

FORLAG FOR PSYKOLOGISK FORSKNINGSMETODE

Metodologiske indblik og udsyn

*festskrift til
Benny Karpatschof*

Metodologiske indblik og udsyn

Festskrift til Benny Karpatschhof

Metodologiske indblik og udsyn

Festskrift til Benny Karpatschhof

Redigeret af
Anders Bech Christiansen, Tytte Hetmar,
Jan Ivanouw og Henriette Kirkeby

Forlag for Psykologisk Forskningsmetode

©Copyright 2008 forlaget og forfatterne.
ISBN 978-87-992738-0-5

Lay-out og opsætning: Jan Ivanouw.

Omslag og grafik: Anders Bech Christiansen
Grafisk konsulent på omslag: Annemette Blomgren Larsen

Sat med fonten Computer Modern
ved hjælp af pakken Memoir i tekstbehandlingsprogrammet L^AT_EX

Udgivelsen er støttet af Psykologisk Institut, Københavns Universitet.

Forlag for Psykologisk Forskningsmetode,
Svankærvej 2
2720 Vanløse

Bogen kan hentes på websiden <http://www.forskningsmetode.dk>

Indhold

| | |
|--|------------|
| Indhold | 5 |
| Indledning | 7 |
| Tytte Hetmar og Jan Rattleff: Metodeundervisningen på Københavns Universitets psykologistudium 1968-2008 | 15 |
| Nini Prætorius: Intersubjectivity, Cognition, and Language. Epistemological conditions and their methodological consequences | 27 |
| Boje Katzenelson: Tingene er, som de er, fordi det er bedst sådan: Om kausalitet og finalitet i sjælelivet og intentionalitetens kilder | 51 |
| Simo Køppe: Det psykofysiske problem – superveniens og emergens | 81 |
| Jens Mammen: Om betingelserne for objektiv erkendelse af kvantitative forhold. En praktisk-realistisk teori | 97 |
| Jan Ivanouw: Statistiske niveauer og Structural Equation Modeling i psykologisk forskning og til beskrivelse af behandlingseffekt | 117 |
| Svend Kreiner: Statistiske modeller – forenklinger, forvanskninger og/eller fortrængninger? | 153 |
| Erik Lykke Mortensen: On the choice between traditional item analysis and item response models in psychology | 179 |
| Guido Makransky: Computer Adaptive Testing: An introduction to the practical considerations related to developing a computer adaptive test for occupational testing | 197 |

| | |
|--|-----|
| Mette Høyer: Kontekst-begrebet i kvantitativ forskning, generaliserbarhed og ekstern <i>invaliditet</i> | 225 |
| John Brodersen: Udvikling af velfunderede, relevante og valide spørgeskemaer til sundhedsvidenskabelig forskning | 241 |
| Tom Teasdale: Intelligenstests og Flynn-effekten i Danmark | 253 |
| Jesper Hermann: Vores indtryks udtryk – et metodesyn på vore “sprog”vaner | 259 |
| Bjarne Sode Funch: Phi for livets skønhed | 273 |
| Tine Nielsen: Universitetspædagogisk praksis og videnskablighed | 283 |
| Forfatterliste | 295 |

Indledning

Dette festskrift er udgivet i anledning af at Benny Karpatschhof fyldte 70 år 22. oktober 2008. Samtidig er det også året hvor Benny har 40 års jubilæum som underviser i forskningsmetode og statistik på Psykologisk Institut ved Københavns Universitet¹.

Benny begyndte at læse psykologi, men skiftede senere over og blev uddannet i statistik. En af Bennys tidlige ansættelser siges at være inkasator i et forsikringselskab, formentlig anset for at være et studierelevant job for en kommende (forsikrings)statistiker. Mere relevant for Bennys senere karriere var dog nok en ansættelse i begyndelsen af 1960'erne i Miltærpsykologisk Tjeneste, og arbejdet som studentermedhjælper hos professor Georg Rasch på Statistisk Institut.

I begyndelsen af 1960'erne var Benny studentermedhjælper på Danmarks Pædagogiske Institut. Uddannet cand. stat. i 1967 blev han ansat på DPI som amanuensis, og fra 1968 på Psykologisk Laboratorium, Københavns Universitet, det senere Psykologisk Institut.

Igennem alle disse år har Benny ihærdigt arbejdet med at forsøge at forstå og sammenfatte psykologisk teori. Hans interesseområder strækker sig fra konkrete psykologiske emner, f.eks. intelligensforskning, til videnskabsteori og filosofi, hvilket blandt andet hans disputats fra 2000, *Human activity: contributions to the anthropological sciences from a perspective of activity theory*, vidner om. Han har også forfulgt denne interesse gennem sit lange virke som medgrundlægger af, medredaktør for og bidragyder til det psykologisk teoretiske tidsskrift *Psyke og Logos*. Han har samtidig igennem alle årene arbejdet med at systematisere og formalisere forskningsmetodologien, tidligt fremstillet i to grundbøger sammen med Jan Rattleff og Svend Kreiner, *Psykologisk Metodelære*, og *Psykologiens Metodelære*², fra henholdsvis 1974 og 1977, og senest i *Udforskning i psykologien. De kvantitative metoder* fra 2006. Benny er blevet karakteriseret som en "polyhistor" og en at de mest tværfagligt orienterede psykologer, vi har i Danmark.

¹I denne elektroniske udgave er der enkelte rettelser i nogle af bidragene i forhold til den trykte udgave

²Bemærk formuleringens tegn på en øget sikkerhed om hvad det hele drejer sig om.

Benny har ofte i sin undervisning samlet kolleger og studenterundervisere omkring sig. I dette skrift gennemgår Tytte Hetmar og Jan Rattleff historien bag et af de vigtigste af disse undervisningsteams, "Metodegruppen".

Benny har en meget stor faglig kontaktflade. Han har gennem alle årene været kendt som en engageret, humoristisk og energisk hjælper. Det har en lang række af studerende og kolleger nydt godt af. Han har ydet forskningsvejledning for en lang række post-graduate studerende og forskere fra andre fagområder, ofte i form af at restrukturere problemer, at gennemføre en begrebsanalyse så nye delbegreber opstår eller gamle undergår en syntese.

Som et eksempel nævner Jens Mammen, der er bidragsyder til dette festskrift, den hjælp han har modtaget fra Bennys, dels under arbejdet med at formulere Weibullfordelingen som alternativ til Raschs model for hundestejlers ventetider³, hvilket omtales i Bennys *Skalainvariante sandsynlighedsmodeller* fra 1971, dels Bennys funktion som leder af "Bennys hjælpeklasse" eller "MAD (Mammens Algebraiske Demonstrationer)" under Jens' arbejde i 1994-98 med at udforme det matematiske grundlag for en teori for sanse- og udvalgs kategorier.

Selv under det alvorligste faglige arbejde er Bennys humor et gennemgående træk. En udgivelse i universitetets skriftserie sammen med Henriette Elkjær om ipsativt scorede personlighedstests hedder således *Yet the bumblebee flies*.

En anden side af Bennys faglige aktivitet er engagement i organisatorisk arbejde, herunder entusiastisk medvirken i forbindelse med utallige nye studieplaner. I den forbindelse har Jan Ivanouw i 1969 sammen med Benny og en gammeldags bordregner med håndsving udviklet prognoser for gennemførelsestider for de studerende, og herunder var Benny i stand til at opstille en model som tog hensyn til individuelle studerende med særlige studievaner og gennemførelsesforløb.

Benny var medstifter af *Psykologisk Selskab for Forskningsmetodologi* i 1999, som siden har afholdt 2-3 arrangementer pr semester, heraf hyppigt med Benny som oplægsholder. Redaktionen af dette skrift udspringer af dette selskab.

En anden side af Bennys virke har bestået i forsøg på ikke kun at forstå verden, men også at forandre den. I den forbindelse har Benny været aktiv i mange sammenhænge, f.eks. inden for psykiatri- og galebevægelsen.

Dette festskrift består af indlæg fra personer med relation til Benny og den side af Bennys faglige interesseområde som vi bredt har sammenfattet i festskriftets titel: *Metodologiske indblik og udsyn*.

Der er blandt teksterne i festskriftet fem hovedtyper.

³som var emnet for Iven Reventlows disputats: *Studier af komplicerede psykobiologiske fænomener*.

Den første tekst, *Metodeundervisningen på Københavns Universitets psykologistudium 1968-2008*, beskriver en af hovedaktiviteterne i Bennys faglige virke, oprettelsen og driften af metodeundervisningen på Psykologisk Laboratorium, det senere Psykologisk Institut, Københavns Universitet. Jan Rattleff og Tytte Hetmar var i mange år en del af denne gruppe og har givet en historisk gennemgang af metodeundervisningen på psykologistudiet gennem de 40 år.

Den anden hovedgruppe af tekster har som samlende tema grundlæggende erkendelsesteoretiske problemer i psykologien. Nini Prætorius' artikel, *Intersubjectivity, Cognition and Language*, er et indlæg i den århundrede år gamle debat inden for filosofien om hvilken status, der kan tillægges vores erkendelse og beskrivelser af mentale og andre "indre" tilstande - såsom vores tanker, emotioner, forestillinger og følelser af smerte. I denne debat, som har fået fornyet aktualitet inden for vore dages bevidstheds- og neurokognitionsforskning, stilles bl.a. følgende spørgsmål: Givet at sådanne tilstande og fænomener ikke er offentligt observerbare, hvordan kan vi da være sikre på, at sådanne fænomener eksisterer, eller at vor erkendelse og beskrivelser af dem har den samme status af sandhed og gyldighed som vor erkendelse og beskrivelser af objekter og sagsforhold i den offentligt observerbare materielle virkelighed, og hvordan kan vi være sikre på, at beskrivelser af sådanne tilstande bruges på samme måde eller at de har de samme implikationer for forskellige personer og sprogbrugere? Eftersom indre mentale tilstande angår fænomener vi ikke kan være fælles med andre om at observere, hvordan går det da til, at vi overhovedet kan lære at tale om dem i et sprog vi er fælles med andre om? I artiklen argumenterer Nini Prætorius for at de grundlæggende betingelser for vores erkendelses og beskrivelsers intersubjektivitet er præcist de samme, hvad enten det drejer sig om erkendelse og beskrivelse af indre mentale tilstande eller om ting i den offentligt observerbare materielle virkelighed. Det vises tillige at på disse betingelser beror selve muligheden for en veldefineret skelnen imellem på den ene side erkendelse og beskrivelse af det der er offentligt og dermed objektivt observerbart, og på den anden det der er subjektivt eller personligt, af fundamental betydning for udviklingen af videnskaberne. I artiklens afsluttende afsnit benyttes argumenterne for de generelle betingelser for vores erkendelses og sproglige beskrivelsers intersubjektivitet til at argumentere for at den skepsis mod brugen af første-persons beskrivelser af mentale fænomener som data i den videnskabelige udforskning af bevidstheden, som er udbredt blandt neurokognitionsforskere, er grundløs.

I sit indlæg, *Tingene er, som de er, fordi det er bedst sådan: Om kausalitet og finalitet i sjælelivet og intentionalitetens kilder*, behandler Boje Katzenelson de psykiske egenskaber. Han argumenterer for at det er disse der bestemmer hvilke biologiske organer der opstår i evolutionens forløb. Han mener for eksempel at det er evnen at kunne se som er årsag til øjnes fremkomst og funktion. Artiklen omhandler det der

efter Aristoteles kaldes finale årsager og deres betydning for tilblivelsen af den humane psykes arts-karakteristiske fundamentale træk set i et evolutionært perspektiv. Dette kan ses som en metodisk vinkel i bredere erkendelsesteoretisk forstand. De finale årsager rummer indre mål som foregriber forholdet mellem subjekt og objekt, og denne foregriben eller intentionalitet er et vigtigt træk ved psykens væsen. I artiklen undersøger forfatteren intentionalitetens psykiske grundlag og ser den således at evolutionshistorien har designet os til i højere grad at være rettet mod og parat til at opfange og optimalt udnytte information fra visse dele af omgivelserne end fra andre.

Simo Køppes indlæg, *Det psykofysiske problem - superveniens og emergens*, behandler nogle fundamentale forhold ved psykens dannelse. I artiklen diskuteres det psyko-fysiske problem på baggrund af teorier om emergens. Specielt diskuteres om begrebet superveniens er en løsning på denne gamle filosofiske problemstilling. Der skitseres en alternativ version som kaldes "omvendt superveniens" som et forsøg på at forene psykens relative ontologiske selvstændighed, emergens, uden nedadgående kausalitet, og uden entydig determinisme.

En tredje samling af tekster vedrører spørgsmål vedrørende måling og anvendelse af statistiske metoder i psykologien.

Jens Mammen tager i sit indlæg, *Om betingelserne for objektiv erkendelse af kvantitative forhold. En praktisk-realistisk teori*, udgangspunkt i erkendelsesteoretiske synspunkter om hvorledes empirismen og rationalismen gennem tiderne har søgt grundlaget for den objektive erkendelse henholdsvis i elementære sanseindtryk og i forudgivne forståelsesformer. Artiklen præsenterer en tredje teori som begrundet objektiviteten i faste strukturer der opstår ved subjekternes møde med omgivelsernes mangfoldighed af objekter og strukturer der hænger sammen med erkendelsesapparatets redskabsmæssige karakter. Med dette udgangspunkt opstiller forfatteren en måleteori inspireret af den danske statistiker Georg Rasch, der i sin teori om specifikt objektive målinger i en matematisk form har præciseret at betingelsen for objektiv erkendelse er at resultatet af den beskrevne vekselvirkning skal være en såkaldt "latent additiv" funktion af redskabs- og objektparametre. Dette vil sige at den ved en matematisk transformation kan omskrives til en addition af reelle tal. Denne betingelse kan generaliseres ud over måling af metriske relationer, og den peger således på et alment vilkår for erkendelsens objektivitet.

Jan Ivanouw opdeler i sit indlæg, *Statistiske niveauer og Structural Equation Modeling i psykologisk forskning og til beskrivelse af behandlingseffekt*, de statistiske metoder i niveauer med stigende kompleksitet, men også med stigende grad af relevans for psykologisk forskning. Han gennemgår den deskriptive statistik, og dens udbygning med statistisk hypotesetestning. Han introducerer de egentlige statistiske modeller med regressionsanalysen som grundmodel og viderefører den i pathanalysen, som muliggør analyse af en struktur af influerende og medierende vari-

able. Med dette udgangspunkt introducerer han anvendelsen af latente variable inden for området structural equation modeling. Disse modeller omfatter både en målingmodel i form af konfirmatorisk faktoranalyse og en strukturmodel der kan beskrive relationen mellem en række latente variable. Disse forudsættes ikke, som ved metoderne på de tidligere niveauer, målt direkte gennem "operationalisering", men kan måles med flere forskellige samtidige metoder. Herunder tages der højde for at disse målinger princippielt er ufuldkomne - dels fordi de er behæftet med måleusikkerhed, dels fordi de oftest vil være systematisk influeret af variable der ikke indgår i modellen, og denne generelle usikkerhed indgår i estimerbar form i modellen. Hvor metoderne fra de tidligere niveauer ofte af psykologer er blevet oplevet som utilstrækkelige og ikke i tilstrækkelig grad i stand til at beskrive den kompleksitet der nødvendigvis indgår i psykologisk tænkning, er pointen i de statistiske modeller på dette niveau at tilbyde analyseredskaber der ligger tæt på den måde forskeren tænker i psykologiske begreber og teorier. Forfatteren sætter sin beskrivelse af de forskellige niveauer af statistiske metoder i relation til statistisk beskrivelse af udviklingsprocesser, herunder behandlingseffektundersøgelser.

Svend Kreiners artikel, *Statistiske modeller - forenklinger, forvanskninger og/eller fortrængninger*, samler en række løse tråde op fra Benny Karpatschofs definition af statistiske modeller som et udtryk for en bevidst forenklet beskrivelse af det objektsystem som man ønsker at undersøge nærmere ved hjælp af statistiske analyse af data. Modellerne er ifølge forfatteren ikke kun udtryk for forenklinger. De er også udtryk for både fortrængninger og forvanskninger, og de metodologiske problemer man står over for når man arbejder med statistiske metoder, er ikke kun et spørgsmål om hvorvidt man har fravalgt for meget eller forkert. De alvorligste problemer er problemerne med tolkningerne af resultaterne af de statistiske analyser som følge af at det ofte er uklart hvilke dele af de statistiske modeller der rent faktisk afspejler forhold i objektsystemet, og hvilke dele af modellerne der er referenceløse egenskaber ved modellerne. Forfatteren trækker bl.a. på sine erfaringer med itemresponsmodeller, Raschmodeller, grafiske modeller og modeller for strukturelle ligninger. Han opfatter imidlertid problemstillingen som generel og ser de metodeproblemer han beskriver som gældende for alle former for statistiske modeller.

