Měření v psychologii

[TODO: odstranit neustále opakování stejných věcí]

Představme si, že zatímco čteme tento text, sedí kousek od nás psycholog a snaží se zjistit, zda jsme se do čtení pustili ze zvědavosti či nudy, zda nás baví, nebo mu nerozumíme. Není jisté, že se mu to podaří, ale jistě se nevzdá a s pomocí standardizovaných testů a dotazníků se pokusí posoudit jiné naše vlastnosti. Psychologové věří, že jsme psychologické entity, které lze popsat, změřit, či dokonce kvantifikovat, a porovnávat s ostatními osobami i se sebou samými v průběhu času. Tento popis pak vytváří sice nedokonalou, ale přesto informativní představu, která se dá použít k predikci.

V moderní společnosti proto plní psychologické měření důležitou úlohu. Učitelé nechávají školáky testovat, zdali nemají dyslexii či nejsou hyperaktivní. Rodiče o dětech zjišťují, jaké mají zájmy a schopnosti. Státy testují celé populace žáků, aby ověřily, co se ve škole naučily. Firmy si najímají psychology, kteří jim mají pomoct najít nejvhodnějšího uchazeče o pracovní pozici. Kognitivní vědci mohou určit, kolik náhodných číslic udržíte v pracovní paměti, rychlost, s jakou mentálně rotujete trojrozměrnými objekty, jak moc ovlivní chyba váš následující výkon. Následně odhadnou vaši obecnou inteligenci, která predikuje střední délku života, očekávaný příjem a pracovní výkonnost napříč dekádami. Studiem změn v psychologickém stavu dokáží psychologové dynamicky reprezentovat naši emocionální pohodu a tak předpovídat, jaký soubor psychologických symptomů nás činí náchylným, nebo odolným proti depresi. Rovněž neutuchající oblíbenost a množství různých kvízů na internetu ukazují, že psychologické testování je běžnou a nedílnou součástí společnosti.

Rozmanitost psychologických charakteristik je ohromující. Psychologové je zjevně rádi měří, protože vykouzlili nesmírný počet testovacích postupů, a říkají, že dokáží měřit atributy jako je inteligence, osobnostní rysy, postoje, kvalita života, spokojenost zákazníka, či různá psychická onemocnění.

Když se zaměříme na specializovanější oblasti psychologie, seznam atributů se ještě prodlouží. V širší doméně měření inteligence jsou zkoumány prostorové, verbální, numerické, emocionální nebo percepční schopnosti respondentů; můžeme hodnotit řadu dovedností, od konkrétních jako je psaní a počítání, až po abstraktní jako je kreativita; z literatury o osobnostní psychologii se dozvíme, že osobnost lze rozdělit podle pěti faktorů: extraverze, neuroticismu, svědomitosti, otevřenosti vůči zkušenosti a přívětivosti, přičemž každý z nich lze dále rozdělit do specifičtějších subfaktorů; v klinické psychologii objevíme rozličné subtypy schizofrenie, dyslexie a deprese, přičemž každá z nich může být vyhodnocena různými způsoby.

Zakladatelé psychologie předpokládali, že měření je pro vědu nutné, a svoje postupy za měření pokládali. Psychologie se ustanovila jako kvantitativní věda, a tato představa nebyla dlouho rozporována ([3]). Ale jak tyto postupy fungují? A splňují podmínky, abychom je mohli považovat za měření? Obecně, v sociálních vědách a především v těch behaviorálních se měřením zabývá disciplína nazývaná psychometrika, která se nejvíce vyvíjí a které se týkají nejzávažnější debaty o *soft* měření ([87]). V předchozí kapitole jsme zmínili hlavní rozdíl mezi *hard* a *soft* měřením: *soft* vlastnosti jsou realizací atributů, kterým chybí přesná teoretická definice a které nejsou propojeny s dalšími atributy; obtíží, které při pokusu o měření psychologických vlastnosti musíme překonávat, je však více.

Nejdříve si připomeňme definici měření: „Měření je konceptuální a experimentální proces, který implementuje morfní přiřazování hodnot vlastnostem a je schopný produkovat informace o předdeklarované vlastnosti s určenou a prokazatelnou úrovní objektivity a intersubjektivity.“

Pro měření v psychologii jsou klíčové tyto části definice, které pro psychologické měření představují problematické oblasti:

* Měření se týká předdeklarované vlastnosti
* Měření zahrnuje experimentální aktivity
* Měření implementuje morfní přiřazování hodnot
* Měření produkuje informace s určenou a prokazatelnou úrovní objektivity a intersubjektivity.

Měření se týká předdeklarované vlastnosti, ale deklarace psychologických atributů je složitá. Není jasný jejich ontologický status („Existuje g faktor?“), ani definice („Co je to vlastně inteligence?“). Často jsou definovány pouze s pomocí kompletní a detailní procedury, kterou je měříme; takový operacionalistický přístup pak nevyhnutelně vede k tomu, že obdobná, ale ne stejná procedura měří jinou vlastnost. Stejně tak obvykle neexistují relevantní atributové teorie, které měřený atribut propojují s dalšími atributy ("Jak přesně inteligence souvisí s exekutivními procesy či pracovní pamětí?").

Měření zahrnuje experimentální aktivity, a to se pochopitelně týká i vyplňování dotazníků či provádění testovacích úkolů. Jenže zatímco u hard měření je kauzalita implikovaná samotným experimentálním prováděním měření pomocí měřicích systémů, a kauzální řetěz je zřejmý z teoretického poznání procesů (změřili jsme teplotu, protože měřený objekt zahřál rtuť v teploměru a ta se následkem toho roztáhla), v psychologii, která se velice často spoléhá na korelační studie, není kauzální vztah mezi měřenou vlastností (např. inteligencí) a hodnotou indikace (např. v podobě hrubého skóru testu) samozřejmý. Když respondent vyplňuje položku, jaké procesy se vlastně odehrávají v jeho mysli? Kauzalita v psychologickém měření proto silně souvisí s jeho validitou ([38]) – pokud nevím, jaká vlastnost způsobila změnu indikace, jak víme, co jsme vlastně změřili?

Měření implementuje morfní přiřazování hodnot, jenže psychologové často předpokládají měření na intervalové škále, ale přitom není vždy jasné, jestli je zvolená škála vhodná; způsoby zpracování dat jsou často zvoleny bez zkoumání, zda relace mezi empirickými vlastnostmi tomu odpovídají.

Měření produkuje informace s určenou a prokazatelnou úrovní objektivity; transdukční chování měřicích přístrojů, tj. psychologických testů, mívá problémy se stabilitou a selektivitou. Psychologické testy zpravidla nejde respondentům předkládat opakovaně, a navíc nejsou vždy schopny zajistit, že výsledek je ovlivněn pouze měřenou vlastností a ne prostředím (Jsou podmínky v místnosti nepříjemné či znervózňující, a jaký to má vliv na výkon?), jinými vlastnostmi respondenta (Odpověděl špatně, protože nemá dost vysokou úroveň znalostí, nebo jen spěchal a špatně si přečetl otázku?) či rozličnými interakcemi (Rozlišuje test mezi respondenty s různou hodnotou měřené vlastnosti stejným způsobem?).

Měření produkuje informace s určenou a prokazatelnou úrovní intersubjektivity, ovšem v psychologickém měření je toto možné pouze do určité míry, neboť není možné kalibrovat testy a dotazníky stejným způsobem jako fyzikální přístroje – neexistuje referenční vlastnost, respektive je vztažená k určité populaci a vůči jiné nemá smysl (jak vypadá referenční vlastnost s hodnotou 100 IQ?).

Každé této problematické oblasti se nyní budeme věnovat podrobněji; ale než se k nim dostaneme, je nutné zmínit spor o kvantitativnost psychologických atributů. Tento spor je nejen historický, ale zároveň stále přetrvává – každá kritika psychologického měření od současných autorů se ho dotýká; a protože se tito autoři zabývají špatnou otázkou, nemohu sám zůstat pozadu.

# Kvantitativnost psychologických atributů

V roce 1932 *British Association for the Advancement of Science* (od roku 2009 *British Science Association*) vytvořila v Yorku komisi, která se měla zabývat kvantitativními odhady senzorických událostí a potom podat zprávu o měřitelnosti souvisejících atributů. Jinými slovy, cílem komise bylo určit, zda se psychologické vyhodnocovací metody dají považovat za měření psychologických atributů, a zda jsou tyto atributy vlastně měřitelné ([66], [X], [X]).

Komise vytvořila dlouhou předběžnou zprávu v roce 1938 ([91]), obsahující výroky pro i proti měření počitkových intenzit. Zakončila tudíž žádostí o znovuustanovení, aby se zvážilo, jestli tyto pohledy jsou či nejsou nesmiřitelné. Tento úkol byl proveden subkomisí, která vytvořila finální zprávu ([92]).

Členové komise byli vybráni ze dvou sekcí Společnosti, ze sekcí A (fyzika) a J (psychologie). Podle svého předsedy se nyní komise nazývá Fergusonova komise (pro obsáhlejší seznam členů komise viz [3]), ale pro naši debatu je nejdůležitějším členem fyzik Campbell, nejvlivnější z fyziků ([66], [X]), jehož přístupem k měření jsme se již zabývali[[1]](#footnote-1) ([TODO: odkaz na kapitolu]) a nyní si ho stručně připomeneme.

Campbell definoval měření pomocí omezení epistemologického procesu zapojeném v přiřazování čísel objektům ([21, 3]). Tvrdil, že měřitelnost lze ustanovit tak, že nejdříve dokážeme, že zkoumaný atribut splňují empirickou relaci uspořádání, a pak buď najdeme fyzickou operaci konkatenace, která dovoluje konstrukci referenční škály měření a se kterou lze provádět měření porovnáním, nebo najdeme nějaký fyzikální zákon, který měřený atribut vyjadřuje jako funkci jiných atributů. ([66], [X])

Představme si typický psychofyzikální experiment o nejmenších znatelných rozdílech. Respondentovi ukážeme nějaký podnět *a* a pak ho požádáme, aby zvolil podnět *b*, který jen sotva dokáže odlišit od a. Je-li φ(*a*) nějaká odpovídající fyzikální intenzita podnětu a, a je-li δφ(*a*) = φ(*a*) − φ(*b*) změna této intenzity nutná, aby vznikl nejmenší znatelný rozdíl v počitku podnětu, je možné vyjádřit δφ(*a*) jako funkci *l* intenzit φ(*a*) pro různé podněty: δφ(*a*) = φ(*a*) − φ(*b*) = *l*(φ(*a*)). V mnoha případech toto vede k hladké křivce, takže můžeme předpokládat, že mezi nimi existuje nějaká empirická relace. ([66])

Na těchto premisách se Komise shodla, ale otázkou bylo, zda tato funkce plně popisuje fakta. ([66])

Psychologové tvrdili, že kvantitativní vyjádření intenzity počitků vznikajících na základě intenzity podnětů je nutné, a lze ho získat – podle Fechnera ([93]) – následovně. Předpokládáme, že nejmenší znatelný rozdíl podnětu δφ vyvolává rozdíl v počitku δψ, a že tyto rozdíly jsou stejné a mohou být vyjádřeny konstantní hodnotou c([94]). Můžeme potom psát:

