwa opisane projekty badawcze ujawniają niewątpliwie złożoność zagadnienia relacji między wynikami badań laboratoryjnych lub ankietowych a efektami występującymi w warunkach naturalnych. Warto jednak, jak sądzę, zastanowić się, dlaczego pomimo tak wielu wątpliwości wysuwanych przez metodologicznie zorientowanych teoretyków zachowań konsumenckich pod adresem wyników badań osiąganych w laboratorium w praktyce badawczej są one nagminnie ignorowane. Doszło już nawet do tego, że w wielu badaniach, np. nad skutecznością reklamy, poprzestaje się tylko na rejestrowaniu deklaracji zakupu reklamowanego produktu, zupełnie ignorując to, czy w naturalnej sytuacji ten produkt w ogóle zwróciły czyjąkolwiek uwagę. W rezultacie takich działań dochodzi do lawinowej nadprodukcji wyników badań i publikacji, które nierzadko spełniają nawet formalne wymogi stawiane pracom naukowym, ale praktycznie i teoretycznie są zupełnie bezwartościowe.

Niestety taka sytuacja często występuje w naukach interdyscyplinarnych. Zachowania konsumentów w naturalny sposób są uwikłane w tak szeroki kontekst społeczno-ekonomiczny, kulturowy, a także psychologiczny związany ze strukturą indywidualnych potrzeb, zainteresowań, ambicji i celów, że standardowe techniki operacjonalizacji zmiennych w ramach jednego eksperymentu uniemożliwiają osiągnięcie zarazem laboratoryjnej precyzji i wysokiej trafności zewnętrznej. Konieczne są zatem próby opracowywania nowych strategii badawczych, które dokładniej odpowiadałyby potrzebom danej dziedziny wiedzy. Te jednak najczęściej są drogie i długotrwałe, ale czy jest lepsze rozwiązanie?



Usprawiedliwiająca moc świadomości ograniczeń osiągniętych wyników

Teraz kilka zdań na temat tzw. retoryki naukowej. Niemal każdy artykuł opublikowany w poczytnych czasopismach marketingowych oraz dotyczących zachowań konsumenckich wieńczy paragraf, w którym autor (lub autorzy) wyrażają swoje wątpliwości i zastrzeżenia (zarówno merytoryczne, jak i metodologiczne) odnośnie osiągniętych przez siebie wyników badań. Zapoznanie się z treścią tych paragrafów ujawnia, że sami badacze zachowań konsumenckich zdają się do pewnego stopnia dostrzegać niektóre trudności metodologiczne i jednym głosem niemal w każdej publikacji sugerują, że słabością wniosków płynących z ich badań jest to, że albo zostały one przeprowadzone w warunkach nienaturalnych (w laboratorium, za pomocą komputera), albo że posługiwali się fikcyjnymi nazwami firm, albo że kategorie produktów, ich opakowania czy reklamy zostały na użytek badań nadmiernie uproszczone, albo że pomiary reakcji osób badanych następowały w znacznie krótszych interwałach czasowych niż ma to miejsce w codziennym życiu konsumenta itd. Interesujące, że w nielicznych tylko pracach wprost wyrażono wątpliwości, co do reprezentatywności próby osób testowanych. Co więcej, w paragrafach tych niezmiennie pojawia się sugestia, że następnym krokiem badawczym, jaki powinien być zrobiony, jest weryfikacja osiągniętych przez nich wyników w warunkach naturalnych. Można by więc oczekiwać, że autorzy tych postulatów podejmą takie próby w nieodległej przyszłości. Niestety, nie udało mi się znaleźć ani jednego przykładu badań, które stanowiłyby kontynuację wyrażonej uprzednio potrzeby rozwijania podjętej już problematyki w warunkach naturalnych. Badania zostały przeprowadzone, grant rozliczony, a publikacja w druku. „I czym by się tu teraz zająć?”