Erik Lykke Mortensen tager i sin artikel, *On the choice between traditional item analysis and item response models in psychology*, udgangspunkt i det forhold at traditionel psykometri versus itemresponseanalyser er et spørgsmål der tit vækker stærke følelser. Det er efterhånden almindeligt accepteret at der er en række fordele forbundet med anvendelse af itemresponsmodeller ved udvikling og analyse af "rating scales" og kognitive tests. Alligevel er anvendelsen af disse metoder ikke uproblematisk, og artiklen diskuterer problemerne med særlig vægt på spørgsmålet om itemresponseanalyser af instrumenter der er udviklet med traditionelle

psykometriske metoder. Forfatteren konkluderer at det sjældent er hensigtsmæssigt at anvende itemresponsemodeller til at analysere instrumenter som er udviklet ud fra helt andre principper, og der argumenteres for en mere fleksibel tilgang og en højere grad af integration af traditionel psykometri og itemresponsmodeller.

Guido Makransky giver i sit bidrag, *Computer adaptive testing: An introduction to the practical considerations related to developing a computer adaptive test for occupational testing*, en introduktion til anvendelsen af IRT-baserede målingsmodeller til computerbaseret testning. Pointen i denne anvendelse er (udover at testningen kan foregå på en PC, eventuelt direkte gennem netadgang) at hver testperson ikke løser det samme sæt opgaver, men at det er muligt under testningen løbende at give testpersonen opgaver der er tilpasset til det færdighedsniveau personen indtil da har udvist.

Mette Høyer fremlægger i sin artikel, *Kontekst-begrebet i kvantitativ forskning, generaliserbarhed og ekstern invaliditet*, nogle overvejelser over begrebet økologisk eller ekstern validitet, samt en problematisering af modsætningen, den eksterne invaliditet. Problemerne illustreres med udgangspunkt i en gennemgang af to high-risk studier. Hun mener at begrebet ekstern invaliditet i bedste fald fremstår uafklaret, i værste fald som meningsløst. Begrebet ekstern validitet giver mening hvis de målte fænomeners afgrænsning også kan måles, altså operationaliseres, i forhold til konkrete empiriske undersøgelser og undersøgelsesmetoder. I modsat fald betegnes med dette begreb egentlig ikke en målings eller et undersøgelsesresultats validitet. I alle tilfælde er begrebet ekstern validitet ikke i sig selv metodisk vejledende da der med dette begreb ofte refereres til den eksterne kontekst for en given undersøgelse. Derfor bør dette begreb erstattes med mere specifikke overvejelser over design for og operationalisering af undersøgelsens genstand og konteksten.

John Brodersen ser i sin artikel, *Udvikling af velfunderede, relevante og valide spørgeskemaer til sundhedsvidenskabelig forskning*, en sammenhængen mellem på den ene side en øget anvendelse af spørgeskemaer til afdækning af befolkningens opfattelser vedrørende sundhedsvæsenet og på den anden det radikale kulturskifte der inden for lægeverdenen er sket væk fra en lægelig paternalisme hen i mod patientautonomi, og et samtidigt skift væk fra en monokausal, statisk, reduktionistisk biomedicinsk videnskabelig tankegang. Han beskæftiger sig med to former for spørgeskemaer som anvendes i medicinsk forskning, generiske og specifikke og beskriver forskellene mellem dem. Han gennemgår de metodologiske forskelle og begrundet at det er nødvendigt at anvende specifikke spørgeskemaer for at opnå valid, reliabel og konsistent måling. I den sidste del af artiklen gennemgår forfatteren hvorledes man udvikler velfunderede, relevante og valide specifikke spørgeskemaer til brug i sundhedsvidenskabelig forskning. To-panel-metoden til oversættelse og tilpasning af fremmedsprogede spørgeskemaer gennemgås. Herudover gennemgås en

metode hvor man ved hjælp af enkelt- og gruppeinterviews kan teste et spørgeskema for forståelighed, relevans og dækningsgrad og dermed sikre sig høj indholdsvaliditet. Afslutningsvis introduceres læseren til statistisk skalavalidering ved hjælp af Raschmodellen, hvor begreberne redundans, gulv- lofteffekt, specifik objektivitet, sufficiens, differentiell itemfunktion og lokal afhængighed kort berøres.

I sin artikel, *Intelligenstests og Flynn-effekten i Danmark*, gennemgår Tom Teasdale Flynn-effekten inden for intelligensforskningen. Han giver et kort rids over udviklingen af intelligenstests gennem det forrige århundrede og fokuserer på den forbedring i prøveresultater, man har fundet gennem perioden, en stigning kendt som "Flynn-effekten". Han omtaler biologiske og sociale teorier vedrørende effekten. Flynn-effekten har været særligt veldokumenteret i Danmark hvor forfatteren har haft et unikt datamateriale til rådighed i form af løbende intelligensundersøgelser af unge danske mænd på session. Den samme prøve, som er udviklet ud fra Raschs principper af Børge Prien, er blevet anvendt i alle årene efter 2. verdenskrig til i dag, og man har i dette materiale kunnet konstatere de stigende resultater gennem årgangene. Man har endvidere kunnet konstatere at denne stigning nu er ophørt.

De følgende indlæg behandler sproglige og æstetiske forholds betydning i psykologien.

Jesper Hermann diskuterer i sin artikel, *Vores indtryks udtryk – et metodesyn på vore "sprog"vaner*, en synsvinkel vedrørende hvorledes sproglige udtryk indgår som selvstændige udtryk for vor erkendelse og vore færdigheder. Han præsenterer det synspunkt at det sproglige ikke fortjener nogen beskrivelsesværdig psykologisk særstatus adskilt fra vore andre færdigheder. Artiklen hævder desuden at vor normale omgang med "det sproglige", ved talen og lytten, ved læsning og skrivning, for det første viser at brug af sproglige udtryk indebærer at vi benytter en række indgroede vaner. For det andet at påstanden om vaner medfører et nyt og anderledes personforankret syn på hvordan det foregår når vi taler og lytter til nogen, skriver eller læser noget, og understreger hermed betydningen af det sproglige i vort dynamiske tankeliv.

I artiklen, *Phi for livets skønhed*, beskriver Bjarne Sode Funch det således at livets mystik giver sig til kende i selv de simpleste former når de er proportioneret efter det gyldne snits princip. I matematikken bestemmes denne proportion ved tallet 1,618... efterfulgt af en uendelig række decimaler og betegnes ved det græske bogstav ϕ . I lyset af gestaltpsykologiens studier over oplevet bevægelse, der interessant nok også betegnes ved bogstavet ϕ , og eksistenspsykologiens paradigme om livets evindelige tilbliven, nyfortolker han Gustave Fechners eksperimentelle studier af den gyldne rektangels skønhed. Han konkluderer at al skønhed hidrører fra de øjeblikke hvor livets bevægelse forenes med lidenskab. Forfatteren vil med bidraget give en hyldest til Bennys lidenskabelige engagement både som forsker, græsrod, teoretiker, tidsskriftsredaktør,

underviser og vejleder.

Med det sidste indlæg af Tine Nielsen afsluttes temanummeret, som begyndte i pædagogikken med en gennemgang af metodeundervisningen ved psykologistudiet ved Københavns Universitet, også i pædagogikken. Tine Nielsen opridser først i sin artikel, *Universitetspædagogisk praksis og videnskabelighed*, nogle af de centrale vilkår for den universitetspædagogiske praksis i Danmark. Derefter diskuterer hun forskningsbaseret undervisning i forhold til videnskabelighed i den universitetspædagogiske praksis. Sidst præsenterer hun et eksempel på hvorledes undervisende forskeres videnskabelige kvalifikationer kan bidrage til en stadig udvikling af den universitetspædagogiske praksis i form af en scholarship of teaching and learning-model og fortæller hvorledes denne er implementeret i adjunktuddannelsen i universitetspædagogisk kompetence ved Copenhagen Business School.

28. oktober 2008

Jan Ivanouw
Anders Christiansen
Tytte Hetmar
Henriette Kirkeby

Metodeundervisningen på Københavns Universitets psykologistudium 1968-2008

Af *Tytte Hetmar og Jan Rattleff*

Her skal berettes om udviklingen i metodeundervisningen på psykologistudiet ved Københavns Universitet 1968-2008, altså 40 år med Benny Karpatschof ved roret.

Historien er dels fortalt ud fra, hvad vi selv husker om de ret mange år, vi var en del af metodelærergruppen, dels ud fra interviews med og små tekster fra nogle af metodelærerne før og efter os. Desuden har vi brugt, hvad vi havde af tekster fra metodekursets mange år, herunder sikkert også nogle som Benny selv har skrevet.

Før Benny var der ikke undervisning i metodelære, men derimod i støttefaget statistik ved Gustav Leunbach. Han havde skrevet en lærebog til undervisningen og var meget dygtig. Derimod var han ikke nogen stor formidler og talte svagt og svært forståeligt. De studerende, der kunne forstå statistikken, måtte hjælpe deres medstuderende. Jan Ivanouw, som senere blev en af de første metodeinstruktører fortæller, at han i perioden 1966-68 her havde sit første faglige job som privatmanuduktør for skiftende hold psykologistuderende.

“Første gang, jeg mødte Benny, var i 1959, da vi begge gik til psykologiundervisning. Vi var vel 100 studerende i Metro-annekset til en pædagogikforelæsning ved daværende professor i pædagogik K. Grue-Sørensen. Jeg sad tilfældigvis ved siden af Benny, vi hviskede sammen og blev flere gange dysset til ro af professoren, hvilket vi syntes var meget pinligt. Det var jo dengang, man rejste sig op, når professoren kom ind i lokalet.”

Jan Rattleff, studiekammerat

I marts-april 1968 blev undervisningen som bekendt afbrudt. Der blev

indledt med en debatuge, hvor alle undervisningstimerne blev brugt til diskussion af undervisning. Og dette førte til en ny epoke på psykologistudiet, også for statistikundervisningen. Benny træder ind på arenaen.

“Vi havde et kursus med Elementær Thomsen, som vi kaldte direktøren for Danmarks Pædagogiske Institut, Erik Thomsen. Kurset hed ”De elementære fags psykologi og pædagogik” og undervisningen foregik på Danmarks Lærerhøjskole i Emdrup. Benny havde nogle voldsomme, faglige diskussioner om ordblindhed med Carl-Aage Larsen, der var professor i didaktik og metodik. Benny opdagede, at psykologi var for vagt et fag, der manglede stringens, så Benny droppede psykologien og begyndte at læse til statistiker i stedet for.”

Jan Rattleff, medstuderende

Indtil 1968 var der således ikke nogen egentlig undervisning i psykologisk metodelære ved psykologistudiet på Københavns Universitet. Der blev undervist i statistik, på et tidspunkt uden i matematik.

Men efter studenteroprøret blev Benny ansat, først som ekstern lektor (i statistik) og senere som lektor i Psykologisk statistik og Metodelære ved Psyko-

logisk Laboratorium på Københavns Universitet. Samtidig fandt man ud af, at man ville ansætte studenterinstruktører til at undervise i statistik på små hold.

Det var jo det, som studenteroprøret blandt andet havde gået ud på: Nye undervisningsformer, især gruppeundervisning. Og Benny lavede fortløbende forelæsningsreferater, som han delte ud. En hel nyskabelse på det tidspunkt.

“Masser af undervisning var simpelthen en forelæsningsrække og en afsluttende eksamen, ikke som metodekurset, hvor lærerne arbejdede sammen og hvor de studerende selv skulle arbejde. Det var revolutionerende. Benny skrev senere en sang, *Metodemanifestet*, om det.”

Jan Rattleff, medudvikler af kurset i Psykologisk metodelære

Ressourcetilførslen i form af instruktører drejede efterhånden faget fra den statistiske oprindelse hen imod en bredere, metodekritisk undervisning - der til tider var endog anti-statistisk. Formålet var således snarere at advare de studerende mod positivismens vildfarelser i almindelighed og statistikkens i særdeleshed end at lære dem statistik.

Til trods for de mange konstruktive bestræbelser på at udvikle metodeundervisningen og kurset i psykologisk metodelære i metodologisk og videnskabsteoretisk retning, fandt kurset ikke sin egen form før Septemberkonferencen i efteråret 1971, hvor alle psykologistudiets eksisterende strukturer imidlertid blev opløst.

Der var ikke enighed i lærergruppen om metodelærens fremtid, hvilket førte til en periode med en undervisning, der nu blot bestod af en række forelæsninger og en eksamen med en, formel, bedømmelse af det metodologiske indhold i en integreret laborantopgave.

Efter nogle år, der på den ene side var præget af disse opløsningstendenser, på den anden side var præget af spændende eksperimenter af pædagogisk og forskningsmæssig art, blev faget reorganiseret, da Jan Rattleff og Svend Kreiner blev knyttet til faget som eksterne lektorer. Der blev desuden knyttet en ny stab af undervisningsassistenter og hjælpelærere til kurset.

Det nye lærerkollegium stod bag planlægning og gennemførelse af et metodekursus under noget kaotiske lokaleforhold i 1973, hvorefter kurset fandt sin form. I løbet af kurset er der – i tilknytning til kurset – blevet skrevet og publiceret en del metodelitteratur (se referencer).

“I efteråret 1969 lavede vi en undersøgelse som oplæg til at forklare tabeller og χ^2 -tests. Vi skaffede fem plakater med meget smukke billeder af samlejstillinger (de kunne bestilles som tillæg til en bog der hed ‘Stillinger’). Dem hængte vi op på tavlen. Derefter delte vi blå og røde sedler ud. Pigerne skulle tage de røde, drengene de blå. Så skulle de skrive et tal fra 1 til 5 med den samlejstilling som de foretrak, og vi talte derefter op og udfyldte 10-feltstabellen, og beregnede χ^2 med tilhørende p -værdi. Desværre husker jeg ikke resultatet. Der var vist en forskel på drenge og piger, men om den var signifikant husker jeg ikke.”

Jan Ivanouw, instruktør

Metodekurset

1974 blev det første, egentlige metodekursus gennemført. Allerede dengang var kurset centreret om tre opgaver, der havde til formål at bibringe de studerende en bred indsigt i psykologiens metodeproblemer: En grundlagskritik af intelligensprøver, en designmæssig vurdering af en eksperimentalspsykologisk undersøgelse og en statistisk analyse af datamaterialet og endelig en kvalitativ opgave i form af en indholdsanalyse af kontaktannoncer.

Efter et meget kærligt afslag fra Suzanne Brøgger om at introducere emnefeltet for den kvalitative opgave, optrådte Benny selv med et oplæg om “Den seksuelle kode i kærlighedsnovellen”.

Efter udflytningen til Amager og ibrugtagning af KUAs forelæsningslokaler med enorm, syntes det, afstand mellem tavle og sal eksperimenterede lærerkollegiet med undervisningsformen. Der blev eksperimenteret med, hvor mange stud. psykker, der maksi-



Jan Rattleff og Benny Karpatschof

malt kunne være i et lokale normeret til 50. Der kunne være forbavsende mange, men de blev meget sure. Desuden blev der eksperimenteret med to og tre parallelle lærerteams. Også på indholdssiden blev der arbejdet, idet questiologi, datalogi og matematisk statistik kom ind i de kvantitative og interviewteknik i den kvalitative opgave. Senere førte dette til opbygning af to selvstændige delkurser, et kvantitativt og et kvalitativt.

“Noget af det bedste ved at være metodelærer i de år var, at man lærte så meget, og ikke kun om metode. En af grundene hertil var ideen om, at vi var lige i lærergruppen - det var jo altså 70-erne. Vi delte f.eks. ansvaret for forelæsningserne mellem os: 6 lærere med hver 2 forelæsninger (Benny har vel taget den første og den sidste, for der var jo 14 gange). Jeg glemmer aldrig mine første rystende forelæsninger som studenterunderviser. Man kan jo forberede sig på, hvad man vil sige; men så stiller de studerende spørgsmål! Og så var Benny der altid, som solidarisk sikkerhedsnet - og man tabte ikke engang ansigt. Dét har jeg virkelig bestræbt mig på at give videre. Hele lærergruppen var i øvrigt altid tilstede ved alle forelæsninger - vi skiftedes også til at få løn.”

Tytte Hetmar, metodelærer

omkring problemformulering (hvor laborantundervisningen bidrog med specifikation af teoretisk referenceramme), indsamlingsplan og analyseplan.



Svend Kreiner og Benny Karpatschof

I 1982 blev kurssets undervisningsressourcer imidlertid drastisk beskåret, de tre undervisningsassistenter fyret og de to eksterne lektorer timetal beskåret. Metodelærergruppen var sprængt, og kurset blev derfor endnu engang reorganiseret. Den kvalitative del blev skåret væk, og tabet af lærerkræfter søgtes erstattet ved rationalisering med inddragelse af den dengang nye og moderne edb-teknologi - og førte blandt andet til udvikling af en konverserede indlæringsmaskine, *Metodesystemet*.