Pokud dále předpokládáme, že takový vztah platí i pro odpovídající diferenciály (s konstantou c' místo c), můžeme psát:

a, za předpokladu, že počitek s nulovou intenzitou odpovídá nějaké prahové hodnotě intenzity podnětu φ0, dosáhneme (vzorec 4):

Najdeme tak funkční vztah h mezi fyzikální intenzitou φ a intenzitou odpovídajícího počitku ψ. Pokud se rozdíly intenzity řídí Weberovým zákonem

kde k je konstantní, získáme Fechnerův zákon

kde a . Fechner tvrdil, že toto poskytuje způsob, jak nepřímo měřit ψ jako funkci fyzikální veličiny, tj. jedná se o odvození měření ψ ze změření φ. To lze vyjádřit pro každý podnět *a*

Ve Fechnerově představe lze počitky nepřímo měřit s pomocí měření odpovídajícího podnětu, díky psychofyzikálním zákonům. Fyzikové ale argumentovali, že možnost matematického provedení transformace podle vzorce 4 nedokazuje měřitelnost intenzity počitku, protože je založena na Fechnerově předpokladu, že každý nejmenší znatelný rozdíl v podnětu je mapována do lidského percepčního systému jako stejná změna v intenzitě počitku. Tento předpoklad ale nelze ověřit, pokud zároveň není možno nezávisle změřit intenzitu počitku. A to není možné, protože pro počitky neexistuje operace konkatenace. ([66])

Komise nikdy nedosáhla jednomyslného souhlasu, ale názor zastánců "tvrdé linie" byl, že postupy, které psychologové používali, nešlo nazývat měřením, neboť nebyli schopni prokázat empirickou smysluplnost matematické operace sčítání psychologických atributů ([3]). Campbellovo pojetí měření kladlo na psychologické atributy příliš přísné podmínky: psychologické atributy nemají žádnou operaci konkatenace, a zároveň je nelze definovat pomocí odvozeného měření. Byla tedy otázka, zdali jsou takové atributy kvantitativní a tím pádem vůbec měřitelné. ([21], [3])

Jenže psychologové, kteří investovali svůj čas do konceptu měření, nemohli nechat stanovisko Fergusonovy komise bez odpovědi. Stevens, jehož škála Sone byla rovněž předmětem jednání komise, proto vytvořil vlastní definici měřen[[2]](#footnote-2). Vzal Campbellovu definici měření jako „přiřazování číslic objektům nebo událostem podle pravidla“, odstranil upřesnění „pravidla“ v podobě konkrétní specifikace empirických podmínek, které musí atribut splňovat, a ještě Campbella drze ocitoval.

Stevensův operacionalistický přístup k měření převládá, ke vzteku Michella ([3]) a jiných ([X]), dodnes. Toto velké přijetí Stevensovy definice měření vede psychology k víře, že měření se sestává výhradně z „přiřazování číslic objektům nebo událostem podle pravidla“. Okamžitý nepříznivý výsledek této definice je, že cokoliv lze chápat jako měření, protože obvykle můžeme vždy najít nebo stanovit nějaké pravidlo, s jehož pomocí číslice mohou být přiřazeny různorodým aspektům věcí ([X]). A dále, ignorujeme fakt, že otázka měřitelnosti zahrnuje empirickou otázku, tj zkoumání morfizmu mezi empirickými a symbolickými relacemi. ([45])

Již víme[[3]](#footnote-3), že axiomatická reprezentační teorie měření tento spor vyřešila: Stevensovy škály byly formalizovány s pomocí teorému jedinečnosti a teorie spojitého měření ukázala, že o kvantitativnosti atributu lze uvažovat i pouze na základě kvalitativních relací (čímž vyvrátila Campbellovo tvrzení o důležitosti konkatenace). Psychologie ovšem nedokázala tyto objevy využít – a jak jsem argumentoval[[4]](#footnote-4), ani nemohla, neboť reprezentační teorie měření není úplnou a prakticky aplikovatelnou teorií měření. Tento paradox – tedy že formální řešení sporu o kvantitativnost psychologických atributů sice existuje, ale nejde použít v praxi – psychology a především psychometriky dodnes rozčiluje.

Michell ([3]) sveřepě kritizuje, že psychologové netestují empirické relační struktury atributů a stále definují škály na základě svých rozhodnutí, nikoliv na základě konzistentních relací v datech – obvykle automaticky předpokládají, že psychické atributy jsou kvantitativní.

„Proč revoluce nenastala?“ zeptal se Cliff ([48]) a upozornil na příčinu toho, proč axiomatický přístup k měření neslavil úspěchy: modely měření v podobě víceparametrických IRT modelů[[5]](#footnote-5) porušují základní předpoklad teorie spojeného měření, tj. že ony dva atributy, které „spojitě měříme“ spolu nesmí interagovat. Praxe ukazuje, že položky psychologických testů od sebe odlišují respondenty s různou úrovní měřeného atributu různým způsobem – což je pak důvodem použití víceparametrických IRT modelů. Druhým důvodem je nevyhnutelná přítomnost chyb v měření a deterministická povaha RTM.

Schönemann ([51]) tvrdí, že neúspěch RTM v sociálních vědách má svůj důvod: modely měření vznikly jako abstraktní formální teorie s tím, že empirická intepretace se nechá na později; jenže v empirických vědách je hlavní způsob zvyšování vědomostí indukce založená na pečlivém pozorování, nikoliv dedukce. Schönemann si dále nejprve protiřečí: navrhuje, že by se psychologové místo snahy využívat matematiku, podobně jako fyzikové, měli soustředit na empirické zkoumání „neurčitostí živých bytostí“ – jenže to nejde bez pomoci právě matematiky. A následně se Schönemann rovnou plete: tvrdí, že protože ve fyzice je většina měření fundamentálních, a ne odvozených od jiných měření, a v psychologii toto neplatí, měli by psychologové na měření zapomentout. Nejenže tak opakuje tak mylný Campbellův argument, ale navíc ukazuje, že má o měření špatné představy – ve fyzice jsou atributy natolik provázané přírodními zákony, že je prakticky většina měření nepřímá či odvozená [41] – příkladem může být opět teplota, základní veličina soustavy SI, která jinak než nepřímo měřit nejde.

Trendler ([45]) je ve svých tvrzeních ještě příkřejší; argumentuje následovně. Měření vyžaduje, aby se dvě vlastnosti buď lišily, nebo byly stejné. To musí být empiricky ustanoveno, buď přímo, či nepřímo. Nepřímé porovnání vyžaduje důkaz, že stejné vlastnosti mají stejné hodnoty v rámci náhodné chyby. To vyžaduje ustanovit sobě rovné vlastnosti nezávisle a pak je použít k porovnání těch měřených. V psychologii nelze docílit tak přesné manipulace a nelze plně kontrolovat systematická zkreslení. Díky tomu nelze objevit kvantitativní strukturu psychologických atributů, a bez ní nelze měřit. Pokud nechceme redukovat pravdu na verifikaci, Trendlerův argument přinejlepším dosahuje nemožnosti epistemického přístupu k pravdě o měřitelnosti; neznamená to, že měřitelnost neplatí ([46]). Podstatnější částí argumentu je tvrzení, že lze měřit pouze atributy s kvantitativní strukturou; přestože operace konkatenace není nutná k dokázání kvantitativnosti, kvantitativnost sama je stále nutnou podmínkou měřitelnosti atributu.

Možná by si tento mylný postoj zasloužil vlastní název: klam podmínky kvantitativnosti (*quantity condition fallacy*)? Fixace psychologů na kvantitativní strukturu atributů je zjevná a pochopitelná, pramení z touhy, aby bylo možno psychologické atributy mezi sebou vztahovat pomocí matematických rovnic podobně jako ve fyzice. Ale již jsme si dokázali[[6]](#footnote-6), že kvantitativní struktura atributu není nutnou ani postačující podmínkou měření. To samozřejmě neznamená, že není třeba zkoumat morfizmus mezi empirickými a symbolickými relačními strukturami – škála, na které atribut měříme, musí mít své opodstatnění. Znamená to ale, že Michellova ([3]) „námitka vůči kvantitativnosti“ (*quantity objection*) je zbytečná. Nezáleží tolik na tom, zdali jsou psychologické atributy kvantitativní, či nikoliv. Problémy psychologického měření spočívají jinde; výše uvedení autoři je rovněž zmínili: velká nejistota měření a obtíže při vytváření referenční škály. Je třeba ovšem systematičtějšího popisu podle frameworku z minulé kapitoly.

# Aplikace konceptuálního frameworku struktury měření

[TODO: předělat, rozvést, doplnit]

Při aplikaci frameworku na psychologické měření, tak abychom popsali proces měření určitého atributu, musíme vyřešit následující úkoly:

* Plánování
  + Určení cílů měření – proč by vlastně mělo být užitečné měřit psychologické atributy?
  + Definice měřeného atributu
  + Vytvoření modelu měření
  + Analýza nejistoty
  + Design referenční škály
  + Design přiřazování hodnot
  + Kalibrace
* Provedení
* Interpretace

## Cíle měření

[TODO: rozvést, přidat plky]

Pro definici cíle měření je třeba obvykle získat informace o následujících pěti oblastech ([61]):

1. Měřený objekt ­– v případě psychologického měření je to obvykle člověk, patřící do určité populace či specifické subpopulace (Děti či dospělí? Zdraví či nemocní jedinci? Určité povolání?).
2. Smysl – vesměs chceme srovnávat měřeného jedince s jinými jedinci, klasifikovat ho do skupiny, nebo srovnávat se sebou samým v jiném čase
3. Atribut – jaký psychologický atribut potřebujeme měřit.
4. Perspektiva – pohled, ze kterého uživatel výsledků měření definuje; to s sebou může nést různé požadavky na přesnost měření, kritéria standardy pro výsledky, atp.
5. Prostředí – empirické prostředí, kde očekáváme, že se měřená osoba vyskytuje.

## Definice měřeného atributu

Je třeba nejen definovat samotný psychologický atribut – což je samo o sobě obtížný úkol, neboť nejsou přímo přístupné (jako např. při měření délky) – ale teoreticky ho propojit i s jinými atributy. Přitom je třeba určit ovlivňující vlastnosti, tj. jaké atributy – jak měřeného objektu či prostředí – či jejich interakce mohou mít vliv na výsledek měření.