Istotność różnic i wielkość efektu eksperymentalnego

Zastosowanie statystyki, a w szczególności wnioskowania statystycznego, w naukach społecznych niemal od samego początku budziło wiele niepokoju, zwłaszcza wśród tzw. humanistycznie zorientowanych badaczy. Na ogół stroniący od nadmiernej formalizacji w nauce, traktują statystykę jako panaceum na wszelkie interesujące ich problemy badawcze, bezkrytycznie przyjmując za dobrą monetę niemal każdy wynik analizy, który przekracza magiczny poziom prawdopodobieństwa błędu równy 0,05 lub mniejszy. Źródła tej postawy i podnoszonej od czasu do czasu dyskusji na temat wartości wnioskowania statystycznego tkwią jednak, jak sądzę, przede wszystkim w niewłaściwym rozeznaniu możliwości i ograniczeń metod analizy i wnioskowania statystycznego. Krótko mówiąc, jeżeli statystykę potraktuje się jako jedno z wielu (a nie jedyne przecież) narzędzie analizy danych gromadzonych w badaniach, to jego wartość jest wyznaczona przez umiejętności posługiwania się nim z uwzględnieniem zarówno wszystkich jego mocnych, jak i słabych stron. Niestety, świadomość metodologiczna wielu badaczy (nie tylko zresztą podejmujących problematykę zachowań konsumenckich) pozostawia wiele do życzenia.

Jednym z najczęstszych błędów popełnianych przez badaczy zachowań konsumenckich na poziomie formułowania wniosków wynikających z ich eksperymentów jest traktowanie każdego rozstrzygnięcia prawdziwości hipotezy z prawdopodobieństwem błędu choćby minimalnie tylko mniejszym niż 0,05 jako bez wątpienia potwierdzającego wyrażoną



w hipotezie relację między zmiennymi (zob. np. Chow 1988; Cohen 1990; Sawyer, Peter 1983). Co gorsza, nierzadko wartość takiego wyniku jest traktowana jako na tyle wysoka, że uprawnia do formułowania ogólnych wniosków, znacznie wykraczających poza twierdzenie zawarte w samej hipotezie. Stanowi także „empirycznie uzasadnioną” podstawę konstrukcji bardziej lub mniej złożonej teorii, której „drobną” słabością jest tylko to, że formułowane na jej podstawie predykcje w praktyce mają niewielką wartość. William Wells (1993) słusznie zauważa, że w badaniach konsumenckich minimalna (kontrolowana lub niekontrolowana) zmiana stymulacji prowadzi do dużych zmian na poziomie reakcji osób testowanych. Przyjęcie zbyt liberalnego poziomu odrzucenia hipotezy zerowej oraz podlegająca wielu zmiennym niezależnym sytuacja eksperymentalna, w którą zaangażowani są tzw. konsumenci, powoduje osiąganie w tej dziedzinie zbyt wielu niewiarygodnych wyników.

Czy jest na to jakaś rada? Jedną z metod pozwalającą na stwierdzenie, co tak naprawdę wykazał badacz, odrzucając hipotezę zerową, jest obliczenie wskaźnika wielkości efektu eksperymentalnego. Wartość tę można zdefiniować jako różnicę między średnimi, podzieloną przez odchylenie standardowe danej cechy w populacji. Po przemnożeniu jej przez 100 interpretuje się ją jako procent wariacji wyjaśnionej przez daną zmienną. W praktyce badawczej wielkość efektu mniejsza niż 0,5, czyli ok. 50% wariacji wyjaśnionej należy uznać za wynik, co do którego powinniśmy zachować daleko idącą ostrożność, nawet jeśli poziom odrzucenia hipotezy zerowej przekroczył magiczną wartość 0,05. Amerykańskie Towarzystwo Psychologiczne w swoim podręczniku, poświęconym standardom w zakresie formy publikowania raportów badawczych w psychologii, zaleca podawanie wskaźników wielkości efektu (omega kwadrat, eta kwadrat lub d Cohena) oprócz wyniku testu i poziomu istotności. Szerzej na temat wielkości efektu piszemy w naszym podręczniku (Francuz, Mackiewicz, 2007).

Statystycznie istotne korelacje

Mit statystycznej istotności ma jeszcze jeden interesujący aspekt. Nieco wnikliwsza analiza relacji między wielkością prób osób badanych a osiąganymi na podstawie statystycznej analizy danych pochodzących z tych prób wskaźnikami istotności różnic lub korelacji ujawnia, że w odniesieniu do wielu metod statystycznych istnieje odwrotnie proporcjonalna zależność między liczbą stopni swobody a odpowiadającymi im poziomami istotności (Chow 1988; Sawyer 1982). Wystarczy tylko rzucić okiem na tablicę statystyczną zawierającą poziomy istotności współczynnika korelacji, żeby uświadomić sobie, że np. już dla 100 stopni swobody współczynnik korelacji równy 0,254 jest statystycznie istotny na poziomie 0,01, chociaż z punktu widzenia oceny jego bezwzględnej wartości wskazuje on na zdecydowanie niską korelację między analizowanymi zmiennymi.