Desuden var der i flere tilfælde samarbejde med lærerne på laborantholdene - de studerendes eksperimentelle øvelseshold - nogle gange ved at metodelærerne gik ind på strategiske tidspunkter

Fra 1985 blev der igen mulighed for at ansætte flere lærere. Tre undervisningsassistenter blev ansat, og metodelærergruppen bestod atter af seks lærere, nemlig en intern og to eksterne lektorer samt de nye undervisningsassistenter.

Nu fulgte en del år med yderligere udvikling af metodekurset. Der blev foretaget ændringer i den praktiske organisering af undervisningen, idet det halve af en årgang fik metode i efterårssemestret og den anden halve årgang i forårssemestret.

Vægten var stadig på de studerendes egne erfaringer med empiriske projekter. Opgavekravene blev imidlertid ændret, således at de studerende kun skulle udarbejde én opgave i form af en spørgeskemaundersøgelse med egen empiri. På denne måde skulle de studerende gøre erfaringer med og forholde sig til alle faserne i den empiriske forskningsproces i et enkelt sammenhængende projektforsløb. Denne periode var derudover i særlig grad præget af den teknologiske udvikling med nye statistikprogrammer mv (se nedenfor).

Indhold, organisering og pædagogik indtil 1996

Sigtet med metodekurset har gennem årene været at introducere de studerende til de forskellige faser i den empiriske forskningsproces fra den første løse ide over problemafgrænsning, teoretiske antagelser om en problemstilling, begrebsafklarings, hypoteseopstilling, operationalisering, design og sampling, dataindsamling, databehandling og analyse til formidling af undersøgelsesresultater.

Da det er vanskeligt at tilegne sig forskningsmetodiske kompetencer ved udelukkende teoretisk undervisning, vekslede metodeundervisningen systematisk mellem teoretisk undervis-

“Tit havde Benny lige udviklet en ny teori i s-toget. Og så begyndte vi mødet i lærergruppen med, at Benny holdt et oplæg om ‘verden og omegn’.”

Jan Rattleff, kollega

ning ved forelæsningerne og undervisning rettet mod de studerendes egne empiriske projekter i holdundervisningen. Der var således en direkte sammenhæng mellem forelæsningerne og holdundervisningen, idet hver ny fase i forskningsprocessen indledtes med en forelæsning med præsentation af begreber og metoder i denne fase, hvorefter de studerende i grupper skulle forholde sig til og gennemføre projektaktiviteter omhandlende netop denne fase. Ved selv at gennemløbe de forskellige faser sideløbende med teoretisk undervisning ville de studerende få den bedste indføring i empirisk arbejde.

Det pædagogiske rationale var naturligvis, at de studerende ved selvstændigt at skulle anvende de præsenterede metodologiske begreber og metoder i de teoretiske oplæg som redskaber ville kunne forholde sig mere konkret til og bedre tilegne sig disse end ved udelukkende teoretisk undervisning - også selv om denne indeholdt mange konkrete eksempler, herunder på andre metoder i hver af faserne, end de studerende selv aktivt stiftede bekendtskab med. Det var deltagernes selvstændige arbejde med anvendelsen af metodebegreber i forskningsprocessens forskellige faser, der var metodekursets særkende.

“*Metodemanifestet*
 (Mel: Se den lille
 kattekillling)
 Nu har vi metodekursus
 Og det er så sjovt
 Blot vi ikke dummer os
 Det syn’s vi er så flovt
 Klogskab kommer kun af fejl
 Man har begået selv
 Det forstanden får fra andre
 Slår den halvt ihjel
 Hvis du derfor syn’s dit
 hoved’
 Er er helt fuld af rod
 Så er du på spor af det
 Du ikke før forstod”

Benny Karpatscof, engang
 i 70’erne (ifølge Tyttes
 hukommelse)

aktadmønstre (med titlen “Pisk”), mobning på kommunale arbejdspladser, plejehjemsbeboeres trivsel og netværk, adfærdsændringer hos kvindelige studerende på Grønjordsskollegiet efter nogle på tidspunktet aktuelle voldtægter, indfødte danskeres og indvandreres opfattelser af hinanden samt mange andre komparative holdningsundersøgelser - selvom vores nabo Politiskolen på et tidspunkt udbad sig et par års fredningstid for psykologistuderendes glubende interesse for at sammenligne politievers og univertetsstuderendes holdninger til alskens emner.

“Benny har jo altid været en genial
 sangskriver. Og Benny lavede ustandseligt
 sange til metodekurset, nogle af dem bliver
 stadig brugt.
 Uforglemmelige er *Jeg ved en ki-i-anden ...*
 og *Om hvordan psykologiske data bli’r til*
 - *det er ikke det, at de bli’r til, det er*
Måden altså de bli’r til på ...
Noget om at hade sin næste er fra
 et år, hvor det overordnede tema for
 metodeprojekterne var *‘Indfødt og fremmed’*”

Tytte Hetmar

faldsproblemer generelt samt problemer med selvsektion ved nogle projekter.

På den anden side medførte undervisningens, og metodeprojekternes, fokus på alle faser i den empiriske forskningsproces, at ingen studerende fra disse årgange var fremmedgjorte i forhold til datafødsel; de vidste ”hvor tallene kom fra”, i modsætning til hvad der er tilfældet ved megen anden statistikundervisning. Desuden gjorde de studerende gennem pro-

I metodeprojekterne skulle de studerende i arbejdsgrupper selv formulere, planlægge og gennemføre et empirisk projekt, typisk en spørgeskemaundersøgelse. De skulle selv forestå udvælgelse og dataindsamling og herefter inddatere og analysere undersøgelsesresultaterne. I metodeopgaverne blev der lagt vægt på, at de studerende kritisk skulle gøre rede for og reflektere over alle faser i den empiriske proces.

De studerende var meget kreative i deres valg af problemstillinger, som for at nævne nogle få eksempler spændte fra undersøgelser af småbørnsforældres syn på pædagogikken i børnehaver, sado-

masokisters seksualpraksisformer og kontaktmønstre (med titlen “Pisk”), mobning på kommunale arbejdspladser, plejehjemsbeboeres trivsel og netværk, adfærdsændringer hos kvindelige studerende på Grønjordsskollegiet efter nogle på tidspunktet aktuelle voldtægter, indfødte danskeres og indvandreres opfattelser af hinanden samt mange andre komparative holdningsundersøgelser - selvom vores nabo Politiskolen på et tidspunkt udbad sig et par års fredningstid for psykologistuderendes glubende interesse for at sammenligne politievers og univertetsstuderendes holdninger til alskens emner.

For kvantitative projekters vedkommende var der imidlertid ét stort metodemæssigt problem i hele denne periode, og det var naturligvis samplingen, da de studerende selv forestod udvælgelsen af respondenter til deres projekter. Udvalgelsesmetoderne resulterede i ikke-randomiserede materialer, ligesom der ikke kunne være styr på bort-

jekterne yderst værdifulde, omend dyrekøbte erfaringer, når de frustreret ofte oplevede at ærgre sig i analysefasen over fejl og mangler i designfasen ("Hvorfor spurgte vi dog sådan", "Hvorfor spurgte vi ikke også om?", "Hvorfor inddrog vi ikke også personalet?" mv). Den måde metodekurset var organiseret på gav således de studerende erfaringer med sammenhæng mellem og konsekvenser af metodevalg i de forskellige faser i forskningsprocessen. Og den slags erfaringer kan man ikke læse sig til.

Teknologisk udvikling

Metodekurset udvikling var gennem årene naturligvis også præget af den generelle teknologiske udvikling. Hvor de studerende i Lampeåret i 1974 måtte gennemføre deres analyser i hånden, altså f.eks. regne ki-i-anden teststørrelser ud med blyant og papir og herefter slå op i tabeller for at få deres p -værdier, blev denne del af analysearbejdet radikalt ændret med adgangen til at bruge elektronisk regnekraft. Disse muligheder forelå selvfølgelig allerede i 70erne, men tilgængeligheden var ikke for menigmand. Det var mainframecomputernes tid, og Københavns Universitet havde en stor afdeling på Østerbro ved navn RECKU (det Regionale Edb Center ved Københavns Universitet, senere UNIC), som man telefonisk kunne komme i kontakt med via modem. RECKU havde også en budtjeneste, som to gange dagligt kørte udskrifter ud til brugerne. Så man planlagde sit arbejde og ventede pænt på at få udskrifter fra RECKUs *High Speed Printer* bragt ud. Personligt, med bud. Efter mainframetiden kom så pc-erne med egne programmer på egen computer, og den besværlige modemtid var slut.

Metodelærergruppen havde i forskningsammenhæng tidligere udviklet statistikpakken IDOL (Individualised Data Organisation Language), som også blev anvendt på metodekurset, og det var denne programpakke, som i begyndelsen af 80erne fik en interaktiv brugergrænseflade kaldet *Metodesystemet*. De studerende kunne

Noget om at hade sin næste

(Mel: Når en pebernøddebager
bager pebernøddekager)

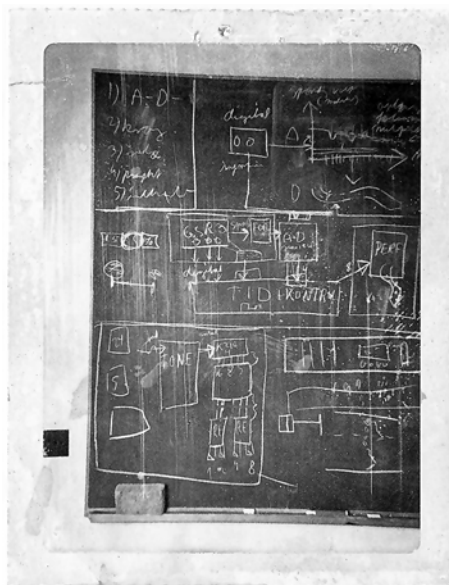
Se en fremmedhader hader
alle fremmede som vader
frække rundt i vore gader,
så at Danmark går i sort,
Islams sære falbelader
gi'r kulturen dybe skader,
hvis de ikke selv forlader
landet, må de sendes bort.

Se en fremmedhader-hader
hader hver en fremmerhader
der forstyrrer stille gader
og politisk helt er sort,
deres nazieskapader
volder landet dybe skader,
der er hjerneblæste spader,
og skal ha' det røde kort.

Se er sjælevender-kender
kender sjælens løse ender,
hvor logikken går på hænder,
sammenblander løgn og sandt,
hvor de vilde projektioner
synger psykens falske toner,
nag til andre mænd og koner
er vort selvhads falske pant.
Benny Karpatschof

(Metodekurset, 1993)

så anvende Metodesystemet fra Amagerterminalens brugerlokaler.



Modelarbejde!

I 1987 begyndte vi at bruge statistikpakken SPSS, som på dette tidspunkt var en DOS-version, som gik ned konstant. Hertil kom, at man hele tiden skulle identificere sig som lovlige bruger, og metodelærerne rendte i disse faser af undervisningen rundt med en såkaldt nøglediskette mellem de studerende i edb-lokalerne på Amagerterminalen, når de blev smidt af systemet for at koble dem på igen.

Efter disse frustende forhold gik vi i en del år over til at bruge MIDOL som var en videreudvikling af IDOL kombineret med grænsefladen Metodesystemet. Begrebsmæssigt var MIDOL stærkere end SPSS, blandt andet var skalatyper integreret i systemet på

en måde, der bevirkede at man ikke kunne udføre statistisk "ulovlige" analyser som f.eks. middelværdiberegninger på ordinale variable eller tests der krævede ratioskalaniveau på variable af en lavere orden. IDOL var nemlig oprindeligt konstrueret, så det opfyldte robotikkens love ("En robot må ikke skade et menneske", "IDOL må ikke skade en forsker" mv.). Men tabeller og grafik i det hele taget var generelt flottere i SPSS.

I 1996 kom SPSS så i en windows-version, hvilket gjorde det statistiske analysearbejde meget nemmere for de studerende. Før denne tid kunne de nemlig kun arbejde på Amagerterminalen på KUA; men nu etableredes der en licensordning, og de studerende kunne derved få SPSS på deres egne pc-er.

Metodeundervisningen efter 1996

I foråret 1997 blev metodeundervisningen radikalt ændret, såvel hvad angår indhold som organisering. Metodeundervisningen placeredes fast i forårssemestret for alle studerende på 2. semester og stadig med én metodeopgave. Men indholdet ændredes. Der havde gennem en tid været debat på studiet om, at de studerende skulle lære mere statistik. Med et uændret timetal kunne der imidlertid ikke indføres mere stof i metodeundervisningen, hvorfor opprioriteringen af mere statistik måtte ske på

bekostning af noget andet. Og så droppedes metodekursets mangeårige praksis med vægt på de studerendes egen empiri, hvilket afspejledes i en ny fagbeskrivelse (for *statistik*undervisningen). I stedet for egen empiri skulle de studerende nu arbejde med udleverede datasæt.

Desuden skete der ændringer på den organisatoriske front med følger for pædagogik og lærersamarbejdet, idet forelæsningserne udskiltes fra holdundervisningen. Undervisningen organiseredes som forelæsninger ved to lektorer, Benny Karpatschof og Tom Teasdale, og øvelseshold ved deltidslærere på andre ugedage.

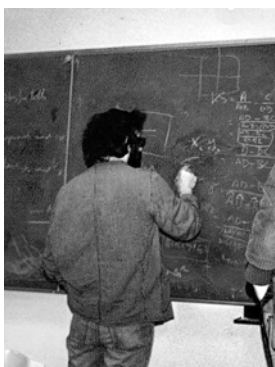
Sammenhængen mellem forelæsninger og holdundervisning var dermed ikke så direkte som i de foregående godt 20 år, hvor metodeundervisningen havde været samlet på en bestemt ugedag med holdundervisningen placeret tidsmæssigt lige efter forelæsningen. Indholdsmæssigt var der naturligvis stadig forbindelse mellem forelæsninger og holdundervisning; men lærersamarbejdet ændrede karakter ved tabet af det ugentlige samvær med muligheder for udvikling og justering af undervisningen undervejs. Samt højtflývende diskussioner og skæg og ballade – vi savnede den tætte kontakt med Benny, såvel fagligt som personligt.



Benny Karpatschof

I 2000 ændredes opgavekravene, således at de studerende skulle aflevere to statistikopgaver, hvor den første var med deskriptiv, den anden med inferentiell statistik. I 2006 indførtes endnu en ændring i opgavekravene, og de studerende afleverede nu 6 statistikopgaver i løbet af semestret, én hver anden uge. Desuden gav de mange mindre opgaver mulighed for, at de studerende prøvede at arbejde med analyse af flere forskellige typer af data.

Hvad er så tabt og hvad er vundet ved denne transformering af det oprindelige metodekursus til opprioritering af statistikken? Det kommer an på, hvad man lægger vægt på. Der er ingen tvivl om vigtigheden af at opprioritere det statistiske stof, de studerende skal have tilegnet sig som kommende cand. psych.er. Så her er klart en gevinst. Hertil kommer, at de studerendes kendskab til forskellige former for psykologisk relevant empiri er udvidet under de seneste ændringer af undervisningen, da de i opgaverne får udleveret datasæt med f. eks. testdata, forløbsdata og eksperimentelle data - og derved prøver at arbejde med andet end spørgeskema-data.



Benny underviser

Imidlertid er undervisningsforløbet meget kompakt og stoftrængslen stor, hvilket dog er et bevidst valg, truffet netop for at øge de statistiske kompetencer hos de studerende.

Hvad der savnes i den nye statistikundervisning er overblikket over hele den empiriske proces fra en ides fødsel til færdigt projekt, og især egne indhøstede metodeerfaringer i den forbindelse. Hertil kommer, at design, operationalisering og dataindsamling, dvs alle de første faser af et empirisk forskningsprojekt, udgør væsentlige emneområder, som de studerende ikke får aktivt kendskab til efter ændringerne af metodeundervisningen til statistikundervisning. Dette er der rådet bod på ved indførelse af faget Psykologiens Undersøgelsesmetoder på 1. semester.

Referencer

Metodelitteratur, skrevet og publiceret i tilknytning til Metodekurset på psykologistudiet ved Københavns Universitet
Her i kronologisk orden.

Karpatschof, B., Kreiner-Møller, S. og Rattleff, J. (1974). *Psykologisk Metodelære*. København: Akademisk forlag.

Karpatschof, B., Kreiner-Møller, S. og Rattleff, J. (1977). *Psykologiens Metodelære*. København: Akademisk forlag.

Simonsen, K. (1978). *IDOLhåndbogen*. Institut for Pædagogik og Psykologi ved Danmarks Lærerhøjskole.

Anonym. (1980). *IDOL-lommelisten*. København: Antropos.

M. E. Todegruppen .(1981). *Om kvalitativ metodelære*. København: Universitetets Psykologiske Laboratorium.

Anonym. (1981). *BABIDOL: EDB-kørsel med IDOL* . København: Universitetets Psykologiske Laboratorium.

Metodegruppen (1983). *Psykologi og Metodelære. Ved metodekursets 10-års jubilæum 11/3-83 på KUA*. København: Antropos.