V minulé kapitole[[7]](#footnote-7) jsme si ukázali příklad s vlivem teploty na měření délky: podle potřeby je možné buď vliv teploty ignorovat (obvykle v situacích, kdy není třeba taková přesnost) nebo modelovat. Obdobným příkladem v psychologickém měření může být rychlost a měření kognitivních schopností. Kognitivní úkol můžeme popsat dvěma parametry: přesnost, tj. jak úspěšně či správně osoba úkol splnila, a rychlost[[8]](#footnote-8), tj. jak dlouho jí to trvalo. Snaha o rychlejší splnění úkolu vede k nižší přesnosti, ale zároveň potřebují schopnější jedinci menší čas ke splnění úkolu ([135]). Některé výzkumy ovšem naznačují, že při různé rychlosti odpovídání se mohou uplatnit různé psychologické procesy ([132, 133]); mezi obtížností úkolu, schopností jedince, rychlosti a přesnosti mohou existovat různé vzájemné interakce ([134]). Měření rychlosti jako ovlivňující vlastnosti tedy může přinést informace o hlavním atributu zájmu, a vzhledem k tomu, že čas je „zavedený“ fyzikální atribut, měření rychlosti plnění kognitivních úkolů je jistě jednou z možných cest, jak budovat teoretická propojení psychologických atributů. Zároveň je zjevné, že takové teoretické propojení „nebude zadarmo“: využití času při měření kognitivní schopnosti vede (podobně jako využití teploty při měření délky) k nutnosti použití komplexnějších modelů (viz. [136]), se všemi nevýhodami s tím obvykle spojenými.

## Model měření

Vybraný model musí reflektovat definici atributu a související teorie. Propojuje hodnoty indikace a měřený atribut; popisuje, jak empiricky získaná data použít k odhadu hodnoty atributu. Zároveň se dá použít i pro úpravu teorie, pokud v průběhu tvorby procesu měření zjistíme, že získaná (testovací) data modelu neodpovídají.

## Analýza nejistoty

[TODO]

## Design referenční škály

Výše popsané téma kvantitativnosti sice není klíčové, ale to samozřejmě neznamená, že lze škálu vybrat bez dostatečných důkazů.

## Design přiřazování hodnot

díky tomu, že psychologické atributy nejsou přístupné, spoléháme se na jejich indikaci v podobě chování respondentů; oním chováním nejčastěji bývá zodpovídání otázek.

## Kalibrace

Aby mohl být test používán jako měřicí přístroj, je třeba ho zkalibrovat. Protože v psychologickém měření obvykle není možné vytvořit standard, spočívá kalibrace ve statistickém odhadu parametrů na kalibračním reprezentativním vzorku určité populace, která je zvolena vzhledem ke stanovenému cíli měření.

◊

[TODO: lepší závěr]

Některým krokům se nyní budeme věnovat podrobněji.

# Definice psychologických atributů

Ve frameworku z předchozí kapitoly je prvním úkolem při definici atributu vytvoření referenční škály; připomeňme si, jaké kroky je třeba splnit, abychom toho dosáhli ([61]):

* Identifikovat třídu objektů vykazující vlastnosti, které jsou instancemi atributu.
* Identifikovat empirické relace mezi objekty, které charakterizují atribut.
* Definovat referenční škálu, tj. určit množinu objektů předpokládaných jako standardy a schopných reprezentovat možné jednotlivé vlastnosti a přiřadit jim hodnoty vlastností tak, aby zachovávaly empirické relace mezi objekty.
* Implementovat navrhnutou referenční škálu.
* Navrhnout alespoň jeden postup pro přiřazení hodnot vlastnostem objektů, které nejsou zahrnuty v referenční škále, konzistentně s existujícími empirickými relacemi.

Je očividné, že u psychologických atributů se při tomto postupu setkáme s řadou obtíží, které mají dvě hlavní příčiny: zaprvé, k psychologickým atributům máme přístup pouze nepřímo, na základě chování jejich nositelů. To by nebyl sám o sobě takový problém – jak už jsem se zmínil, ve fyzice je většina měření nepřímá. „Problémem“ je, že lidé disponují jedním z nejméně vysvětlených fenoménů: vědomím, které je jednak výsledkem velice komplexního systému procesů, a hlavně si díky němu lidé mohou uvědomit, že jsou měřeni, a uchovat si na to vzpomínky.

Jakým způsobem tedy vybrat standardy, když měřenými objekty jsou lidé? Uvažme např. pořadovou škálu inteligence: tu by šlo sestavit vytvořením několika různě těžkých položek, s jejichž pomocí bychom porovnávali respondenty. Pokud respondent nedokáže na *i*-tou položku odpovědět alespoň v polovině případů, kdy mu ji předložíme, přiřadíme mu hodnotu inteligence *i* − 1. Tento postup je samozřejmě nepoužitelný a slouží pouze k ilustraci těžkostí v měřením psychologických atributů. Za prvé, vyžaduje, aby před každým opakovaným zodpovězením položky došlo k mind-washingu[[9]](#footnote-9) respondenta, jinak nemůžeme získat poměr správných odpovědí. Za druhé, přepokládáme, že položky lze používat u všech lidí, ale to není možné kvůli tomu, že lidé nemluví všichni stejným jazykem, nepocházejí ze stejných kulturních podmínek, atp. S prvním problémem se lze vypořádat s pomocí statistických metod – předpokládáme nějakého modelového, idealizovaného jedince s daným vzorcem odpovědí. Ovšem druhý problém je (prozatím) neřešitelný – psychologické škály lze vytvořit pouze pro určité populace ([143]).

Vzhledem k těmto těžkostem s definicí psychologických atributů se proto musíme ptát: Co vlastně psychologické atributy jsou? A můžeme o nich vůbec říkat, že jsou reálné (aspoň v nějakém významu toho slova)? A pokud ano, co to znamená? ([96])

Na obecné úrovni existují psychologické atributy jako emergentní rysy bytostí majících vědomí ([96], [X], [X]), a načrtnutí toho, co představuje specifický atribut, může být předpokládáno, že bude určeno ne-intencionální (např. biologickou) podstatou, nebo se může týkat intencionality[[10]](#footnote-10). Dále, aby tyto atributy byly měřitelné v relevantní populaci, musí existovat rozdíly v atributu v této populaci. ([96])

Vědomí je esenciálně subjektivní, a tudíž se zdá být mimo doménu činností zaměřených na získávání objektivních znalostí. Toto vedlo mnohé k závěru, že buď je psychologická věda nemožná, nebo že to může pouze věda zabývající se pozorovatelným. Nicméně, jak argumentuje [104], toto si zaměňuje objektivitu existence věci (ontologickou objektivitu) s objektivitou znalostí (epistemickou objektivitou).

Podle [104], věc je ontologicky objektivní, pokud její existence nezávisí na tom, zda ji prožívá nějaká bytost mající vědomí. Věc je ontologicky subjektivní, pokud její existence závisí na vědomí: když zažívám bolest, elektrochemická reakce neuronů je ontologicky objektivní, ale ontologicky subjektivní pocit bolesti existuje pouze, dokud ho cítím – jeho existence závisí na tom, zdali ho prožívám.

Toto vysvětlení objektivní a epistemické objektivity a subjektivity implikuje, že do formulace vědeckého realismu musí být přidáno doplnění. Hlavní tvrzení vědeckého realismu je, že jedním z cílů vědy je získávat pravdivá tvrzení o světě. To je tvrzení epistemické objektivity: pravdy o světě, které se snažíme dozvědět, by neměly záviset na názorech nebo jiných idiosynkraziích těch, kdo se dovídají. Toto je možné, i když je určitý předmět vědy sám ontologicky subjektivní; je možné získat epistemicky objektivní vysvětlení ontologicky subjektivního fenoménu. Z pohledu vědecké filozofie je to nutná podmínka, aby byla psychologická věda možná. ([96])

Takže pokud chceme měřit psychologické atributy, musíme popřemýšlet nad jejich ontologií. Tato ontologie musí být vypracována před tím, než se pokusíme o nich získat nějaké znalosti. Skutečný proces vědeckých objevů často funguje jako bootstrapping mezi vývojem teorie a empirickou studií – jak ukázal příklad s měřením teploty –, takže asi není důvodné u této ontologie vyžadovat naprostou jednoznačnost před tím, než ji začneme studovat. Nicméně, otázky týkající se ontologie atributu nelze svěřovat epistemologii (např. operacionalizací definice atributu jako čehokoliv, co odpovídá latentní proměnné v latentním modelu). ([96])

Největším problémem psychologického měření je častá absence dobře vyvinutých teorií o atributech. Místo toho jsou položky často konstruovány nejlepšími odhady na základě jakékoliv dostupné teorie, ale také na základě intuice (co se zdá být rozumné?), tradice (jak byly konstruovány podobné testy?) a konformitě (co si myslí moji kolegové?). Role atributové teorie je často podceňována. Měření musí být založeno na substantivní teorii a zároveň vést k vytváření a testování dalších teorií. ([90])

Dobře definovaná teorie umožňuje výzkumníkovi formulovat hypotézy, které předpovídají, která chování se vztahují k jiným jako koreláty atributu zájmu, jak se chování vztahují k chováním typickým pro jiné atributy, a jak manipulace typického chování může ovlivnit jiná typická chování. Tyto hypotézy mohou zahrnovat skupinové rozdíly vzhledem k naměřeným hodnotám, změny v měření v průběhu času a interní struktury měřicích přístrojů, ale rovněž i procesy, které byly aktivovány položkami a vedly k odpovědím. ([90])

Přestože detaily psychologické úrovně popisu (jako příčiny a následky jak mezisubjektové, tak vnitrosubjektové variace atributu) mohou být neznámé, tato neznalost detailů struktury atributu nesmí být překážkou výzkumu její role ve funkčních systémech. Stejně tak to ani nutně nečiní tvrzení o její existenci bezvýznamné. Takovéto *black-box* termíny jsou empirické konstrukty, např. koncept genu jako jednotky dědičnosti, byl původně funkcí bez struktury, odvozené pouze z pozorování dědičnosti; struktura byla doplněna později, což dovolilo existenci genu potvrdit nezávisle na fenoménu, kvůli kterému byly postulovány, aby ho vysvětlily. ([96])

## Problém redukce

Jsme nejen psychologické entity, ale zároveň jsme i entity biologické. Jsme tvořeni miliardami neustále interagujících buněk: máme zhruba 1100 gramů mozkové tkáně ([X]) s 86 miliardami neuronů ([X]). Můžeme tak změřit řadu neurologických atributů: množství šedé hmoty v různých oblastech mozku; mapování, jak bílá hmota tyto oblasti propojuje; kvantifikaci jak kortikální subsystémy synchronizují svoji aktivitu v průběhu času; či lokalizované změny v úrovních neurotransmiterů.

[95] ukázali z aktivity vizuálního kortexu dekódovat, s pomocí vážené lineární kombinace aktivačních vzorců voxelů, podobu podnětů, a to s více jak 90% přesností. Pro takto nízkoúrovňový mentální proces se zdá být přijatelné, že aktivační vzorce vizuálních oblastí mozku mohou sloužit jako měření prezentovaného podnětu. Oproti tomu, změřením třeba objemu mozku sotva změříme inteligenci – byť mezi těmito dvěma atributy existuje slabá korelace ([X]). Mezi těmito situace mi existují esenciální rozdíly, které ukazují na obtíže s intepretací neurologických měření jako reprezentace psychologických atributů.