Analiza liczebności prób osób testowanych w badaniach konsumenckich ujawnia, że z małymi tylko wyjątkami badania te mają charakter niemal masowy. Na podstawie 20 losowo wybranych publikacji zamieszczonych w *Journal of Consumer Research* ustaliłem, że przeciętnie grupy osób testowanych liczą ponad 250, a niekwestionowani rekordziści prowadzą badania na grupach liczących do 25 000 osób testowanych (zob. np. Zielskie,



Henry 1980). Prowadząc badania na tak licznych grupach, autorzy tych eksperymentów starają się, jak sądzę, zwiększyć wiarygodność osiąganych wyników. Zapominają jednak o tym, że przede wszystkim reprezentatywność, a nie liczebność grup osób testowanych stanowi podstawę generalizacji. Zwłaszcza na gruncie tzw. statystyk małych prób.

Przyczynowa interpretacja korelacji

Sprawa przyczynowej interpretacji wyników analizy korelacyjnej w badaniach zachowań konsumenckich stanowi wyjątkowo irytujący wątek dla każdego badacza, który przeszedł choćby tylko podstawowy kurs statystyki opisowej. W tym miejscu pozwolę sobie pominąć szczegóły efektów analizy metod interpretacji korelacji w konkretnych pracach z omawianego zakresu i ze względów już czysto dydaktycznych przedstawię sugestywny przykład, który być może pozwoli innym badaczom na ostrożniejsze interpretowanie wskaźników korelacji.

Zastanówmy się chwilę nad korelacją zachodzącą między paleniem papierosów a zawałem serca. Jest faktem – i co do tego nie ma wątpliwości – że u osób, które palą papierosy, znacznie częściej diagnozuje się zawał serca niż u niepalaczy. Oznacza to, że korelacja między liczbą wypalanych papierosów dziennie albo długością palenia (np. w latach) a wystąpieniem zawału serca jest bardzo wysoka. Ale czy z tego koniecznie wynika, że palenie jest źródłem choroby? Otóż, absolutnie, nie.

Jest bowiem możliwe, że częściej po papierosy sięgają ludzie, którzy mają jakieś wrodzone predyspozycje do zawału. Oznaczałoby to, że bezpośrednią przyczyną zawału są wrodzone predyspozycje, zaś palenie stanowiłoby tylko czynnik uboczny, niekoniecznie będący jego bezpośrednim powodem.

Może być również i tak, że istnieje jakiś inny czynnik, np. stres lub niepokój, który z jednej strony skłania do palenia papierosów, a z drugiej niekorzystnie wpływa na pracę naczyń wieńcowych i serca, powodując jego zawał. W tym wypadku palenie i zawał serca byłyby po prostu symptomami określonego stanu psychicznego (trzeciego czynnika), same pozostając względnie niezależne od siebie.

Warto również pamiętać o tym, że sama procedura doboru osób do badań może wpłynąć na ostateczny wynik badania zależności między paleniem papierosów i zawałem serca. Wysoka korelacja między tymi zmiennymi może bowiem wynikać z takiego doboru osób badanych do próbek, że znajdują się w nich np. tylko ci, którzy mają problemy z sercem i dużo palą, oraz ci, którzy mają „serce jak dzwon” i w ogóle nie palą papierosów.

O elastyczności testów t i F

Na koniec warto przypomnieć jeszcze jedną sprawę, wokół której narosło trochę nieporozumień. W podręcznikach do statystyki w różny sposób podchodzi się do przestrzegania założeń parametrycznych testów statystycznych. Rzecz idzie zwłaszcza o dwa najpopularniejsze testy, czyli t Studenta i F , stosowany w analizie wariancji. Na szczęście współczesna statystyka podsuwa nie tylko narzędzia służące do weryfikowania hipotez badawczych, ale również techniki testowania tych narzędzi. Niezwykle użyteczne pod tym względem są



tw. metody *Monte Carlo*, które pozwalają na sprawdzenie, w jakim stopniu dany test jest elastyczny (*robust*) na odstępstwa danych od jego założeń. Odpowiadają one np. na pytanie, jak zmienia się wynik testu parametrycznego w zależności od większych lub mniejszych odchyień danych od normalności rozkładu analizowanej cechy w populacji.

