

## Metanauka poznawcza: nowe podejście do badania teorii

Marcin Miłkowski<sup>1</sup>

*Polska Akademia Nauk, Instytut Filozofii i Socjologii*

<https://orcid.org/0000-0001-7646-5742>

### Streszczenie

W artykule przedstawiono argumentację na rzecz większego nacisku na teoretyzowanie w badaniach naukowych w świetle trwającego kryzysu wiarygodności w psychologii. Choć istotną rolę odgrywają wiarygodne dowody eksperymentalne, prerejestracja, rygor metodologiczny i nowe podejścia obliczeniowe do modelowania, postęp naukowy opiera się również na właściwie funkcjonujących teoriach. Roli teoretyzowania w psychologii nie rozumiemy jednak wystarczająco, a to może prowadzić do kolejnych kryzysów. Teorie nie powinny być postrzegane jako czcze spekulacje czy proste uogólnienia indukcyjne.

W tym kontekście autor wprowadza podejście zwane „metanauką poznawczą”, mające badać czynności i wytwory oceny praktyki naukowej. Badanie to ma przebiegać zarówno jakościowo, jak i ilościowo, poprzez analizę dyskursu naukowego z wykorzystaniem technik przetwarzania języka.

Teorie analizowane są jako artefakty poznawcze wspierające procesy poznania w nauce. Ta perspektywa ujawnia odmienne role różnych teorii, a badanie tych ról, wraz z ich epistemicznymi zaletami i wadami, pozwala uchwycić normy teoretyzowania. Autor wzywa do zmiany kultury badawczej, by odpowiednio uwzględniano różnorodne zastosowania odmiennych typów teorii i systematycznie wspierano postęp naukowy.

**Słowa kluczowe:** kryzys teoretyczny, metanauka poznawcza, artefakt poznawczy, wartość teoretyczna, kryteria epistemiczne

Wiele sugeruje, że mamy do czynienia z kryzysem zaufania w psychologii i dziedzinach pokrewnych. Wspomnieć należy tu problemy z replikowalnością, generalizowalnością, kumulatywnym charakterem badań oraz spójnością

---

<sup>1</sup> Adres do korespondencji: [marcin.milkowski@ifispan.edu.pl](mailto:marcin.milkowski@ifispan.edu.pl).

rozumienia w psychologii i dziedzinach pokrewnych (Baker, 2016; Boekel i in., 2015; Hughes, 2018; Ioannidis, 2005; Manninen i in., 2018; Nosek i in., 2022; Open Science Collaboration, 2015; Simmons i in., 2011; Yarkoni, 2022;). Chociaż wśród przyczyn kryzysu wskazuje się wątpliwe praktyki badawcze, oszustwa, niewłaściwe wybory metodologiczne i słabą interpretowalność badań wynikającą z zakłóceń w komunikacji naukowej (Hensel, 2020; Miłkowski i in., 2018), wielu autorów sugeruje, że źródłem kryzysu są w istocie niejasny status i funkcja teorii w badaniach, przez co jest to kryzys teoretyczny (Carsel i in., 2018; Hensel i in., 2022; Hughes, 2018; Irvine, 2021; Klein, 2014; Levenstein i in., 2023; Muthukrishna i Henrich, 2019; Oberauer i Lewandowsky, 2019; Smaldino, 2017; Szollosi i Donkin, 2019; Young, 2016; odmienne zdanie przedstawiają Trafimow i Earp, 2016). Aby jednak ocenić prawdziwość tej tezy, musimy lepiej zrozumieć, czym są teorie i do czego powinny służyć. Mimo że klasyczna filozofia nauki koncentrowała się na teoriach, nadal nie rozumiemy ich roli w praktyce naukowej.

Metanauka poznawcza bada czynności i wytwory oceny praktyk naukowych, także tych zachodzących w naukach psychologicznych. Podejście to pozwala zrozumieć, w jaki sposób praktycy sami rozumieją działalność naukową. Dzięki temu można zanalizować to, co Flis (2019) określa mianem „epistemologii tubylczej”. Kluczową cechą metanauki poznawczej jest wykorzystanie zarówno jakościowych, jak i ilościowych metod badawczych, wywodzących się z różnych dyscyplin, w tym z technik przetwarzania języka naturalnego. Dzięki dostępności dużych zbiorów danych (*big data*) możliwe jest obecnie badanie różnych cech dyskursu na temat reprezentacji teoretycznych w psychologii. Dotyczy to nie tylko jawnych wzmianek na ich temat, lecz także niejawnych relacji semantycznych uwikłanych we wzorcach użycia języka. Dyskurs naukowy dotyczący wyboru teorii odwołuje się do założeń normatywnych związanych z celami badań naukowych, co sugeruje, że tego rodzaju badania mogą prowadzić do wniosków o charakterze opisowym i normatywnym.

W kolejnej części tekstu dokładniej przedstawiam podejście metanauki poznawczej do teorii. Następnie wysuwam argumenty na rzecz istnienia wielu rodzajów teorii w naukach psychologicznych, co uzasadnia przyjęte tutaj podejście. Teorie te odgrywają różne role i powinny być oceniane w świetle specyficznych dla siebie kryteriów epistemicznych wyboru teorii. Na zakończenie przedstawiam argumentację na rzecz tezy, że prawdopodobną przyczyną kryzysu jest mylenie różnego rodzaju teorii oraz lekceważenie przyjmowanych założeń teoretycznych. Walka z kryzysem polegać więc może na wspieraniu kultury badawczej prowadzącej do poprawnego rozwoju teorii.

## **Podejście metanauki poznawczej**

Celem metanauki poznawczej jest systematyczne rozumienie i doskonalenie refleksji nad praktyką naukową. Jej metody wywodzą się z filozofii nauki w praktyce (Callebaut, 1993), kognitywistyki nauki (Langley i in., 1987; Nersessian, 2008; Thagard i Findlay, 2012), ilościowej metanauki (Schooler, 2014)

oraz humanistyki cyfrowej (Moretti, 2000), w tym cyfrowej filozofii nauki (Lean i in., 2021; Pence i Ramsey, 2018; Thagard, 1993). Powyższe metody mogą służyć do analizy teorii w różnych dziedzinach psychologii, prowadząc do opisu oraz – w ostatecznym rozrachunku – adekwatnego normatywnie ujęcia różnych rodzajów, funkcji, a także wad i zalet teorii. Teoretyzowanie jest praktyką poznawczą (Callebaut, 2013; Chang, 2017), podejmowaną przez badaczy w różnorodnych celach, także praktycznych, np. klinicznych czy politycznych. Jednakże różnorodne role odgrywane przez teorie w badaniach naukowych mogą sprawiać, że samo pojęcie teorii pozostaje dla wielu autorów mało uchwytne (Gorelick, 2011).

W moim ujęciu teoria jest rodzajem artefaktu poznawczego, czyli obiektem służącym do „przechowywania, przedstawiania lub modyfikowania informacji w celu realizacji funkcji reprezentacyjnych, i mającym wpływać na poziom wykonania czynności poznawczych” (Norman, 1991, s. 11). Treści teorii nie da się w pełni wyrazić w kategoriach danych, eksperymentalnych czy obserwacyjnych. Teorie dostarczają natomiast perspektywy na badane zjawiska. Wiele artefaktów poznawczych może tworzyć układy kierujące praktykami badawczymi. Artefakty te mogą pełnić różne funkcje i być przystosowane do określonych potrzeb w danej dziedzinie (Miłkowski, 2022).

Funkcje reprezentacyjne teorii warto analizować w ramach obejmujących je rozproszonych mechanizmów poznawczych (Afeltowicz i Wachowski, 2015; Giere i Moffatt, 2003; Hutchins, 1995; Osbeck i Nersessian, 2014; Zhang i Norman, 1994). Co więcej, rodzaje teorii można wyodrębnić poprzez analizę ich swoistych funkcji, co prowadzi do funkcjonalizmu w odniesieniu do teorii. Badanie ich funkcji obejmuje też możliwe dysfunkcje. Wynika stąd, że cele metanauki poznawczej są nie tylko opisowe, lecz także normatywne. Jeśli np. jednym z zadań teorii jest klasyfikowanie zjawisk do wyjaśnienia, wówczas teoria, która nie klasyfikuje zjawisk poprawnie, jest pod tym względem wadliwa. Ogólnie rzecz biorąc, metanauka poznawcza ma zapewnić perspektywę normatywną służącą do oceny artefaktów poznawczych, w tym także teorii, a także podnieść jakość teoretyzowania w nauce.

Przy założeniu tak szerokiego rozumienia teorii możemy teraz skupić się na tym, jaki jest związek między kryzysem zaufania a brakiem właściwych teorii w psychologii, na co często zwraca się uwagę (Fiedler, 2017; Muthukrishna i Henrich, 2019; Oberauer i Lewandowsky, 2019). Dlaczego niewłaściwe teorie miałyby tu w ogóle mieć jakiegokolwiek znaczenie?

Dosyć oczywista odpowiedź brzmi następująco: teorie ograniczają hipotezy wysuwane przez badaczy. Wartość predykcyjna hipotezy zależy w dużej mierze od prawdopodobieństwa *a priori*, że ta hipoteza jest prawdziwa; jeśli większość testowanych przez nas hipotez jest fałszywa, wówczas należy oczekiwać, że duży odsetek empirycznie potwierdzonych teorii będzie również fałszywy. Skoro zaś teoria powinna ograniczać zbiór poważnie rozważanych hipotez, narzuca się myśl, że kryzys teorii stoi u podstaw kryzysu replikacji i generalizacji (Hensel i in., 2022). Racjonalna ocena teorii może być zaburzona, jeśli rozważa się tylko ograniczoną liczbę konkurencyjnych teorii (Almaatouq i in., 2022; Dellsén, 2020).

Teorie mogą też niewystarczająco ograniczać hipotezy predykcyjne lub eksplanacyjne (Bird, 2021; Button i in., 2013; Fiedler, 2017; Muthukrishna

i Henrich, 2019). Na przykład Bird podkreśla, że w wielu dziedzinach należy spodziewać się nieudanych replikacji. Jednak taka interpretacja nieudanych replikacji nie uwzględnia w pełni różnic w poziomach replikacji między stosunkowo podobnymi polami badawczymi, a także w ramach tych samych pól badawczych, przez co jest niedostatecznie precyzyjna, żeby takie różnice dało się odpowiednio zrozumieć (Autzen, 2021)<sup>2</sup>.

