

ILUZE OBJEKTIVITY V PSYCHOLOGII

The illusion of objectivity in psychology

Mattias Desmet

Ghent University, Ghent, Belgie; mattias.desmet@ugent.be

Psychoterapie. 13 (1), 6–21; ISSN 1802-3983

ABSTRAKT:

Téměř po jednu dekádu je akademická psychologie v krizi. Sužovány metodologickou nedbalostí, replikace výzkumných poznatků příliš často selhávají a objevilo se i několik případů podvodu. Tento článek argumentuje, že jedna z hlavních příčin krize v psychologii je často nepovšimnuta a nachází se na úrovni metod měření. Zatímco metodické části nesčetného množství studií tvrdí, že validita použitých měřících nástrojů je „přijatelná“, „dobrá“ či dokonce „výborná“, každá podrobná analýza vede k hlubokému skepticismu. Po mnoho let použití čísel poskytuje psychologii auru vědecké přesnosti, ale při bližším prozkoumání spíše staví psychologii do rizika, že se stane pseudovědou. Tento článek ilustruje velmi hmatatelně, jakým způsobem ovlivňují proměnné, které nemají být měřeny, výsledky měření, a jak je kvůli nim většina typů statistického usuzování neefektivní. Shrnuje, že návrat k případovým studiím, přehodnocení narativního a kvalitativního popisu a paradigma zkoumání zaměřené na kvantifikaci formálních charakteristik jazyka, může lépe vyladit psychologické výzkumné metody na komplexní a dynamickou povahu předmětů jejich zkoumání. Spíše takto můžeme přispět ke skutečnému překonání replikační krize, než vkládat naděje do optimalizace klasických metod měření.

Klíčová slova: replikační krize, validita a reliabilita psychologického měření, nomotetické a idiografické měření

ABSTRACT:

For about a decade, academic psychology has been in crisis. Plagued by methodological sloppiness, the replication of research findings all too frequently fails and several cases of outright fraud have surfaced. This paper argues that one of the major causes of the crisis in psychology often goes unnoticed and is situated at the level of measurement methods. While the method sections of myriads of research papers claim that the validity of the measurement instruments used is ‘acceptable’, ‘good’ or even ‘excellent’, every thorough analysis leads to profound skepticism. For many years, the use of numbers has given psychology an aura of scientific exactitude, yet upon closer examination, it rather puts psychology at risk of becoming a pseudo-science. This paper illustrates in a very tangible way how a variety of variables that are not intended to be measured impact on measurement outcomes and render most types of statistical inference ineffective. It concludes that a return towards single case research, a re-appraisal of narrative and qualitative description, and a measurement paradigm centered on quantification of

formal characteristics of language might attune psychological research methods better to the complex and dynamic nature of its objects under study. Rather than pinning our hopes on the optimization of classical measurement methods, investing time and energy in such a reorientation might contribute to a true overcoming of the replicability crisis.

Key words: the replicability crisis, validity and reliability of psychological measurement, nomothetic and idiographic measurement.

Od druhé světové války se akademická psychologie téměř exkluzivně spoléhala na kvantitativní výzkum s co největšími vzorky, což mělo údajně přinášet „objektivní“ a „zobecnitelné“ poznatky. Ale více a více se ukazuje, jak problematický tento výzkum je. Téměř před deseti lety, po publikování několika zpráv o širším nedostatku replikability vědeckých poznatků, rozsáhlé metodologické nedbalosti a sérii případů zřejmého podvodu (například Allison, Brown, George a Kaiser, 2016; Baker 2016), se psychologie ponořila do krize. Rozsah této krize byl takový, že někteří významní výzkumníci, kteří ji studovali, jako například Ioannidis (2005), došli k závěru, že „většina publikovaných vědeckých poznatků je chybná“. Systematická rešerše Fanelliho (2009) došla ke konzervativnímu odhadu, že až v 72 % publikovaných výzkumů byly poznatky nějakým způsobem záměrně zkresleny a Baker (2016) ukázal, že neúspěšné replikace jsou spíše pravidlem než výjimkou. Je jistým způsobem ironické, ale nezbytné zmínit, že i tato čísla se podstatně různí napříč publikacemi. Jinými slovy: po-

znatky o omezené replikabilitě samy postrádají replikovatelnost. To samo o sobě možná nejlépe ilustruje hloubku této krize.

Při procházení literatury analyzující tuto krizi stojí za povšimnutí, že příčiny jsou téměř vždy přičítány nedostatku přísnosti při aplikaci vědeckých metod, například v takových faktorech jako ledabylost vý-

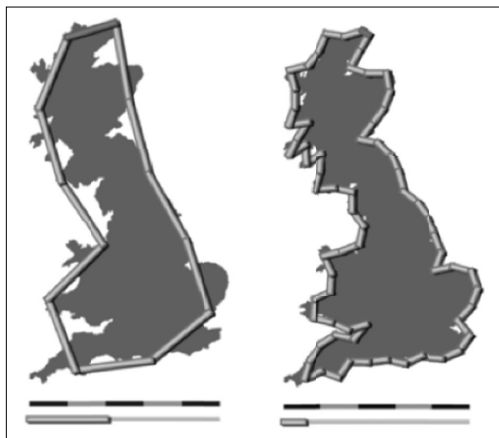


Mattias Desmet

Je profesorem klinické psychologie na Fakultě psychologie a pedagogiky na Univerzitě Ghent v Belgii. Jeho nejznámější knihou je „The Pursuit of Objectivity in Psychology and Lacan’s Logic of Subjectivity: A Walk on the Graph of Desire“.

zkumníků nebo neochota periodik publikovat replikační studie. Mně se ale navzdory tomu zdá, že je také problém v metodách samotných. Alespoň jeden důležitý přispívající faktor, jímž je problematická validita psychologických měření, tedy zůstává téměř naprosto nerozpoznán. V nejlepší případě se v člancích analyzujících krizi uvádí tento faktor na okraji (například Asendorpf et al., 2013, s. 111, viz „zvýšte reliabilitu měření“).

V tomto článku budu polemizovat jak z logicko-konceptuálních, tak z empirických důvodů, že na první místě je to právě



Obrázek 1. Naměřená délka pobřeží Británie je přibližně 2 400 km s měřítkem 200 km (levá strana obrázku) a 3 400 km (cca) s měřítkem 50 km (pravá strana obrázku). Obrázek získaný od Van de Sande (2004)

problém měření, co kompromituje vědecký status psychologie, a že není přehnané říct, že budoucnost akademické psychologie jednoznačně záleží na tom, jak úspěšně či neúspěšně si s ním poradí. Argumentují, že jediný způsob, jak vyřešit tento problém, je navrátit se k případovým studiím a interpretovat měření v kontextu narativních popisů pacientova globálního subjektivního fungování. Argumentační linie prezentovaná v tomto článku je stručným souhrnem mé nedávno publikované knihy nazvané Hledání objektivit v psychologii (Desmet, 2018).