Otázka potom zní, jakým způsobem propojit psychologické a neurologické explanatorní úrovně? Jde tedy o problém redukce – jak jsou propojené jednotlivé explanatorní úrovně vědy. [98] navrhli hierarchii explanatorních úrovní podle toho, jak velkými, vysokými, komplexními či abstraktními (sociální skupiny) či malými, nízkými, jednoduchými či konkrétními (subatomární částice) entitami se zabývají. Entity na vyšší úrovni se pak skládají z těch na nižší úrovni – věda se pak sestává ze základních jednotek, které postupně vytváří hierarchii věd. Tato zjednodušující představa pak udává fyzikální vědy jako nejabstraktnější, přes biologické vědy až po sociální vědy; sociologie se zabývá skupinami lidí, psychologie mentálním fungováním jedince, neurověda studuje jeho mozek, který se skládá z biochemických prvků, a tak dále až k subatomárním částicím.

[99] formalizoval požadavky, které musí být splněny, aby teorie určité úrovně mohla být redukována na teorii nižší úrovně. Na každé explanatorní úrovni můžeme najít pravidelnosti, které vysvětlují interakce jejích entit, a musíme proto objevit způsob, jak tyto zákonitosti přeložit do zákonitostí nižších úrovní. Tyto přemosťující zákony (*bridge laws*) by nám dovolili analyticky přeložit jakoukoliv vědeckou zákonitost „nahoru“ či „dolů“. Mnoho takových zákonů neexistuje, ale představa, že psychologické procesy lze popsat s pomocí pochodů v mozku je velice lákavá.

[100] a [101] tuto představu oprávněně kritizovali argumentem mnohonásobné realizace. Tato teze tvrdí, že určitý mentální stav, jako je „cítit bolest“ může být realizován mnoha fyzickými způsoby. Není důvod, aby to bylo pouze aktivací C-vláken ([X]): jiné druhy neuronů (např. v jiných živočiších) nebo dokonce nebiologické substráty mohou plnit podobnou funkci. Pokud tedy lze mentální stav realizovat více vzájemně disjunktními způsoby, není jasné, jak přemosťující zákony popíší toto *one-to-many* mapování. Obecně můžeme říct, že vlastnosti a události na vyšších úrovních mohou být realizovány mnoha způsoby na nižších úrovních: vlastnost být hodinami může být realizována mnoha způsoby, protože hodiny jsou definovány tím, co dělají, než jaká struktura jim umožňuje to dělat ([96]).

Jaký tedy popsat vztah mysli a fyzického světa? Máme tu vědeckou představu světa, který je ve své podstatě fyzický (skládající se z elementárních částic) a všechno v tomto světě, včetně bytostí mající vědomí, není nic než tyto fyzické částice spojené do komplexnějších celků. Na druhou stranu vědomí prožíváme z pohledu první osoby, který nám říká, že máme tělo *i* mysl.

Historickým pojetím vztahu mysli a těla je karteziánský dualismus. Descartes ([102]) dospěl k názoru, že zatímco fyzikální věci zabírají trojrozměrný prostor a mentální nikoliv, musí se jednat o odlišné substance. Carteziánský dualismus tvrdí, že mentální a fyzické operuje v odlišných ontologických sférách/doménách, ačkoliv není jasné, jak mentální ovlivňuje fyzické, a naopak. ([96])

Alternativou dualismu je monismus: vše co existuje, existuje v jednom časoprostoru. [101] navrhl pohled, známý jako nereduktivní fyzikalismus, který zachovává ontologickou i epistemologickou autonomii vyšších explanatorních úrovní. Nemůžeme tvrdit, že jelikož je mozek nutný pro existenci psychologických procesů, je všechna psychologie redukovatelná na popisy mozku; i přes četné objevy není jasné, jak konceptualizovat roli neurověd a biologické psychologie a jejich vztah k psychologickým konstruktům. Na druhé straně extrému je představa eliminativního materialismu, že *folk* psychologie neexistuje, a tím i její koncepty jako vzpomínky, představy, přání, a budou nakonec nahrazeny neurovědami ([37]).

Pokud člověk přijme silný fyzikalismus, není důvod, proč se při redukci zastavit na úrovni mozku; neurologické procesy lze kauzálně redukovat na proteiny, molekuly, atomy, atd. Z pragmatického hlediska se na úrovni mozku zastavujeme, protože „to funguje“; pragmatický redukcionismus říká, že zastavení je v pořádku, pokud je empiricky odůvodněné, nikoliv na základě apriorních tvrzení. Dávat ontologickou přednost biologické úrovni explanace, bez toho aniž bychom ustanovili koherentní framework redukce může mít nepříznivé důsledky. ([37])

Konvenční uvažování považuje vztah mezi vlastnostmi nižšího a vyššího řádu za takový, že neurologické atributy jsou příčinou pozorovatelných rozdílů na psychologické či behaviorální úrovni: drogy ovlivňují kognitivní schopnosti, zranění může ovlivnit osobnost, u řady neurologických onemocnění fyziologické změny předcházejí kognitivním. Ovšem existují i důkazy o opačném působení: trénování určitých dovedností vede ke změnám v mozku. A možná je i určitá reciprocita, tj. vzájemné vlivy z neurologických vlastností k psychologickým a naopak: injekce testosteronu ovlivní chování, ale určité chování může vést i ke zvýšené produkci testosteronu. Kauzalita tedy může působit mezi neurologickými a psychologickými procesy oběma směry. ([37])

Protože dualismus je v současnosti překonaný ([X]), za zmínku stojí tři pojetí vztahu mysli a těla: identita, supervenience a emergence.

### Identita

Teorie identity tvrdí, že psychologické procesy a atributy jsou identické jejich neurologickým realizacím: to, co prožíváme, je přesně reflektováno v odpovídajícím neurologickém stavu mozku. Výhoda této teorie tkví v relativně neproblematickém přiřazení kauzality mentálním událostem: protože mentální událost či stav je identický určité neurální realizaci v jakémkoliv čase, má stejnou kauzální sílu jako tento neurologický stav, kterému je identický. ([37])

Teorie identity má více podtypů. *Type-type* identita stanovuje, který typ neurální aktivity (aktivace c-vláken) je identická typu psychologického stavu (pociťování bolesti). Vlastnost X je *type* identická vlastnosti Y, když platí tvrzení: X právě tehdy, když Y.

Méně přísná varianta, *token-token* identita stanovuje, že psychologický stav v určitém čase je identický neurální realizaci v tomto čase, ale že nemusí existovat určitá pravidelnost v neuronových realizacích (tj. pro stejný typ mentálního stav se mohou lišit, pro různé výskyty tohoto stavu). V daném časovém okamžiku pro osobu v psychologickém stavu X, tento stav X je identický k nějakému neurologickému stavu Y, který stav X instanciuje. Ale neexistuje žádné omezení na konzistenci toho vztahu: u té samé osoby může být v jiném časovém okamžiku psychologický stav X identický jinému neurologickému stavu Y2. Tato pozice nenese žádnou ontologickou váhu, a lze na ni nahlížet jako na antidualistické ukotvení psychologických stavů ve fyzické realitě. ([37])

Pokud redukujeme na nižší úroveň, každý *type* identický model se stane *token* modelem z nutnost, protože lidé nemají naprosto, do poslední molekuly či atomu, stejné mozky. ([37])

Alternativou těmto dvěma extrémům může být *type-token* teorie identity. Ta tvrdí, že pro každého jedince existuje type identický vztah mezi psychologickým stavem X a nějakou neurologickou realizací Y. Tj. kdykoliv je v psychologickém stavu X, je v neurologickém stavu Y, a naopak, ale tento stav Y se může lišit napříč lidmi. Příkladem může být hypotetická existence „buněk podnětu S“ – aktivovány jsou pouze určité konstelace neuronů, pro tento specifický podnět S, ale nezávisle na vizuální podobě tohoto podnětu (lingvistický podnět, či obrázek podnětu, ale nikoliv obrázek jiného podnětu). Tento shluk buněk nemůže být anatomicky identifikován takovým způsobem, aby byl jeho popis zobecnitelný napříč jedinci: ti liší se v umístění a struktuře tohoto shluku. Ale může platit, že všichni lidé, či lidé často vystavovaní tohoto podnětu, mají nějakou konstelaci buněk, která se takto chová, tj. jejichž aktivace může být považována za měření latentní proměnné „percepce podnětu S“. ([37])

### Supervenience

Davidson ([107]), Kim ([103]) a jiní ([X]) argumentovali, že mentální vlastnosti mohou supervenovat na fyzických. Koncept supervenience určuje, že vyšší úrovně abstrakce jsou determinovány nižšími; její definice je následující: o atributu X můžeme říct, že supervenuje atribut nižšího řádu Y, pokud nemůže existovat rozdíl v X (IQ osoby A je vyšší než osoby B) bez rozdílu v Y (vlastnosti mozku osoby A se liší od vlastností mozku osoby B). Tudíž, přítomnost Y-rozdílu je nutnou, ale ne postačující podmínkou pro přítomnost X-rozdílu. Tedy pokud není rozdíl v Y, není rozdíl ani v X. Tato relace je postačující podmínkou, abychom ji mohli nazývat superveniencí. Atributy nižšího řádu při supervenienci nutně nesdílí všechny charakteristiky atributu vyššího řádu: atomy, ze kterých se skládá umělecké dílo, nejsou krásné, a neurony nejsou inteligentní. Atributy vyšší úrovně supervenují na nižších kauzálně asymetrickým způsobem. ([37, 96])

Existuje víc variant supervenience; Kim ([103]) definoval silnou supervenienci pro atribut vyšší úrovně A a atribut nižší úrovně B následovně: Nutně, pro každé x a y, pokud x a y sdílejí všechny vlastnosti v B, potom x a y sdílejí všechny vlastnosti v A, tj. nerozlišitelnost v B zahrnuje nerozlišitelnost v A. Vztah supervenience je asymetrický, protože struktura nižší úrovně se může lišit, zatímco vyšší úroveň zůstává stejná (protože rozdíly nižšího řádu jsou nutné, ale ne postačující, pro rozdíly vyššího řádu). V kontextu vědomí to znamená, že mentální stav supervenuje na fyzickém stavu, pokud nemůže být změna v mentálním stavu bez nějaké změny v neurofyziologickém stavu. Supervenující atributy mohou být realizovány mnoha způsoby atributy, na kterých supervenují, což implikuje, že stejné rozdíly v mentálním stavy nemusí korespondovat stejným rozdílům v neurálním stavu napříč osobami a okolnostmi. Kvůli této asymetrii dává Kim kauzální prioritu realizacím nižšího řádu: neurologické indikátory kompletně determinují kauzální vlastnosti systému. Supervenience je konzistentní s *many-to-one* mapování z nižších do vyšších vlastností, ale ne s izomorfismem, což vylučuje *type-type* identitu.