Replikacja może być też utrudniona przez niewłaściwe ujęcia teoretyczne obserwowanych zjawisk (Eronen i Bringmann, 2021; Levenstein i in., 2023), a także z powodu mylenia opisów zjawisk z ich wyjaśnieniami (Scheel i in., 2020). Gigerenzer (1998) zauważył, że psychologowie chętnie posługują się surogatami właściwych teorii: pseudowyjaśniającymi hasłami, redyskrypcjami, niejasnymi dychotomiami oraz dopasowywaniem do danych. Krótko mówiąc, kryzys może na wiele sposobów wiązać się z niewłaściwym użyciem teorii.

Badania psychologiczne raz wydają się niedostatecznie teoretyczne, a innym razem wręcz przeciwnie. Z jednej strony badacze w wielu poddziedzinach psychologii zajmują się niemal wyłącznie zbieraniem danych eksperymentalnych i obserwacyjnych, poszukując efektów, nawet jeśli te ostatnie nie powinny być przedmiotem wyjaśnień (Cummins, 2000; Fried, 2020; van Rooij i Baggio, 2021). Z drugiej strony zdaje się, że wystarczająco atrakcyjna teoria niekiedy zastępuje badaczom obserwacje. Ba, wiele istniejących teorii wydaje się nieobalalnych, gdyż pasują one niemal do każdego odkrycia empirycznego (Frankenhuis i in., 2022; Miłkowski i Litwin, 2022; Roberts i Pashler, 2000; Szollosi i Donkin, 2019). W zamian wielu obrońców określonej teorii skupia się na jej potwierdzeniu, zaniedbując konieczność porównywania jej z teoriami konkurencyjnymi (Greenwald i in., 1986). Przeszarżałe teorie są też rzadko, jeśli w ogóle, eliminowane z praktyki badawczej (Ferguson i Heene, 2012).

Brak odpowiednich teorii może prowadzić do wielu innych konsekwencji, nie tylko do niskiej replikowalności i generalizowalności. O ile psychologia będzie nadal tak ateoretyczna, o tyle prawdopodobne jest, że jej wyjaśnienia okażą się *ad hoc*, a sama nie będzie kumulatywna (Erđin, 2021; Newell, 1973). Według wielu autorów kluczowym problemem jest brak jedności teoretycznej w psychologii, która dzieli się na odrębne i niespójne podejścia czy paradygmaty (Hughes, 2018). Jednym z możliwych zagrożeń związanych z brakiem jedności teoretycznej, jak się argumentuje, jest to, że rozdrobnione badania są rozproszone i niezintegrowane, a więc pozbawione systematycznych powiązań między zjawiskami (Bower, 1993; Goertzen, 2008; Staats, 1986; Young, 2016; ale zob. Dale i in., 2009; Matthews, 2020; Zittoun i in., 2009; w sprawie przeglądu zob. Gaj, 2016).

Co więcej, bez wyraźnie określonej teorii nie sposób ustalić, które wyniki naukowe zależą od podstawowych twierdzeń teorii, a które od jej dodatkowych założeń (Cooper i Shallice, 1995; Miłkowski i in., 2018). A nawet gdy psychologiczna teoria generuje przewidywania, są one zwykle niedookreślone (Fried, 2020; Meehl, 1967). Ta teoretyczna nieokreśloność może faktycznie służyć strategicznym celom badaczy, chociaż sama jest szkodliwa dla postępu naukowego

---

<sup>2</sup> Dziękuję jednemu z recenzentów za wskazanie tej pozycji literaturowej.

(Frankenhuis i in., 2022). Jednakże rodzi to pytanie, kiedy teorie są właściwe i do jakich celów służą.

Na szczęście, aby oceniać teorie w kategoriach artefaktów poznawczych, można skorzystać z utrwalonego już w filozofii nauki pojęcia wartości teoretycznej (*theoretical virtue*), które w latach 70. XX wieku stało się wpływowe w metodologii. Pojęcie to może posłużyć do wyjaśnienia związków między poszczególnymi cechami teorii a możliwymi przyczynami kryzysu. Wartości teoretyczne przywołał Thomas Kuhn (1977), który wskazał pięć wartości teoretycznych istotnych dla badań naukowych w ogóle: trafność empiryczną, spójność, zakres (ogólność), prostotę i płodność. Co ciekawe, systematyczność, chociaż niewątpliwie jest to istotna cecha nauki (Hoyningen-Huene, 2013; Rescher, 1979), podobnie jak zdolność do projektowania powtarzalnych eksperymentów, nie tylko nie występuje na liście Kuhna, lecz także brak jej w późniejszych badaniach nad cnotami teoretycznymi (Keas, 2018; Longino, 1996; McMullin, 2010; Schindler, 2018). Jednak nawet gdy badacze mózgu i umysłu uznają znaczenie cnot teoretycznych (np. Muthukrishna i Henrich, 2019), rzadko zastanawiają się nad konkretnymi funkcjami i właściwościami teorii.

Nie przecząc istotnej roli wartości zidentyfikowanych przez Kuhna w praktyce naukowej, należy zauważyć, że standardy metodologiczne mogą się różnić w zależności od konkretnej dziedziny badań (Laudan, 1984). Na przykład w modelowaniu obliczeniowym praktyczna obliczalność (Dror i Gallogly, 1999; Frixione, 2001; van Rooij, 2008; van Rooij i Baggio, 2021) jest jedną z takich cnot. Przetłumaczalność terapeutyczna do praktyki klinicznej jest innym przykładem z psychologii klinicznej. Jedna lista koniecznych i wystarczających właściwości, jakimi powinna odznaczać się każda zadowalająca teoria, nie musi istnieć; właściwości te zależą od funkcji teorii.

Co więcej, Kuhn, obstając, że wartości teoretyczne przypominają inne wartości, poszedł zbyt daleko. Podczas gdy przynajmniej niektóre wartości można swobodnie wybierać, nie dotyczy to wartości epistemicznych (Norton, 2021). Są to, jak podkreśla Norton, w istocie tylko epistemiczne kryteria wyboru teorii. W ramach prezentowanego tu stanowiska kryteria te wyprowadzane są z funkcji różnego rodzaju teorii jako artefaktów poznawczych i nie można ich swobodnie przyjmować bez zaburzania ich kognitywnych ról. Oznacza to, że zalety i wady teorii należy oceniać przede wszystkim przez pryzmat funkcji teorii. W następnej części artykułu proponuję więc wstępną funkcjonalną taksonomię teorii, związaną z ich odpowiednimi zaletami i wadami. Szczegółowe badanie empiryczne poszczególnych rodzajów teorii i ich wartości wykracza, niestety, poza ramy niniejszego artykułu, ale moim nadrzędnym celem jest tutaj uzasadnienie podejścia metanauki poznawczej do kwestii teoretycznych.

## Rodzaje, funkcje i wartości teorii

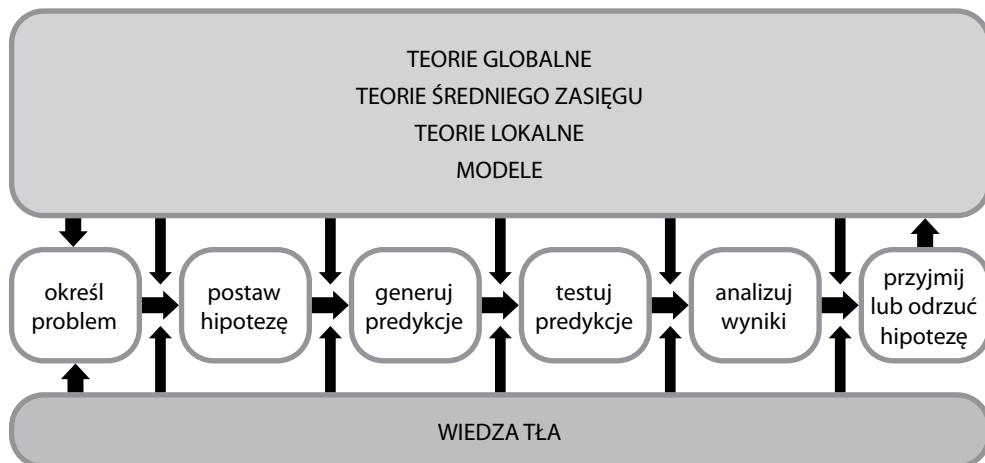
Założenia teoretyczne wykraczające poza dane eksperymentalne lub obserwacyjne są wszechobecne w całym procesie badawczym. W tej części proponuję

wstępną taksonomię teorii opartą na wcześniejszych ustaleniach metodologów, wzbogaconych analizą dyskursu naukowego. Taksonomia ta jest wstępna i ma na celu uzasadnienie dalszych systematycznych ich badań z perspektywy meta-nauki poznawczej. Niniejszy artykuł jest zaledwie pierwszym krokiem w stronę lepszego rozumienia różnych rodzajów teorii – nie tylko w badaniach psychologicznych, lecz także w ogóle w szeroko rozumianych naukach o mózgu i umyśle. Stawiam sobie na celu przeprowadzenie argumentacji uprawdopodobniającej twierdzenie, że istnieje wiele rodzajów teorii, z których każdy ma swoje własne standardy oceny.

Aby lepiej zilustrować te tezy, rozróżnię kilka rodzajów teorii, które można powiązać z uproszczonym przebiegiem badania confirmacyjnego (zob. rysunek 1). Głównym celem w tej części tekstu jest pokazanie, że istnieje więcej niż jeden rodzaj teorii. Chociaż bez wątplenia istnieją też inne rodzaje teorii (np. teorie lokalne i teorie średniego zasięgu, teorie normatywne i opisowe), skoncentruję się wyłącznie na przebiegu teoretyzowania na kilku etapach badań empirycznych. Oczywiście, w rzeczywistości przebieg ten nie musi być chronologiczny (np. narzędzia analityczne mogą oczywiście istnieć wcześniej), ale takie uproszczenie ułatwia prezentację idei.