LOGICKO-KONCEPTUÁLNÍ REFLEXE

Dovolte mi začít své úvahy zdůrazněním faktu, že měření je typ popisu, neboli *deskripce*. Deskripce se dá shrnout jako přeložení „surových“ fenoménů či „objektů“ našeho zkoumání do výzkumných dat, a toho lze docílit různými způsoby.

V prvé řadě je zde *narativní* deskripce, ve které člověk popisuje objekt zkoumání

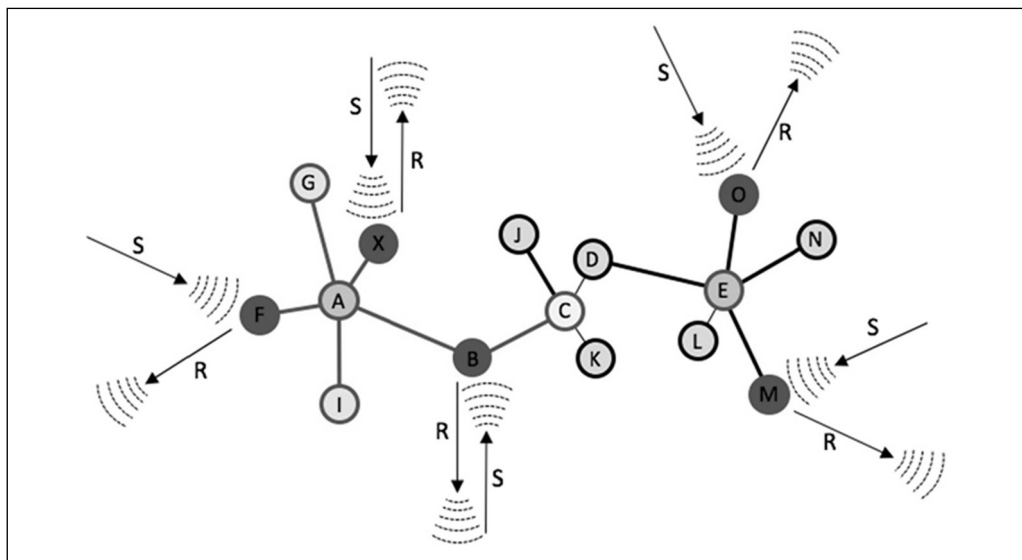
pomocí slov. Ačkoliv na první pohled postrádá úroveň exaktnosti a sofistikovanosti požadovanou ve vědě, je jasné, že hrála významnou roli v podstatě v každém velkém vědeckém úspěchu, ať už vezmeme práci Koperníka (1543), Linné (1753), Newtona či kteréhokoliv jiného přelomového myslitele.

Za druhé, můžeme popisovat s pomocí obrazů. Historické příklady zahrnují například detailní kresby lidské anatomie Vesalia (1543), kresby a fotografie zvířat v práci Tinbergena (například 1948), a rentgenové krystalografické obrázky (Wilkins a Randall, 1953), na základě kterých Watson a Crick (1953) objevili dvojitou šroubovici DNA.

Zatřetí máme *kvantitativní* deskripce, neboli *měření*. Měření objektu se dá shrnout jako jeho situování na jednodimenzionální škále, která sestává z jednotky měření násobené reálným číslem (Joint Committee for Guides in Metrology, 2012). To implikuje, přísně vzato, že pouze jednodimenzionální vlastnosti objektu mohou být měřeny. Například, nemůžeme měřit *auto an sich*, neboli s pomocí auta samého. Musíme napřed určit jednodimenzionální vlastnost, jakou je například délka či šířka, zvolit stabilní, neměnnou jednotku měření, a potom sečíst počet jednotek, který má vlastnost objektu, již měříme. Důležité je, že i když se s použitím selského rozumu zdá, že měření poskytuje kvazi-absolutní zhodnocení objektu, ve skutečnosti je velmi relativní k použité jednotce. Mandelbrot (1967) toto ilustroval velmi uchopitelným způsobem ve svém článku zvaném „Jak dlouhé je pobřeží Británie?“ (Obrázek 1). Když se podíváme na pobřeží Británie, nevidíme rovnou čáru, dokonce ani hladkou, zakřivenou čáru, ale vlnitou, meandrovou čáru. Když ho měří-

me v jednotkách 200 kilometrů, jak lze vidět na levé části snímku, délka této čáry je 2 400 kilometrů, zatímco při jednotce 50 kilometrů je délka 3 400 kilometrů. Mandelbrot zde snadno ukazuje, že „s tím, jak je použita škála menší, měřená délka pobřeží neomezeně roste“ (Gleick, 1987, s. 96). Richardson (1961) doložil, že toto skutečně vede k praktickým problémům. Prošel encyklopedie

vidíme u slunečnic, mušlí, tornád a galaxií. První tři typy deskripce se ukazují být relativní a nepřesné – narativní deskripce může být provedena mnoha různými kombinacemi slov, fotografie či obrázky mohou být provedeny z mnoha různých perspektiv a měření vždy závisí na použité škále. Nicméně se zdá, že matematická deskripce se blíží absolutnu. Matematická forma a její



Obrázek 2. Síťová koncepce psychického aparátu

ve Španělsku, Portugalsku, Nizozemsku a Belgii a zjistil, že uvedená délka společných hranic se různí zhruba o 20 % kvůli rozdílným škálám, které byly použity.

Za čtvrté, kromě narativní, obrazové a metrické deskripce máme ještě deskripce matematickou (např. Prusinkiewicz a Fowler, 1995, s. 163), která je často přehlížena nebo zaměňována s měřením. Tento typ deskripce sestává z určení vzorce daného jevu. Například Fibonacciho posloupnost určuje spirálovitou křivku, která ukazuje poměr zlatého řezu a odpovídá přesně řadě křivek v přírodě, například těm, které

určující vzorec jsou unikátní a neměnné. To dává matematické deskripce speciální auru a postavení, připomínající Platonovu teorii univerzálních forem, jako by toto bylo tím posledním doladěním, nasazující korunu vědecké práci.

To, který typ deskripce je vhodný, záleží na charakteristice objektu, který se snažíme popsat. Když se vrátíme zpět k psychologii, měli bychom si klást otázku: Jaké jsou objekty psychologie a jaké vlastnosti mají?