Anonym. (1984). *Metodenoter 1984: Om kvantitativ analyse af kausal-sammenhænge*. København: Universitetets Psykologiske Laboratorium.

Karpatschof, B., Kreiner, S. og Rattleff, J. (1985). *Metodenoter*. København: Universitetets Psykologiske Laboratorium.

Johansen, J., Karpatschof, B. og Rattleff, J. (red.) (1987). *Kommunikation og opgaveløsning i en interaktiv grænseflade*. Psykologisk forskningsrapport nr. 7, ISSN 0107-3060. København: Universitetets Psykologiske Laboratorium.

Rattleff, J. (red.) (1988). *Metodenoter-1*. København: Universitetets Psykologiske Laboratorium.

Rattleff, J. (red.) (1988). *Metodenoter-2*. København: Universitetets Psykologiske Laboratorium.

Rattleff, J. (red.) (1989). *Metodenoter*. København: Universitetets Psykologiske Laboratorium.

Metodegruppen. Rattleff, J. (red.) (1990). *Metodekompendium*.

Rattleff, J. (red.) (1992). *Introduktion til kvantitative metoder i psykologisk forskning*. København: Semi-forlaget.

Christensen, L. H. og Kristensen, T. D. (1993). *Brugervejledning til Middel*. (3.ed.). København: Semi-forlaget.

Karpatschof, B. (2006). *Udforskning i psykologien. De kvantitative metoder*. København: Akademisk forlag.

Intersubjectivity, Cognition, and Language. Epistemological conditions and their methodological consequences

Af Nini Praetorius

In our normal everyday encounters with other people it is generally assumed that the knowledge we have and the language we use to describe and communicate about things and events in the physical and social world of which we are part is *intersubjective*, that is, it is shared by the people with whom we may communicate and co-act. Indeed, this assumed intersubjectivity of cognition and language would seem to be a *precondition* for any co-action and linguistic communication to take place among people about things which exist in the so called “outer”, publicly observable physical and social world. Arguably, it is a precondition for our very notion of a publicly observable physical and social world, i.e. a world that may be observed and described objectively and truthfully from a so called *third-person* view.

Whereas scepticism is no longer in fashion when it comes to the possibility of so-called *third-person* description and cognition of publicly observable reality being true or objective – indeed, not only our everyday encounters but also our scientific investigations rely on this possibility – there has for centuries been an extensive debate within philosophy, and more recently within the scientific studies of mind and consciousness, concerning the status of the description and cognition of things which do not exist in a publicly observable and describable world, such as our so called “internal” mental states. The uncertainty and scepticism are expressed in questions such as these: Given that mental states such as thoughts, emotions and feelings of pain, are not publicly observable, but may only be known or experienced by the persons who have them, how then can we be certain that these states exist and may be observed in the same way by different people? Furthermore, in view of the lack of public

criteria or standards, how can we be certain that assertions put forward about such states have the same implications and use for different persons and language users? Or be certain that we are using language in the same way – or indeed using the *same* language – as we do when talking about and describing things in public material reality? Indeed, given that internal states are not the sort of things which are publicly *shareable*, how do we ever come to learn and talk about them in a language we do share?

That solutions to these problems have serious consequences, not only for the possibility of scientific studies of mind and consciousness but for a science of psychology in general, is obvious when we consider that a crucial part of the knowledge a person has of himself, of his mind, body and acts - and, not the least, *of his mind and body being his*, and *of his acts being acts intended and controlled by himself* - rests on observations and experiences to which only the person himself has access. No one except the person himself will ever be able to experience what it is like to be the person he is *in the way he experiences* it, or to know what it is like to have his perceptions, thoughts, or feelings of pain in his body *in the way he does*, because no one except the person himself can see with his eyes, think his thoughts or feel his pain.

To this uniqueness of our personal experience must be added the differences in our knowledge and conception of ourselves, of others and the world in which we find ourselves, due, for example, to differences in our upbringing, training, education or cultural background. And yet, neither the knowledge of persons which is uniquely personal, nor the differences in knowledge and conceptions of themselves, of others and the world, due to their different backgrounds, can be said to be *private*. For, despite these differences in our personal experiences and conceptions, to be a person is something fundamentally social. Indeed, no one can be a person, and thereby someone who may realise that he or she is uniquely different from other persons, *without* other persons from whom he or she may differ - nor without having possibilities of determining *how* one differs from others - whether such differences concern one's notions and experiences of things in material reality, or one's inner feelings, thoughts, beliefs, emotions or pains. Indeed, it would seem that no one could be a person, that is, someone who *differs* and *knows* that one differs from others, without being able to communicate and talk with others about *how* one differs from them.

In what follows an attempt will be made to argue that conditions for intersubjective communication exist which apply equally for knowledge and description of things in publicly observable material reality and for our non-publicly observable mental or internal states. They are conditions on which rely the possibility of persons to communicate about their knowledge and experience, and conditions for determining and discussing their individual differences concerning their knowledge and experience – be it of that which is publicly observable or of that which is uniquely ob-

servable to themselves, such as their thoughts, emotions and feelings of pain. By the same token they are conditions on which relies the very possibility of persons developing together criteria or standards to determine and distinguish between what is publicly observable and what is personal, and by which, so it will be argued, any notion about private cognition and private languages are rendered untenable. Furthermore, it will be argued that despite the significant observational differences applying to the cognition and description of things existing in public material reality and to the cognition and description of our internal or mental states, we have to assume that the language in which we talk about both these things and states, is a language which necessarily relies on and is part of the very same language we use to describe things in public material reality.

In the concluding section I shall outline the consequences of the arguments put forward about the general conditions of intersubjectivity of cognition and language for the current debate within consciousness studies of the feasibility, methodologically as well as epistemologically, of using first-persons reports from subjects about their non-publicly observable mental states as evidence in scientific research.

But first, in the section which follows I shall present the problems concerning knowledge and description of internal or mental states as set out within traditional philosophy. To this end an analysis will be carried out of the position on these problems defended by Wittgenstein. It is the aim of this analysis to show why it has to be assumed that, despite not being publicly observable, our knowledge and experience of such states are as amenable to intersubjective discourse as are things in the publicly observable world.

Conditions for description of mental or internal states

One of the significant differences in the conditions for the cognition and description of things in publicly observable reality and of our non-publicly observable internal states concerns the *procedures* for determining the truth and correct application of descriptions of such things and states. In the case of descriptions of publicly observable things, for example, we will be able to take part in a procedure for determining whether the implications of the descriptions hold true for the things in question, and thus to observe the result of a test as to the correct application of the description. However, in the case when somebody says e.g. that he has a pain in his finger, no such public procedure exists for determining whether a “pain-description” is a true description of what the person feels in his finger. For only he knows and may observe what he feels in his finger.

But given that no public procedure exists for determining the correct

use of our description of internal states such as feelings of pain, how then can we be sure that when different people talk about pains, they are talking about the same kind of "thing"? How, more precisely, *can I be sure that when I use the term 'pain', I use this term to refer to the same kind of "thing" as others do when they use the same term?* Or, conversely, *when other people use the term 'pain' do they then use this term to refer to the same kind of "thing" as I do when I use the same term?*

This would seem a perfectly sensible question. However, in his classical "Beetle in the box" example, in which Wittgenstein attempts to make clear what this question entails, it seems to be a question to which, for obvious reasons, no sensible answers may be given. Thus, Wittgenstein writes:

If I say of myself that it is only from my own case that I know what the word "pain" means - must I not say the same of other people too? And how can I generalise the one case so irresponsibly?

Now someone tells me that he knows what pain is only from his own case! - Suppose everyone had a box with something in it; we call it a "beetle". No one can look into anyone else's box, and everyone says he knows what a beetle is only by looking at his beetle. - Here it would be quite possible for everyone to have something different in his box. One might even imagine such a thing constantly changing. - But suppose the word "beetle" had a use in these people's language? If so it would not be used as the name of a thing. The thing in the box has no place in the language-game at all; not even as something: for the box might even be empty. - No one can 'divide through' by the thing in the box; it cancels out, whatever it is. (Wittgenstein 1945/1953, para. 293, p. 100.)

In the analysis which follows an attempt will be made to show that the question about 'pain', posed in this *radical* sense, is not only a question which cannot be answered, but more importantly, it is also a question which cannot be *asked*. For it is a question in which the very condition for putting it forward is itself questioned - thereby debarring it of any sensible meaning.

Let me begin by noting that Wittgenstein does not only *suppose* that the word pain "has a use in people's language" - he knows it for a fact. Indeed, when he or anyone else asks the question, "do I use the term 'pain' to refer to the same "thing" as others do when they use the term", he and they are obviously asking this question in a language of which the term 'pain' is part - and thus is a term which is supposed to have a meaning and use which he and other people with whom he shares the language know *in virtue* of being speakers of that language, just as they know to

what the term may be correctly applied, i.e. some particular unpleasant sensations somewhere in one's body. If Wittgenstein did not know this - and did not *presuppose* that all other speakers of the language knew this - neither he, nor they would have any idea what he is asking about, and no further discussion of the question would seem possible.

Now, it has to be admitted that we may have all kinds of difficulties in giving adequate accounts in words of the pain we may feel, and that we often feel uncertain about the choice of appropriate terms. Is this pain, for example, a sharp, piercing, dull, shooting, tender, searing pain, or is it a nagging or stabbing pain? Indeed, we may have great difficulties in conveying precisely to others the suffering we endure when we are in pain. However, these difficulties of *adequately* describing pains are not relevant to, nor do they invalidate the point just made. Indeed, these difficulties of adequate descriptions of pains, and discussions about them, could not take place *unless* people discussing them had a concept of and a term for pain which they shared, and knew of what it may be used correctly to refer to.¹

We may contend then that Wittgenstein's discussion of the "language game" of giving expressions to pain relies on this knowledge and these presuppositions, and hence on the assumption that sensations and feelings of pain are the sort of "things", which do indeed exist and which we may use language to refer to. Furthermore, it relies on the assumption that other people would use the term 'pain' to refer to the feeling of pain that he may have in his body - could they feel what he feels - and hence that he may indeed generalise from himself to others, when they use the term 'pain' - just as they may generalise from their use of the term to his and the use of the term by others.

However, contrary to the presuppositions on which his whole discussion of the question of 'pain' rely, Wittgenstein suggests, for the sake of argument, that because his feeling of pain, and those of others, are not publicly observable, his feelings of pain may be completely different from others and, thus, he as well as others may be using the term 'pain' to

¹ To spell it out, behind questions such as "do we mean the same thing when we talk about 'pain' - or, for that matter about other mental or internal states such as 'memory' or 'recognition' - lies the assumption that such terms are no more randomly applied to mental or internal states of human beings than 'oak' and 'ash' are randomly applied to trees. Indeed, to take such questions seriously is already, albeit implicitly, to endorse certain assumptions of *how language functions in communicative contexts*. Thus, it is implied and assumed that it is possible for language users together to identify - if only rudimentarily - what they are talking about (e.g. some particular states or properties of our mind, or some sensations felt somewhere in our body), and that, on the basis of such common consent, it is possible to investigate whether the implications of those terms are in fact the same for everyone - and whether we in fact use these terms to refer to the same sort of "things". Conversely, such questions cannot be asked in any sensible way, nor may these terms be "mentioned" without or independently of how language and its terms are used to refer to actual things or events.

refer to completely different things – or to none at all. Indeed, he invites us to assume that in the language he shares with others it would be perfectly sensible so to say and suggest. However, it does not make sense to suggest that we may talk about, let alone *determine* any individual differences in our feelings of pain or in our use of the term ‘pain’, *unless* the implication of the term ‘pain’ is shared by everyone involved, and unless, furthermore, this term is used to refer to the same sort of “thing”. Without these presuppositions, any discussion of the question of ‘pain’ disintegrates into nonsense.

In summary, it would seem that the question of ‘pain’ as set out by Wittgenstein is *obtuse* in the sense that putting it forward presupposes that we know the meaning and use of the terms of the question and also to what it may be correctly applied. But then we are asked to forget or ignore this presupposed knowledge, and to pretend that it is immaterial for a discussion of what the terms ‘pain’ may be correctly used to refer to - or whether it may be used correctly to refer to anything at all. But it is a question which can only be asked and discussed granted we have already learned a language which we may use to talk about pain, and thereby granted pain to be the kind of “thing” which exists as something we may talk *correctly* about and *refer* to. Hence, if we do ask this question, the answer is logically implied: an affirmation would be redundant, while a denial would be contradictory.

Now, if the same sceptical question had concerned the use of the terms ‘cups’ or ‘neurons’ instead of ‘pains’, it would have been obvious why such a question would not make sense. Indeed, it would probably have been so obvious that we would hesitate to ask it. For if we did ask this question about the use of the terms ‘cups’ and ‘neurons’, we might as well ask the same question about all other terms in our language - and we would be well on the road to asking whether we can be sure that we may use terms in our language to refer to and talk correctly about *any* objects in material reality. However, although in *particular* cases we may be in doubt as to whether a particular term may be correctly used to refer to some particular thing, i.e. a thing which has been identified in a shared public world, we cannot doubt that as language users taking part in this discussion, we do know (other) correct terms for the thing (i.e. those forming part of the identification of the thing), nor doubt that we know how to use *these* terms correctly. One cannot doubt the necessity of these conditions for settling the question under discussion - unless, of course, one has been seriously contaminated with scepticism, and mistakenly assumes that one may get away with using language to question the very possibility of using language to talk correctly about anything.

However, similar conditions seem to apply to the question of ‘pain’ – and by extension to questions concerning other internal or mental states such as emotions, thoughts and belief - in the sense that scepticism as

to whether we may use language to talk correctly – or at all - about such states, and make reference to them, presupposes that we know the meaning of terms for these states, and that together with other language users we may determine what they may be correctly used to refer to. So rather than questioning the existence of such internal or mental states and the possibility of having knowledge about them and of correctly describing them and being able to communicate about our experiences of such states, this questioning logically rests on the assumption of both their existence and this possibility. Indeed, without these assumptions concerning the *intersubjectivity* of both our experiences of, and of our use of terms to talk about and refer to our non-publicly observable internal or mental states, neither ordinary everyday communication nor philosophical discussions about such states would be possible.²

In the section which follows I shall further clarify what it means that our cognition and use of language is *intersubjective*, just as I shall show that the very same assumption of intersubjectivity is a necessary condition for knowledge, descriptions and communication of both that which is and that which is not publicly observable and shared.

The intersubjectivity of public and personal knowledge and experiences

It has to be admitted that it is somehow puzzling that pains and other internal states, which are only directly observable to the persons who have them, and are not *observationally* shareable by others, are nevertheless things which we may communicate about in a language we *do* share with other persons. So, apparently, shareability in the sense of being *publicly* observable and known cannot be a condition for the possibility of communication among persons about things which only they may directly observe, experience and know about.

In the case of observing, experiencing and communicating about objects in material reality, our situation is arguably significantly different. Take for example two people sitting on either side of a table with cups and plates, a teapot, a bowl of sugar and a bottle of milk. All these things exist in a shared public world and are perfectly observable to both persons; they may together determine the things on the table and also

² I am not saying that the fact that we have a language with terms for both objects in reality and pains in bodies *proves* the existence of objects in reality and pains in bodies, nor that all and every concrete statement we put forward about either objects or pains are always or infallibly correct. What I am saying is that we cannot begin to discuss or investigate language and the use of language to talk about such things as objects or pains without assuming, generally, that both objects in reality and pains in bodies exist as things that we may have knowledge of and use language to talk correctly about. One cannot take part in this kind of philosophers' discussion without committing oneself epistemologically.

whether the descriptions they put forward about them are correct. And yet, it could be argued that how these things are observed and appear to them from their different vantage points on either side of the table is different, i.e. due to the fact that the parts and features of the things which are directly observable to the one, are not the same as those which are so observable to the other. However, this does not present any serious difficulties since, first, it is part of our knowledge of things having been identified as particular things, that they will appear differently when e.g. we move around and look at them from different vantage points - and that, generally, things do appear differently when observed with regard to different possibilities of observations and action. And these differences of perception and experience do not represent any serious problems since, secondly, none of the particular ways of perceiving and experiencing the things on the table, and none of the descriptions by either person of their experiences of the cups, plates, etc., are unique to him or her. Indeed, it is *assumed* that they are not, just as in general any person and language user will assume that if other people could look at the things from his or her vantage point(s), they would observe what he or she does, and report that they perceive the same features and properties of the things, and describe what they perceive as he or she does. *If we could not count on this assumption, communication and action between persons about things in the world would be impossible.*

But there are numerous other ways in which the knowledge and description of persons concerning things in publicly observable reality may differ. Just think of the differences due to our different background, education, previous history of experiences, and the opportunities to observe and describe such things which are or have been available to us. Examples are *legion* – I only have to think of the knowledge I have of aeroplanes, their construction and how to fly them compared to that of a pilot. Or, conversely, think of the knowledge I have as a clinical psychologist about the transference phenomena occurring between client and therapist during psychotherapy compared to that of an aeroplane pilot, who has not encountered such phenomena, and who does not have the language and terms to describe them that I have. And yet, despite the fact that our knowledge of these and numerous other matters are not exactly the same, and probably never will be, we are in fact able to make available and to communicate the knowledge that each of us has about aeroplanes and transference problems and those other matters, and thus to share our knowledge of what in this respect is personal to each of us. But if the condition for the intersubjectivity and shareability of knowledge and description in the actual case as well as between persons in general, is not and cannot be that persons have exactly the same knowledge and experience of things, or the same possibilities of describing things, since this condition is only rarely if ever met due to their different background, education, history of experiences, and so forth, on *what* then relies this

intersubjectivity of knowledge and description of persons?