### Rysunek 1

*Wszystkie etapy badań są przeniknięte przez założenia teoretyczne*



Zacznijmy więc od elementarnego kroku w każdym badaniu eksperymentalnym lub obserwacyjnym: uznania czegoś za problem. Wymaga to już pewnego zaangażowania teoretycznego. W szczególności wyjaśnienia nie dotyczą ani surowych danych (zob. również Aronova i in., 2017; Gitelman, 2013), ani w ogóle samych danych. Przedmiotem wyjaśnień są raczej zjawiska (Bogen i Woodward, 1988) czy też modele danych (Suppes, 1962). Mimo że te ujęcia są do pewnego

stopnia odmienne, ich podstawowe założenie jest takie samo: wyjaśnienia nie dotyczą danych jako takich; wyjaśnia się obciążone teoretycznie dane o zjawiskach. Podobnie jest w przypadku predykcji wyprowadzonych z teorii: rzadko kiedy, jeśli w ogóle, chcemy przewidywać np. błędy pomiaru (obecne w danych eksperymentalnych).

*Teorie danych* pełnią przede wszystkim funkcję opisową, określając przedmiot naszych badań. Chociaż obejmują one terminy teoretyczne lub pojęcia (takie jak „kompetencja gramatyczna”, która sama w sobie nie jest obserwowalna, ponieważ zdolność do analizowania struktury gramatycznej potencjalnie nieskończonej liczby zdań w języku nie jest obserwowalna), teorie te wskazują również istotne fakty charakteryzujące desygnaty tych pojęć, w szczególności dotyczące ich prawidłowości. Natomiast same zbiory okazów lub przypadków mogą być w pełni obserwowalne i nie niosą za sobą żadnego zaangażowania teoretycznego.

Teorie danych (lub zjawisk) same w sobie nie są jednak wyjaśniające ani predykcyjne – a przynajmniej nie jest to ich podstawowym celem. W przypadku złożonych teorii danych można mówić o *teoriach klasyfikacyjnych* (Leonelli, 2016), których przykładem są podręczniki diagnostyczne psychiatrii (American Psychiatric Association, 1980) czy ontologie poznawcze (Poldrack i in., 2011). Ten rodzaj teorii określa również sposób wyodrębniania i identyfikowania mechanizmów odpowiedzialnych za zjawiska (Craver, 2007, 2009). Prawdopodobnie to właśnie brak właściwych teorii danych stanowi jedną z przyczyn trwającego kryzysu (Eronen i Bringmann, 2021; Levenstein i in., 2023). Jedną z hipotetycznych przyczyn kryzysu wiarygodności psychologii jest również mylenie teorii danych z teoriami wyjaśniającymi (Scheel i in., 2020).

Na następnym etapie formułuje się hipotezę. Podstawy *formułowania hipotez* mogą opierać się na heurystykach związanych z daną teorią czy tradycją badawczą. W kognitywistycznych tradycjach badawczych do generowania hipotez często służą zawodne heurystyki (Fiedler, 1991; Miłkowski, 2019). Heurystyki stosowane do generowania hipotez mogą również opierać się konkretnych metodach lub narzędziach stosowanych przez badaczy (Gigerenzer, 1991, 1992; Greenwald, 2012; Nickles, 2018). Te czysto *heurystyczne teorie* mogą funkcjonować obok właściwych teorii wyjaśniających lub predykcyjnych.

Chociaż filozofowie nauki i badacze znają przykłady teorii służących wyłącznie do predykcji, często zakładają, że właściwym zadaniem teorii jest wyjaśnianie. A choć teorie czysto predykcyjne (np. teorie detekcji sygnałów) mogą nie zapewniać bogatych wyjaśnień, uzasadnione jest określanie ich mianem „teoria”, ponieważ wykraczają one poza dane.

Pochopne byłoby np. odrzucenie wszystkich teorii detekcji sygnałów jako narzędzia predykcji w psychologii, ponieważ podstawowym celem takich teorii jest przewidywanie nieznanych jeszcze danych obserwowalnych. Takie teorie mogą być więc przydatne do wielu różnych celów, nawet jeśli nie prowadzą do zadowalającego wyjaśnienia. Ponadto poważnym błędem jest założenie, że samo przewidywanie nieznanych jeszcze danych świadczy o mocy wyjaśniającej teorii (Shmueli, 2010).

Psychologia dopiero zaczyna rozpoznawać potencjał teorii predykcyjnych, które są oceniane za pomocą metod takich jak walidacja krzyżowa, a nie przez

tradycyjne sprawdzanie istotności statystycznej. Teorie te mogą zapewnić bardziej uogólnialne wnioski niż lokalne teorie wyjaśniające (Yarkoni i Westfall, 2017). Co ciekawe, badacze w praktyce odróżniają „rozumienie predykcyjne” (*predictive understanding*) i „rozumienie wyjaśniające” (*explanatory understanding*). Można to pokazać za pomocą narzędzi przetwarzania języka, analizując korpus DOAJ, zawierający artykuły z czasopism w otwartym dostępie, dostępny za pośrednictwem zaawansowanego oprogramowania korpusowego SketchEngine (Kilgarriff i in., 2014)<sup>3</sup>. Jednak nawet filozofowie, którzy podkreślają wartość predykcji, tacy jak Broadbent (np. Broadbent, 2018), czasami błędnie utożsamiają rozumienie z rozumieniem wyjaśniającym. Natomiast w ramach metanauki poznawczej uznaje się, że te dwa rodzaje rozumienia mogą się różnić. Niemniej jednak niektóre teorie wyjaśniające mogą zostać „przekonwertowane” na predykcyjne (Shmueli i Koppius, 2011). Proces taki jest jednak złożony, a w jego ramach tym bardziej uwidacznia się, że teorie predykcyjne i wyjaśniające odgrywają odmienne role w badaniach naukowych.

Czasami ogólne teorie nie tyle bezpośrednio generują testowalne hipotezy, ile raczej motywują lub ograniczają modele zjawisk. Chociaż modele są względnie niezależne od teorii, pełnią rolę mediatorów między teorią a rzeczywistością (Morgan i Morrison, 1999). Mimo że związek między teorią a jej modelem może być bardziej lub mniej luźny, ogólne teorie mogą systematycznie motywować modele. Są to teorie *motywujące modele*.

Hipoteza lub model może wymagać dalszej operacjonalizacji, aby można było wygenerować testowalne predykcje. *Operacjonalizacje* te nie są koniecznie elementami nadrzędnej teorii wyjaśniającej, ale mogą zakorzenić się w badaniach. Na przykład u niemowląt do badania procesów uwagowych wykorzystuje się fiksjację wzroku. Ruchy oczu i fiksjacja są więc wykorzystywane do projektowania warunków eksperymentalnych. Co więcej, na następnym etapie procesu badawczego potrzebne są dodatkowe informacje do zaprojektowania eksperymentu. W grę wchodzi tu nie tylko typowe protokoły eksperymentalne, które różnią się od siebie w zależności od dziedziny (Sullivan, 2009), lecz także narzędzia pomocnicze (np. służące do poprawnego wnioskowania statystycznego). Chociaż operacjonalizacje nie dostarczają nowych przewidywań, są one wystarczająco ważne, aby nie zacierać różnic między nimi przy próbach replikacji badań. Operacjonalizacje mają umożliwić empiryczne (obserwacyjne lub eksperymentalne) badanie zjawisk opisanych przez teorie danych. Operacjonalizacja jest poprawna czy też cechuje się wysoką trafnością teoretyczną o tyle, o ile wybrana przez badacza zmienna obserwowalna odzwierciedla obecność odpowiedniego konstruktów teoretycznego. Skoro większość zjawisk badanych przez psychologię nie jest bezpośrednio obserwowalna, jakość operacjonalizacji ma ogromne znaczenie dla trafności i rzetelności badań (Hughes, 2018).

Po zebraniu i wstępnym przetworzeniu danych wykorzystuje się wreszcie narzędzia analizy statystycznej i inne metody analityczne wraz z oryginalną teorią, aby ustalić, czy dowody empiryczne potwierdzają lub obalają hipotezę zerową.

---

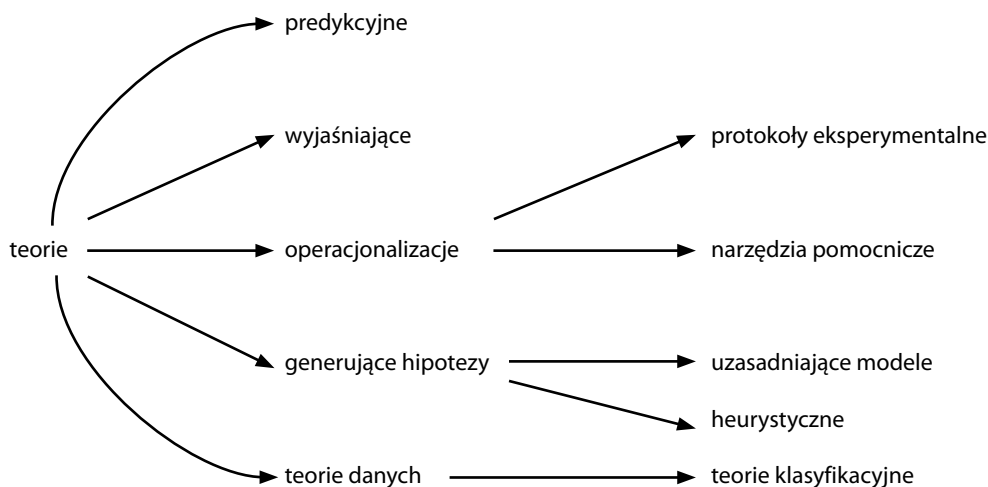
<sup>3</sup> Dane źródłowe udostępniam pod adresem: <https://osf.io/df3ru/>.



Te pomocnicze czy też analityczne narzędzia mogą mieć kolosalny wpływ na wyniki badań empirycznych. Ostatnie badania wskazują, że ten sam zbiór danych może zostać analizowany w rozbieżny sposób przez różne zespoły badawcze, co przyczynia się do teoretycznej elastyczności w naukach o umyśle i mózgu (Botvinnik-Nezer i in., 2020; Silberzahn i in., 2018). Rysunek 2 ilustruje wstępną taksonomię opisaną powyżej.