Na psychické objekty můžeme pohlížet z mnoha různých úhlů pohledu, ale za účelem této úvahy se budu odkazovat na síťový kon-

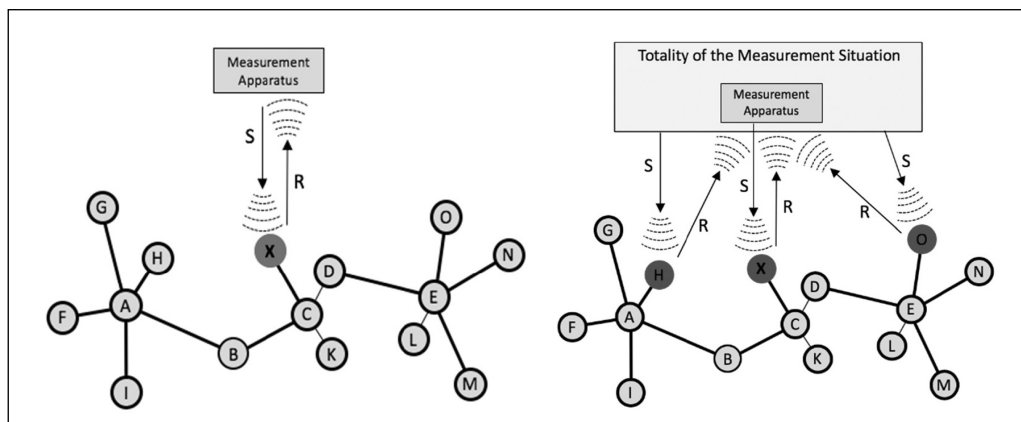
cept (Obrázek 2). Psychický systém může být koncipovaný jako síť navzájem propojených lingvistických uzlů, které jsou aktivovány celou řadou spouštěčů. Když spouštěč aktivuje uzel, aktivace se šíří po síti a aktivuje sérii dalších uzlů. Skrz tento proces je význam přiřazený ke spouštěči velmi komplexním a pro každého jedince specifickým způsobem. Například to, že vidíme králíka, může spustit řetězec asociací, které vedou ke vzpomínkám na to, jak jsme měli vlastního králíka jako děti, nebo na pitvu během hodiny biologie, ale také může vést k přezdívice, kterou měla naše přítelkyně, nebo k sousedovi, který zabíjel králíky krutým způsobem.

Proces šíření aktivace a tvorby významu je nejen komplexní a pro každého specifický, ale také vysoce nestálý a nepředvídatelný. Je například zřejmé, že drobné rozdíly v úrovni přichodícího stimulu mohou vyvolat radikální změny v úrovni vzniku významu. My všichni to známe z každodenního života: i jen lehký úsměv nebo drobné nervózní zakašlání může radikálně změnit celý způsob, jakým člověk interpretuje cizí smýšlení.

Můžeme se nyní, na základě těchto výchozích úvah, podívat na deskriptivní metody v psychologii. Začneme tím, že se podíváme na historický rozpor mezi nomotetickým a idiografickým paradigmatickým. Před druhou světovou válkou bylo nomotetické paradigma dominantní v akademické psychologii v Americe; idiografické v kontinentálním akademickém prostředí, zejména v Rakousku a Německu. Během nacistického režimu idiografické paradigma ustoupilo a nomotetické paradigma rychle nabralo na síle a stalo se dominantní v celém Západním akademickém světě. Toto změnilo povahu akademické psychologie a zároveň také metody vědecké deskripce.

Idiografické paradigma považovalo metrickou deskripci za nápomocnou, ale jen pokud byla měření situována v kontextu idiografických informací v každém v každém jednotlivé případě. Vezměme například tento citát Williama Sterna (1921, s. 3–4, citováno podle Lamiel, 2009, s. 72), vynálezce inteligentního kvocientu: „*Pro dotyčného zkoušeného testy poskytnou číslo, na jehož základě může být zkoušený umístěn někde na kvantitativní škále, ale které skrývá věci kvalitativně zvláštní pro daného jedince. Výsledky přímého pozorování nelze kvantifikovat, ale umožňují kvalitativní zpřesnění psychologického profilu. Ze všech těchto důvodů musí být metody přímého pozorování zkoušeného vždy používány k doplnění zkušebních metod a první musí být vyvinuto s tou stejnou péčí jako to druhé.*“ Tato základní myšlenka byla široce sdílena valnou většinou vědců před druhou světovou válkou, včetně předních behavioristů jako byli Watson, Skinner a Pavlov. Poslední zmiňovaný například tvrdil, že kvantitativní data o slinění jeho psů by byla k ničemu a nemohla by být správně interpretována, kdyby člověk nebyl obeznámen s celkovou povahou každého jednotlivého psa (Todes, 2014).

Nomotetické paradigma zkoumalo psychologické procesy naprosto odlišným způsobem. V prvé řadě spoléhalo téměř výhradně na metrickou deskripci; narativní deskripce byla považována za méně „objektivní“ a méně „přesnou“. Za druhé, v zájmu kontroly výzkumníkovy subjektivity byl proces měření standardizován tak přísně, jak to bylo možné. Za třetí, aby se zaručila zobecnitelnost poznatků, měření bylo aplikované na tak velkém vzorku, jak jen bylo možné. Začtvrté, protože souhrnné idiogra-



Obrázek 3. Klasické explicitní a implicitní metody měření vycházejí z předpokladu, že izolované testovací stimuly aktivují izolované aspekty asociativních sítí participantů (vlevo); logicko-konceptuální uvažování naznačuje, že postupy měření se skládají z komplexní a variabilní sady podnětů, které aktivují složitou a proměnlivou sadu aspektů asociativní sítě

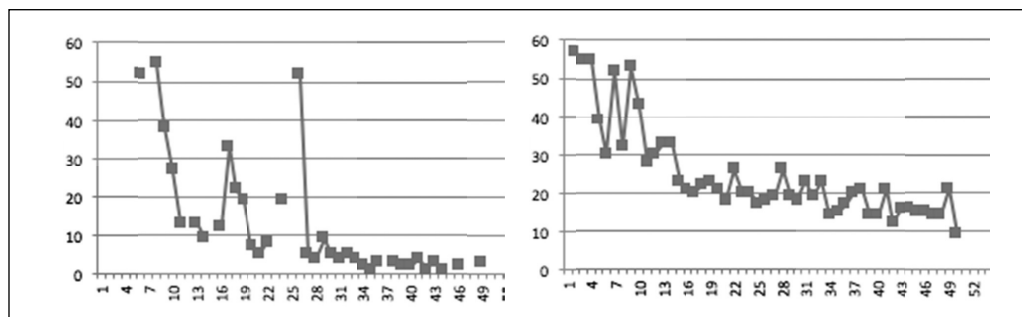
fické deskripce jediného subjektu mohly být provedeny pouze narativním způsobem, byly považovány za nevyhnutelně subjektivní, a byly proto zcela vynechány. Měření proto přestalo být interpretováno v kontextu narativních dat o jednotlivých subjektech, jak se dělo v idiografickém paradigmatu. Namísto toho bylo interpretováno standardizovaným způsobem, který přepokládal, že ten samý výsledek měření znamená totéž u různých participantů.