Symulacje przeprowadzone za pomocą metod *Monte Carlo*, pozwalają na wyciągnięcie wniosku, że testy t i F zasadniczo można stosować nawet wtedy, gdy rozkład analizowanej cechy nie jest normalny. Nawet w przypadku znacznego odchylenia od normalności wyniki testów są porównywalne. Inaczej jednak jest w odniesieniu do postulatu, zgodnie z którym wariancje w porównywanych grupach nie powinny istotnie różnić się od siebie. Okazuje się bowiem, że obydwa testy są znacznie mniej elastyczne na nierówne wariancje w grupach niż na odchylenia danych od rozkładu normalnego.

Podsumowując, warto więc pamiętać, że:

- ◆ Testy t i F są elastyczne w odniesieniu do założenia o normalności rozkładu badanej cechy w populacji. W praktyce, jeśli próbki są większe niż 12 elementów, to na wynik testu nie ma wpływu to, czy pochodzą z populacji o rozkładzie normalnym czy innym niż normalny (por. Good, Hardin, 2003).
- ◆ Testy t i F są mało elastyczne wobec zróżnicowania wariancji w porównywanych próbach. Najbezpieczniej jest zatem przyjąć zasadę, że nie należy ich stosować wtedy, gdy po sprawdzeniu jednorodności wariancji (np. testem Levene'a) nie można utrzymać hipotezy o ich jednorodności.
- ◆ Odporność testów t i F na odstępstwa od normalności i jednorodności wariancji w dużym stopniu zależy od wielkości próbek. Im większa jest różnica między liczebnościami porównywanych zbiorów, tym większe jest ryzyko popełnienia błędu I lub II rodzaju.

Podsumowanie

Celem niniejszego szkicu była krytyczna analiza strategii badawczych stosowanych w eksperymentach dotyczących zachowań konsumenckich. Ujawnienie zasygnalizowanych wyżej trudności, uchybień czy wręcz poważnych błędów metodologicznych pojawiających się na etapie formułowania pytań badawczych, analizy gromadzonych danych oraz wyprowadzania z nich wniosków wydaje się kluczowe z punktu widzenia konieczności podniesienia rzetelności wiedzy dotyczącej zachowań konsumenckich. Szczególna odpowiedzialność za wyniki praktycznej realizacji projektów badawczych w ramach omawianej dziedziny wiedzy ciąży na metodologach tych nauk szczegółowych, które stanowią podstawę konstrukcji interdyscyplinarnego forum badań marketingowych. Szczegółowy przegląd literatury przedmiotu ujawnia zasadniczą trudność tych badań związaną z brakiem względnie spójnej koncepcji metodologicznej. W zależności od swoich uprzednich doświadczeń i wiedzy (np. psychologicznej, ekonomicznej lub socjologicznej) badacze zachowań konsumenckich albo ortodoksyjnie stosują metody charakterystyczne dla tych dziedzin wiedzy lub też różnych ich działów, albo też ignorują jakiegokolwiek postulaty



metodologiczne i popełniają niemal „szkolne” błędy, które dyskwalifikują merytoryczne wyniki ich działań.

Kończąc, chciałbym raz jeszcze podkreślić, że dziedzina zachowań konsumenckich wydaje się szczególnie interesującym polem doświadczalnym, w którym ogniskują się niemal wszystkie istotne problemy poruszane w różnych naukach społecznych. Z założenia miała ona stanowić forum wymiany myśli między przedstawicielami różnych ośrodków naukowych, reprezentantami różnych poziomów władzy zarówno lokalnej, jak i ogólnonarodowej, jak również przedstawicielami zarówno prywatnego sektora działalności komercyjnej, jak i organizacji lub fundacji, które nie są nastawione wyłącznie na działania dochodowe. Bez względu jednak na jej dotychczasowe osiągnięcia stanowi ona wyraźną próbę wyjścia naprzeciw wyzwaniom współczesnego świata, w którym odpowiednio szeroko rozumiane kupowanie i sprzedawanie różnego rodzaju dóbr materialnych i duchowych stanowi istotną część aktywnie spędzanego czasu przez ludzi każdego dnia. Warto więc, jak sądzę, podjąć wysiłek zmierzający do podniesienia wartości osiąganych na jej gruncie wyników badań naukowych, do czego – mam nadzieję – przyczynią się również uwagi i refleksje przedstawione w niniejszym szkicu.