## Rysunek 2

*Wstępna taksonomia rodzajów teorii w badaniach confirmacyjnych*



Nie da się ukryć, że przedstawiona tutaj taksonomia jest dopiero wstępnym zarysem tego, co wchodzi w grę w badaniach psychologicznych. Przykładowo, można odróżnić teorie wyjaśniające i predykcyjne mające opisywać wartości średnie dla grup od teorii opisujących pojedynczych badanych, a także teorie normatywne od opisowych. Tutaj za cel postawiłem sobie uzasadnienie dalszych systematycznych badań empirycznych nad reprezentacjami teoretycznymi, które już występują w praktyce naukowej. Zarazem to, co zostało ujęte w tej taksonomii, powinno wystarczająco dobitnie świadczyć o tym, że w nauce mamy do czynienia z dużą różnorodnością niezależnych reprezentacji teoretycznych: chociażby dane eksperymentalne dotyczące uwagi u niemowląt same w sobie nie zmuszą nikogo do skorzystania z bayesowskiej statystyki jako narzędzia pomocniczego. Niektóre wybory teoretyczne można oczywiście obronić empirycznie, wskazując, że prowadzą do rzetelnych predykcji (np. tak można uzasadnić czas fiksacji wzroku jako operacjonalizację uwagi), ale wykraczają one poza dane obserwowalne.

Zauważywszy pewną jedność w całej tej wielości – wszystkie te artefakty poznawcze wykraczają swoją treścią poza to, co jest dostępne w danych obserwacyjnych lub eksperymentalnych – warto docenić, jak bardzo różnią się one pod

względem sposobu oceny metodologicznej. Nie wszystkie rodzaje teorii należy oceniać przez pryzmat tych samych kryteriów epistemicznych. Innymi słowy, nie należy mechanicznie narzucać stosowania wobec nich wszystkich pięciu wartości teoretycznych wymienionych Kuhna (1977). Zaczniemy od teorii danych: nie muszą być one owocne, tj. systematycznie prowadzić do nowych odkryć, np. poprzez generowanie pytań badawczych i heurystyk (Ivani, 2019). Nawet teorie klasyfikacyjne nie muszą generować pytań badawczych: podręcznik diagnostyczny zaburzeń psychicznych ma pomóc klinicyście w (statystycznie rzetelnych) diagnozach, a nie w badaniu ich przyczyn. Oczywiście, to mile widziane dodatkowe zalety takiej teorii klasyfikacyjnej, lecz nie są one niezbędne do jej właściwego funkcjonowania w zakresie klasyfikowania tylko określonych objawów jako wskazujących na poszczególne patologie.

Co ciekawe, często zakłada się znacznie większą zawodność teorii formułujących hipotezy niż innych rodzajów teorii, tj. nie muszą cechować się one wysoką empiryczną trafnością. Szczególnie często zakłada się zawodność teorii heurystycznych. Łatwo przytoczyć jeden przykład: tradycja badań nad poznaniem ucieleśnionym w kognitywistyce posługuje się eksperymentalną heurystyką, która sugeruje, że przynajmniej część wariacji wyników w zadaniach poznawczych można wyjaśnić czynnikami cielesnymi – takimi jak morfologia ciała czy stan fizjologiczny – nawet jeśli ta heurystyka mogłaby w większości przypadków prowadzić na manowce (Miłkowski, 2019). Uznaje się, że ta heurystyka jest mimo to cenna, ponieważ może prowadzić do większej liczby odkryć empirycznych: teorie heurystyczne muszą być owocne, w przeciwieństwie do teorii klasyfikacyjnych. Teorie motywujące modele mogą być nietrafne pod innym względem: mogą dostarczać jedynie ogólnego schematu do budowy modeli. Można sądzić, że np. teorie poznania oparte na przetwarzaniu predykcijnym nie są wystarczająco szczegółowe, aby były w stanie generować konkretne przewidywania, lecz wystarczają, by inspirować budowanie modeli obliczeniowych ogromnej liczby zjawisk eksperymentalnych (Litwin i Miłkowski, 2020).

Załóżmy, że sklasyfikowaliśmy zjawisko, sformułowaliśmy hipotezę, a nawet zbudowaliśmy model obliczeniowy. Czas na zaprojektowanie eksperymentu: czy operacjonalizacja zjawiska w eksperymencie powinna być zewnętrznie spójna z poprzednią? Wcale nie. Operacjonalizacje często się spektakularnie różnią od siebie i z biegiem czasu ulegają poprawie. Oznacza to, że choć powinny być spójne z innymi założeniami teoretycznymi stosowanymi w tym samym układzie eksperymentalnym (i ogólnym kontekście badawczym), to nie muszą być szczególnie spójne z tym, co twierdzono wcześniej. Zauważmy, że teorie klasyfikacyjne zwykle różnią się w znacznie mniejszym stopniu: zmiany w podręcznikach diagnostycznych zwykle wprowadza się latami, a ich zatwierdzenie wymaga powołania komitetu, nie zaś tylko opublikowania pojedynczego artykułu w czasopiśmie, ponieważ są one znacznie głębiej zakorzenione (Eronen i Bringmann, 2021). Podobnie teorie wyjaśniające muszą zachować większą spójność z ogólną tradycją badawczą niż szczegółowe operacjonalizacje. Poprawa protokołu eksperymentalnego wydaje się najbardziej elastyczną cechą badań eksperymentalnych.

Spróbujmy teraz ocenić siłę wyjaśniającą hipotezy sformułowanej przez stworzenie kompletnej pełnej teorii wyjaśniającej. Czy powinniśmy dążyć do

uzyskania jak najszerszego jej zakresu? Niekoniecznie. Teorie wyjaśniające nie muszą być maksymalnie ogólne. Na przykład tzw. mikroteorie lub teorie dotyczące wyników poszczególnych badanych niekoniecznie muszą generalizować się na pozostałych badanych, których historia uczenia się różni (Newell i Simon, 1972). Ponadto teorie wyjaśniające nie muszą być predykcyjne, a teorie predykcyjne nie muszą być wyjaśniające (cecha ta nie jest uwzględniona przez Kuhna z oczywistych powodów: zdaje się, że zakładał, iż wszystkie teorie są teoriami wyjaśniającymi).

Powinno być już jasne, dlaczego jedną z podstawowych przyczyn kryzysu teorii może być mylenie różnych rodzajów teorii. Jest tak z dwóch istotnych względów. Po pierwsze, różne rodzaje teorii należy oceniać przy użyciu różnych kryteriów. Po drugie, brak zrozumienia, że kompletny opis zjawiska psychologicznego wymaga opracowania kilku, do pewnego stopnia niezależnych, reprezentacji teoretycznych, może prowadzić do niepełnego komunikowania wyników, co z kolei może obniżyć prawdopodobieństwo replikacji. Opisy teoretyczne zjawisk są kluczowe w każdej próbie replikacji badania w innym układzie eksperymentalnym, ponieważ to one określają, które czynniki powinny być kontrolowane przy projektowaniu warunków eksperymentalnych i warunków kontrolnych. Co ważne, nawet w stosunkowo bliskich dziedzinach badawczych typowe protokoły eksperymentalne służące do badania podobnych zjawisk mogą znacznie się różnić, co może utrudniać integrację ich wyników eksperymentalnych (Sullivan, 2009).

Krótko mówiąc, różne rodzaje teorii należy odróżniać dlatego, że nie powinny być one oceniane w świetle tych samych standardów. Pełnią one swoiste funkcje, a rozdzielne traktowanie różnych teorii może faktycznie przyczyniać się do większej odtwarzalności i replikowalności badań. Może się to również przyczyniać do zdrowej kultury badawczej.

## W stronę kultury teorii

Skomplikowana maszyna wielu ząbających się teorii, występująca w każdym badaniu naukowym zwykle obejmuje skomplikowaną sieć reprezentacji naukowych, kierujących procesem badawczym. Chociaż wchodzące w grę teorie mogą być osobne, zwykle pozostają ze sobą splecione. Na przykład osobne hipotezy teoretyczne mogą służyć do wyjaśniania tego samego zjawiska, podczas gdy odwołują się do tej samej teorii danych.

Weźmy spektakularny przykład doniesienia na temat niepowodzenia replikacji efektu wyczerpania ego (Hagger i in., 2016). Przed tym niepowodzeniem opublikowano niemal 600 badań, w których opisywano zaobserwowanie efektu wyczerpania ego. Wszystkie one miały wspólną podstawową teorię danych: uważano, że eksperymentalnie wywołuje się pewien rodzaj zmęczenia umysłowego, co miało później zmniejszać zdolności badanych do samokontroli w kolejnych zadaniach. Jednak badania replikacyjne przeprowadzone przez wiele laboratoriów jednocześnie ujawniły efekt zerowy, co powinno osłabić wiarę w trafność teorii danych (zwłaszcza że protokół eksperymentalny został zatwierdzony przez

twórcę teorii wyczerpania ego). Efekty zerowe zwykle uważa się za niewymagające wyjaśnienia, lecz w tym wypadku podważone zostały różnorodne hipotezy dotyczące natury „energii umysłowej” (lub poziomów glukozy) mające tłumaczyć występowanie tego efektu. Oznacza to, że hipotezy wyjaśniające wyczerpanie ego, nawet jeśli zostały postawione niezależnie od teorii danych, mają sens tylko wtedy, gdy teoria danych zdaje próbę. Można by sądzić, że mogłoby to doprowadzić do upadku pierwotnej teorii wyczerpania ego i zaniku popularności protokołu eksperymentalnego (Vadillo, 2019).

W tym przypadku wydaje się jednak, że efekt zerowy wymaga wyjaśnienia. W szczególności powinniśmy móc określić, czy zbieżne wyniki 600 badań można wyjaśnić po prostu jako wyniki fałszywie dodatnie, podobnie jak ogromnie wiele innych podobnych wyników w psychologii (Simmons i in., 2011), i czy brak publikacji wyników negatywnych wynika tylko z efektu szuflady (Rosenthal, 1979), w rezultacie którego efekty zerowe są rzadko publikowane.