### Emergence

Další teorií je emergence, která tvrdí, že vlastnosti vyšší úrovně abstrakce se vynořují či nastávají kvůli komplexním kombinacím nižších vlastností (inteligence neexistuje na úrovni neuronů, ale existuje, protože miliardy neuronů spolu interagují). Obecný popis emergence je takový, že emergentní fenomén je konceptualizovaný jako nastávající na makro úrovni, v kontrastu k mikroúrovňovým komponentám a procesům, z kterých vyvstává. Mozek může příkladem takové komplexní struktury sestávající se ze stavebních bloků, které spolupracují komplexním způsobem. Silná emergence tvrdí, že emergentní fenomén má kauzální sílu a ontologický status, který nemůže být plně redukován na jeho konstituující komponenty. Např. vědomí nemůže být, zdá se, plně vysvětleno na základě jednotlivých neuronů. Emergentní perspektiva může být pro psychologické vlastnosti vhodná, pokud vezmeme v úvahu dimenzi času: vlastnosti vyšší úrovně mohou vyvstávat s postupujícím vývojem mozku. ([37])

## [Výzkum psychologických procesů, simulace a modely, atd.]

[TODO: rozšířit]

Přestože tedy doposud nemáme jasnou odpověď na to, jak spolu souvisí mentální a neurologická úroveň, [37] navrhl způsob, jak jejich vztah modelovat. Data, která se týkají psychologických atributů, nazýváme P-indikátory: jedná se o psychologická měření, jako jsou řešení úlohy X, výběr odpovědi C, počet objektů udržených v pracovní paměti, atd. Data, které se týkají neurologických atributů, nazýváme N-indikátory: elektrická měření kortikální aktivity (EEG), měření rychlosti zpracování, BOLD signály či fyziologické indikátory jako hustota šedé hmoty, objem mozku, úrovně neurotransmiterů, atd.

Konkrétní podoba propojení N a P indikátorů a atributů v modelu závisí na konkrétní testované teorii redukce. Přestože využití při samotném měření je doposud sporné, jak experimentální technika pro další výzkum má podobné modelování jistě potenciál.

◊

Jak poznamenal [56], nic vás nedonutí ocenit lidskou inteligenci tak jako poznání, že vytvořit počítač uvažující jako my je nesmírně náročné. Možnosti, jak studovat mozek, jsou omezené; nemůžeme v reálném čase sledovat aktivitu velkého množství konkrétních neuronů. Posledních 25 let se vědci proto snaží *simulovat* mozek tak, aby tato simulace do určité míry byla biologicky odpovídající. Simulace ve velkém měřítku jsou pochopitelně velice drahé (kvůli vysoké výpočetní náročnosti), a umožňují získávat znalosti spíše o výkonosti simulovaného systému, než o neurobiologickém systému mozku, který simulace představuje ([112]).

Protože neexistuje jednotná teorie fungování mozku, úspěšné modely tudíž řeší spíše dílčí úlohy ([112]). Zajímavé je sledovat, jak se v posledních letech exponenciální měrou zvyšuje schopnost komputačního řešení složitých problémů, které jsou pro člověka jednoduché, ale je složité je popsat algoritmicky ([110, 56]): rozpoznávání obrázků, hraní Go, či řízení aut. Především úspěch aplikace neuronových sítí je pozoruhodný. Neuronové sítě jsou komputační modely, které se inspirují biologickou podobou neuronů. Neuronové sítě se skládají z jednotlivých jednotek, vzájemně propojených tak, aby si předávaly signály. Určité neurony jsou určeny jako vstupy, další jako výstupy, a ty zbylé různými způsoby propojují vstupy s výstupy tak, aby došlo k žádoucí transformaci. Z matematického hlediska neuronové sítě fungují tím, že aproximují komplexní matematické funkce jednoduššími. ([X])

Velmi úspěšné jsou rekurentní neuronové sítě, které jsou místo vektorů schopné zpracovávat sekvence vektorů, a tudíž jsou schopné se naučit programy mající stav; hodí se tedy k modelování mentálních procesů, jako je čtení a vytváření textu a obrázků ([113]).

Využití neuronových sítí v psychometrice není příliš rozšířené, protože sice dovolují modelovat nelineární vztahy, ale vyžadují větší množství dat; zároveň není jasná jejich intepretace ([111]).

◊

[TODO: lepší přechod]

## Validita psychologického měření

Pokud je měřená vlastnost na vysoké úrovni abstrakce, může být velice těžké zjistit, zdali skutečně měříme to, co chceme. Vzhledem k tomu, jak je obtížné vytvoření a ověření teorií psychologických procesů, má u psychologického měření tato otázka validity zásadní význam.

Pro validitu psychologického měření jsou důležité dva rysy procesu měření, jak jsme ho doposud popisovali. Prvním je závazek k vědeckému realismu: měřené atributy objektivně existují a můžeme se pokoušet získávat o nich informace.

Druhým rysem je přítomnost kauzálních relací v procesu měření: měření je experimentální aktivita a měřený objekt, měřicí systém i prostředí jsou vzájemně kauzálně propletené ([39], [146]). Mezi měřenou vlastností a indikační vlastností (jejíž hodnotu, indikaci, využíváme ke stanovení hodnoty měřené vlastnosti) musí nutně existovat kauzální vztah: pohyb elektronů v ampérmetru způsobí vychýlení ukazatele, pohybující se molekuly předají kinetickou energii teploměru, jehož měřící kapalina se pak roztáhne, fotony dopadající na fotorezistor snižují jeho odpor.

V posledních letech této vlastnosti procesu měření si všimli i někteří psychometrikové ([38], [21]) a vytvořili tak odpovídající definici validity psychologického měření: test je validní pro měření atributu právě tehdy, když a) atribut existuje, a b) změny atributu kauzálně produkují změny ve výsledku měřicí procedury.

S tím souvisí i kritika dřívějších přístupů k validitě.

Validita není komplexní, fasetovaná, nebo závislá na nomologických sítích – je to jednoduchý koncept: test je validní pro měření atributy právě tehdy, když atribut existuje a variace v tomto atributu kauzálně vytváří variace výstupu měřící procedury. Základní tvrzení je tedy ontologické, a logicky předchází procesu měření, protože je to nutná podmínka, aby bylo měření možné; toto tvrzení je nicméně odlišné od samotného procesu měření. Sjednocení všech procedur týkajících se ověřování této vlastnosti testů, oněch různých „validit“, testování vhodnosti, atd., nejsou validitou, ale *validizací*. ([38])

Měření je kauzální koncept, s jasně daným směrem svět ↝ nástroj; nemůže být tedy založen pouze na korelaci. Kriteriální validita – která využívá korelací – vede k tomu, že test je „validní“ pro spoustu věcí – šance potkat nulovou korelaci je téměř nulová. Dále – pokud by měly dva testy perfektní korelaci, znamenalo by to, že měří stejnou věc pod dvěma názvy. Korelace je rovněž statistika závislá na populaci – a proto by byla i validita na ní založená. ([38])

◊

Je těžké nedefinovat kauzalitu necirkulárním způsobem; stále neexistuje jednotná definice tohoto konceptu, či teorie usuzování o kauzalitě, které dokáží zvládnout všechny možné případy a protipříklady. Důsledkem toho je určitý „kauzální pluralismus“, metafyzický, konceptuální i důkazový; a i kdyby opravdu existovala jediná představa o kauzalitě, může existovat mnoho způsobů, jak kauzální vztah rozpoznat. ([106])

Místo toho, abychom příčiny definovali jako události ve smyslu věcí, které nastanou v určitý čas na určitém místě, je užitečné dovolit, aby příčinou mohla být vlastnost (červená barva látky vyprovokovala býka) či fakt, a nutně událost změny vlastnosti (látka se nestala červenou, byla červená a bude červená i po té, co se býk rozběhne). ([106])

Příčina musí v čase předcházet následku; kauzalita má specifikovaný směr: X ↝ Y a X ↜ Y je něco jiného. Pokud uvažujeme o jednotlivých případech a data obsahují pozorování systému v čase, časové pořadí je jasné. Jiné metody usuzování o kauzalitě, která nevyžadují data o plynutí času, zahrnují další znalosti, aby našly časové pořadí a odlišily následek od příčiny. ([106], [21])

Druhé kritérium je, že možné příčiny mění pravděpodobnosti svých následků; pozorované změny obvykle doprovázejí manipulaci v příčinách, ale ne *vice versa*. ([106], [21])

Další vlastnosti kauzality nejsou pro její definici podstatné, ale stojí za zmínku. Kauzalita je tranzitivní: X ↝ Y ∧ Y ↝ Z ⇒ X ↝ Z. Korelace mezi příčinou a následkem může nastat i z jiných důvodů než je kauzalita, ale pokud jedna příčina způsobuje následek, lze očekávat korelaci. Nakonec, příčina se liší od následku; nic nezpůsobuje sebe nebo část sebe. ([21])

◊

[TODO: předělat]

U psychologického měření je často kauzální propojení všech objektů a osob zapojených do měření je nejasné, a navíc obsahuje různé zpětné vazby; opakované měření ovlivňuje měřený objekt[[11]](#footnote-11): respondent si zapamatuje odpovědi, testováním si natrénuje měřenou vlastnost, či je mu prozrazen původně skrytý smysl měření. Psychologickým atributem navíc nelze obvykle manipulovat a tak kauzalitu experimentálně zjišťovat tím nejjednodušším způsobem (jak je to možné v přírodních vědách).

Popis mentálních procesů stojících za psychologickým měřením samozřejmě vyžaduje rozvoj atributových teorií, které popíší na úrovni jedince, co se děje „mezi“ atributem a testovým skóre. Je sotva možné konstruovat uvažování testovaného respondenta pomocí ne-kauzálního přístupu. Položky mají mnoho společných vlastností, a změna některé – např. písma, konkrétní formulace otázky, změna číslic na slova vyjadřující čísla… – by neměla ovlivnit výsledek, protože tyto vlastnosti nehrají kauzální roli ve vyvolání konkrétního procesu odpovídání. [21]

Je proto pochopitelné, že nedostatečně rozvinuté psychologické teorie označuje jako největší problém psychometriky řada autorů ([38], [3], [90], [X], [X]).

Je složité vytvářet a testovat teorie o atributech, ke kterým nemáme přímý epistemický přístup.

Tvrzení „atribut X způsobuje skór Y“ musí mít podmínky, za kterých je pravdivé. Tyto podmínky závisí na analýze termínů atributu, skóru a kauzality, a empiricky testovatelných implikacích jejich intepretací. Z jediného vztahu atribut ↔ skór nemůže usoudit na směr kauzality; teorie je vzhledem k datům podurčená. Validace mezisubjektové kauzality vyžaduje méně důkazů než vnitrosubjektová – tato kauzalita znamená, že obdržíme stejné výsledky napříč jedinci i napříč různými okolnostmi u jednoho jedince. Empirické důkazy pro toto se hledají těžko: vnitrosubjektový efekt nemusí existovat, či funguje jinak než mezisubjektový. Musíme tedy používat metodologie, které zahrnují dimenzi času, či manipulaci formátu položek. Nejlepší důkazy o validitě se sestávají z různých druhů studií, které potvrzují jednu intepretaci. [21]

K tomu všemu jsou pochopitelně potřeba vhodné statistické modely.

# Modely psychologického měření

[TODO: lepší úvod]

Pro modelování psychologického měření potřebujeme model, který dokáže modelovat proměnné, které nelze pozorovat přímo, a který dokáže modelovat kauzální vztahy. Takové podmínky splňuje strukturní modelování (*structural equation modeling*, SEM) a modely spadající do této kategorie, jako jsou modely s latentní proměnnou.