It relies no doubt on the fact that, apart from differences in our knowledge, experience and background, we do share a substantial amount of knowledge and descriptions of the world in which we live and act, of the things with which we may act, of ourselves and of the persons with whom we may co-act. However, to say so does not of course add anything to our notions of 'shared knowledge and description', that is, knowledge and descriptions of things which have been available to shared inspection. Nor does it suffice to account for how knowledge of things which may only have been available to ourselves, and which in this sense is personal, is amenable to description and intersubjective communication. It does not do so unless it is assumed that the vantage points, backgrounds, and situations we may be in, are *in principle* shareable, and thus that other persons *could be* or *could have been* in the same situations. Nor does it suffice independently of assuming that, granted other people had had the same background, or had been in the same situations that we ourselves have been or are in, then they would have the knowledge of the situation and the things that we have, and describe them the way we do. However, this suffices, indeed it will *have* to suffice to say that it is fundamental to the cognition and experience of persons that, although other people may not be in our situations, and may not have, or may not have had, exactly the same experiences and knowledge that we have or have had, they would – *could* they be, or *had* they been, in our situations. Likewise, it suffices, and will *have* to suffice to say that to be language users and to share a language with other persons *logically* implies and presupposes that other language users, granted that they could be in our situation and have the experience, knowledge, background, points of view etc. that we have, would use language to describe what we experience, know of, etc., in those situations as we do. Or, they would consent that the descriptions we put forward about our experience and knowledge are correct and correctly applied.³

Now, if we can agree that these presuppositions must be fundamental to the cognition, use of language and communication of persons, and indispensable for any meaningful discussion among persons about what they know and how they describe what they know, I think we shall also have to agree that this intersubjectivity of cognition and language relies on a notion of 'truth' which implies that what is true or false, correct

³ The importance of this intersubjectivity of human cognition, communication and co-action, becomes clear when we consider that a substantial part of our knowledge of both material reality and of the societies in which we live and co-act with others, does not rely on first hand personal experiences, but rather is knowledge we have adopted or acquired from others. In this sense knowledge of reality thus acquired resembles knowledge communicated to us by others about their mental states, and in the sense, furthermore, that both kinds of knowledge relies on and presupposes the conditions of intersubjectivity of cognition and use of language outlined above.

or incorrect, is also true or false, correct or incorrect for other persons. That is, this intersubjectivity relies on a notion of 'truth' which *logically* implies and presupposes a notion of 'others'. (I shall expand on this point in a later section.) This assumed intersubjectivity must be the rock bottom, the very point of departure from which any discussion about our knowledge and description must be based and proceed - whether such discussions concern our knowledge and description of thing in publicly observable reality, or our internal states, such as our emotions, thoughts or feelings of pain.⁴

In what follows it will be argued that the intersubjectivity of human cognition, language and of the notion of 'truth' is a condition for the possibility of rigorously determining and thus of distinguishing between knowledge and descriptions of that which is publicly observable and that which is not - and hence for people together to develop conventions, criteria or standards for the validity and objectivity of their knowledge and description of things and events in publicly observable reality. And it will be argued, furthermore, that on these conditions relies the possibility of characterizing and determining individual differences and variations in the cognition and description of different persons - both concerning that which is publicly observable and shared and that which is *personal*.

Conditions for distinguishing what is publicly observable from what is not

It is well known that considerable variations may exist in how things are described by different persons, and by the same person in different situations, not only when the things concerned are "internal" states and events, but even things in material reality. Not only our intentions, purposes and possibilities of observation and action, but even our moods and temperaments may at times determine what we experience, and how we experience and describe ourselves and the rest of reality. So, would it not then be reasonable to reserve the notion of 'true descriptions and assertions', and 'knowledge of what exists objectively' for cases in which no such individual differences and variations exist, and to reserve the terms 'exist' and 'determinable' for things and events about which no uncertainty prevails - because they belong to what is materially and publicly observable?

This solution has been attempted, notably by the logical positivists and by the radical and logical behaviourists in both psychology and philosophy, who aimed to establish a firm epistemological basis for scientific research. In this pursuit they argued that only *that* exists objectively, and

⁴ As argued extensively elsewhere (Praetorius 2000), this assumed intersubjectivity of cognition, language and of the notion 'truth' can neither be proved nor doubted without being conceded, and hence will have to be taken granted as a *principle*.

hence can be the object of scientific research, which may be determined by rigorous public criteria and standards, and that only knowledge and description of what has been thus determined and observed, may be said to be meaningful and true. In effect, any determinate notions of the truth and meaning of statements and knowledge would have to derive from observations fulfilling such criteria and determinations. Consequently, what cannot be subject to rigorous public scrutiny and observation fulfilling such standards and criteria does not exist, nor can description of things which cannot be so observed be true; hence, descriptions and the existence of knowledge of such things may be discarded as nonsensical.

However, against such arguments we only have to consider that the very process by which we identify and determine a situation, and what in this situation is materially and publicly observable, presupposes that something is the case or true about the situation and things being determined and observed, which is also the case or true to other people – *in casu* the people involved in the observation. Thus, it is not because situations exist or may be arranged, in which things are publicly observable, and which we may come to agree to describe in particular ways, that the notion “emerges” of what - for everyone involved - is true or correct about things and situations. It is the other way around - for no such determinations of correctness of knowledge and description of things and situations could be agreed upon, let alone be arranged and function as criteria or standards for correctness, unless it was presupposed that *when* arranging and determining these criteria or standards, we *already* have a concept of ‘truth’ which we know how to use correctly; a concept of ‘truth’, furthermore, which is such that what is true or false, correct or incorrect, is also true or false, correct or incorrect for others.

Thus, the point, so easily overlooked, is that even in a situation in which the things and events being described are publicly observable and identifiable, it is logically implied and presupposed of descriptions put forward and being understood by others, and of these others, being in the same situation in which we are, that they will have the knowledge we have about these things and events, and will describe them as we do. In general, it is presupposed that our notion of ‘correct’ or ‘true’ is such that what is true or correct, is also true or correct for others. *This* presupposition granted, it is possible for persons - in spite of the differences and variations which exist in their knowledge, background, opportunities for observation and action - to arrange conditions and develop criteria and standards for what may count as correct, *objective* descriptions and *publicly* existing things, that is, possible to *determine and distinguish between conditions under which such criteria or standards apply, and in which they do not*. Furthermore, it is *because* of this presupposed intersubjectivity that it is possible to determine individual differences in the cognition and descriptions of different persons – both concerning that which is publicly observable and that which is only observable to

the persons themselves, such as their mental or internal states. In other word, it is because *both* the knowledge shared by different persons *and* the knowledge of persons which is personal, are perfectly sensible issues of intersubjective discourse *that it is possible to distinguish between what is publicly observable and shared and what is personal*.

Let me illustrate this point by giving the following example. One of my friends tells me: “I am terribly depressed; everything looks so grey and colourless - even the trees and flowers look grey and colourless”. Now, could it not reasonably be argued that at least in this case we are not talking about a “public” issue, but rather of something “private”, and also that it is a situation in which it would make no sense to maintain that my friend is still using language correctly? Is it not a situation in which any well-defined notions of correct or true assertions has been suspended? Not at all. For one thing, I do understand what my friend is saying. I am perfectly able to communicate with him about his - in this case - curious experience of the colours of trees and flowers. But a condition for maintaining that I understand what he is saying, and for communicating with him about his curious experiences of the colours of trees and flowers is, naturally, that he still uses language correctly when talking about his experiences of these things. That is, it is a condition that he knows the correct implications of terms for various colour categories, and that he knows how to apply them correctly. And it is a condition that what he is talking about is something he may refer to, and about which true and false assertions may be made, i.e. his curious experience of the colours of trees and flowers. Thus, it is a condition that both of us are still using the same language - indeed the very same language that he and I use under normal circumstances to talk about quite ordinary everyday matters; and it is a condition that we are together able to determine *what* he is talking about. If we are able so to do, and thus able to talk about his experiences, however curious, it has to be maintained that he is using language correctly when describing his experiences.

However, it is quite clear that his description of the colours of the trees and flowers is not of *general validity*, and I do not take his description as an attempt on his part to produce descriptions of general validity - i.e. descriptions which would be correct under normal everyday conditions of observation. But an important part of the message he is trying to convey to me - and which I understand - is precisely that his situation is *not* normal, i.e. that his depression affects his perception and description of things in ways which differ from how he normally perceives and describes them. A fact of which he himself is perfectly - and probably painfully - aware.

So, although my friend may feel eminently on his own with his unusual and personal experiences during his depression, neither his experiences nor his descriptions of them are *private*. They are perfectly understandable to others because it is presupposed - by us and by him - that he is

using language when describing what he experiences as others would - could they be in his situation and experience what he does. Indeed, our communication about what he experiences relies on the presupposition that what to him is the case or true about his experiences would also be the case or true for others, had they his experience.

Let me conclude my discussion of this example by saying that it shows that individual differences in the cognition and description of reality of different persons may indeed exist and be determinable. However, it also shows that a condition for these differences between persons and their cognition and description to exist, is that persons and language users, despite such differences, share a vast amount of knowledge and correct descriptions of reality. And it shows, furthermore, that the possibility of determining and of talking correctly about such differences relies on the presupposition that, when we describe what we know of or experience, we use language to describe it as others would, if they had our knowledge and experience; however, this in its turn both presupposes and implies that the notions of ‘correct’ and ‘true’ of persons are such that what is correct or true, is also correct or true to others. If these presuppositions and implications concerning the intersubjectivity of our cognition and language did not apply, no personal differences in knowledge and experiences could exist, nor any possibility to determine or talk sensibly about such differences - whether these differences concern what is or what is not publicly observable. Conversely, granted these presuppositions, and hence that situations, points of views and observations are in *principle* shareable, it is possible for persons to determine and distinguish between individual differences and variations in their cognition and description concerning both what is and what is not publicly observable, and to determine under what circumstances such individual differences occur. And it is possible *therefore* for persons together with others to develop criteria and standards with which rigorously to determine and distinguish between first-person and third-person observation and description, so crucial for the possibility of developing sciences.

In the section which follows I shall clarify what it means that the intersubjectivity of cognition and language relies on a notion of “truth” which logically implies a notion of “others”.

Implications of the logical relation between the notion of “truth” and “others”: the impossibility of private cognition and languages

Suppose my friend’s condition has deteriorated; he now tells me: “I have experiences and knowledge about some particular “things”, but what I take to be true or correct about them, is not true or correct for others, - or I cannot be certain that it is”. Now, what could he possibly mean by that?

Could what he says mean, for example, that if others could experience and know of the “things” that he experiences and knows of, then what for them would be the case or true about them, is not the same as what for him is true or the case about them? Well, if that is what he means, he is obviously contradicting himself, for in that case it would be impossible to ascertain that “the things” being experienced and known by others concern the same “things” as those experienced and known by my friend. Or, could it be that he means that these “things” of which he knows and experiences, are in principle inconceivable to others, because the notion of true and correct in the case of his unique experiences and knowledge of these “things” is different from the notion of true and correct which others have, and which he himself has in other cases, for example when communicating to others his unique experiences and knowledge. In other words, does he mean that this concept of true or correct, which applies to his experiences and knowledge about these particular “things”, is a concept which is special to him, in the sense: *private*, and consequently, that his experiences and knowledge of these “thing” are equally private?

Now, for such a claim to have any bearing - even for himself - it would seem to require that he be able to account for how his “private” concept of true and correct differs from the one he shares with others, i.e. account for it in the language in which the claim is put forward. But if he could do that, his “private” concept of true and correct would *not* be private, nor inconceivable to others. Likewise, in order to maintain that what he experiences and knows to be the case or true about these particular “things” is not the case or true for others, would seem to require that he be able to determine what would be the case or true for others about these “things”, and how it differs from what is the case or true to him - again in the language in which the claim is made. However, if he could do that, then what to him is true or false about the “things” he experiences would be perfectly conceivable for others, and what he knows about them would be perfectly expressible in terms of the language which he shares with others.

So, we may conclude that my friend is either contradicting himself or talking nonsense - or both. This would have been immediately obvious if instead he had said “I have discovered that the statement, “”, is true or correct, but it is not true or correct for others, or I cannot be certain that it would be true or correct for others”. What is obvious is that he mistakenly thinks that one could share a language with other people, and also that in this language the notions of ‘true’ and ‘correct’ could be different for different people.⁵

I think we shall have to agree that for the same reason that nobody

⁵ There is of course the possibility that what he means is merely that he is not sure of the correct implications and application of the statement - but that is a quite different matter.

may claim to possess private knowledge or a private language, and thus a private notion of truth, no such private language or knowledge may be ascribed to others - neither *in toto*, nor in part. For, how could we justifiably ascribe a private language or knowledge to others without being able to account for both such a language and knowledge in a language which is *not* private, and into which this knowledge and language must somehow be translatable? This, I think, suffices to show that *one cannot know for oneself what others might not in principle know as well*; that is to show, once again, that to be a user of a language one shares with others, means that one cannot know what is true or correct to say about a “thing”, which may not be true or correct to others, could they experience and know what one knows about the “thing”.

What it means that our notion of “truth” is fundamentally social

The previous section argued that the intersubjectivity of cognition and use of language of persons to describe what they know about themselves, others and the world around them relies on a notion of “truth” which logically implies a notion of “others”, and hence on a notion of “truth” which is fundamentally social. Now, it would seem to be almost self-evident that to be able to communicate about what one knows, implies knowing that one’s knowledge, categories, conceptual systems and descriptions are indeed “inter-personal”, i.e. that one shares such categories, conceptual systems and descriptions with others. Although I may know of and describe things which others do not (yet) know of, or know different things about them than others do, to know and to say so necessarily implies that if others had the possibilities of observing and describing the things that I have, then others would know what I know about them and describe them the way I do. In other word, what to me is true or false would also be true or false to others.

However, the notion that the ‘truth’ of cognition and language of persons is fundamentally social does *not* mean that it relies on the fact that persons may come to agree on and make conventions, and develop criteria, standards or rules for what may “count” as objective and true knowledge and descriptions of the things or situations in the world in which they find themselves. On the contrary, no social conventions, agreement or criteria about the truth and objectivity of knowledge and correctness of descriptions could be established among persons, unless *prior* to establishing such conventions, agreement and criteria, they had together determined and identified *both* things and situations to which these conventions and criteria apply, and *therefore*, unless they already had a notion of ‘truth’ in which it is presupposed that what is true or correct is also true and correct for others. Hence, to say that our notion

of ‘truth’ is fundamentally social does not mean that the notion of ‘truth’ is a social phenomenon, i.e. a product of socially agreed practice. On the contrary, social phenomena and practice, including the development of conventions, criteria, rules or agreement on how to use language and its terms, depend on notions of ‘correct’ or ‘true’ and ‘incorrect’ or ‘false’ which are inherently shared.

Consequences for developmental psychology

However fundamental – and almost embarrassingly banal – the presuppositions concerning the intersubjectivity of cognition, language and the notion of truth of persons may seem, it has been widely overlooked within philosophy of mind and consciousness. According to the traditional assumption, shared by many philosophers even today, we all start out as “Cartesian subjects”, having knowledge and experience of the content of our own mind, i.e. our sense data, perceptions, thoughts, emotions and feelings. From this supposed private, though certain knowledge and experiences “from our own cases” it is believed to be possible to work towards true knowledge of the nature of what causes this content and the rest of the objective order of reality, including other persons, and to develop a language in which we may talk with others of this knowledge and experience. This same assumption also seems to inspire and lie behind functional models and accounts of the cognition and use of language of people currently being developed within Cognitive Science – be those models computational *cum* representational or connectionist. However, it would seem to be a insoluble problem for such models to account for how the cognition of an individual – formed in “splendid solipsistic isolation”, and with no notion of the intersubjectivity of its own cognition and that of others – would ever come to accord with the cognition of others. And it would be impossible without such a notion to explain, moreover, how they come to distinguish that part of their knowledge concerning themselves and the world which is *personal*, from that which is *public* available and shared by others.