Jinými slovy: nomotetické nástroje měření – jako například Q třídění, implicitní asociační testy, hodnocení výzkumníka a dobře známé sebeposuzovací dotazníky – všechny vycházejí z obecného předpokladu, že testy měří jeden neměnný objekt, jednu psychickou charakteristiku, kterou zamýšlíme měřit, která neměnně způsobuje měřený výsledek neohledně na různé kontexty, různé participanty, různé body v čase, v nichž je test administrován, atd. Toto můžete vidět na levé straně Obrázku 3. Test spouští aktivitu v jedné psychické charakteristice, a reakce na test následně vychází z této charak-

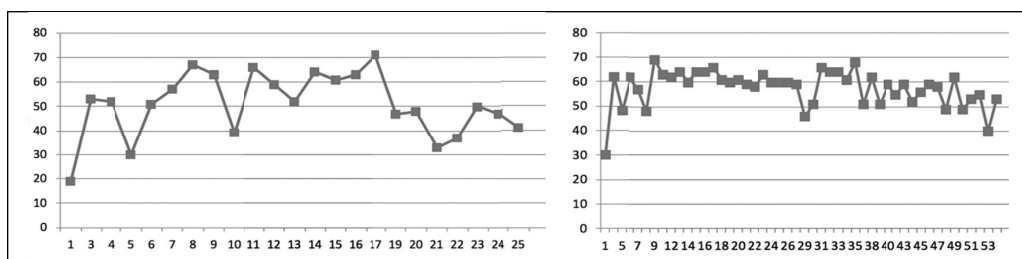
teristiky. Je zřejmé, že jde o silný předpoklad, který je do značné míry neslučitelný se složitou a dynamickou povahou procesů vzniku významu. Spíše očekáváme situaci, která je znázorněna na pravé části Obrázku 3, konkrétně to že aktivace šíří se v síti jako důsledek testovacích podnětů aktivuje velmi komplexní a proměnlivou sadu uzlů, která nevyhnutelně způsobí, že test bude citlivý k mnoha faktorům, jež nejsou určeny k měření.

EMPIRICKÉ ARGUMENTY

Nejsme první, kdo vznesl takové podezření. V empirickém výzkumu bylo hojně ukázáno, že stejný postup měření může u různých participantů měřit naprosto různé věci. Mohl jsem se zde zmínit psychometrické studie, ale místo toho předvedu některé konkrétní příklady z mého vlastního výzkumného programu na univerzitě v Ghentu. Můj výzkumný tým studuje psychoterapeutické postupy a náš postup shromažďování údajů je takový, že umožňuje analýzu jak na úrovni vzorku, tak na úrovni



Obrázek 4. Skóry dvou pacientů v nástroji Inventory of Interpersonal Problems, která naznačují zlepšení interpersonálního fungování v průběhu psychodynamické psychoterapie



Obrázek 5. Skóry dvou pacientů v nástroji Inventory of Interpersonal Problems, která naznačují zhoršení interpersonálního fungování v průběhu psychodynamické psychoterapie

jednotlivých pacientů (podrobný popis naší strategie shromažďování dat viz Meganck et al., 2017). Během terapií pacientů se shromažďují kvantitativní údaje o procesu a výsledcích a současně se pořizují audionahrávky všech terapií, aby bylo možné provést i kvalitativní analýzu. Tohoto projektu jsem se účastnil jak jako výzkumník, tak jako terapeut a tento postup jsem použil na terapii 30 pacientů ve své vlastní klinické praxi.

Obrázek 4 znázorňuje vývoj skóre dvou mých pacientů v průběhu 50 terapeutických sezení v testu Inventory of Interpersonal Problems (Horowitz, Alden, Wiggings a Pincus, 2000), dotazník, jehož celkový skóre měří obecné interpersonální potíže a obavy. Grafy obou pacientů jsou velmi podobné, ukazují v obou případech postupně klesající křivku, skóre v IIP je asi 60 na počátku léčby

a přibližně 0 na konci. Grafy tak naznačují, že míra distresu postupně během léčby klesá, což naznačuje, že oba pacienti prošli srovnatelnou a úspěšnou terapií.

Analýza narativních informací z audionahrávek terapie však ukázala, že proces nebyl vůbec srovnatelný. U prvního pacienta, který je znázorněn na levé straně obrázku, narativní údaje potvrdily skutečně zlepšené interpersonální fungování. Zatímco vztahový život tohoto pacienta souhrnně vykazoval na začátku léčby destruktivní vzorec, audionahrávky sezení i následné rozhovory s externím výzkumníkem potvrdily, že do konce terapie začal nový, uspokojivý romantický vztah, že začal lépe vycházet se svými kolegy a přáteli a že ostrá agrese a nenávisť vůči rodičům se zmírnila.

Zcela odlišný příběh se však odehrál v případě druhé pacientky, reprezentované na pra-

vé straně obrázku. Na počátku terapie tato pacientka uvedla, že je ve velmi frustrujícím interpersonálním prostředí, pokud jde o vztahy s jejím partnerem, přáteli a rodinou. Během terapie jsem neměl dojem, že by se věci zlepšily. Naopak, stále více a více se izolovala od jiných lidí, většinu času se stahovala do své studentské místnosti. Když skončila terapie, vykreslil jsem si graf prezentovaný na obrázku a byl jsem velmi překvapen, když jsem viděl klesající trend v dotazníkových skórech, který naznačuje, že se její úroveň vztahů podstatně zlepšila. Nejprve jsem se radoval, ale pak jsem začal být podezřívavý a rozhodl jsem se, že se jí přímo zeptám, proč v IIP uváděla nižší míru interpersonálních potíží.

Abych věci trochu přiblížil, zmíním se, že IIP typicky obsahuje položky jako „Je těžké pro mě věřit jiným lidem“, „Je těžké pro mě neříkat ne ostatním lidem“, „Je těžké abych se připojil k ostatním lidem“, „Je těžké pro mě někomu říct, aby mě přestal obtěžovat“ atd. Když jsem se jí zeptal, co myslí tím, že odpovídá téměř nulově na všech těchto položkách, odpověděla, že už nemá problémy se všemi těmito věcmi, *protože už se s lidmi prostě nesetkává*. „Je pro vás těžké důvěřovat jiným lidem?“ „Ne, já už nemusím věřit ostatním lidem, protože už s nimi nejsem v kontakt.“ „Je pro vás těžké říkat ne ostatním lidem?“ „Vůbec ne, už nejsem konfrontovaná s žádostmi jiných lidí.“ Je tedy zřejmé, že zatímco výsledek postupu měření, jak je znázorněn na grafech, byl u obou pacientů stejný, psychická realita, kterou má znázorňovat, byla zcela odlišná. Obrázek 5 nám poskytuje další zajímavý příklad. Stejně jako v předchozích grafech ukazuje vývoj ve skóre IIP u dvou pacientů v téže studii. Pokud prostě považujeme

validitu měření za daný fakt, dospějeme k závěru, že oba pacienti na konci terapie vykazovali více interpersonálních potíží než na začátku, a proto se léčba u obou pacientů nezdařila. Audionahrávky nicméně jasně ukázaly, že tomu tak není. Oba pacienti byli zasláni na psychoterapii praktickým lékařem, protože měl v jejich případech podezření, že série jejich somatických stížností souvisí se záchvaty paniky a úzkostí. Během prvních sezení se mnou oba odmítli souhlasit s tím, že jejich somatické příznaky mají něco společného s psychologickými problémy, a proto v dotaznících skórovali nízko. Během následujících deseti sezení však postupně začali připouštět, že jejich somatické symptomy nějak souvisí s jejich psychickými problémy, a proto se skóre dotazníků začalo zvyšovat. Následně v dalších sezeních, s tím, jak se uváděné problémy staly centrem terapeutické práce, se skóre pomalu začalo snižovat a oba pacienti se na konci terapie cítili podstatně lépe než na začátku. Dotazníkový skóre však nikdy neklesl tak nízko jako na začátku terapie. Bez ohledu na to, jak úspěšná je terapie, zřejmě nemůže soupeřit s mechanismem popření, pokud jde o snížení skóre stížností!