SEM se sestává z grafické a matematické reprezentace směrových statistických asociací mezi měřenými a latentními proměnnými. Tato reprezentace implikuje specifickou kovarianční strukturu, která se dá testovat vhodnými daty – zda je pozorovaná kovarianční struktura konzistentní s kovarianční strukturou odpovídající specifikovaným lineárním vztahům. Modely s latentními proměnnými se velice hodí pro psychologická měření, neboť psychologické atributy nemůžeme pozorovat přímo. Rozlišujeme dvě základní široké třídy modelů: formativní a reflektivní. ([114, 37, 21])

[TODO: popis reflektivních a formativních modelů]

Modely s latentními proměnnými fungují následujícím způsobem. Předpokládá se, že psychologický atribut má určitou strukturu; tento předpoklad by měl být podpořený teorií a ověřený testováním vhodných reprezentačních axiomů[[12]](#footnote-12). Struktura může být kategorická, pořadová či spojitá. Na těchto nepozorovatelných, latentních proměnných statisticky závisí množina proměnných pozorovaných, které mohou být rovněž kategorické, pořadové či spojité. ([21])

V psychometrice se nejčastěji předpokládá atribut s kvantitativní strukturou, a tudíž latentní proměnná v modelu je spojitá. Model vztahuje pozorovanou odpověď na položku *U* k latentní proměnné θ pomocí funkce ([38]):

Kde g je spojovací funkce; *E*(*Uij*) je interpretováno jako očekávaná hodnota odpovědi respondenta *i* na položku *j*, nebo jako očekávaný poměr odpovědí na položku v populaci respondentů s pozicí θ*i* na latentní proměnné; α*j* a β*j* jsou regresní koeficienty parametry položky. ([38], [97])

Model nelze testovat kvůli přítomnosti latentní proměnné přímo, pouze nepřímo díky implikacím pro sdružené rozdělení pravděpodobnosti odpovědí na položky; odpovědi položky jsou nezávislé, podmíněné pouze latentní proměnnou; máme-li vektor *K* odpovědí označený jako *Yi*, s pozorovanými hodnotami *yi* = (*yi*1, *yi*2,… *yiK*) respondenta *i* s hodnotou latentní proměnné θ*i*, platí ([97]):

### Raschův model

([97])

Nejzákladnějším a nejrozšířenějším IRT modelem je Raschův model, jednoparamertrový logistický model. Má tedy kromě latentní proměnné jediný parametr obtížnosti položky a logistickou spojovací funkci. Pravděpodobnost správně zodpovědné položky je proto udávána jako:

Parametr obtížnosti bk udává bod na škále latentní proměnné, který odpovídá pravděpodobnosti 50 %, že respondent odpoví správně. Důležitou vlastností modelu je, že charaketritické funkce položek s různými obtížnostmi se navzájem nekříží. To znamená, že pro položky stejné zvýšení v úrovni latentního rysu vede ke stejné zvýšení pravděpodobnosti úspěchu – položky diskriminují mezi respondenty stejně.

[TODO: graf charakteristických křivek]

### Dvouparametrový model

([97])

Tento model přidává diskriminační parametr:

Důsledkem je, že charakteristické funkce položky mají proměnlivý sklon a položky se tak nevztahují k latentní proměnné všechny stejně. Čím je vyšší hodnota diskriminačního parametru, lépe je položka schopná rozlišit mezi nízkými a vysokými úrovněmi hodnotami latentní proměnné. Diskriminace položky se vztahuje k její obtížnosti: položka s vysokou rozlišovací schopností je užitečná v užší oblasti, která odpovídá oblasti obtížnosti, než položka s nízkou rozlišovací schopností.

Existuje rovněž probit varianta modelu:

Kde je kumulativní distribuční funkce normálního rozdělení a je funkce hustoty pravděpodobnosti normálního rozdělení.

Výraz je někdy uváděn jako , kdy je obtížnost uváděna na stejné škále jako latentní proměnná (podobně jako v Raschově modelu) a představuje bod s 50% pravděpodobností správné odpovědi.

### Tříparametrový model

([97])

Dvouparametrový model lze rozšířit přidáním parametru uhádnotelnosti tím, že zavedeme nenulovou spodní asymptotu charakteristické funkce položky:

### [Bayesovský přístup]

[TODO]

### [Informační funkce]

[TODO]

◊

[TODO: sjednotit s popisem kauzality výše]

To, že je odpověď na položku realizací náhodné proměnná, lze interpretovat dvěma způsoby ([38], [21], [97]):

* Stochastický subjekt – Funkce odpovědi na položku určuje pravděpodobnost, s jakou respondent s určitou pozicí na latentní proměnné odpoví na danou položku. Vyžaduje myšlenkový experiment s brainwashingem. To předpokládá kauzální relaci na vnitrosubjektové úrovni.
* Mezisubjektový pohled – Soustředí se na charakteristiky populace, a nevyžaduje tak brainwashing. Pravděpodobnostní rozložení odpovědí na položky, podmíněné latentní proměnnou, je koncipováno jako opakované vzorkování z populace se stejnou hodnotou latentní proměnné; funkce odpovědi na položku určuje očekávání odpovědí na položku v populaci s určitou pozicí na latentní proměnné. Očekávaná hodnota tak může být brána jako subpopulační průměr. To předpokládá kauzální relaci na mezisubjektové úrovni.

Mezisubjektový model latentní proměnné nemá jak popsat vnitrosubjektové procesy, které vedou k odpovědím. Vnitrosubjektový model kauzality by šel vytvořit pro takové proměnné, které jsou schopné empirické konkatenace, nebo pokud bychom dokázali, že mezisubjektová a vnitrosubjektová struktura modelů je totožná. To je ovšem další předpoklad – nelze tvrdit, že model měření platí pro každého respondenta; rozhodně nelze tvrdit na základě mezisubjektových dat, že stejná struktura platí pro všechny jedince. ([38], [21])

Vhodná teorie může popsat, jakým způsobem latentní proměnná může vysvětlovat jak mezisubjektové, tak vnitrosubjektové rozdíly. Takovými psychologickými konstrukty může být nálada či úzkost. Je tedy možné uvažovat, zda dimenze rozdílů v  náladě osoby v průběhu času a dimenze rozdílů této osoby a ostatních lidí jsou jedna a ta samá. Takový konstrukt nazýváme lokálně homogenní, ale v psychologii nebude příliš častý, protože předpokládáme řadu individuálních odlišností: je možné, že v průběhu času budou různí lidé vykazovat jiné vzorce změn: pro někoho může být úzkost jednodimenzionální, pro někoho dvoudimenzionální. Lokálně heterogenní konstrukty tak mají stejnou dimenzionální strukturu mezisubjektově, ale odlišnou na úrovni jedinců. ([38])

Lokálně irelevantní konstrukty jsou ty, které předpokládáme stabilní. Pokud je např. osobnostní rys stabilní, nebude vykazovat žádnou strukturu; kovariance opakovaných měření nebude souhlasit s latentním modelem, protože jediný rozptyl bude chybový rozptyl. Latentní proměnná by pak fungovala pouze s mezisubjektovou variabilitou, nikoliv s vnitrosubjektovou. ([38])

Jak konceptualizovat realistickou ontologii latentních proměnných, pokud je nemůžeme nalézt na úrovni jedince? Může se jednat o emergentní vlastnost populace skládající se jedinců. To neznamená, že neexistuje vztah mezi procesem na úrovni jedince a mezisubjektovou latentní proměnnou – pokud by tato na vnitrosubjektovém procesu nějakým způsobem nezávisela, nemohl by existovat zdroj mezisubjektové variability. Otázkou pak je, které procesy generují které mezisubjektové latentní struktury? Např. jaké druhy procesů jsou kompatibilní s teorií inteligence? ([38])

[TODO: přidat nějaký závěr]

# Design referenční škály

Přestože pro měření psychologických atributů není nutné, aby měly kvantitativní strukturu, použití určitého druhu škály je možné jen za správných podmínek. Psychologové často (prakticky téměř vždy) používají k popsání psychologických atributů intervalovou škálu. Michel ([117]) proto oprávněně tvrdil (a možná si to myslí dodnes), že psychometrika je patologická věda, a to z toho důvodu, že psychologové netestují hypotézy, že jimi „měřené“ atributy mají kvantitativní strukturu. A co víc, tento svůj omyl schválně ignorují. [52] dokonce tvrdí, že právě proto je tolik výsledků výzkumů jen „pomíjivým popisem situaci, které již nikdy nenastanou a které lze popřít téměř jakoukoliv replikací výzkumu“.

Problémem použití axiomatického reprezentačního přístupu je, že je deterministický, a neposkytuje způsob, jak se vypořádat s chybou ([X], [X], [X]); to vedlo některé psychology k úvaze, že RTM nemá praktické uplatnění ([X], [X], [X]). Přístup latentních proměnných řeší problém chyby měření tím, že předpokládá, že data obsahují šum, a modeluje očekávané hodnoty pozorovaných proměnných. Tím překonává obtíže s vytvořením deterministického relačního systému. Model s latentními problémy schovává relační systém pod „nepozorovatelnost“ proměnných, tudíž fakt, že relační systém nelze vytvořit není dále potřeba řešit ([21]). Navíc, Raschův model má unikátní vlastnosti, které dovolují následující úvahu.

Borsboom a Mellenbergh ([118]) tvrdí: psychologové nejsou líní, ale hypotézu o kvantitativnosti nelze testovat samostatně. Raschův model je pravděpodobnostní variantou spojeného měření, protože latentní proměnná představující hodnotu atributu a obtížnost položky jsou nezávislé, a struktura modelu je díky tomu aditivní. Předpoklad kvantitativnosti je pak jedním z předpokladů tohoto modelu, a pokud je model úspěšně testován na empirických datech (tj. projde nějakým testem vhodnosti), můžeme pak onen předpoklad považovat za odůvodněný.

Tato myšlenka není nová, podobnost Raschova modelu s teorií spojeného měření byla zmiňována prakticky od první publikace obojího ([116, 115]). Bond a Fox ([119]) dokonce svoji knihu o Raschově modelu doplnili podtitulem „fundamentální měření v sociálních vědách“; jakoby jakýkoliv model otestovaný testem vhodnosti automaticky zajišťoval kvantitativnost latentní proměnné a tím měřeného atributu, jak tvrdí třeba Embretson a Reise ([121]).

Faktem ovšem je, že Raschův model bude ukazovat vhodnost, i když měřený atribut bude mít pouze pořadovou strukturu[[13]](#footnote-13). Testy vhodnosti modelu obecně trpí problémem podurčenosti – existuje mnoho způsobů, jak mohou vzniknout data, která mají stejnou strukturu, jaká odpovídá hypotetizovanému modelu ([123], [38], [X]). Odhadnutí Raschova modelu prostě netestuje jednotlivé axiomy spojeného měření ([120], [123]).