George Herbert Mead was among the first within psychology to provide substantial arguments against the assumption that individuals living mentally in isolation from others, could ever develop or acquire knowledge about themselves, others and reality of the kind and in the way that we humans actually do, and in particular to argue that no such individual could be aware of itself – since, according to Mead, to be aware of oneself is to “look” at oneself from the standpoint of another (Mead, 1934). In view of the arguments presented so far, it seems that we are now able to strengthen Mead’s original claim. If we suppose with Mead that to be aware of oneself requires being able to “look” at oneself from the standpoint of another, it has to be conceded, that this “looking” both

presupposes and implies that if one could be in the position of other persons and look at oneself as they do, then one would see and come to know oneself the way others do. Without the presupposition that what, from the position of others is known by them to be the case or true about me, would also be the case or true for myself, if I could be in their situation and have their knowledge, there would be no sense in talking about seeing oneself from the point of view of others. Thus, a notion of truth which is such that what is true or the case is also the case for others, and hence a notion of truth which logically implies a notion of others, must necessarily *precede* the possibility of anyone “looking” at himself from the standpoint of another; it is not something which may be acquired by such “looking”, nor by imagining such “looking”.

Apart from these logical reasons, there seems to be empirical support from developmental psychology for assuming that for an individual, say a child, to become aware of himself and of others as persons, and of acquiring notions of his or her own mind and cognition and those of others, it is necessary that he or she be *received* and *understood* by others as a person. Thus, empirical research of early mother-infant communication seems to show that a child’s ability to develop knowledge about himself, reality and other people around him, and later on to acquire a language to communicate this knowledge and that of others, depends on the mother’s (or other care-person’s) indefatigable effort and willingness to understand and see the child’s behaviour as being *intentional*. And it seems to rely on the mother’s efforts to understand, not only what goes on “inside” the child, but to interpret the child’s reactions to her, and his action with things, as expressing his knowledge about things, and his attempts intentionally to act upon them (for an excellent account of the development of early mother-infant interaction and communication, see Bruner, 1983).

Empirical investigations also seem to show that the child up to a certain age – presumably due to an over-generalisation of what he sees as his mother’s apparent unlimited knowledge about his experiences, intentions, needs and actions – believes that others are in the same situation as he himself and share his point of view, and the knowledge he has about things in those situations – and even that they may “have” his thoughts and feelings. Thus, the child seems to over-generalise the fact that “what he knows may also be known by others” to mean that others do indeed find themselves in exactly the same situation as he does, and having the same knowledge he has, and having access to his thoughts and feelings. Only later on does the child learn that other people may perceive the same situation from points of view which are different from his. And only later does the child realise that part of himself, his feelings and thoughts, are only directly observable to himself, and also that this part of him is what makes him uniquely *him*, being a person both physically and mentally distinct from other persons.

If we go back to the assumption held by most philosophers even today, namely that the child starts with “private” and “subjective” knowledge about himself and the world, it would now seem that this assumption turns the issue on its head. For, if what empirical investigations of the child’s initial development of knowledge seems to indicate is correct, and what from a logical point of view must necessarily be the case, the child does not start with private knowledge “from his own case”, but with knowledge of which it is assumed by the child that it is indeed shared by others – i.e. by his mother. The problem, it would seem, is rather to account for how the child later in his development comes to appreciate that, although what he knows may indeed be known and shared by others, others may not be in exactly the same situation as he himself. Thus, the problem seems to be to account for how the child begins to learn to appreciate the notion of ‘different points of view’, and how he begins to learn and appreciate the difference between what is and what is not observable to him as opposed to others, i.e. that the knowledge one may have of a situation and of oneself may be *personal*.⁶

However, what cannot be accounted for nor explained, but which has to be presupposed and taken for granted, is that for a child to be able to learn this from other people in the community in which he grows up, and be able to take part in their “forms of life”, the child must have a notion of ‘truth’ which is such that what is the case, true or false, is also the case, true or false for others. What is lacking in epistemological approaches which start from the position of the individual alone set against the rest of the world – be they generic constructive approaches, or biological or computational functional approaches – is not just a social context of others, which enables the individual to confront and compare his knowledge with the knowledge of others for the purpose of determining, for example, whether his knowledge is in accordance with theirs, and thus may “count” as objective or true, or whether it relies on one’s subjective dreams, illusion or imagination. What is lacking is precisely a notion of ‘truth’ which logically implies and presupposes a notion of ‘others’, which makes it possible for persons together to develop procedures for determining the objectivity and truth of their knowledge, and for everyone to compare his knowledge with the knowledge of others. That is, an intersubjective notion of ‘truth’ which makes it possible for someone to *be* a person, i.e. someone who may share an incredible amount of knowledge with other people, but who also has knowledge about himself and the world which is uniquely personal – and thus is someone who also differs from other persons; someone with whom we may agree – and disagree.

⁶ For thorough investigations of this development, see Tomasello and Rakoczy (2003)

Consequences for a science of consciousness: Conclusion

Ever since the start of experimental psychology in the late 19th century, concerns has been raised about the use of first-person reports by subjects in psychological investigations to report about their experiences, feelings, thoughts, emotions and other mental “goings-on”. The concerns centre around two main questions, both of which cast doubt about the scientific status of psychological investigations using first-person reports. The first question, as formulated in present day discussions among researchers within cognitive neuroscience, is

1. How can correlational studies of brain processes and concurrent conscious mental states be considered *scientific*, given that the descriptions of brain processes rely on rigorously developed methods of observation and precise, well defined terms to describe such processes and structures, whilst observations and reports by subjects about their conscious mental states lack this rigour of observation and precision of description?

The second question is

2. how can descriptions and reports of something to which only the subjects have access and may observe, i.e. their conscious mental states, be considered as data for scientific research, given the requirement for such research of public criteria to determine *what* is being observed, and third-person validation of the correctness of descriptions of such states?

The problems raised in (1) and (2) about the use of first-person reports in scientific research, although often confounded, involve two distinctly different issues, the clarification and solution of which requires distinctly different approaches. Thus, whilst the concerns raised in (1) concerns *methodological issues* about the conditions of subjects (and experimenter) reliably to observe and describe phenomena of a first-person nature, the concerns raised in (2) involve *epistemological* issues about the very possibility of such phenomena being amenability to scientific scrutiny.

Concerns about (1), the methodological issue, has been systematically addressed and solutions attempted by a number of cognitive scientists.⁷ The solutions suggested typically include the development of adequate

⁷ Ericsson and Simon 1993, Varela 1996, Vermersch 1999, Roy et al. 1999, Lutz and Thompson 2003, Jack and Roepstorff 2002, Overgaard, 2001, Galagher, 2003, Ericsson 2003. For extensive discussions on these matters I refer the reader to the special issues of Journal of Consciousness Studies, *Trusting the Subject*, Part 1 (2003), and Part 2 (2004).

conditions and training of subjects to observe and be aware of their “mental goings-on”, as well as of appropriate vocabularies and terms by which to describe them. The assumption is that, granted this pre-experimental training as well as adequate conditions to observe and describe their conscious mental states, then experimental subjects will be able to learn how to observe their mental states during the experiments and to use these terms both reliably and correctly to describe them - as well as “what it is like” to have them. Indeed, with such methodological techniques firmly in place, research using introspective data have proven to measures up to the methodological requirements which apply for scientific research in general, such as the possibility of precise description of experimental conditions and reproducibility of data - that is, the same introspective reports may be reproduced under the same experimental conditions either by the same subject or by different subjects.

Concerns about (2), the epistemological issues, i.e. the lack of third-person validation of first-person phenomena as well as reports by subjects about them, is a much deeper problem than the first, the methodological problem, in that the very *status* of first-person reports about such phenomena as scientific evidence is called into question. Unlike the first problem, this is not a problem that may be solved by the employment of appropriate procedures and methods for first-person observation and description of conscious mental states. Concerns over this absence of third-person validation have given rise to "resistance to introspective evidence"(Jack and Roepstorff 2002), in particular among cognitive researchers versed in the methods of the natural sciences – and underlie the contention of Dennett that, "First-person science of consciousness is a discipline with no methods, no data, no results, no future, no promise. It will remain a fantasy"(Dennett 2001; quoted in Galagher 2003). The consequence of this view for cognitive research seems clear - for if the absence of third-person validation disallow introspective reports as evidence, then it is hard to see how the scientific study of consciousness could get off the ground at all.

A more optimistic view, now shared by many cognitive scientists, has recently been put forward by Bernard Baars (Baars 2003). According to this view the solution of the problems of validation of first-person phenomena and reports about them is actually being made available by third-person data from within cognitive neuroscience. Thus, in “How Brain Reveals Mind”, he argues that although “it is still true that brain evidence has greater credibility than subjective reports, no matter how reliable”, new brain studies of a wide variety of conscious states show “increasing convergence between subjective experience and brain observations”; indeed, such studies “supports the central role of consciousness” (ibid.). However, the argument that evidence from observations of the brain vindicates the existence and role of conscious states, as well as the reliability of the subjective reports about them, suffer in my view

from circularity in that the *correlational* studies cited by Baars of brain states and concurrent conscious states all *depend* on and hence *presuppose* both the existence of the conscious states being investigated, and that the subjects' introspective reports about them may be relied upon to be in general correct and reliable. Put differently, to the extent that the experimental determination of neuronal mechanisms giving rise to or underlying conscious states necessarily relies on the reliability of introspective reports by subjects about their conscious states during the experiments conducted, it can never be an empirical *outcome* of such experiments that these conscious states actually exist and that the subject's reports about them are in fact reliable. Thus, rather than being vindicated by empirical research, both the existence and the reliability of the reports by subjects about the conscious states under investigation is taken for granted, indeed, this assumption is the very point of departure for all empirical studies within cognitive neuroscience - without which there could be no such scientific studies of conscious mental states at all.

Appeal to data from this kind empirical research - whether present or future - would not convince the sceptics among cognitive scientists and philosophers holding the views of Dennett. Nor may their scepticism concerning the use of introspective reports as scientific evidence, i.e. on the grounds that what these reports are about is not publicly accessible, be overcome by the fact that these data are reproducible - *any more* than a sceptic concerning the stories he reads in a newspaper would be convinced about their truth by repeated reading. Since their scepticism is of an epistemological nature, i.e. it concerns the very possibility of knowledge and description of phenomena which are not publicly observable to be reliable and true, such scepticism may only be refuted by appeal to principled epistemological reasons.

The arguments put forward earlier in this paper about the general conditions for intersubjectivity of cognition and description would seem to represent just such principled reasons. Thus, it was argued, firstly, that everyday communication about our internal or mental states in general, as well as any philosophical discussion - and indeed scepticism - as to both their existence and the status of our cognition and descriptions of them, necessarily rests on the presupposition that as users of the language we share with others, we know correct implications of terms for such states, just as we know their correct application. Secondly, it was argued that the conditions for intersubjectivity of knowledge, description and communication about our mental states are the same as those which apply for knowledge, description and communication about things which are publicly observable. This condition is expressed by the general presupposition applying in both cases, namely, that what is true or correct about the things or states we observe and experience would also be true or correct to others, could they observe and experience the things or states that we do - and vice versa. Put differently, if other

people could be in our situation, and had the same possibilities of observing the things and states that we do, then they would observe what we do, and use language to describe them as we do - and vice versa. This assumption, and hence that knowledge and descriptions of what is only first-personal is just as amenable to intersubjective discourse as are knowledge and descriptions of that which is third-personal and publicly observable, is *logically* necessary for developing criteria and standards with which rigorously to determine and distinguish between first-person and third-person observation and description, and hence to distinguish what is *publicly observable and what is not*, so crucial for the possibility of developing sciences.

According to these arguments there would seem to be no principled epistemological reason why first-person verbal reports by experimental subjects about their internal or mental states may not be considered as valid data for scientific research.⁸ For, although these arguments do not render reports of such states open to third-person validation, the point still stands - namely that these states may be just as amenable to *reliable* observation and *correct* descriptions as are observations and descriptions of publicly observable things or states – and hence in this respect may fulfil the requirement of scientific evidence. Since this is an assumption which, as argued above, is a necessary condition for any rigorous determinations of and hence any well defined distinctions between what is publicly accessible and what is not, it must be among the very basic assumptions on which science itself is founded.⁹

References

Baars, B. (2003). How Brain Reveals Mind: Neural Studies Support the Fundamental Role of Conscious Experience. *Journal of Consciousness Studies* 10: 100-114.

Bruner, J. (1983). *Child's Talk. Learning to Use Language*. London: Norton and Company.

Dennett, D. (2001). The fantasy of first-person science. Nicod Lectures. Private circulation.

⁸ This said, it does not, of course, mean that no *practical* problems may exist in observing and giving adequate or satisfactory descriptions of our internal or mental states. How these problems may be overcome, i.e. the problems referred to as the *methodological* problems of introspective reports, are thoroughly discussed by, among others, the researchers and volumes referred to in footnote 8 of this paper.

⁹ Thanks to a reader of this paper for pointing out that Erwin Schrödinger presented similar arguments in 1935 in his discussion of the conditions on which scientific knowledge within physics are based (Schrödinger 1935).

Ericsson, K.A. (2003). Valid and Non-Reactive Verbalization of Thoughts During Performance of Tasks. *Journal of Consciousness Studies* 10: 1-18.

Ericsson, K.A. and Simon, H.A. 1993. *Protocol analysis: Verbal reports as data*. (Revised ed. 1st ed. 1984). Cambridge Mass: MIT Press.

Galagher, S. 2003. Phenomenology and Experimental Design: Towards a Phenomenologically Enlightened Experimental Science. *Journal of Consciousness Studies*. 10:85-99.

Jack, T. and Roepstorff, A. (Eds) (2004). Trusting the Subject, Part 2. Special Issue of Journal of Consciousness Studies, 11, No. 7– 8.

Jack, T. and Roepstorff, A. (Eds) (2003). Trusting the Subject, Part 1. Special Issue of Journal of Consciousness Studies, 10, No. 9 – 10.

Jack, T. and Roepstorff, A. 2002. Introspection and Cognitive Brain Mapping: From Stimulus-Response to Script-Report. *Trends in Cognitive Science*.

Lutz, A. and Thompson, E. (2003) Neurophenomenology. *Journal of Consciousness Studies* 10:31-52.

Mead, G.H. (1934). *Mind, Self, and Society*. Chicago: University of Chicago Press.

Overgaard, M. 2001. The Role of Phenomenological Reports in Experiments on Consciousness. *Psychology* 12 (29).

Praetorius, N. 2000. *Principles of Cognition, Language and Action. Essays on the Foundations of a Science of Psychology*. Dordrecht/Boston/London: Kluwer Academic Publishers.

Roy, J. M. , Petitot, J., Pachoud, B. and Varela, F. J. 1999. *Naturalizing phenomenology*. Stanford, CA: Stanford University Press.

Ryle, G. 1949. *The Concept of Mind*. London: Hutchinson.

Schrödinger, E. 1935. "Quelques remarques au sujet des bases de la connaissance scientifique", *Scientia*, LVII: 181-191.

Tomasello, M. and Rakoczy, H. 2003. What Makes Human Cognition Unique? From Individual to Shared to Collective Intentionality. *Mind & Language* 18: 121-147.

Varela, F.J. 1996. Neurophenomenology: A Methodological Remedy to the Hard problem. *Journal of Consciousness Studies* 3: 330-50.

Vermersch, P. (1999). Introspection as Practice. *Journal of Consciousness Studies* 6: 17-42.

Wittgenstein, L., (1945/1953). *Philosophical Investigations*. Ed. G.E.M. Anscombe and R. Rees. Oxford: Blackwell.

Tingene er, som de er, fordi det er bedst sådan: Om kausalitet og finalitet i sjælelivet og intentionalitetens kilder

Af *Boje Katzenelson*

Hvorfor kan vi se? Sådan kan børn finde på at spørge troskyldigt, eller snedigt, hvorfor hvorfor dit og hvorfor dat? Fordi vi har øjne, svarer vi måske og da næsten lige så ureflekteret, for skønt svaret er rigtigt, befatter det sig kun med den halvdel af sagen der henviser til et eksisterende biologisk organ. Svaret indregner ikke dette organs tilblivelseshistorie, der i høj grad drejer sig om en psykologisk funktion, nemlig det at kunne se, synet, se-heden, se-kunnen.

Evolutionshistorisk betragtet er det nemlig ikke øjet, der er årsagen til at (nogle) organismer kan se. Det forholder sig omvendt: Det at kunne se er årsagen til øjets både tilblivelse og form. Det er organets psykologiske egenskab, se-kunnen, der har været bestemmende for dets biologiske tilblivelse og nuværende anatomi og fysiologi. Den psykiske disposition og kapacitet, vi kalder visuel perception, er evolutionært blevet til fordi egenskaben har vist sig fordelagtig for pågældende art at være udstyret med, og derfor er den blevet selekteret. Denne psykiske egenskab og funktion må have et somatisk organ at virke i og ved, og derfor er det biologiske øje blevet til. Øjets væsen er imidlertid psykisk: Til-at-se. Hvis øjet var et selvstændigt legeme og ikke kun ét organ blandt andre, ville synet, se-heden eller se-kunnen være dets psyke.