Taková pozorování jsou pro nomotetickou psychologii velmi problematická a v našich datech nebyla vůbec výjimečná. Ukazují přímým způsobem, že musíme pochopit příběh pacienta, jak jej pacient vypráví během sezení, abychom pochopili, co měření znamená. Měření nejsou jakousi nadřazenou informací, která by odhalila něco, co jsme se nedozvěděli prostým posloucháním našeho pacienta, ne, je to spíše naopak: potřebujeme jednoduše naslouchat našim pacientům, abychom věděli, co jejich výsledky měření znamenají. Zdá se tedy, že badatel, který

používá nástroje psychologického měření, často připomíná dítě, které hraje s dětským mikroskopem: dítě zamyšleně kouká do přístroje, ale když si myslí, že se nikdo nedívá, rychle se podívá pod mikroskop, aby mohlo skutečně vidět, na co se dívá.

Mohl bych poskytnout nekonečnou řadu konkrétních příkladů empirického výzkumu, které ukazují, že měření nelze interpretovat standardizovaným způsobem. Měření jsou citlivá na četné faktory, které nejsou cíleně měřené. Všechny tyto faktory ovlivňují měřicí přístroj různými způsoby při různých zkušebních příležitostech a jako takové narušují ideál standardizovaného měření.

Po vyslechnutí tohoto všeho může zastánce nomotetického paradigmatu pokrčit rameny a odvětit: „No samozřejmě, co jste očekávali? Žádné měření není perfektní. Jen se ujistěte, že vzorky jsou dostatečně velké a všechny tyto chyby se vyrovnají.“ Rád bych odpověděl: „To záleží, pokud je chyba měření příliš vysoká, pak žádný vzorek nemůže být dostatečně velký.“

Studie o shodě napříč různými metodami umožňují odhadnout, kolik chyb je v nomotetických měřeních a jaký to má dopad na statistické závěry. Takové studie vycházejí z prosté, ale zajímavé otázky: Pokud změříte stejnou psychologickou charakteristiku pomocí různých měřících metod – např. sebesposuzovacího dotazníku, hodnocení výzkumného pracovníka a implicitního měření – jak se budou výsledky podobat? Metastudie ukazují, že shoda mezi různými metodami v psychologii, vyjádřená v Pearsonových korelacích, se pohybuje někde mezi 0,00 a 0,45.

Dovolte mi, abych zkusil tyto korelace více přiblížit. Předpokládejme, že tesař měří okna vašeho domu pomocí tří různých měření. Například skládacím metrem,

svinovacím metrem a laserovým měřením vzdálenosti, a uvádí sérii měření v tabulce 1.

Nemyslím si, že byste chtěli, aby vám tento tesař dělal okna ve vašem domě. Nicméně konvergence mezi jeho měřeními je 0,45, což se zhruba rovná horní hranici shody mezi metodami v psychologickém výzkumu. Někteří vědci interpretují nedostatek konvergence ani ne tak jako indikátor chyby měření, ale spíše jako indikaci, že různé metody měří *různé aspekty* psychických charakteristik (Meyer et al., 2001, s. 145). Zatímco tato interpretace může být za určitých podmínek přesná, nic se zatím nezměnilo na tom, že obdržela malou nebo žádnou empirickou podporu. Nikdy například nebylo prokázáno, že konvergence je vyšší u měření, která posuzují více specifické aspekty atributu.

Pozorování, že měřicí metoda má dopad na výsledek měření, sama o sobě není typická pro psychologii. Bell (1987, s. 166) uvedl, že měření v kvantových pokusech „*musí být považováno za společný výrobek „systému „a [měřicího] aparátu“*“. Toto bylo také ilustrováno příkladem Mandelbrota, o kterém jsme se právě zmínili, který ukazuje, že měřená délka britského pobřeží se liší, když se používají různá měřítka.

A přesto je v psychologickém výzkumu něco zvláště problematického. V psychologii výzkumy zaměřené na test-retestovou reliabilitu ukazují, že měření se dokonce i při stejném měřítku výrazně liší. Test-retestová realibilita explicitních osobnostních měření se pohybuje například v rozmezí 0,80 až 0,50 (např. Gnambs, 2014; Viswesvaran a Ones, 2000); implicitní měření osobnosti mezi 0,30 a -0,10 (např. Bosson et al., 2000, str. 636; Asendorpf, Banse a Mücke, 2002, s. 382). Podle vzor-

Tabulka 1

Tři série truhlářova měření stejného okna, které korelují 0,45

	Měření získána Metodou 1	Měření získána Metodou 2	Měření získána Metodou 3
Okno 1	71 palců	51 palců	24 palců
Okno 2	39 palců	79 palců	59 palců
Okno 3	63 palců	87 palců	51 palců
Okno 4	39 palců	67 palců	83 palců
Okno 5	12 palců	39 palců	8 palců
Okno 6	47 palců	31 palců	63 palců
Okno 7	43 palců	59 palců	24 palců
Okno 8	12 palců	35 palců	4 palců

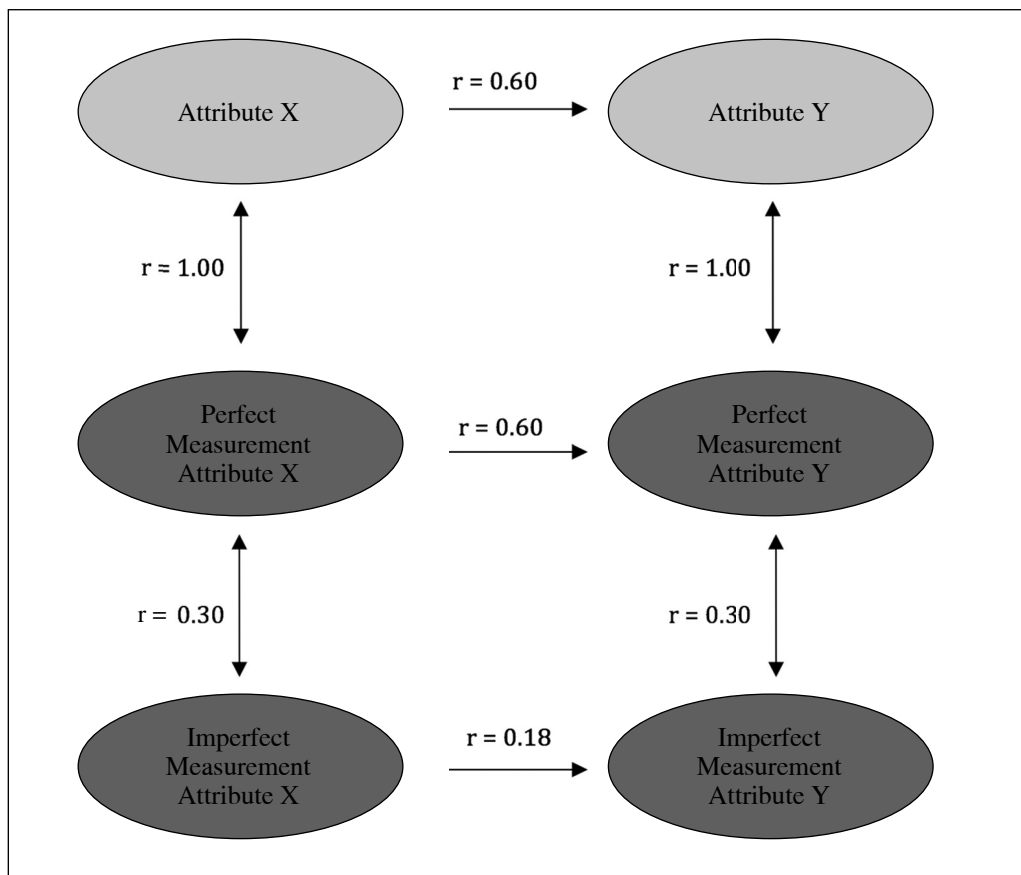
ce, že vysvětlený rozptyl se rovná korelaci na druhou, reliability testu-retestu tak nízké jako 0,30, ukazují, že měření obsahují alespoň 90% chyby. To nám přináší velice pesimistický závěr týkající se validity.