Často je také zdůrazňováno ([125], [126]), že proponenti Raschova modelu se nesnaží najít model, který odpovídá datům, ale hledají data, která odpovídají modelu, a toho jsou schopni jenom díky tomu, že Raschův model obsahuje chybu. Michel ([125]) toto nazývá „Raschův paradox“ – pokud bychom odebrali chybu, dostaneme Guttmanův ordinální model ([X]); ale odebrání chyby přeci nemůže snížit přesnost měření. Pánové se v tomto pletou; fenomén stochastické rezonance[[14]](#footnote-14) ukazuje, že chyba *může* přesnost zvyšovat, a to, podle některých autorů, i v psychologickém měření ([127]). Samozřejmě není apriorně jasné, proč by se stochastická resonance, fenomén na neurologické úrovni, měla vztahovat i na makroúroveň odpovědí na položky ([122]); závěrem však je, že argument „Raschova paradoxu“ neobstojí.

Kyngdon ([120]) přidává ještě filozofickou námitku: přestože Raschův model empiricky sedící na data matematicky odpovídá axiomům spojeného měření, nelze ho za případ spojeného měření považovat. Když odhadneme model, získáme relační struktury odhadnutých parametrů osob a odhadnutých parametrů položek, které jsou seřazeny pomocí pravděpodobností správných odpovědí určitou osobou na určitou položku. Obě tyto struktury jsou symbolické – jenže spojené měření vyžaduje, aby spojitě měřené atributy byly empirické; tento argument lze odmítnout, pokud neuvažujeme o spojeném měření jako o zástupci reprezentační teorie měření, ale o „pouhém“ matematickém nástroji ([43], [124]).

Michel ([54]) prosazuje právě takový postoj k teorii spojeného měření: jedná se o sadu matematických teorémů, které se dají použít v jakémkoliv kontextu, kde se vyskytují takové struktury, které teorie popisuje. Jak se vztahuje k psychologickým atributům, je třeba objevit; psychologové musí najít situace, ve kterých se vyskytují dvojice atributů, které spolu neinteraktivně interagují, a pak tyto situace podrobit testování. Teorie spojeného měření je nicméně teorie o atributech, ne teorie o datech. Data vždy obsahují určitou chybu, a protože spojené měření se chybou nezajímá, není to teorie o pozorováních; jinými slovy, je to teorie o struktuře, není to teorie o kauzálních procesech.

Pokud tedy chceme Raschův model použít jako model atributů odpovídajících spojenému měření, jak provedeme testy axiomů?

Především, potřebujeme atributovou teorii, na jejímž základě můžeme tvrdit: a) pro jakoukoliv dvojici položek *j* a *k* nejenže je *j* obtížnější než *k*, díky určitým rysům položek, a navíc b) pro jakoukoliv čtveřici položek *g*, *h*, *j* a *k*, rozdíl obtížností mezi *g* a *h* je *r*-krát (kde *r* je reálné číslo) větší než rozdíl v obtížnosti mezi *j* a *k*, díky určitým rysům položek ([54]). Pro samotné testy axiomů, které ověří tato tvrzení atributové teorie, navrhlo několik autorů ([128], [129], [130], [131]) postup, který se nyní popíšeme.

Jedná se o Bayesovský postup, který spočívá v tom, že axiomy spojeného měření jsou stochasticky vynucené pomocí navrhovacího rozdělení (*jumping distribution*) Metropolisova–Hastingsova algoritmu.

Předpokládejme, že máme dva psychologické atributy: schopnost *B* a obtížnost položky *D*; *B* = {*b1*, *b2*, *b3*,…} a *D* = {*d1*, *d2*, *d3*,…} jsou neprázdné a . Relace ≽ je totální uspořádání definované na kartézském součinu *B* × *D*. Systém *E* = 〈*B*, *D*, ≽〉 je Raschovský aditivní spojený systém, pokud ≽ splňuje tyto podmínky:

* Neostré uspořádání
* Jednoduché vyrušení
* Dvojité vyrušení
* Řešitelnost
* Archimédovská podmínka
* Nezbytnost
* Symetrie

Pokud sytém E je Raschovským aditivním spojeným systémem, pak existují funkce ϕB a ϕD do systému takové, že pro všechna *a*, *b* ∈ *B* a *p*, *q* ∈ *D*

Pokud a jsou jiné funkce splňující výše uvedený výraz, existují reálná čísla α > 0, β*B* a β*D* taková, že

a

Relace ≽ je uspořádání „spojeného efektu“ schopnosti osoby a obtížnosti položky, a dá se vyjádřit jako pozorovaný poměr správných odpovědí. Tato data o poměru správných odpovědí se dají použít k testování výše uvedených axiomů (tedy těch, které jdou empiricky testovat). Tudíž, k ustanovení toho, zda mají atributy *B* a *D* kvantitativní strukturu, jsou použita empiricky získána data, která tak nejsou použita k odhadu parametrů Raschova modelu. Množiny *B* = {*b1*, *b2*, *b3*,…} a *D* = {*d1*, *d2*, *d3*,…} představují úrovně empirických atributů a neodpovídají tak Θ = { θ*1*, θ*2*, θ*3*,…} a Β = {β*1*, β*2*, β*3*,…}, což jsou množiny numerických parametrů.

Nyní si představme, že máme nějaké osoby, kteří zodpovídají položky testu. Osoby a položky můžeme roztřídit a seřadit do tabulky, tak že dostaneme N skupin osob podle toho, kolik položek zodpověděli správně (seřazených od nejméně úspěšné skupiny po nejvíce úspěšnou), a M skupin položek podle toho, kolikrát byly zodpovězeny správně (seřazených od nejtěžších po nejlehčí). Skupiny osob a položek představují úrovně atributů schopnosti a obtížnosti (jak bylo popsáno výše):

Můžeme vytvořit matici *B* × *D*, která obsahuje *true* hodnoty poměrů správných odpovědí Θ. Tyto pravé hodnoty samozřejmě neznáme; ale máme k dispozici data, ze kterých vytvoříme tři *B* × *D* matice:

kde *y* je počet správných odpovědí a *Y* je celkový počet odpovědí. Matice *C* je tedy maticí správných odpovědí, *T* je matice celkových počtů odpovědí, a je poměr správných odpovědí, tedy *maximum likelihood* odhad Θ. Tyto matice lze znázornit jako tabulku:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Nejtěžší |  |  |  | Nejlehčí |
|  | *d1* | *d2* | *d3* | … | *dM* |
| *b1* |  |  |  | … |  |
| *b2* |  |  |  | … |  |
| *b3* |  |  |  | … |  |
| … | … | … | … | … | … |
| *bN* |  |  |  | … |  |

Axiomy jednoduchého a dvojitého vyrušení testujeme na všech možných 3 × 3 submaticích, které lze z *B* × *D* matice o rozměru *N* × *M* vytvořit. Celkový počet takovýchto submatic je

Cílem je na základě dat získat odhad Θ vždy pro vybranou 3 × 3 submatici. Posteriorní rozdělení podmíněné daty lze vyjádřit Bayesovým teorémem:

kde je priorní rozdělení, které vynucuje omezení uspořádání implikovaná axiomy jednoduchého a dvojitého vyrušení. Toto rozdělení omezuje tak, že leží ve vlastní podmnožině :

*Likelihood* je součinem nezávislých binomických rozdělení:

To předpokládá nezávislost dat podmíněně na parametrech v .

Kvůli jmenovateli posteriorního rozdělení je třeba jejímu odhadu použít MCMC algoritmus, konkrétně variantu Metropolisova–Hastingsova algoritmu. Priorní rozdělení , které vynucuje omezení uspořádání, je implementováno pomocí specifických návrhových rozdělení M–H algoritmu.

Kandidátská hodnota v kroku *t* je vybírána z návrhového rozdělení, které má obecnou podobu rovnoměrného rozdělení:

Počáteční hodnota pro *l* je 0, pro *r* je 1. Hodnoty a implementují jednoduché vyrušení, a dvojité vyrušení, ale to pouze u buněk, u kterých je to relevantní. Pokud , je místo použita 1.

Protože konkrétní hodnoty a jsou závislé na kroku *t*, znamená to, že všechny buňky v 3 × 3 submatici jsou aktualizovány současně[[15]](#footnote-15), postupujeme po sloupcích *i* a řádcích *j*:

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Tj. buňky aktualizujeme v pořadí (1, 1), (1, 2), (1, 3), (2, 1),… (3, 2), (3, 3). Pro všechny buňky platí omezení dle jednoduchého vyrušení; dvojité vyrušení se vztahuje pouze na buňky (1, 3) a (3, 1), a to pouze ve specifických případech, jak si povíme později.

Obecně platí pro buňku (*i*, *j*) v kroku *t*:

Tyto hodnoty pro výše uvedené návrhové rozdělení implementují axiom jednoduchého vyrušení; navrhovaná kandidátská hodnota tak bude spadat do intervalu omezeného hodnotami okolních buněk tak, aby uspořádání odpovídalo podmínkám axiomu.

Pro buňku (1, 3) v kroku *t* navíc platí:

A pro buňku (3, 1) v kroku *t* platí:

Tyto hodnoty pro výše uvedené návrhové rozdělení implementují axiom dvojitého vyrušení[[16]](#footnote-16); pokud jsou splněny nerovnosti, je interval, do kterého může spadat kandidátská hodnota dále omezen.

Jakmile je zvolena kandidátská hodnota , spočítáme pravděpodobnost jejího přijetí[[17]](#footnote-17):

Pak postupujeme klasicky podle M–H; vybereme náhodné číslo a zvolíme hodnotu nového vzorku:

Po dokončení vzorkování sestrojíme 95% intervaly věrohodnosti pro jednotlivé z *Θ*. Pokud platí , tak jsme pro buňku (i, j) nenašli narušení uspořádání, na základě kterého by šlo odmítnout axiomy jednoduchého či dvojitého vyrušení. V ideálním případě bychom takto otestovali všech 3 × 3 submatic; protože to však může být výpočetně příliš náročné, můžeme vybrat jen ty submatice, jejich 9 buněk spolu sousedí, nebo vybrat náhodný vzorek ze všech možných submatic.

Podpora takto popsaného modelu implikuje, že položka z > položka y > položka x, kde relace > znamená „je lehčí než“ a skupina c > skupina b > skupina a, kde relace > znamená „má více schopností než“. Protože „jednoduchost“ položky je množství schopnosti nutné pro její správní zodpovězení, členové množin *B* a *D* jsou v podstatě instancemi stejného atributu (např. nějaké kognitivní schopnosti).

[TODO: odkaz na budoucí simulační studii]

Přestože výše popsaná metoda poskytuje jisté možnosti, jak testovat předpoklad kvantitativnosti u psychologických atributů, má stále několik nedostatků. Nemůžeme testovat axiom řešitelnosti (neboť vyžaduje, abychom byli schopni „vytvořit“ položku či osobu pro doplnění trojice[[18]](#footnote-18)) a archimédovský axiom (neboť mezi každými dvěma úrovněmi atributu schopnosti i obtížnosti bychom museli být schopni najít další). Rovněž se ukazuje, že podmínka jednodimenzionality, stejně jako neinteraktivnosti, dvou atributů, se ve skutečných situacích těžko zajišťuje ([TODO: doplnit příklady: více procesů při odpovídání na položku, interakce např. u rychlosti a kvality odpovědi]).