Den slags betragtninger gjorde Aristoteles sig (her i min version) for knap to et halvt tusinde år siden, og de hænger sammen med hans begreb om årsager og virkninger. Redegørelsen herfor er nødt til at overspringe disse begrebers filosofiske forudsætninger hos Aristoteles, som har at gøre med tanker over forholdet mellem stof og form, potentialitet og aktualitet, bevægelse, kategorier m.m.

Reverens

Denne antologi har metode som overordnet tema. Metode forstået i snævrere forstand som empirisk undersøgelsesdesign er ikke min favoritdisciplin, så det holder jeg mig fra. Derimod jeg nok kan ræsonnere over metode i bredere forstand som teoretisk tilgangsvinkel el.lign. Sådanne betragtninger indeholder artiklen da også, men de er ikke dens hovedsigte, med mindre den bærende evolutionære tankefigur kan siges at være en metodevinkel i bredeste forstand som når Descartes åbner Vestens nyere filosofihistorie med bogtitlen *Discours de la méthode* eller Gadamer tre hundrede år senere udsender *Wahrheit und Methode*. Her betyder metode vel noget i retning af erkendelsesmatrice. Sådant opfatter jeg selv artiklens bærende evolutionære tankefigur, men vil ikke argumentere nærmere for det her.

Metode eller ej, så ville jeg være ked af ikke at kunne deltage i et festskrift til Benny Karpatschof, af to grunde. Dels fordi jeg da ikke kunne kvittere for vort mangeårige samarbejde; vi stiftede i 1980 sammen med andre tidsskriftet *Psyke & Logos*, naturligvis med Benny som initiativtager, dernæst redigerede vi det sammen (og sammen med andre) i 18 år indtil jeg fratrådte; senest har vi redigeret bogen *Klassisk og moderne psykologisk teori* (Hans Reitzel 2007). Dels ville det bedrøve mig ved denne vending i Bennys liv ikke at kunne yde min hyldest til den institution i dansk psykologi, som hans intelligens, kundskaber, virketrang og personlighed har udgjort i så mange år og som har spredt generøse ringe af indsigter, rummelighed og stille livsglæde omkring sig i vort ellers ofte nok bornerte akademi.

Grundlaget

Trods de mange års samarbejde tillige med anden faglig samkvem siden vort første møde i 1968, har Benny og jeg ingenlunde samme opfattelse af sjælens dybere dele og dermed heller ikke af psykologiens genstandsfelt. I et af sine seneste arbejder skriver han at "... de træk ved mennesker, som er så generelle, at vi har dem til fælles med andre dyrearter, falder ... ind under biologien, ikke psykologien" (Karpatschof, 2006, 15). Og videre om Galilei at "... han undersøgte objektive kvaliteter ved rene objekter. Entiteter, der ikke bare besidder objekt karakter i relativ, men absolut forstand. Det er jo netop denne forudsætning, som ikke er opfyldt i psykologien, og som kun kan honoreres ved en særlig objektiveringsmetodologi over for de marginale psykologiske fænomener, som er knyttet til vort biologiske udstyr" (ibid., 135-36).

Det er jeg ikke enig i. Jeg mener nærmest modsat at nævnte psykologiske fænomener ikke er marginale, men centrale (Katzenelson, 1989). Mere præcist kan forskellen mellem os desangående belyses ved evolutionsspsykologiens skelnen mellem to typer faktorer til forklaring af psykiske

egenskaber: De nære og de fjerne eller henholdsvis de proksimale og de distale eller ligefrem yderste (*proximate and ultimate*). Proksimale faktorer er de umiddelbart forekommende årsager til adfærd, som mennesker i nogle tilfælde er bevidste om og på anfordring kan begrunde og redegøre for. De distale faktorer udgøres derimod af bagvedliggende årsager i skikkelse af de stabile dispositioner og kapaciteter, som vor art er anlagt på som følge af dens fylogenetiske evolutionshistorie og som sjældent er bevidste og begrundelige i gerningsøjeblikket. (De stabile dispositioner og kapaciteter der er knyttet til individets ontogenetiske biografi, fx personlighedstræk, er ikke til behandling her).

De to typer faktorer optræder sandsynligvis ofte (måske altid) i fællig, men må ikke sammenblandes, da de repræsenterer forskellige forklaringsplaner eller -niveauer. Ét er seksuel lyst her og nu, et andet spørgsmålet om hvorfor og hvordan vi overhovedet er erotisk inklinerede. Én ting er kulturelle normer og regler som forklaring på en given adfærd, en anden hvorfor og hvordan vi som art betragtet er anlagt på norm- og regelfølgende adfærd (Katzenelson, 2007).

Det er de distale dybtliggende evolutionært tilblevne biopsykiske faktorer eller psykens fundamentsten, denne artikel behandler. De yderste ting. Dem der er, som de er.

Årsag og virkning

Begrebet årsag betød for Aristoteles alle de nødvendige, men hver for sig utilstrækkelige, betingelser der kræves for at kunne redegøre fuldt for en tings væsen. Disse betingelser klassificerer han i fire typer årsagsforklaringer: Den virkende, den materielle, den formale og den finale, som alle må bringes i anvendelse i livsvidenskaberne, hvis vi skal hele vejen rundt i vor virkelighedsforståelse.

Kausalitet og finalitet

1. Den virkende årsag (*causa efficiens*) henviser til de ydre faktorer, som sætter en forandringsproces i gang ved at overføre en kraft af en slags fra ét fænomen, A, til et andet fænomen, B, og dermed ændre sidstnævnte. Årsagen A går forud for virkningen B, og der er en nødvendig forbindelse mellem A og B på den måde, at virkning B ikke kan forekomme uden forudgående årsag A. Derfor kan en virkning ikke være sin egen årsag. A og B er både logisk og empirisk uafhængige af hinanden (Wright, 1971, 93) eller står med østerbergs ord (1966) i et udvendigt forhold til hinanden.

De tre øvrige årsagsforklaringer har med tingenes beskaffenhed eller indre egenskaber at gøre.

2. Ved den materielle årsag (*causa materialis*) forstås de virkninger, der afhænger af tingenes iboende muligheder eller potentialer. Det gør en forskel på virkningen om den forårsagende kraft overføres fra en billardkø til en billardkugle af træ eller til en gummibold.
3. Den formelle årsag (*causa formalis*) henviser til de potentialer, der er givet ved en bestemt formstruktur. Det gør en forskel på virkningen om den forårsagende kraft overføres til en rund eller en elliptisk billardkugle.

Hvis den materielle og formelle årsagsforklaring (2) og (3) opfattes som aspekter ved den virkende årsag (1), kan de alle tre sammenfattes til spørgsmålet om forandringsprocessers forudgående og efterfølgende betingelser og udgør tilsammen **kausale** årsagsforklaringer på tingenes væsen.

4. Fra denne kausale adskiller sig radikalt den **finale** eller teleologiske eller mere generelt og nutidigt udtrykt den funktionelle årsagsforklaring (*causa finalis*), der drejer sig om det mål (*telos*) som forandringsprocessen er stilet mod og normalt resulterer i.

Denne sidste finale årsagsforklaring var den vigtigste for Aristoteles. Den endnu ikke aktualiserede, men kun som mulighed eller potentiale foreliggende, psykologiske egenskab se-kunnen er den forårsagende kraft, som stiler mod at udvikle og resultere i et biologisk organ, et øje, til at virkeliggøre sig i, hvorved dette biologiske organ tjener den psykiske funktion at kunne se. Foruden ydre kausale årsager, som også gør sig gældende - som nævnt må alle fire årsagsforklaringer ifølge Aristoteles inddrages, hvis fuld redegørelse for tingenes væsen skal opnås - består årsagen til øjets udvikling og form også af det indre mål, se-kunnen, som øjet er til for og resultatet af eller hvis funktion det tjener. Udtrykt i evolutionær kognitionspsykologisk nutidssprog (Marr, 1982): Det biologiske organ eller apparatur, der er ansvarlig for den psykologiske egenskab visuel perception, opstod evolutionært og erhvervede sin form eller arkitektur og indre egenskaber som løsning på bestemte tilpasningsproblemer såsom virkelighedsorientering. Den rigtige måde at forstå disse mekanismer på er derfor at undersøge den slags problemer, som de blev designede til at løse. Forståelsen vil da tage skikkelse af spørgsmålet: Hvad sker der med og ved apparatet, hvis en bestemt funktion skal lykkes?

Bemærk at begrebet mål ikke er identiske med objekt (hvilket nogle psykologiske retninger mener). Mål betyder et foregrebet andet forhold mellem subjekt og objekt end det foreliggende og dermed en anden psykisk tilstand hos subjektet end den aktuelle, en anden tilstedeværelses- eller funktionsmåde. Se senere om foregribelse.

Den virkende og den virksomme bevægelse

Causa finalis indebærer, at et fænomen ikke primært er betinget ved et andet fænomen uden for sig selv i rum og forud for sig selv i tid, omend det formentlig ofte eller måske altid også er medbetinget heraf - alle fire årsagsforklaringer er jo nødvendige i livsvidenskaberne - men primært er betinget i og ved sig selv. Det finale eller teleologiske årsagsprincip er hos Aristoteles (såvel som hos forgængeren Platon) et begreb om årsag som indre væsen, ikke som i kausale årsagsforklaringer et begreb om årsag som ydre betingelse for en begivenhed (Johansen, 1988, 105).

Den finale årsag forløber også fra A til B og ikke omvendt, ligesom den kausale gør det, deri er der ingen forskel mellem de to årsagsforklaringer. Det er den henholdsvis ydre og indre betingethed, der er den afgørende forskel mellem det kausale og det teleologiske årsagsprincip, og dermed også to forskellige slags bevægelser: Den virkende og den virksomme (Engelsted, 1989 II, 6, 70).

Den virkende bevægelse forløber i kausale årsagskæder og gælder muligvis uindskrænket for det uorganiske stof, skønt kvantefysikken måske har vanskeligheder ved at måtte nøjes med kausale forklaringer på visse fænomener, såvel som i den organiske materie der ikke er udstyret med psyke, hvilket groft sagt vil sige planteriget. Objektet eller stimulus påvirker organismen, hvorefter organismen reagerer som denne årsags virkning. Tilstandsændringer i det med psyke bestykkede organiske stof kan derimod også skyldes en virksom eller selvvirksom bevægelse forløbende som subjektets teleologiske årsag (indre væsen) relationelt til objektet. Objektet eller stimulus kommer da ikke før organismens reaktion, tværtimod opsøger og inddrager eller aktiverer organismen sit objekt, således at organisme og objekt udgør momenter ved hinanden i tilstandsændringen. Den virksomme bevægelse er foregribende rettet mod at bringe et manglende objekt til veje, hvilket den virkende kausale bevægelse ikke er. Den teleologisk eller finalt forårsagede bevægelse er selvaktivt rettet mod at realisere sit iboende mål (årsag som indre væsen) i fremtiden ved hjælp af et objekt som middel for dette mål, hvad den kausale årsag ikke er. Ud over at blive på-virket af virkende stimuli, eftersøger øjets se-kunnen også selvvirksomt synlige objekter brugelige som middel til mål.

Fordi mennesket besidder en bevidst psyke og ikke kun en psyke slet og ret, kan denne almene psykiske foregribelse eller fremtidsrettethed hos os være forskellig fra andre organismers. Bevidsthedens både refleksive og prærefleksive planer gør (med fx Heideggers ord) mennesket til det væsen, hvis eksistentialitet det er at være en mod fremtiden rettet for-tidig nutid, og denne rettethed eller intentionalitet kan undertiden være bevidst artikuleret. Det er imidlertid af mindre interesse her. Det teleologiske eller finale årsagsprincip henviser ikke primært til menneskets bevidste og eventuelle frit viljede ønsker om og hensigter med et fremtidigt mål, som det søger at realisere (Katzenelson, 1996). I dagligdagens

psykologi er det almindeligt at udnævne bevidst hensigten eller ønsket til den årsag, der sætter handlingen i gang som virkning. Det kan den videnskabelige psykologi ikke nøjes med, for så kan der spørges videre til hvilken årsag hensigten eller ønsket på deres side har? I hverdagspsykologien vil det spørgsmål ofte blive besvaret ved at komme op med endnu en hensigt eller et ønske nummer 2 forud for og som bevirkende årsag til hensigt eller ønske nummer 1, og vi er i gang med en såkaldt uendelig regres der blot finder på flere betegnelser for det samme.

Den videnskabelige psykologi må kunne levere mere solide ankerpunkter og opsporer derfor ændringsprocessens årsag to steder. Det ene er kausale årsagsforklaringers virkende bevægelse. I sin konsekvens kan det føre til reduktionistiske antagelser. Den anden er teleologiske årsagsforklaringers selvvirksomme bevægelse på emergent højere virkelighedsplaner end dem, hvori kausalitetsprincipperne eneheresker. Det vil sige ved iboende mål, der hos mennesket undertiden ledsages af bevidste hensigter, ønsker og lignende mentale forekomster. (Se neden for om reduktionisme og emergens). Det teleologiske årsagsprincip for psykens væsen indbefatter menneskets bevidsthed, men denne er ikke primær for en videnskabelig psykologi der eftersøger årsager og ikke kun grunde og begrundelser. Psykologisk betragtet var Marx derfor ikke dyb, da han anskueliggjorde forskellen mellem dyret og mennesket med byen og muremesteren. Byen bygger bare sin kube, kan ikke andet, medens muremesteren har en hensigt med sit virke og forud planlægger husets arkitektur. Den forskel kan selvsagt være til stede og være vigtig, men den er ikke dyb i betydningen videnskabelig forklaring på de yderste ting. Den udgør højst en filosofisk-psykologisk forklaring på de nære ting.

Den slags muremesterforklaringer kalder Hegel for en "ydre" eller "subjektiv" årsagsteologi, der risikerer at løbe ud i en endeløs regres (en "slet uendelighed" kalder Hegel det). I stedet herfor sætter Hegel med Aristoteles en "indre" eller "objektiv" teleologi (Hartnack, 1990, 100), hvor målet og midlerne eller processerne mod målet ikke er uafhængige af hinanden, men er givet i og med hinanden i en dialektisk relation som hinandens både forudsætninger og resultater. Marx skulle hellere have sagt, at mennesket er anlagt på at hule og huse sig snarere end at kube sig, ligesom øjet er anlagt på se-kunnen snarere end hudtrykssensibilitet. Derfor er mennesket Platonisk sagt delagtig i beboelsens ide på nogle måder, der er forskellig fra byens delagtighed heri. Men det dybe er, at både byens og muremesterens virksomhed yderst er styret af samme type årsagsprincip, nemlig et iboende mål der har forskellige udtryksformer og tager forskellige midler i brug for at virkeliggøre, hvad det forlods er anlagt på at opnå. At have til hensigt og beslutte sig for at planlægge bygningen af et hus og kunne begrunde hvorfor er i denne sammenhæng mindre interessant for Hegel, fordi hans "indre" teleologiske forklaringsprincip er objektivt og ikke subjektivt. Det vil sige at det er en alment gældende lovmæssighed, der er principielt uafhængig af individu-

elle beslutninger, men som selvfølgelig konkret kun kommer til udtryk i individuelle foreteelser.

Funktionalitet og cirkularitet

En af naturvidenskabens indvendinger mod teleologiske eller finale årsagsforklaringer er, at de forekommer at være cirkulære. Se-kunnen er årsag til øjets fremkomst i evolutionshistorien, og således fremkommet kan øjet nu se, fordi se-kunnen er dets væsen og opfylder dets iboende mål. Dette væsen er en fugl, fordi det kan flyve, og det kan flyve fordi det er en fugl.

Sandt nok, men sådan er det nu engang. Fugle kan eller har kunnet eller har været på vej til at kunne flyve (for ikke alle nulevende fugle kan flyve, som bekendt), fordi det har vist sig evolutionært adaptivt for disse væsener at besidde denne dispositionen og kapacitet. Flyve-kunnen er flyvelegemets psyke, ligesom se-kunnen er øjets. Det kaldes som nævnt også en funktionel årsagsforklaring: Et fænomens væsen forstås ved den funktion det udfylder, ved dets til-at (se eller flyve), og funktionen forstås ved dets væsen.

Den slags cirkulært funktionelle årsagsforklaringer har ikke været populære i naturvidenskaberne siden Galileo. For at bryde cirklen har man krævet lineære kausale årsagsforklaringer. Imidlertid *er* evolutionsforklaringer funktionelle og dermed (i forskellig udstrækning) teleologisk cirkulære. Det er der ikke noget at gøre ved. Det må man slet og ret acceptere overalt, hvor man befatter sig med evolutionære udviklingsforklaringer på at noget er, som det er, som følge af dets tilblivelse. En adfærdskapacitet eller –disposition er blevet til, fordi den har vist sig mere fordelagtig end ufordelagtig for denne art i denne biotop, hvorfor den blev selekteret. Det, som den er til, har bevirket at den er til og har formet dens til-at, *and that's that*.