Všechny uvedené studie jsou zveřejňovány ve vysoce hodnocených periodických, což ještě ztěžuje pochopení, proč většina výzkumníků ve svých výzkumech tvrdí, že validita použitých nástrojů je přijatelná, dobrá, nebo dokonce vynikající. Při bližším zkoumání se ovšem ukazuje, že se v psychologii stalo zvykem hodnotit validitu s odkazem na *validitu průměrného měření*. Validita testu se tedy považuje za „přijatelnou“, pokud je validní jako průměrný test. Je zřejmé, že se jedná o pochybný zvyk, zejména v oblasti, kde je průměrná validita poněkud špatná. Hodnocení by se mělo provádět na základě matematické normy vztahující se k dopadu chyby na statistickou inferenci a dosažené závěry. Dnes se však takové hodnocení odehrává pouze při posuzování chyby výběru (kde je vypočítána síla testu); chyba měření a nedostatek validity jsou skutečně zmíněny, ale poté už se jim nikdo nevěnuje. To vyvolalo ironickou poznámku Schmidta a Huntera (1996, s. 199), zda je „pravda, že vědci

stačí pouze se zmínit o chybě měření, aby zmizela“.

Jednou jsem provedl matematický test o dopadu chyby. Již v roce 1904 varoval Spearman výzkumnou komunitu o tomto problému a vytvořil svůj slavný vzorec umožňující vypočítat vliv chyby měření na korelační výzkum. Použil jsem jak tento vzorec, tak matematickou simulační studii, abych ukázal, že v psychologii bude skutečná korelace dosahující hodnot 0,60 pozorována pouze jako nevýznamná korelace 0,18 kvůli chybě měření. Obrázek 6 toto znázorňuje schematickým způsobem. Nebudu zde zacházet do podrobností, jelikož je to čistě matematická záležitost, a z téhož důvodu nebudeme zkoumat důvody, proč si myslím, že strategie vyvinuté v psychologii k řešení chyby, jako je modelování strukturálních rovnic, z velké části selhávají. Zájemci o podrobnější diskusi mohou číst Desmet (2018).

Pozorování, že zjištěná korelace se ani zdaleka neblíží skutečné korelaci, ukazuje, že je naivní myslet si, že dopad mnoha zdrojů chyb rozptylu bude vyvážen pouze tím, že budou vzorky dostatečně velké; pokud je chyba měření příliš vysoká, statistická inference prostě nefunguje. To je jen logické.



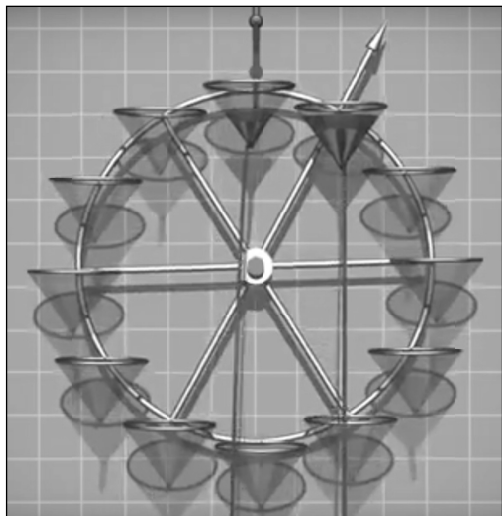
Obrázek 6. Vztah mezi skutečnými a pozorovanými korelacemi mezi atributy X a Y, v případě, že se spoléháme na nedokonalé měření, které koreluje 0,30 s dokonalými měřeními atributů X a Y

To všechno dokládá něco, co Kalton a Schuman uvedli před třiceti lety. [V psychologii] „selhání toho, aby data odpovídaly teorii, obvykle zpochybňuje měřící nástroj stejně, jako teorii“. A naopak: „i kdyby se pozorované vztahy shodovaly s jejich teoretickými předpověďmi, tato shoda není jasným potvrzením... může to být například artefakt souboru použitých měřicích přístrojů – „efekt metody““ (Kalton & Schuman, 1982, s. 43). Jinými slovy, psychologická měření obsahují příliš mnoho chyb, aby bylo možné je použít k testování teorie.

ŘEŠENÍ

Problém měření skutečně odhaluje širší problém v akademické psychologii, který spočívá v tom, že mezi povahou zkoumaných předmětů a metodami výzkumu existuje obecný nesoulad. Náš popis psychického aparátu jako sítě, ve které se objevuje význam, nám ukazuje, že psychický systém je v každém ohledu komplexním dynamickým systémem. Takové systémy, díky vlastnostem jako je citlivost k počátečním podmínkám, splývání mezi systémem a prostředím a hierarchické organizaci

a historičnosti, nelze nikdy popsat výlučně kvantitativně. Jak jsem poznamenal na začátku této přednášky, abychom měřili plodným způsobem, musíme zredukovat objekt na jednorozměrné charakteristiky; kom-