[TODO: lepší přechod]

# Nejistota psychologického měření

[TODO: rozvést]

Rozeznáváme tři hlavní zdroje nejistoty měření[[19]](#footnote-19):

* Definiční nejistota
* Přístrojová nejistota
* Interakční nejistota

## Definiční nejistota

[TODO: rozvést]

Definiční nejistota vzniká, když máme k dispozici pouze určité množství informací určené definicí měřené veličiny. Není tak vždy jasné, nakolik výsledek měření popisuje měřený atribut. Proč tomu tak je, jsme si ukázali v předchozích kapitolách.

## Přístrojová nejistota

[TODO: rozvést]

Přístrojová nejistota je nejistota vzniklá použitím určitého přístroje; v psychologickém měření může zdánlivě splývat s definiční nejistotou, díky přílišnému spoléhání se na operacionalistický definice: pokud k měření určitého psychologického atributu existuje pouze jeden test, těžko usuzovat, jaký podíl celkové nejistoty je způsobem právě tímto testem.

Přístrojová nejistota má řadu zdrojů:

* Malý počet položek
* Použité technologie (papír vs. počítače)
* Interakce přístroje s prostředím (např. odlesky slunce na lesklém displeji)
* Chyby administrátora (kterého považujeme za součást měřicího systému)

Velkou část těchto chyb lze vyřešit důslednou standardizací.

## Interakční nejistota

[TODO: rozvést]

vznikající díky interakci mezi měřeným objektem a měřicím systémem. V případě psychologického měření má tato interakce velkou důležitost, neboť měření obvykle nelze opakovat. A pokud existují alternativní formy testu, stále mohou vznikat systematické rozdíly v naměřených hodnotách; důvody tohoto retestového efektu popisuje [137]:

* Rozdílná motivace při opakovaném testování
* Snížení nervozity z testové situace
* Zvýšení znalostí o testu
* Zapamatování si položek testu
* Zvýšení úrovně měřené schopnosti

◊

[TODO: závěr]

# Design přiřazování hodnot

[TODO: v podstatě technické záležitosti]

## Měřicí přístroj

[TODO: dodělat]

### Počítačové adaptivní testování

[TODO]

### Využití počítačů

[TODO: rozvést]

K využívání počítačů – či obecně zařízení[[20]](#footnote-20) – je mnoho důvodů:

* Snadnost zaznamenávání dat, jejich statistická analýza a automatické generování zprávy
* Možnost zaznamenávat i jiná data než odpovědi na položky (např. čas, sledování pohybu očí, vzorce používání ovládacích prvků, atp.)
* Variabilní formát položek
* Počítačové adaptivní testování

Zároveň je nepraktické, aby test, standardizovaný pro určitý počítač, byl administrovaný pouze na stejném druhu tohoto zařízení – dá se těžko předpokládat, že psychologové-administrátoři budou mít k dispozici vždy stejné zařízení, na kterém byl test kalibrován. Proto je nutné vzít v úvahu všechna úskalí, která testování s pomocí počítačů obnáší. Konkrétně se jedná o následující problémy (dle Naglieri et al., 2004, Širůček, 2010):

1. Neekvivalence grafických podnětů. „Stejné“ položky mohou na respondenta působit jinak na papíře a na displejích. Rovněž displeje samotné se u různých zařízení liší v řadě parametrů: velikost uhlopříčky, rozlišení, hustota pixelů, kalibrace barev či obnovovací frekvence. To činí velice obtížným, či přímo nemožným plně standardizovat grafickou podobu položek.
2. Specifika prohlížečů. Každé zařízení má jiný software, od operačního systému po webový prohlížeč, které se liší v nesčetném množství detailů.
3. Kompetence respondenta ovládat zařízení, na kterém test probíhá. Pokud uživatel neovládá práci s počítačem, negativně to ovlivní jeho výkon.
4. Odlišnosti ovládání. Různá zařízení lze ovládat mnoha způsoby: klávesnicí, myší, touchpadem, nebo dotykovým displejem; každý vede ke kvalitativně odlišnému zpracovávání – jinak se vybírá odpověď myší a jinak dotykovým displejem, což může hrát roli především v testech měřících čas. Ani způsob ovládání tak nelze standardizovat napříč zařízeními.

Tyto problémy jsou vesměs řešitelné třemi způsoby.

Prvním je zahrnout proměnné, mající případný vliv na přesnost měření, do bayesovského modelu diagnostiky odlišného fungování položek (*differential item functioning*, DIF) – to se týká především kompetencí respondenta, ale i rozličných vlastností zařízení (např. rozlišení a velikosti displeje).

Druhou možností je zahrnout vlastnosti zařízení do modelu měření. Tedy rozhodnout, které vlastnosti ovlivňují parametry položek a které nikoliv, a zda či jak tyto vlastnosti zařízení modelovat. Např. u „obyčejného“ dotazníku nejspíš na konkrétní grafické podobě nebude záležet; opačným extrémem naopak může být metoda založená na vnímání barev. Stejně tak zatímco text lze efektivně zpracovat a chápat i na malém displeji (Dillon, Richardson & McKnight, 1990), u grafických podnětů toto platit nemusí.

Třetí variantou je vypořádat se s rozdíly mezi zařízeními po technické stránce, důkladnou standardizací. Především by měl být využíván princip transparentní imediace (viz Bolter & Grusin, 2005): práce s testem musí být přirozená a bezprostřední, což klade velké nároky na UX. Inspiraci můžeme najít u tvůrců internetového obsahu (webových stránek, aplikací, apod.). Web developeři vcelku úspěšně řeší problém, jak se vyrovnat s různými vlastnostmi zařízení tak, aby jimi vytvářený obsah byl stále stejně použitelný, pomocí tzv. *responsive design* – rozvržení prvků obsahu se pružně přizpůsobuje použitému zařízení, přičemž tato změna není zcela jednoduchá a automatizovaná, ale spočívá v optimálním a adaptivním využití parametrů použitého zařízení (viz Marcotte, 2010). Podobně je řešitelný i problém s dostatečnou kompetencí uživatele testu – tu lze ověřit přidáním položek, které by kontrolovaly respondentovy dovednosti s ovládáním používaného zařízení. V posledních letech se navíc rozšířilo ovládání pomocí dotykových displejů, které je užitečné díky své intuitivnosti – je vhodné pro rychlou necvičenou interakci (Taveira & Choi, 2009), a tedy na jednoduchou manipulaci s prezentovaným podnětovým materiálem.

[TODO: závěr]

# [Implikace]

[TODO: rozvést]

Jakým způsobem tedy vytvořit validní měření psychologických atributů tak, aby splnilo všechny náležitosti definice z minulé kapitoly?

Psychologové a psychometrikové se historicky soustředili na kvantitativnost psychologických atributů; jak jsme si ukázali, toto téma není klíčové: pro měření není důležité, jakou má měřený atribut strukturu, ale zda je teoretický předpoklad určité struktury empiricky ověřený, tj. lze dokázat morfizmus. V posledních letech se pro tento účel rozvíjí řada Bayesovských metod pro testování pořadových axiomů.

Problém tedy není ve struktuře atributů, ale v atributech samotných, přesněji v neexistenci dostatečně podrobných atributových teorií. Ty se samozřejmě rozvíjí pomalu, neboť je těžké je *experimentálně* ověřovat. V psychologickém měření je proto klíčovým odhalování psychologických procesů; jak jsme si ukázali, běžné IRT modely nejsou pro testování vnitrosubjektové kauzality vhodné. Při „běžném“ experimentu manipulujeme se nezávislou proměnnou a sledujeme, jaký to má vliv na proměnnou závislou ([45]). Psychologickým atributem však většinou manipulovat nemůžeme; nelze zvýšit či snížit inteligence člověka, či ovlivnit efektivitu jeho pracovní paměti. Vědecký pokrok je samozřejmě možný i bez experimentů, astronomie i teorie evoluce se bez nich obešly. [47] To ale vyžaduje rigoróznost na mnohem vyšší úrovni, než jaké jsou – zdá se – psychologové schopni, jak – mimo jiné – ukazuje nedávná „kauza“ s replikovatelností ([138]). Částečně je to jistě tím, jak podotkl Borsboom ([49]), že psychologové nejsou schopni používat aktuální matematické nástroje psychometriky – a komplexnější modely psychologického měření matematické schopnosti vyžadují ještě více

V každém případě máme k dispozici dvě alternativní metodologie ([21], [X], [X]):

* Opakovaná měření – tj. sledování jedince v průběhu času.
* Manipulace formátem položky.

Protože druhá možnost je finančně a časové méně náročná, budu se v další kapitole věnovat právě jí, tedy automatickému generování položek. Výhoda tohoto přístupu k tvorbě testů je, že máme k dispozici informace o tom, jak byla určitá položka vytvořena, tj. na čem závisí její parametry, což umožňuje lépe zkoumat kauzalitu a tudíž validitu ([108]). Máme tak možnost systematicky variovat obtížnost položek

[TODO: závěr]

# Závěr

[TODO]

# Bibliografie

[TODO]

1. Viz kapitolu X. [↑](#footnote-ref-1)
2. Viz kapitolu X. [↑](#footnote-ref-2)
3. Viz kapitolu X. [↑](#footnote-ref-3)
4. Viz kapitolu X. [↑](#footnote-ref-4)
5. Viz kapitola X. [↑](#footnote-ref-5)
6. Viz kapitolu X. [↑](#footnote-ref-6)
7. Viz. kapitolu X. [↑](#footnote-ref-7)
8. [↑](#footnote-ref-8)
9. [TODO: Mr. Brown] [↑](#footnote-ref-9)
10. Intencionalita je definována jako schopnost mysli vytvářet reprezentace. Mentálních stavy mají vlastnost být o něčem nebo se na něco zaměřovat; vypovídají „o něčem jiném“, a zahrnuje to tudíž naděje, strachy, přání, názory, nejenom plány či záměry ([96]). [↑](#footnote-ref-10)
11. Toto se samozřejmě děje i ve fyzikálním změření: použití teploměru změní teplotu měřeného objektu, ale v prakticky zanedbatelné míře. [↑](#footnote-ref-11)
12. Viz kapitoly X a X. [↑](#footnote-ref-12)
13. [TODO: odkaz na simulace] [↑](#footnote-ref-13)
14. [TODO: definice] [↑](#footnote-ref-14)
15. Tj. najednou vzniká devět Markovových řetězců. [↑](#footnote-ref-15)
16. Pokud bychom hodnoty a nepoužil a testovali pouze jednoduché vyrušení, ověřovali bychom, zda mají atributy alespoň pořadovou strukturu. [↑](#footnote-ref-16)
17. V praxi samozřejmě používáme log-likelihood, kvůli lepším výpočetním charakteristikám; použití likelihood vede k součinům mnoha hodnot menších než jedna, a brzy bychom narazili na omezenou přesnost reprezentace takových čísel. [↑](#footnote-ref-17)
18. Viz. kapitolu X. [↑](#footnote-ref-18)
19. Viz kapitolu X. [↑](#footnote-ref-19)
20. Slovo počítač dle mého názoru příliš evokuje klasický desktop – proto budu dále používat termín zařízení, kterým zahrnuji nejen desktopy, ale i laptopy a různé handheld přístroje, jako jsou tablety, telefony či e-inkové čtečky. [↑](#footnote-ref-20)