Bibliografia

1. Anderson D. R. (1985). Online Cognitive Processing of Television. W: L. F. Alwitt, A. A. Mitchell (red.), *Psychological Processes and Advertising Effects* (s. 177-199). Hillsdale, NJ: Erlbaum.
2. Anderson D. R., Field D. E. (1991). Online and Offline Assessment of the Television Audience. W: J. Bryant, D. Zillmann (red.), *Responding to the Screen: Reception and Reaction Processes* (s. 199-216). Hillsdale, NJ: Erlbaum.
3. Armstrong J. S. (1991). Prediction of Consumer Behavior by Expert and Novices. *Journal of Consumer Research*, 18, 251-256.
4. Calder B. J., Phillips L. W., Tybout A. M. (1982). The Concept of External Validity. *Journal of Consumer Research*, 9, 240-244.
5. Chow S. L. (1988). Significance Test or Effect Size? *Psychological Bulletin*, 103, 105-110.
6. Cohen J. (1990). Things I Have Learned (So Far). *American Psychologist*, 45, 1304-1312.
7. DeLozier M. W., Woodside A. (1978). *Marketing Management: Strategies and Cases*. Columbus: Charles E. Merrill Publishing Company.
8. Francuz P., Mackiewicz R. (2007). *Liczby nie wiedzą skąd pochodzą. Przewodnik po metodologii i statystyce nie tylko dla psychologów*. Lublin: Wydawnictwo KUL.
9. Good P. I., Hardin, J. W. (2003). *Common errors in statistics (and how to avoid them)*. Hoboken: John Wiley and Sons Inc.
10. Hawkins S. A., Hoch S. J. (1992). Low-Involvement Learning: Memory without Evaluation, *Journal of Consumer Research*, 19, 212-225.



11. Hoch S. J. (1988). Who Do We Know: Predicting the Interests and Opinions of the American Consumer. *Journal of Consumer Research*, 15, 315-324.
12. Hovland C. I., Lumsdaine A. A., Sheffield F. D. (1949). *Experiments on Mass-Communications*. Princeton, NJ: Princeton University Press.
13. Jacoby J. (1976). Consumer Research: Telling It Like It Is. W: B. B. Anderson (red.), *Advances in Consumer Research* (t. 3, s. 1-11). Ann Arbor, MI: Association for Consumer Research.
14. Jacoby J. (1978). Consumer Research: A State of the Art Review. *Journal of Marketing*, 42, 87-96.
15. Kollat D. T., Blackwell R. D., Engel J. F. (1972). The Current Status of Consumer Behavior Research: Developments during the 1968-1972 Period. W: M. Venkatesan (red.), *Proceedings of the Third Annual Conference of the Association for Consumer Research* (s. 576-593). College Park, MD: Association for Consumer Research.
16. Kurtz D. L., Dodge H. R., Klompmaker J. E. (1988). *Professional Selling*. Plano: Business Publications, Inc.
17. Lichtenstein, D. R., Burton, S., Karson, E. J. (1991). The Effect of Semantic Cues on Consumer Perceptions of Reference Price Ads. *Journal of Consumer Research*, 18, 380-391.
18. Lynch J. G. (1982). On the External Validity of Experiments in Consumer Research. *Journal of Consumer Research*, 9, 225-239.
19. Olney T. J., Holbrook M. B., Batra R. (1991). Consumer Responses to Advertising: The Effects of Ad Content, Emotions, and Attitude toward the Ad on Viewing Time. *Journal of Consumer Research*, 17, 440-453.
20. Peterson R. C., Thurstone L. L. (1933). *Motion Pictures and the Social Attitudes of Children*. New York: Macmillan.
21. Pratkanis A. R., Greenwald A. G. (1985). A Reliable Sleeper Effect in Persuasion: Implications for Opinion Change Theory and Research. W: L. F. Alwitt, A. A. Mitchell (red.), *Psychological Processes and Advertising Effects* (s. 157-173). Hillsdale, NJ: Erlbaum.
22. Sawyer A. G., Peter J. P. (1983). The Significance of Statistical Significance Tests in Marketing Research. *Journal of Marketing Research*, 20, 122-133.
23. Sears D. O. (1986). College Sophomores in the Laboratory: Influences of a Narrow Data Base on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology*, 51, 515-530.
24. Wells W. D. (1993). Discovery-oriented Consumer Research. *Journal of Consumer Research*, 19, 489-504.
25. Wentz W. B. (1979). *Marketing*. St. Paul: West Publishing Company.
26. Wilkie W. L. (1994). *Consumer Behavior*. New York: John Wiley & Sons, Inc.
27. Zielskie H. A., Henry W. A. (1980). Remembering and Forgetting Television Ads. *Journal of Advertising Research*, 20, 239-243.