Niemniej jednak zakwestionować można nawet niepowodzenie replikacji w wielu laboratoriach, np. poprzez przeprowadzenie kolejnej replikacji w wielu laboratoriach z innym protokołem eksperymentalnym (Dang, 2016; Vohs i in., 2021). Główny zwolennik koncepcji wyczerpania ego, Roy Baumeister, śmiało twierdzi, że jest to najlepiej zreplikowane zjawisko w całej psychologii społecznej (Baumeister i Tice, 2022). Istotnie, za niepowodzeniem może kryć się rzekoma odmienność zadania eksperymentalnego od protokołu używanego w oryginalnych setkach badań. W tym przypadku zwolennik koncepcji twierdzi, że zjawisko jest rzeczywiste, ale interwencja eksperymentalna powinna być wykonana w sposób odzwierciedlający jego metody badawcze. Innymi słowy, teoria danych miałaby pozostawać trafna, tylko protokół eksperymentalny niewłaściwy. Krytycy twierdzą jednak, że wywołanie określonego rodzaju zmęczenia umysłowego w bardzo krótkim czasie może mieć niewielki lub wręcz żaden wpływ na przetwarzanie umysłowe. Innymi słowy, są sceptyczni co do tego, czy procedury pomiarowe używane w protokole cechuje jakakolwiek trafność teoretyczna.

Ten krótki przykład pokazuje, że za każde niepowodzenie replikacji można obwiniać kilka reprezentacji teoretycznych. Obejmują one teorie danych, gdzie zjawisko może być artefaktem wątpliwych praktyk badawczych; operacjonalizacje, w tym protokoły eksperymentalne z ich swoistym wyborem zmiennych zależnych i niezależnych; oraz pomocnicze narzędzia statystyczne, które często są przedmiotem kolejnych analiz. Są wreszcie teorie wyjaśniające, pod kątem których projektuje się eksperymenty, dostarczające czynników wyjaśniających wraz z możliwymi nowymi predykcjami eksperymentalnymi do przetestowania w przyszłości. Wszystkie one są częściowo lub całkowicie niezależne, nawet jeśli używane razem, a niepowodzenie replikacji może być niewystarczająco diagnostyczne, aby wskazać, gdzie jest pies pogrzebany.

Próby replikacji mogą spełznąć na niczym z powodu oszustwa lub innych wątpliwych praktyk badawczych, niewłaściwych wyborów metodologicznych, utrwalanych przez tendencyjność publikowania wyników (włącznie z efektem szuflady na wyniki negatywne), czy nawet z powodu niewystarczająco szczegółowego opisu oryginalnego badania w publikacjach (np. często brakuje kwestionariuszy samoopisu stosowanych w wielu badaniach psychologicznych). Jeśli

jednak pominiemy te czynniki, to po niepowodzeniu replikacji na placu boju pozostaną problemy związane z teorią. Innymi słowy, reprezentacje teoretyczne odgrywają kluczowe role w badaniach naukowych, a gdy nie uda się replikacja, ich trafność musi być sprawdzona.

Nasuwa się myśl, że aby przezwyciężyć obecny kryzys wiarygodności, należy stworzyć teorie pozbawione niedookreśleń i niedopowiedzeń. Aby wyniki badawcze mogły przetrwać próbę czasu, badacze muszą *explicite* sformułować kilka teorii, w tym teorie danych i ważne operacjonalizacje. Dzisiaj wiele z tych teorii cechuje się niedookreślonością i jest zaledwie naszkicowana. Możliwe, że ogólne zasady psychologiczne uważane są za zbyt banalne, by je wypowiadać (Hensel i in., 2022). Jednak to systematyczny rozwój wielu teorii może pomóc zacząć zażegnawać kryzys wiarygodności.

Następnym krokiem jest oczywiście opracowanie katalogu zalet i wad przysługujących reprezentacjom teoretycznym oraz dogłębna dyskusja nad nimi. Ta dyskusja już się zaczęła: proponuje się dzisiaj już nie tylko procedury minimalizowania możliwego wpływu wątpliwych praktyk (takich jak przemilczane odrzucenie danych niezgodnych z hipotezami pierwotnymi jako wartości oddalonych), lecz także wymaga się udowadniania poprawności metod. Jednak poprawne powinny zarówno metody, jak i teorie. Poważnym problemem z obecną wersją teorii wyczerpania ego jest to, że łączy ona poziomy glukozy z efektem wyczerpania ego, a tymczasem poziomy glukozy nie mają prawa tak drastycznie zmieniać się w stosunkowo krótkich odstępach czasu (Vadillo i in., 2016). Innymi słowy, teoria wydaje się całkowicie niezgodna z (dużo bardziej empirycznie solidnymi) teoriami fizjologicznymi metabolizmu glukozy. Tymczasem bez poziomu glukozy jako zasobu do wyczerpania podczas zadań samokontroli teoria wyczerpania ego zatracą treść i staje się mglista.

Od lat toczy się dyskusja na temat zalet (i odpowiadających im wad) przypisywanych powszechnie teoriom wyjaśniającym (Keas, 2018), nawet jeśli pomija się w niej niektóre kryteria stosowane w praktyce (takie jak systematyczność). Nadal brakuje nam tego rodzaju dyskusji w odniesieniu do innych rodzajów reprezentacji teoretycznych, w tym protokołów eksperymentalnych. Jednak w podejściu metanauki poznawczej, opartym na narzędziach przetwarzania języka naturalnego, łatwo daje się wskazać przymiotniki przypisujące typowe właściwości protokołom eksperymentalnym w dużym zbiorze literatury (ponownie skorzystamy z korpusu DOAJ). Są to przymiotniki: „identical”, „suitable”, „specific”, „similar”, „consistent” i „necessary” („identyczny”, „odpowiedni”, „specyficzny”, „podobne”, „spójny” i „niezbędny”). Istnieją użycia przymiotnika „identical”, które wyrażają wartość poznawczą: podczas odtwarzania wyniku eksperymentu należy skorzystać z identycznego protokołu eksperymentalnego (jest to konieczny warunek odtwarzalności); inne użycia po prostu wyrażają konstatację, że dla wielu badań zgłoszonych w artykule lub serii artykułów użyto tego samego (identycznego) protokołu, co nie jest związane z wartościami teoretycznymi. W tym artykule brak miejsca, aby omówić szczegóły, więc pozostawię tu tylko sugestię, że tego rodzaju dowody empiryczne mogą pozwolić lepiej zrozumieć i udoskonalić formę obecnych praktyk badawczych. W szczególności należy nie tylko sprawdzać trafność protokołu, lecz także jego interoperacyjność i otwartość.

Intersubiektywna trafność protokołu eksperymentalnego wiąże się z jednym z nawracających problemów w psychologii, a mianowicie z intersubiektywną trafnością jej teorii danych. Na przykład powracają obawy dotyczące intersubiektywnej rzetelności kryteriów diagnostycznych zaburzeń psychicznych. Jednym z rewolucyjnych aspektów DSM-III był nacisk położony na rzetelność, prawdopodobnie ze względu na dostrzegane problemy z elastycznością diagnostyki (Kawa i Giordano, 2012; Wilson, 1993). Rzekomy kryzys rzetelności diagnostyki został poniekąd wyolbrzymiony, gdyż słynne badania Rosenhana (1973), który dowodził, że diagnozy w instytucjach psychiatrycznych są całkowicie arbitralne i przedwczesne, okazały się oszustwem naukowym (Cahalan, 2019; Scull, 2023). Zwiększenie rzetelności diagnoz to przykład eliminacji niepożądanego elastyczności teoretycznej. Na potrzeby tego artykułu wystarczy zauważyć, że DSM-III był reakcją na coś, co uznano za kryzys, i że realizacja planu zwiększenia intersubiektywnej rzetelności diagnoz jest rzeczywiście powodem do chwały tej teorii nosologicznej.

Każdą teorię danych, w tym DSM, można oceniać przez pryzmat tego, czy przyczynia się ona też do poprawności procedur pomiarowych, a nie tylko do kategoryzacji. Poprawne pomiary mają co najmniej dwie kluczowe cechy: są precyzyjne i zbieżne (Isaac, 2019). O precyzji świadczy to, że wielokrotnie wykonywane procedury pomiarowe konsekwentnie dają te same wyniki (w zbliżonych granicach błędu). Notabene, w korpusie DOAJ można zauważyć, że ta właściwość jest przypisywana pomiarowi za pomocą takich terminów jak „reproducible” (*odtwarzalny*) i „repeatable” (*powtarzalny*), co jasno wskazuje, w jaki sposób może wiązać się z reprodukowalnością i replikowalnością badań. W przypadku diagnoz oznacza to, że powinny pozostawać precyzyjne w różnych okresach i u wielu diagnostów. W praktyce klinicznej po przejściu z systemu DSM-I lub DSM-II na DSM-III stwierdzono znaczny wzrost precyzji, rozumiany w kategoriach zwiększenia intersubiektywnej zgodności ferowanych ocen (Di Nardo i in., 1983). Znacznie jednak bardziej skomplikowane jest osiągnięcie zbieżności, która polega na uzyskiwaniu tego samego wyniku za pomocą różnych procedur. Jest tak dlatego, że sprzeczne założenia teoretyczne stojące za tymi procedurami mogą rzutować na diagnozy. Uzyskanie zbieżności mogłoby uzasadniać realistyczne czy obiektywne podejście do wyników diagnoz i wystarczałoby, aby udowodnić trafność teoretyczną. Niestety, pozostaje to nadal w sferze mrzonek.

Ponadto pojmowanie ocen opartych na DSM-III w kategoriach pomiaru, podobnie jak wielu innych sądów jakościowych, jest poniekąd na wyrost, gdyż opierają się one na skalach nominalnych lub porządkowych, które nie mają wszystkich cech przestrzeni metrycznych stosowanych w procedurach pomiarowych w innych dziedzinach (jeśli chodzi o analizy DSM w kategoriach psychometrycznych, zob. np. Borsboom, 2008). Choć mało prawdopodobne wydaje się, że cała psychologia zdoła upodobnić się do psychofizyki i będzie korzystać z przestrzeni metrycznych (zbiorów elementów z zdefiniowaną na nich metryką odległości), to jednak możliwy jest pewien postęp w tym zakresie. Można np. twierdzić, że podejmowane od niedawna próby budowania nosologii opartej na sieciach objawów (Borsboom i in., 2018; Borsboom i Cramer, 2013) są krokiem w tym kierunku.