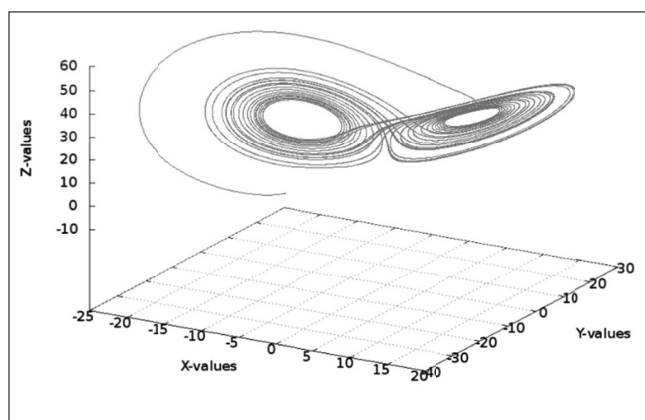


Obrázek 7. Lorenzovo chaotické vodní kolo (obrázek získaný od Leys, Ghys a Alvarez, 2003)

plexní dynamické systémy jsou však ve své podstatě vícerozměrné. To lze pozorovat velmi konkrétně v situaci měření. Například: Pokud konfrontujete particpanta s položkou na dotazníku, která se ptá: „Cítíte se depresivně?“, odpověď na tuto položku nejenže odráží „množství deprese“ v této osobě, ale také do jisté míry celou v průběhu života vytvořenou síť účastníka, která vstupuje do hry při interpretaci této položky (Rosenzweig, 1933, str. 337). Bez ohledu na to, jaké testovací podněty se používají k měření jednoho jednorozměrného aspektu systému, člověk také nevyhnutelně měří systém ve své celistvosti. Proto jakýkoliv důkladný popis ukazuje, že v určitém okamžiku musí být popsán celý systém a jediný způsob, jak to udělat, je narativní

způsob, a to na úrovni jediného případu. To je důvod, proč klinici mohou a musí hrát zásadní roli v psychologickém výzkumu: bez klinických znalostí je téměř nemožné psát takové vyprávění o subjektivitě. Částečně zde zmiňuji, že toto vše platí také pro komplexní dynamické systémy v jiných oborech, jako je ekonomika, biologie, chemie nebo fyzika. Také v těchto oblastech jsou komplexní dynamické systémy zkoumány případ od případu, vždy s použitím narativního popisu po boku metrického popisu. To činí ještě pozoruhodnějším fakt, že psychologie, která se zabývá komplexními dynamickými systémy par excellence, se rozhodla výhradně spoléhat na metrický popis na úrovni vzorků.

Kromě interpretace měření narativním způsobem na úrovni jednotlivého případu, jsem přesvědčen, že celá metodika měření by měla být koncipována jiným způsobem. Je nad rámec této přednášky proniknout do všech detailů. Pro ty, kteří by měli zájem, odkazuji znovu na svou knihu (Desmet, 2018), kde navrhuji, že kvantifikace naturalistického jazyka, jakým se mluví například v terapeutických sezeních, může nabídnout lepší perspektivu než typy měření, které se nyní používají. Jakožto konečný krok vědeckého procesu lze kvantifikaci jazyka popsat matematickým způsobem. Matematický popis je v psychologii zřídka vyhledávaný. Edward Lorenz, americký matematik a klimatický vědec, toto ilustroval na pěkném příkladu s jeho takzvaným chaotickým vodním kolem. Jeho virtuální příklad vidíte na obrázku 7. Zatímco chování kola je zcela chaotické, bez jakékoliv pravidelnosti. (Což je docela úžasné, i když necháte kolo otočit tisíc let, nikdy přesně nezopakuje stejný pohyb, jako kdyby si pamatovalo všech-



Obrázek 8. Lorenzův atraktor (obrázek získán z Reed, 2014)

na svá minulá chování.) Nuže, bez ohledu na to, jak chaoticky toto kolo se chová, Lorenz (1963) ukázal, že je nicméně striktně určen sadou tří nelineárních rovnic a že toto určení může být viditelné reprezentováním řešení rovnic v tom, co matematici nazývají “fázový prostor”. Skutečně ohromujícím způsobem se objevila esteticky velkolepá matematická figura, která se stala známou jako Lorenzův atraktor (viz Obrázek 8). Gleick (1987, s. 135) říká o fázovém prostoru a matematických objektech v něm: „*Fázově prostorové obrazy fyzických systémů odhalily vzorce pohybu, které byly jinak neviditelné, tak jako infračervená krajinná fotografie může odhalit vzory a detaily, které existují mimo dosah vnímání.*“

Stejně jako Lorenz by měli psychologové mít důvěru, že také pod chaosem manifestní subjektivní zkušenosti je možné nalézt přísný řád a že tento řád může být vizualizován ve fázovém prostoru.

ZÁVĚR

Bylo by fascinující pokračovat dále, ale je na čase, abych došel k závěru. Abych

shrnuł tuto přednášku: Souhlasím s Toomelou (2007, s. 19), že „*posledních 60 let výzkumu v psychologii, zdá se, zbloudilo z cesty*“. Myšlenka na dosažení „objektivního“ či „zobecnitelného“ poznání prostřednictvím nomotetického měření se ukázala jako velmi problematická. Pokus nomotetické psychologie kontrolovat subjektivitu výzkumného pracovníka výlučně spoléháním se na standardizovaný metric-

ký popis vzorků na úkor narativního popisu jednotlivých případů selhal. Dobře zdokumentovaná literatura o nedbalosti ve výzkumu a podvodu v psychologii odhaluje, že subjektivita výzkumných pracovníků byla spíše zcela rozpoutána. Jediná subjektivita, která byla skutečně vyloučena z návrhů výzkumu, je subjektivita účastníků, tj. to, co mělo být předmětem vědecké psychologie. Souhlasím s Valsinerem (2012, s. 203, citováno ve Schwarz, 2014, s. 235), že v psychologii „*Nemůže být dosaženo objektivity jinak než přes subjektivitu*“. Komplexní a dynamická povaha subjektivity klade specifické požadavky na metodologii výzkumu, a kromě kvantitativního popisu vyžaduje také narativní a matematický popis. Souhlasím také s Molenaarem a Valsinerem (2005, s. 2), že psychologická věda by měla pokročit především zkoumáním jednotlivých případů: „*Takže – existují dvě cesty k zobecněnému poznání [v psychologii] – tradiční... nomotetická [cesta]... Za jedinečností jsou základní univerzální procesy, které musí být objeveny. Proto zdůrazňujeme novou značku univerzální vědy, která se zabývá*

jednotlivými, samostatnými, autonomními, na souvislostech závislými systémy v biologii, psychologii, sociologii a antropologii. Jednotlivé případy – systematicky analyzovány – jsou primárními empirickými předměty pro dosažení obecných poznatků.“ Hledání řešení této metodologické situace je ještě naléhavější, pokud se zamyslíme nad zřejmým významem psychologie jako vědy. Z mnoha hledisek by mohla být nadcházející éra érou psychologie. A to není zbožné přání psychologa; spíše vědců z jiných oborů, převážně v první polovině dvacátého století, kteří narazili na význam psychologie. Merleau-Ponty a Husserl jako filozofové, Mauss jako etnolog, Lévi-Strauss jako antropolog a další, všichni popsali hluboký dopad subjektivní zkušenosti na jevy, které byly dříve koncipovány pouze z hlediska materialistického mechanismu; jevy jako vnímání, fyziologické reflexy, bolest a motorické chování. A dopad psychické zkušenosti se ještě více rozšiřuje. Kvantová fyzika prokázala, že akt pozorování, který nevyhnutelně musí být koncipován v psychologických termínech, určuje umístění subatomických částic v prostoru a čase. Materiální jevy se ukázaly být výrazně ovlivněny subjektivitou, nebo, jak uvedl Bertrand Russell v roce 1922, ještě před hlavními objevy kvantové fyziky: „*Všechny naše údaje, a to jak ve fyzice, tak v psychologii, podléhají psychologickým příčinným zákonům... V tomto ohledu je psychologie blíže tomu, co skutečně existuje.*“ Kromě toho, vsudypřítomnost účinků placebo v lékařských zákrocích a použití hypnotické anestézie v chirurgii, někdy v tak drastickém zásahu jako operace s otevřeným srdcem, jednoznačně ukazují, že psy-

chologie není jen intelektuálně fascinující věc, ale že také může mít obrovský praktický význam, jehož potenciál si teprve začínáme uvědomovat. Pokud psychologie splní své povinnosti, pokud sehraje zásadní roli, kterou je předurčena hrát, přispěje ke všem vědeckým oblastem. V souladu s Husserlem se domnívám, že úspěch nebo neúspěch psychologie v překonání jejích metodologických problémů nebude menší než rozhodující pro západní kulturu a civilizaci. Zda se lidstvo bude dívat zpět na tradici osvícenství jako žalostné selhání nebo jako jeden z nejdůležitějších úspěchů, bude záviset v první řadě na vědě o subjektivitě. Neboli, slovy Husserla, „*psychologie je skutečně rozhodující pole*“ (Husserl, E., 1937, s. 208). Proto musí najít odvahu čelit svým problémům a měnit se.

ZDROJE

- Allison, D. B., Brown, A. W., George, B. J., & Kaiser, K. A. (2016). Reproducibility: a tragedy of errors. *Nature*, 530, 27.
- Asendorpf, J. B., Barse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: the case of shy behavior. *Journal of Personality and Social Psychology*, 83, 380–393.
- Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., & Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, 27, 108–119.
- Baker, M. (2016). Is there a reproducibility crisis? *Nature*, 533, 452–454. doi: 10.1038/533452a
- Bell, J. (1987). *Speakable and unspeakable in quantum mechanics*. Cambridge University Press: Cambridge.
- Bosson, J. K., Swann, W. B., Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79, 631–643.

- Copernicus, N. (1543). *De revolutionibus orbium coelestium*. Online available at: [http://www.geo.utexas.edu/courses/302d/Fal_2011/Full%20text%20-%20Nicholas%20Copernicus,%20_De%20Revolutionibus%20\(On%20the%20Revolutions\),_%201.pdf](http://www.geo.utexas.edu/courses/302d/Fal_2011/Full%20text%20-%20Nicholas%20Copernicus,%20_De%20Revolutionibus%20(On%20the%20Revolutions),_%201.pdf)
- Desmet, M. (2018). The pursuit of objectivity in psychology. Gent: Borgerhoff & Lamberigts.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *Plos One*, 4, art. nr. e5738.
- Gleick, J. (1987). *Chaos: making a new science*. Penguin Books: London.
- Gnamb, T. (2014). A meta-analysis of dependability coefficients (test-retest reliabilities) for measures of the Big Five. *Journal of Research in Personality*, 52, 20–28.
- Horowitz, L. M., Alden, L. E., Wiggins, J. S., & Pincus, A. L. (2000). *Inventory of Interpersonal Problems*. London: The Psychological Corporation.
- Husserl, E. (1937). *The crisis of the European sciences and transcendental phenomenology. An introduction to phenomenology*. Northwestern University Press: Evanston.
- Ioannidis, J.P.A. (2005). Why most published research findings are false. *PLoS Med*, 2, e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Joint Committee for Guides in Metrology (2012). *International vocabulary of metrology: basic and general concepts and associated terms (VIM)*. Online available at: <http://www.bipm.org/en/publications/guides/vim.html>
- Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: a review. *Journal of the Royal Statistical Society Series A*, 145, 42–73.
- Lamiell, J. T. (2003). *Beyond individual and group differences. Human individuality, scientific psychology, and William Stern's critical personalism*. Sage Publications: Thousand Oaks.
- Linnaeus, C. (1753). *Systema Naturae*. Online available at: https://www.kth.se/polopoly_fs/1.199546!/Menu/general/column-content/attachment/Linnaeus-extracts.pdf
- Lorenz, E. N. (1963). Deterministic non-periodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141.
- Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractal dimensions. *Science*, 156, 636–638. <https://doi.org/10.1126%2Fscience.156.3775.636>
- Meganck, R., Desmet, M., Bockting, C., In-slegers, R., Truijens, F., et al. (2017). The Ghent Psychotherapy Study (GPS) on the differential efficacy of supportive-expressive and cognitive behavioral interventions in dependent and self-critical depressive patients. *Trials*, 18, 126.
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dries, R. R., et al. (2001). Psychological testing and psychological assessment. *American Psychologist*, 56, 18–165.
- Molenaar, P. C. M., & Valsiner, J. V. (2005). How generalization works through the single case: a simple idiographic process analysis of an individual psychotherapy. *International Journal of Idiographic Science*, 1, 1–13.
- Prusinkiewicz, P., & Fowler, D. R. (1995). Shells in three dimensions. In Meinhardt, H. (ed.) *The Algorithmic Beauty of Sea Shells* (pp. 163–182). Springer: Berlin.
- Richardson, (1961). The problem of contiguity: an appendix of statistics of deadly quarrels. *General Systems Yearbook*, 6, 139–187.
- Rosenzweig, S. (1933). The experimental situation as a psychological problem. *Psychological Review*, 40, 337–354.
- Russel, B. (1921). *The Analysis of Mind*. Gutenberg Ebook.
- Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: lessons from 26 research scenarios. *Psychological Methods*, 1, 199–223.
- Stern, W. (1921). Richtlinien für die Methodik der psychologischen Praxis. Beihefte zur Zeitschrift für angewandte Psychologie, 29, 1–16.
- Tinbergen, N. (1948). Social releasers and the experimental method required for their study. *The Wilson Bulletin*, 60, 6–51.
- Todes, D. P. (2014). *Ivan Pavlov: a Russian life in science*. Oxford University Press: Oxford.

- Toomela, A. (2007). Culture of science: strange history of the methodological thinking in psychology. *Integrative Psychological and Behavioral Science*, 41, 6–20.
- Valsiner, J. (2012). *A guided science. History of psychology in the mirror of its making.* Transaction Publishers: New Brunswick.
- Vesalius, A. (1543). *De humani corporis fabrica libri septem.* Online available at: <https://ceb.nlm.nih.gov/proj/flash/vesalius/vesalius.html>.
- Viswesvaran, C., & Ones, D. S. (2000). Measurement error in “Big Five Factors” personality assessment: reliability generalization across studies and measures. *Educational and Psychological Measurement*, 60, 224–235.
- Watson, J. D., & Crick, F. H. C. (1953). Molecular structure of nucleic acids. *Nature*, 171, 737–738.
- Wilkins, M. H. F., & Randall, J. T. (1953). Crystallinity in sperm heads: molecular structure of nucleoprotein in vivo. *Biochimica and Biophysica Acta*, 10, 192–193.
- Došlo do redakce a přijato k publikaci jako zvaný text dne 5. 12. 2018*