Chociaż żadnemu podręcznikowi diagnostycznemu w psychopatologii nie można przypisywać pełnego sukcesu, przyjmowane w tym artykule podejście

proceedzi do konstatacji, że rzetelność jest koniecznym, choć niewystarczającym kryterium epistemicznym wyboru teorii danych. Eliminuje ona jedno źródło kryzysu, którym jest elastyczność – zarówno procedur pomiarowych, jak i opisów teoretycznych. Jednak daleko stąd do całkowicie satysfakcjonującej teorii.

Jest tak nie tylko dlatego, że zagadnienia pomiarowe nie są jedyną przyczyną trwającego kryzysu. Chociaż postęp w zakresie pomiaru może rzeczywiście przyczynić się do postępu teoretycznego, postęp teoretyczny może przekraczać nasze zdolności pomiarowe (Bringmann i Eronen, 2016). Przecież Galileusz nie był stanie dokonywać rzetelnego pomiaru czasu (Koyré, 1953). Można twierdzić, że symulacje obliczeniowe w psychologii stanowią przykłady precyzyjnych teorii (Anderson, 2007; Marr, 1982), których pełna ocena empiryczna jest niezwykle trudna – tym bardziej że nie mamy odpowiednio szczegółowych dowodów neurologicznych, aby ocenić niektóre z ich tez. W rzeczywistości zaś postęp teoretyczny może opierać się po prostu na lepszej pojęciowej reprezentacji wyjaśnianych zjawisk. Tak było w przypadku krytyki Chomsky’ego (1959) wymierzonej w podejście Skinnera do języka. Wskazując, że użytkownicy języka potrafią zrozumieć i wytworzyć nieskończenie wiele poprawnych gramatycznie zdań (co określa się mianem „produktywności” lub „twórczości”), zaproponował on wiarygodną teorię obserwowalnych danych językowych dotyczących naszego zachowania. Jego główny argument sprowadzał się do tego, iż teoria Skinnera nie jest w stanie uchwycić owej produktywności. Nawet jeśli niektóre aspekty jego krytyki można podważyć (MacCorquodale, 1970), teoretyczna konstatacja, że użycie języka przez ludzi cechuje się produktywnością, pozostaje jednym z głęboko zakorzenionych w dalszych badaniach założeń.

Zwiększenie roli rozwoju teoretycznego w badaniach psychologicznych wymaga zarówno małych, jak i większych kroków. W tym tekście zwracałem uwagę tylko na postęp w teoriach danych, pomijając teorie wyjaśniające czy predykcyjne. Potrzebny jest wszakże postęp we wszystkich typach teorii. Wymaga to zbiorowego wysiłku, ale brak kumulatywności w refleksji teoretycznej prowadzi do fragmentacji teoretycznej i braku współpracy naukowej w zakresie rozwoju teorii. Teorie traktuje się jak cudze szczoteczki do zębów (Mischel, 2008). Aby sobie z tym poradzić, potrzebna jest współpraca, w tym także współpraca rywalizacyjna, nad kwestiami teoretycznymi (Cowan i in., 2020; Del Pin i in., 2021), oprócz systematycznych przeglądów. Musimy dopuścić do rozkwitu tysiąca kwiatów, ale teorie powinny być też eliminowane, gdy stają się przestarzałe. Do tego potrzebny jest wyraźniejszy nacisk na ocenę teorii.

## Podsumowanie

Moim celem było uzasadnienie nowego podejścia w badaniach nad nauką: metanauki poznawczej, zwłaszcza w czasach rosnącego poczucia istotności kryzysu w psychologii. Ujęcie teorii jako artefaktów teoretycznych, których treści nie można sprowadzić do danych doświadczalnych ani obserwacyjnych, otwiera nową perspektywę badań nad reprezentacjami teoretycznymi, których role nie ograniczają

się jedynie do wyjaśniania, ale obejmują także opisywanie zjawisk, przewidywanie, operacjonalizowanie i formułowanie nowych hipotez. Takie szerokie rozumienie teorii musi zostać uzupełnione stosowaniem odpowiednich (często odmiennych) kryteriów normatywnych do oceny tych reprezentacji w praktyce naukowej.

Bez odpowiednio opracowanych teorii nie sposób liczyć na stworzenie kompleksowych koncepcji teoretycznych zjawisk psychologicznych i rozwiązywanie licznych problemów omawianych w debatach metanaukowych. Kryzys wiarygodności wymaga złożonej odpowiedzi. Jednym ze środków zaradczych jest większy nacisk na teorie i ich ocenę. Do tego niezbędna jest zmiana kultury badań. Ta zmiana powinna prowadzić do opracowania wspólnych i bardziej rygorystycznych kryteriów epistemicznych służących do oceny teorii.

Ze względu na ograniczenia długości tekstu mogłem tylko pobieżnie potraktować kwestię odrębności ról i odpowiednich zalet (i wad) różnych rodzajów teorii. Dalsze moje prace, tak jakościowe, jak i ilościowe, skoncentrują się na konkretnych rodzajach reprezentacji teoretycznych, co ma przybliżyć nas do bardziej systematycznego rozumienia teorii. Metanauka poznawcza stawia sobie za cel stworzenie ogólnych ram teoretycznych, czerpiąc inspiracje z filozofii nauki, ale odbiegając nieco od praktyki badań nad nauką i techniką oraz kognitywistyki nauki, które zwykle skupiają się na badaniu poszczególnych laboratoriów czy przypadków historycznych. Dzięki wykorzystaniu dużych zbiorów danych metody humanistyki cyfrowej umożliwiają „daleką lekturę” (Moretti, 2000), co ułatwia generalizację wychodzącą poza indywidualne studia przypadków (Piper, 2020). Podczas gdy kognitywiści dogłębnie badają funkcjonowanie artefaktów poznawczych w różnych konkretnych przypadkach, nie sposób uzyskać ogólnych wskazań normatywnych bez systematycznego i wszechstronnego zbadania typów funkcjonalnych teorii. Bez zastosowania technik przetwarzania języka przekracza to ludzkie możliwości. Badania z wykorzystaniem tych technik będą jednak tematem przyszłych prac mojego zespołu.

## Dostępność danych

Źródłowe konkorancje i kolokacje z korpusu DOAJ, wygenerowane w systemie SketchEngine (w formacie CSV) dostępne są do pobrania pod adresem: <https://osf.io/df3ru/>

## Bibliografia

- Afeltowicz, Ł., Wachowski, W. (2015). How Far we Can Go Without Looking Under the Skin: The Bounds of Cognitive Science. *Studies in Logic, Grammar and Rhetoric*, 40(1), 91–109. <https://doi.org/10.1515/slgr-2015-0005>
- Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J., Watts, D. J. (2022). Beyond Playing 20 Questions with Nature: Integrative Experiment Design in



- the Social and Behavioral Sciences. *Behavioral and Brain Sciences*, 1–55. <https://doi.org/10.1017/S0140525X22002874>
- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (wyd. 3). American Psychiatric Association.
- Anderson, J. R. (2007). *How Can the Mind Occur in the Physical Universe?* Oxford University Press.
- Aronova, E., Oertzen, C. von, Sepkoski, D. (red.). (2017). *Data histories*. University of Chicago Press.
- Autzen, B. (2021). Is the replication crisis a base-rate fallacy? *Theoretical Medicine and Bioethics*, 42(5), 233–243. <https://doi.org/10.1007/s11017-022-09561-8>
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature News*, 533(7604), 452. <https://doi.org/10.1038/533452a>
- Baumeister, R. F., Tice, D. M. (2022). Ego Depletion is the Best Replicated Finding in All of Social Psychology. *Scholarly Journal of Psychology and Behavioral Sciences*, 6(2), 686–688. <https://doi.org/10.32474/SJPBS.2021.06.000234>
- Bird, A. (2021). Understanding the Replication Crisis as a Base Rate Fallacy. *The British Journal for the Philosophy of Science*, 72(4), 965–993. <https://doi.org/10.1093/bjps/axy051>
- Boekel, W., Wagenmakers, E.-J., Belay, L., Verhagen, J., Brown, S., Forstmann, B. U. (2015). A purely confirmatory replication study of structural brain-behavior correlations. *Cortex*, 66, 115–133. <https://doi.org/10.1016/j.cortex.2014.11.019>
- Bogen, J., Woodward, J. (1988). Saving the Phenomena. *The Philosophical Review*, 97(3), 303. <https://doi.org/10.2307/2185445>
- Borsboom, D. (2008). Psychometric perspectives on diagnostic systems. *Journal of Clinical Psychology*, 64(9), 1089–1108. <https://doi.org/10.1002/jclp.20503>
- Borsboom, D., Cramer, A., Kalis, A. (2018). Brain disorders? Not really... Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences*, 1–54. <https://doi.org/10.1017/S0140525X17002266>
- Borsboom, D., Cramer, A. O. J. (2013). Network Analysis: An Integrative Approach to the Structure of Psychopathology. *Annual Review of Clinical Psychology*, 9(1), 91–121. <https://doi.org/10.1146/annurev-clinpsy-050212-185608>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., [...] Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Bower, G. H. (1993). The fragmentation of psychology? *American Psychologist*, 48(8), 905–907. (1994-00003-001). <https://doi.org/10.1037/0003-066X.48.8.905>
- Bringmann, L. F., Eronen, M. (2016). Heating up the measurement debate: What psychologists can learn from the history of physics. *Theory & Psychology*, 26(1), 27–43. <https://doi.org/10.1177/0959354315617253>
- Broadbent, A. (2018). Prediction, Understanding, and Medicine. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 43(3), 289–305. <https://doi.org/10.1093/jmp/jhy003>

- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365. <https://doi.org/10.1038/nrn3475>
- Cahalan, S. (2019). *The Great Pretender*. Grand Central Publishers.
- Callebaut, W. (1993). *Taking the naturalistic turn or how real philosophy of science is done*. University of Chicago Press.
- Callebaut, W. (2013). Naturalizing Theorizing: Beyond a Theory of Biological Theories. *Biological Theory*, *7*(4), 413–429. <https://doi.org/10.1007/s13752-013-0122-2>
- Carsel, T., Demos, A. P., Motyl, M. (2018). Strong scientific theorizing is needed to improve replicability in psychological science. *Behavioral and Brain Sciences*, *41*, e123. <https://doi.org/10.1017/S0140525X1800078X>
- Chang, H. (2017). VI—Operational Coherence as the Source of Truth. *Proceedings of the Aristotelian Society*, *117*(2), 103–122. <https://doi.org/10.1093/arisoc/aox004>
- Chomsky, N. (1959). Review of Verbal Behavior by B. F. Skinner. *Language*, *35*(1), 26–58.
- Cooper, R. P., Shallice, T. (1995). Soar and the case for unified theories of cognition. *Cognition*, *55*(2), 115–149. [https://doi.org/10.1016/0010-0277\(94\)00644-Z](https://doi.org/10.1016/0010-0277(94)00644-Z)
- Cowan, N., Belletier, C., Doherty, J. M., Jaroslawska, A. J., Rhodes, S., Forsberg, A., [...] Logie, R. H. (2020). How Do Scientific Views Change? Notes From an Extended Adversarial Collaboration. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *15*(4), 1011–1025. <https://doi.org/10.1177/1745691620906415>
- Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford University Press.
- Craver, C. F. (2009). Mechanisms and natural kinds. *Philosophical Psychology*, *22*(5), 575–594. <https://doi.org/10.1080/09515080903238930>
- Cummins, R. (2000). “How does it work” versus “what are the laws?”: Two conceptions of psychological explanation. W: F. Keil i R. A. Wilson (red.), *Explanation and Cognition* (s. 117–145). MIT Press.
- Dale, R., Dietrich, E., Chemero, A. (2009). Explanatory Pluralism in Cognitive Science. *Cognitive Science*, *33*(5), 739–742. <https://doi.org/10.1111/j.1551-6709.2009.01042.x>
- Dang, J. (2016). Commentary: A Multilab Preregistered Replication of the Ego-Depletion Effect. *Frontiers in Psychology*, *7*, 1155. <https://doi.org/10.3389/fpsyg.2016.01155>
- Del Pin, S. H., Skóra, Z., Sandberg, K., Overgaard, M., Wierchoń, M. (2021). Comparing theories of consciousness: Why it matters and how to do it. *Neuroscience of Consciousness*, *2021*(2), niab019. <https://doi.org/10.1093/nc/niab019>
- Dellsén, F. (2020). The epistemic impact of theorizing: Generation bias implies evaluation bias. *Philosophical Studies*, *177*(12), 3661–3678. <https://doi.org/10.1007/s11098-019-01387-w>
- Di Nardo, P. A., O'Brien, G. T., Barlow, D. H., Waddell, M. T., Blanchard, E. B. (1983). Reliability of DSM-III Anxiety Disorder Categories Using a New Structured Interview. *Archives of General Psychiatry*, *40*(10), 1070–1074. <https://doi.org/10.1001/archpsyc.1983.01790090032005>

- Dror, I. E., Gallogly, D. P. (1999). Computational analyses in cognitive neuroscience: In defense of biological implausibility. *Psychonomic Bulletin & Review*, 6(2), 173–182. <https://doi.org/10.3758/BF03212325>
- Erdin, H. O. (2021). Appraisal of certain methodologies in cognitive science based on Lakatos's methodology of scientific research programmes. *Synthese*, 199, 89–112. <https://doi.org/10.1007/s11229-020-02612-4>
- Eronen, M. I., Bringmann, L. F. (2021). The Theory Crisis in Psychology: How to Move Forward. *Perspectives on Psychological Science*, 16(4), 779–788. <https://doi.org/10.1177/1745691620970586>
- Ferguson, C. J., Heene, M. (2012). A Vast Graveyard of Undead Theories: Publication Bias and Psychological Science's Aversion to the Null. *Perspectives on Psychological Science*, 7(6), 555–561. <https://doi.org/10.1177/1745691612459059>
- Fiedler, K. (1991). Heuristics and Biases in Theory Formation: On the Cognitive Processes of those Concerned with Cognitive Processes. *Theory & Psychology*, 1(4), 407–430. <https://doi.org/10.1177/0959354391014002>
- Fiedler, K. (2017). What Constitutes Strong Psychological Science? The (Neglected) Role of Diagnosticity and A Priori Theorizing. *Perspectives on Psychological Science*, 12(1), 46–61. <https://doi.org/10.1177/1745691616654458>
- Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, 29(2), 158–181. <https://doi.org/10.1177/0959354319835322>
- Frankenhuis, W. E., Panchanathan, K., Smaldino, P. E. (2022). Strategic ambiguity in the social sciences. *Social Psychological Bulletin*. <https://www.psycharchives.org/en/item/e5bb9192-80a4-4ae4-9cda-5d144008196e>
- Fried, E. I. (2020). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry*, 31(4), 271–288. <https://doi.org/10.1080/1047840X.2020.1853461>
- Frixione, M. (2001). Tractable competence. *Minds and Machines*, 11, 379–397.
- Gaj, N. (2016). *Unity and Fragmentation in Psychology: The Philosophical and Methodological Roots of the Discipline*. Taylor, Francis Ltd.
- Giere, R. N., Moffatt, B. (2003). Distributed Cognition: Where the Cognitive and the Social Merge. *Social Studies of Science*, 33(2), 301–310. <https://doi.org/10.1177/03063127030332017>
- Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, 98(2), 254–267. <https://doi.org/10.1037/0033-295X.98.2.254>
- Gigerenzer, G. (1992). Discovery in Cognitive Psychology: New Tools Inspire New Theories. *Science in Context*, 5(2), 329–350. <https://doi.org/10.1017/S0269889700001216>
- Gigerenzer, G. (1998). Surrogates for Theories. *Theory & Psychology*, 8(2), 195–204. <https://doi.org/10.1177/0959354398082006>
- Gitelman, L. (red.). (2013). *“Raw data” is an oxymoron*. The MIT Press.
- Goertzen, J. R. (2008). On the Possibility of Unification: The Reality and Nature of the Crisis in Psychology. *Theory & Psychology*, 18(6), 829–852. <https://doi.org/10.1177/0959354308097260>

- Gorelick, R. (2011). What is theory? *Ideas in Ecology and Evolution*, 4, 1–10. <https://doi.org/10.4033/iee.2011.4.1.c>
- Greenwald, A. G. (2012). There Is Nothing So Theoretical as a Good Method: *Perspectives on Psychological Science*, 7(2), 99–108. <https://doi.org/10.1177/1745691611434210>
- Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review*, 93(2), 216–229. <https://doi.org/10.1037/0033-295X.93.2.216>
- Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., [...] Zwiener, M. (2016). A Multilab Preregistered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science*, 11(4), 546–573. <https://doi.org/10.1177/1745691616652873>
- Hensel, W. M. (2020). Double trouble? The communication dimension of the reproducibility crisis in experimental psychology and neuroscience. *European Journal for Philosophy of Science*, 10(3), 44. <https://doi.org/10.1007/s13194-020-00317-6>
- Hensel, W. M., Milkowski, M., Nowakowski, P. (2022). Without more theory, psychology will be a headless rider. *Behavioral and Brain Sciences*, 45, e20. <https://doi.org/10.1017/S0140525X21000212>
- Hoyningen-Huene, P. (2013). *Systematicity: The Nature of Science*. Oxford University Press.
- Hughes, B. M. (2018). *Psychology in crisis*. Palgrave.
- Hutchins, E. (1995). *Cognition in the wild*. MIT Press.
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLOS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Irvine, E. (2021). The Role of Replication Studies in Theory Building. *Perspectives on Psychological Science*, 16(4), 844–853. <https://doi.org/10.1177/1745691620970558>
- Isaac, A. M. C. (2019). Epistemic Loops and Measurement Realism. *Philosophy of Science*, 86(5), 930–941. <https://doi.org/10.1086/705476>
- Ivani, S. (2019). What we (should) talk about when we talk about fruitfulness. *European Journal for Philosophy of Science*, 9(4), 1–18. <https://doi.org/10.1007/s13194-018-0231-7>
- Kawa, S., Giordano, J. (2012). A brief historicity of the Diagnostic and Statistical Manual of Mental Disorders: Issues and implications for the future of psychiatric canon and practice. *Philosophy, Ethics, and Humanities in Medicine: PEHM*, 7, 2. <https://doi.org/10.1186/1747-5341-7-2>
- Keas, M. N. (2018). Systematizing the theoretical virtues. *Synthese*, 195(6), 2761–2793. <https://doi.org/10.1007/s11229-017-1355-6>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., [...] Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1, 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Klein, S. B. (2014). What can recent replication failures tell us about the theoretical commitments of psychology? *Theory & Psychology*, 24(3), 326–338. <https://doi.org/10.1177/0959354314529616>
- Koyré, A. (1953). An Experiment in Measurement. *Proceedings of the American Philosophical Society*, 97(2), 222–237.

- Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. The University of Chicago Press.
- Langley, P., Simon, H. A., Bradshaw, G. L., Żytkow, J. M. (1987). *Scientific discovery: Computational explorations of the creative processes*. MIT Press.
- Laudan, L. (1984). *Science and values: The aims of science and their role in scientific debate*. University of California Press.
- Lean, O. M., Rivelli, L., Pence, C. H. (2021). Digital Literature Analysis for Empirical Philosophy of Science. *The British Journal for the Philosophy of Science*. <https://doi.org/10.1086/715049>
- Leonelli, S. (2016). *Data-centric biology: A philosophical study*. The University of Chicago Press.
- Levenstein, D., Alvarez, V. A., Amarasingham, A., Azab, H., Chen, Z. S., Gerkin, R. C., [...] Redish, A. D. (2023). On the Role of Theory and Modeling in Neuroscience. *Journal of Neuroscience*, 43(7), 1074–1088. <https://doi.org/10.1523/JNEUROSCI.1179-22.2022>
- Litwin, P., Miłkowski, M. (2020). Unification by Fiat: Arrested Development of Predictive Processing. *Cognitive Science*, 44(7), e12867. <https://doi.org/10.1111/cogs.12867>
- Longino, H. E. (1996). Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy. W: L. H. Nelson i J. Nelson (red.), *Feminism, Science, and the Philosophy of Science* (s. 39–58). Springer Netherlands. [https://doi.org/10.1007/978-94-009-1742-2\\_3](https://doi.org/10.1007/978-94-009-1742-2_3)
- MacCorquodale, K. (1970). On Chomsky's review of Skinner's Verbal Behavior. *Journal of the Experimental Analysis of Behavior*, 13(1), 83–99. <https://doi.org/10.1901/jeab.1970.13-83>
- Manninen, T., Aćimović, J., Havela, R., Teppola, H., Linne, M.-L. (2018). Challenges in Reproducibility, Replicability, and Comparability of Computational Models and Tools for Neuronal and Glial Networks, Cells, and Subcellular Structures. *Frontiers in Neuroinformatics*, 12, 20. <https://doi.org/10.3389/fninf.2018.00020>
- Marr, D. (1982). *Vision*. W. H. Freeman and Company.
- Matthews, G. (2020). Against consensus: Embracing the disunity of personality theory. *Personality and Individual Differences*, 152, 109535. <https://doi.org/10.1016/j.paid.2019.109535>
- McMullin, E. (2010). The Virtues of a Good Theory. W: M. Curd i M. Psillos (red.), *The Routledge Companion to Philosophy of Science* (s. 561–571). Routledge. <https://doi.org/10.4324/9780203744857.ch53>
- Meehl, P. E. (1967). Theory-Testing in Psychology and Physics: A Methodological Paradox. *Philosophy of Science*, 34(2), 103–115. <https://doi.org/10.1086/288135>
- Miłkowski, M. (2019). Fallible Heuristics and Evaluation of Research Traditions. The Case of Embodied Cognition. *Ruch Filozoficzny*, 75(2), 223–236. <https://doi.org/10.12775/RF.2019.031>
- Miłkowski, M. (2022). Cognitive Artifacts and Their Virtues in Scientific Practice. *Studies in Logic, Grammar and Rhetoric*, 67(3), 219–246. <https://doi.org/10.2478/slgr-2022-0012>
- Miłkowski, M., Hensel, W. M., Hohol, M. (2018). Replicability or reproducibility? On the replication crisis in computational neuroscience and sharing only relevant detail.

- Journal of Computational Neuroscience*, 45(3), 163–172. <https://doi.org/10.1007/s10827-018-0702-z>
- Milkowski, M., Litwin, P. (2022). Testable or bust: Theoretical lessons for predictive processing. *Synthese*, 200(6), 462. <https://doi.org/10.1007/s11229-022-03891-9>
- Mischel, W. (2008). The Toothbrush Problem. *APS Observer*, 21(11). <https://www.psychologicalscience.org/observer/the-toothbrush-problem>
- Moretti, F. (2000). Conjectures on World Literature. *New Left Review*, 1, 54–68.
- Morgan, M. S., Morrison, M. (1999). *Models As Mediators*. Cambridge University Press.
- Muthukrishna, M., Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, 3, 221–229. <https://doi.org/10.1038/s41562-018-0522-1>
- Nersessian, N. J. (2008). *Creating scientific concepts*. MIT Press.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. W: W. G. Chase (red.), *Visual information processing* (s. 283–308). Academic Press.
- Newell, A., Simon, H. A. (1972). *Human Problem Solving*. Prentice-Hall.
- Nickles, T. (2018). TTT: A Fast Heuristic to New Theories? W: D. Danks i E. Ippoliti (red.), *Building Theories: Heuristics and Hypotheses in Sciences* (s. 169–189). Springer. [https://doi.org/10.1007/978-3-319-72787-5\\_9](https://doi.org/10.1007/978-3-319-72787-5_9)
- Norman, D. A. (1991). Cognitive Artifacts. W: J. M. Carroll (red.), *Designing Interaction: Psychology at the Human-Computer Interface* (s. 17–38). Cambridge University Press.
- Norton, J. D. (2021). *The Material Theory of Induction*. University of Calgary Press.
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., [...] Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, 73(1), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Oberauer, K., Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5), 1596–1618. <https://doi.org/10.3758/s13423-019-01645-2>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251). <https://doi.org/10.1126/science.aac4716>
- Osbeck, L. M., Nersessian, N. J. (2014). Situating distributed cognition. *Philosophical Psychology*, 27(1), 82–97. <https://doi.org/10.1080/09515089.2013.829384>
- Pence, C. H., Ramsey, G. (2018). How to Do Digital Philosophy of Science. *Philosophy of Science*, 85(5), 930–941. <https://doi.org/10.1086/699697>
- Piper, A. (2020). *Can We Be Wrong? The Problem of Textual Evidence in a Time of Data*. Cambridge University Press.
- Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., [...] Bilder, R. M. (2011). The Cognitive Atlas: Toward a Knowledge Foundation for Cognitive Neuroscience. *Frontiers in Neuroinformatics*, 5. <https://doi.org/10.3389/fninf.2011.00017>
- Rescher, N. (1979). *Cognitive systematization: A systems-theoretic approach to a coherentist theory of knowledge*. Basil Blackwell.

- Roberts, S., Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107(2), 358–358.
- Rosenthal, D. L. (1973). On Being Sane in Insane Places. *Science*, 179(4070), 250–258. <https://doi.org/10.1126/science.179.4070.250>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Scheel, A. M., Tiokhin, L., Isager, P. M., Lakens, D. (2020). Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science*, 16(4), 744–755. <https://doi.org/10.1177/1745691620966795>
- Schindler, S. (2018). *Theoretical virtues in science: Uncovering reality through theory*. Cambridge University Press.
- Schooler, J. W. (2014). Metascience could rescue the ‘replication crisis.’ *Nature*, 515(7525), 9. <https://doi.org/10.1038/515009a>
- Scull, A. (2023). Rosenhan revisited: Successful scientific fraud. *History of Psychiatry*, 0957154X221150878. <https://doi.org/10.1177/0957154X221150878>
- Shmueli, G. (2010). To Explain or to Predict? *Statistical Science*, 25(3), 289–310. <https://doi.org/10.1214/10-STS330>
- Shmueli, G., Koppius, O. R. (2011). Predictive Analytics in Information Systems Research. *MIS Quarterly*, 35(3), 553. <https://doi.org/10.2307/23042796>
- Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., Awtrey, E., [...] Nosek, B. A. (2018). Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological Science*, 1(3), 337–356. <https://doi.org/10.1177/2515245917747646>
- Simmons, J. P., Nelson, L. D., Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956-797611417632>
- Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. W: R. R. Vallacher, S. J. Read i A. Nowak (red.), *Computational Social Psychology* (wyd. 1, s. 311–331). Routledge. <https://doi.org/10.4324/9781315173726-14>
- Staats, A. W. (1986). Unified Positivism: A Philosophy for Psychology and the Disunified Sciences. *Theoretical, Philosophical Psychology*, 6(2), 77–90. <https://doi.org/10.1037/h0091427>
- Sullivan, J. A. (2009). The multiplicity of experimental protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese*, 167(3), 511–539. <https://doi.org/10.1007/s11229-008-9389-4>
- Suppes, P. (1962). Models of Data. W: E. Nagel, P. Suppes i A. Tarski (red.), *Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Congress* (s. 252–261). Stanford University Press.
- Szollosi, A., Donkin, C. (2019). Neglected Sources of Flexibility in Psychological Theories: From Replicability to Good Explanations. *Computational Brain, Behavior*, 2(3–4), 190–192. <https://doi.org/10.1007/s42113-019-00045-y>
- Thagard, P. (1993). *Computational philosophy of science*. MIT Press.

- Thagard, P., Findlay, S. (2012). *The cognitive science of science: Explanation, discovery, and conceptual change*. MIT Press.
- Trafimow, D., Earp, B. D. (2016). Badly specified theories are not responsible for the replication crisis in social psychology: Comment on Klein. *Theory & Psychology*, 26(4), 540–548. <https://doi.org/10.1177/0959354316637136>
- Vadillo, M. A. (2019). Ego depletion may disappear by 2020. *Social Psychology*, 50, 282–291. <https://doi.org/10.1027/1864-9335/a000375>
- Vadillo, M. A., Gold, N., Osman, M. (2016). The Bitter Truth About Sugar and Willpower: The Limited Evidential Value of the Glucose Model of Ego Depletion. *Psychological Science*, 27(9), 1207–1214. <https://doi.org/10.1177/0956797616654911>
- Van Rooij, I. (2008). The Tractable Cognition Thesis. *Cognitive Science*, 32(6), 939–984. <https://doi.org/10.1080/03640210801897856>
- Van Rooij, I., Baggio, G. (2021). Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science*, 16(4), 682–697. <https://doi.org/10.1177/1745691620970604>
- Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q. F., Finley, A. J., Ainsworth, S. E., [...] Albarracín, D. (2021). A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect. *Psychological Science*. (Sage CA: Los Angeles, CA). <https://doi.org/10.1177/0956797621989733>
- Wilson, M. (1993). DSM-III and the transformation of American psychiatry: A history. *The American Journal of Psychiatry*, 150(3), 399–410. <https://doi.org/10.1176/ajp.150.3.399>
- Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, 45, e1. <https://doi.org/10.1017/S0140525X20001685>
- Yarkoni, T., Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, 12(6), 1100–1122. <https://doi.org/10.1177/1745691617693393>
- Young, G. (2016). *Unifying Causality and Psychology*. Springer International Publishing. <https://doi.org/10.1007/978-3-319-24094-7>
- Zhang, J., Norman, D. A. (1994). Representations in Distributed Cognitive Tasks. *Cognitive Science*, 18(1), 87–122. [https://doi.org/10.1207/s15516709cog1801\\_3](https://doi.org/10.1207/s15516709cog1801_3)
- Zittoun, T., Gillespie, A., Cornish, F. (2009). Fragmentation or Differentiation: Questioning the Crisis in Psychology. *Integrative Psychological and Behavioral Science*, 43(2), 104–115. <https://doi.org/10.1007/s12124-008-9083-6>

---

## Podziękowania

Autor serdecznie dziękuje za uwagi Witoldowi Henslowi, Szymonowi Miłkosiowi, Przemysławowi R. Nowakowskiemu, a także dwóm recenzentom *Przeglądu Psychologicznego*, Gustawowi Nilsonne'owi i Svenowi Arendowi Ulptsowi za pomocne recenzje.