Den retro-progressive metode

Hvis man ikke kan lide det, må man opgive at befatte sig med evolution eller med historiske årsagsforklaringer overhovedet, for en vis cirkularitet er uundgåelig i den slags historisk forskning, der sammenligner nutid og fortid. Man undersøger kildemateriale til belæg for, at dette og hint har eksisteret og fundet sted på denne og hin måde, hvor ud fra man slutter sig til mulige forklaringer på, hvorfor det nu ser ud og opfører sig som det gør. Man går fra da til nu.

Men hvis der kun foreligger sparsomme eller vanskeligt udgrundelige fund fra fortiden, hvis det ikke er soleklart hvad der har eksisteret og fundet sted, og det er desværre ofte nok tilfældet i evolutionsforskningen, må denne fremgangsmåde suppleres med den modsatte fra nutiden til fortiden. Det var (formentlig) hvad Marx mente med det udsagn, at

menneskets anatomi er nøglen til abens anatomi. Tror man at mennesket nedstammer fra aben, kan man undersøge abens anatomi og fra denne fortid trække tråde til forklaringer på menneskets nutidige anatomi. Imidlertid er nutidsmennesket også forskellig fra aben eller rettere fra de hominider vi nedstammer fra, men som alle er uddøde og derfor ikke kan studeres direkte, hvortil kommer at de fossile fund fra hine tider er forsvindende få. Det historiske kildemateriale er m.a.o. tyndt. Derfor må man også analysere menneskets nutidige indretning for at kunne begribe eller i det mindste indkredse, hvad man skal efterspore i fortiden hos de fossile hominiderester som nutidsmennesket er udviklet fra. Det skal jo eftersøges, man går ikke og falder over det, hvorfor man med udgangspunkt i det aktuelt eksisterende i en eller anden udstrækning må gøre sig antagelser eller hypoteser om, hvad man skal lede efter i det historisk forgangne. Og i samme udstrækning man gør det, er man stedt i en eller anden slags cirkularitet, der forudsætter det, man leder efter.

Denne dobbeltbevægelse i den slags historisk erkendelse, der sammenligner fortid og nutid, kaldes undertiden forholdet mellem historie og logik eller konstitution: Bagudrettet eller retrogradt fra nutiden gør man sig begrebslige antagelser ("konstitution") om vor arts fortidigt fremadrettede eller progressive tilbliven alt som slægt efter slægt er rullet frem på jord, hvorefter man eftersøger empirisk evidens for det ("historie"). Sådanne retro-progressive scenarier studier af, hvordan vi er nu, retter søgelyset bagud mod de forhold i fortiden, som kan have afstedkommet at vi fremad er blevet til dem, vi er i dag. Man kan ikke spørge til nogets tilblivelse uden allerede at have forudsat visse antagelser om dets nuværende eksistensform, hvilket vil sige at have konstitueret det begrebsligt forud for den empiriske undersøgelse af dets historie. Man kan kun forstå en tilblivende udvikling i det tilbagekastede lys fra udviklingens foreløbige slutresultat, ligesom man omvendt kun kan forstå det foreløbige slutresultat i det fremkastede lys af dets tilblivelse eller udviklingshistorie. Udviklingshistorien og udviklingsresultatet kan kun forstås i lyset af hinanden, fordi udviklingshistorien og udviklingsresultatet dialektisk forudsætter hinanden.

Den cirkel er der ingen vej uden om, kun ind igennem. Bestræbelsen må derfor være at øge cirkelns omfang og lødighed gennem hyppig og systematisk vekslen mellem tilblivelse og resultat samt mellem baglæns begrebsliggørelse (konstitution) og forlæns empiri (historie), hele tiden stedt i fare for blot at finde og bekræfte det man allerede har gjort sig antagelser om.

Ekskurs om emergens og reduktion

Som gennemgået lancerede Aristoteles fire årsagsprincipper. Man kan muligvis tale om et femte, det emergente.

Emergensprincippet hævder at der undertiden fremkommer nye egen-

skaber, som udgør et højere kompleksitetsniveau end tidligere og som ikke eller ikke kun kan forklares ud fra de forudgående niveauer, hvori de har deres oprindelse. Liv består fx af egenskaber som genetisk information i arvemassen, stofskifte og evnen til at reproducere sig selv. Disse egenskaber er opstået fra det livløse eller uorganiske stof, men er andet og mere end dette og lader sig ikke eller ikke kun forklare ved de samme lovmæssigheder, som gælder for det uorganiske stof, og denne opstandelse kan da kaldes emergent.

Tanken har trådt tilbage til Hegels naturfilosofi (1817, paragraf 376, her efter Prigogine & Stenger, 1984, 135), der ordner naturen i et system af historisk tilblivne hierarkiske niveauer af stigende kompleksitet og med kvalitative forskelle mellem niveauerne. Hvert niveau er betinget af det forudgående, som det overskrider og "ophæver", dvs. hæver op i sig, for derpå selv at blive betingelsen for et nyt efterfølgende og højere kompleksitetsniveau. På hvert niveau fremkommer en ny helhed, der forudsætter og er afhængig af de gamle niveaus delelementer, men integrerer disse i en ny logik med nye lovmæssigheder, som delene ikke selv rummer og ikke kan forklare (ibid., 221). Derfor må man have forskellige videnskaber til at undersøge hvert sit niveau. "Mere er anderledes", som denne tankefigur om planhøjere kompleksitetsstigninger lakonisk blev sammenfattet af fysikeren P. W. Anderson (Nørretranders, 1993, 417). De nye planer er anderledes end de gamle og kan ikke reduceres hertil. Bevidstheden kan eksempelvis forstås som en emergent egenskab ved psyken, der både ligger under for de lovmæssigheder ved psyken, som gælder for alt liv udstyret med psyke langt ned i dyrerækken, men desuden tager disse i brug til dannelsen af nye og mere komplekse mentale virksomheder og egenskaber.

Irreducibiliteten skyldes, at skønt det stadig er de gamle lovmæssigheder der ligger til grund for de nye, organiserer de nye lovmæssigheder de gamle på en netop ny måde. Hvor de gamle lovmæssigheder før var enerådende, er de nu blevet til komponenter i og instrumenter for nye. Mere er anderledes, den ny helhed er mere end de gamle elementer, de gamle lovmæssigheder er blevet de nye lovmæssigheders tjenere samtidig med, vel at mærke, at de gamle bibeholder deres oprindelige principper. Sagt på en anden måde med Polanyi (1969) sætter de nye principper eller logikker nogle randbetingelser for de gamle princippers operationsfelt. De nye principper regulerer de gamle princippers funktionsmåder inden for randbetingelsernes rammer og gør herunder brug af de gamle principper til at realisere nye egenskaber og virksomheder.

Som et andet eksempel på emergens på det biologiske og psykiske plan kan nævnes den finale årsagsforklaring, som må være opstået af den kausale sammen med nye virksomheder og egenskaber og derpå overlejrer denne. Det nye planhøjere teleologiske årsagsprincip udsletter ikke kausalitetsprincippet, det lader sig ikke gøre, men finaliteten "ophæver" kausaliteten eller tager den i sin tjeneste til at realisere sig selv med.

Og når det organiske liv bliver udstyret med psyke, indtager de psykiske "organer" pladsen som primær årsagsforklarer over for de biologiske i den forstand, at en psykisk egenskab som se-kunnen er årsag til det biologiske øjes opkomst og indretning. Og psykologien indtager pladsen som riget i livsvidenskabernes midte, uanset at de andre endnu ikke har opdaget det, for når simple organismer i evolutionens løb udstyres med en psyke, er psyke og liv blevet til to sider af samme sag. Psyken er da ikke længer blot en udvækst på det organiske liv, men evolutionens kerne (se-kunnen som årsag til øje), siger McDougall (1933/1960) og Spencer (1881).

Bestræbelsen på at reducere alle fænomener til deres mindste elementer og studere disse i deres mest elementære samvirken er ikke forkert. Det kan kaldes metodisk reduktionisme, og uden denne kommer viden-skaben ingen vegne. Fejlen ved den bestræbelse, der da kunne kaldes ontologisk reduktionisme, består i antagelsen om, at reduktionen af alting til elementer og enkle lovmæssigheder også indebærer muligheden for her ud fra at kunne rekonstruere alle fænomener i universet, som Anderson siger (Nørretranders, 1993, 418). På grund af emergensen følges kendskabet til de fundamentale enheder og deres love imidlertid ikke af evnen til at kunne regne ud, hvordan virkeligheden i alle henseender er indrettet. Man kan ikke beskrive virkeligheden udtømmende på nogen af niveauerne, og man kan ikke konstruere sig frem til de højere niveauer ud fra de lavere. Viden om atomernes egenskaber og lovene for deres bevægelser indebærer eksempelvis ikke muligheden af at redegøre for, hvorfor og hvordan en elefant drikker vand, fordi den uhyre komplekse ansamling af atomer og atomare rørelser, som udgør en vanddrikkende elefant, rummer egenskaber og lovmæssigheder der ikke findes på lavere atomare kompleksitetsniveauer. Når to eller flere elementer kombineres på et højere integrationsniveau, er ikke alle den ny helheds egenskaber forudsigelige resultater af elementerne (Mayr, 1976, 369).

Emergens er selvsagt vigtig, men jeg betragter det som et beskrivelsesprincip og ikke et forklaringsprincip. Som årsagsforklaring må emergente egenskaber som alle egenskaber forklares ved en kombination af kausale og teleologiske principper. Jeg henviser i øvrigt til Køppes værk (1990) om virkelighedens niveauer.

Selektion

Psyken realiserer sit mål via et objekt ("objekt" er samlebetegnelsen for alting). Psyken er derfor rettet mod dette noget andet end sig selv (formentlig med undtagelse af områder i vort følelsesliv, men det er lige meget her).

Intentionalitet

Denne rettethed kaldes ofte intentionalitet. Det gjorde fx Brentano (1874, I, 124-25), da han i gæld til bl.a. Aristoteles forsøgte at fastlægge en skelnen mellem det fysiske og det psykiske. I modsætning til det fysiske har det psykiske ifølge Brentano et indhold, og indholdet består i det som psyken er rettet mod. I synet er noget set, i ønsket er noget ønsket, i tanken er noget tænkt, i forestillingen er noget forestillet osv. Det kun tænkte eller forestillede har en in-eksistens, siger han med et udtryk hentet fra skolastisk filosofi, hvilket vil sige at psykens rettethed kan foregribe et noget før dette foreligger sanseligt.

Intentionaliteten er et stort kapitel i psykologi- og filosofihistorien, så stort at end ikke dets konturer kan skitseres her. Det gør ikke noget, for intentionaliteten som sådan er ikke artiklens emne. Det er intentionalitetens kilder, der skal eftersøges her, intentionalitetens udspring og grundlag.

Selvaktiv foregribelse

Den intentionale foregriben eller anticipation retter organismen i højere grad mod visse dele af omgivelserne snarere end andre, nemlig dem den er blevet til ved gennem sin evolutionshistorie i tusinder og millioner år og som den så at sige har optaget i sig, eller er en genspejling af (Katzenelson, 1989) og som har beredt organismen til både at være disponeret for og kapabel til at opsøge og opfange visse slags information end snarere andre. Når vort øje betragter verden, tror vi umiddelbart at vi opfatter alt, hvad der sker derude, men det er forkert. Vore øjne er designede til kun at percipere information inden for snævre grænser af det elektromagnetiske bølgespektrum, og selv her inden for registrerer vor se-kunnen for det meste kun en beskeden del af den faktisk foreliggende information (Buss, 2009, 47-49), nemlig den vi forlods er indstillet på at opfange.

Selv om organismer også bliver påvirket af tilfældig stimulation, opfatter de derfor nok så ofte det, de allerede er i færd med at se efter eller lytte sig ind på. Det, organismer møder, er hyppigt ikke en tilfældig påvirkning fra et tilfældigt objekt, men resultatet af dels en virksom eller selvaktiv rettethed mod og foregribelse af objektet og dels en parathed til at registrere det, begge dele designet af pågældende arts evolutionshistorie. Det er intentionalitetens kilde.

Intentionalitet er således ikke en egenskab, der er forbeholdt menneskets bevidsthed. Den optræder langt ned i dyrerækken. Hvis fænomenet psyke træder ind på historiens scene samtidig med at nogle organismer bliver selvbevægelige, hvilket der kan argumenteres for (fx Katzenelson, 1989), er intentionaliteten lige så gammel som psyken selv.

Indramning

Denne evolutionært udviklede forhåndsparethed i organismens rettethed mod verden, som forlods lader nogle objekter komme mere til syne end andre og dermed afgrænser, begrænser og styrer perceptionen mere eller mindre, udtrykkes undertiden på den måde, at perceptionen er indrammet eller "rammesat" (*constrained*), og sådanne begrænsninger og styringer gør sig gældende for alle psykiske dispositioner og kapaciteter, ikke kun de visuelle.

Det er dette såvel indrammede som anticiperende og selvfrekaldte forhold til objektet tillige med den faktisk tilgængelige information i det, der primært bestemmer psykens indhold (Neisser, 1976, 20). Informations tilgængelighed er således almindeligvis et selektivt fænomen. Organismer opsøger og opfatter det, de gennem evolutionens selektionspres har haft brug for at opsøge og opfatte, og som derfor fremtræder som netop "brugbar" (Heidegger, 1927) eller med "ydelseskarakter" (Gibson, 1979, *affordance*); "opfordringskarakter" hedder det hos Lewin (1926) og "tjenlighedsbestemmelse" hos Rubin (1949) (uden at jeg lægger disse forfattere evolutionsforklaringer i penne, de står for egen regning). Hos Maturana & Varela (1987) er det en ide om organisme-omverdensforholdet som kendetegnet ved "informativ lukkethed" eller som nævnt rammesatte begrænsning.

Selektion ude fra og ind

Skønt alle medlemmer af en art deler samme genpulje, adskiller de sig også fra hinanden. Bortset fra enæggede tvillinger har to individer af arten ikke arvet identiske egenskaber. Disse variationer i medfødte egenskaber bevirker, at individerne er forskelligt egnede eller tilpassede til livet i artens habitat. Nogle er født mere egnede end andre og har derfor større sandsynlighed for at formere sig med afkom, som deler deres fordelagtige egenskaber. Denne såkaldt naturlige udvælgelse eller selektion af de bedst egnede individer medfører, at deres egenskaber genhyppighed øges i populationen generation efter generation, indtil ændrede forhold i habitatet bevirker, at tilsvarende ændrede egenskaber udvikles og reproduceres efter samme recept – eller at arten uddør.

Det er kort fortalt essensen af evolutionsbiologiens principper for, hvordan og hvorfor den naturlige selektion ude fra gennem lange evolutionsforløb resulterer i, at indre egenskaber enten bortfalder eller sætter sig og bliver genetisk overført til næste generation som et artstypisk lager af medfødte biologiske og biopsykiske dispositioner og kapaciteter, i evolutionspsykologien ofte kaldet "mekanismer".

Derefter, efter at evolutionens naturlige selektion over lange tidsrum har etableret indre egenskaber, selekteres der fortsat i den aktuelle her og nu adfærd ude fra på den måde, at i interaktionen mellem organisme og

omgivelser vælger omgivelserne så at sige mellem de indre egenskaber ved at tillade nogle blandt dem at komme til udtryk snarere end andre eller at give nogle af dem gunstigere udfoldelsesvilkår end andre. Det er ikke kun påvirkninger fra omgivelserne, der bestemmer organismens reaktion, det afgøres også af hvad organismen både kan og må levere i kraft af sin evolutionært tilblevne arkitektur (Fodor, 1980). Påvirkningen ude fra giver anledning til, at organismen billedlig talt gennem søger sit indre lager af biopsykiske mekanismer og stiller de af dem til rådighed for påvirkningen, som er de bedst egnede under de foreliggende omstændigheder (uanset hvor ilde egnede de kan se ud). De aktuelle omgivelser selekterer således i det sjælestof, som er der i forvejen og er blevet til gennem utallige års naturlig selektion, og vækker noget af det snarere end noget andet til live. Den menneskelige organisme rummer et vældigt repertoire af medfødte biopsykiske mekanismer, som ligger afventende parat til at blive taget i anvendelse i omgivelser, der selektivt tænder deres gener og slukker dem igen (Gazzaniga, 1992, 3-6).

Det er "miljøets" ene rolle i arv-miljø diskussionen. Den anden er at tilslibe de indbyggede mekanismer på bestemte måder, at efterprogrammere de medfødte programmer, men den hører ikke hjemme her (se Katzenelson, 1994b).

Inde fra og ud

Efter at evolutionens naturlige selektion ude fra har etableret de indre egenskaber, selekteres der i den aktuelle adfærd også inde fra og ud på den måde, at i interaktionen mellem organisme og omgivelser foregriber og udvælger organismen som anført sine objekter og dermed sit miljø. Således beskriver Mead denne selektionsproces (1934/1967, 6):

"En handling er en impuls, som opretholder livsproccessen gennem en udvælgelse af bestemte former for stimulering [objekter], som livsproccessen har brug for. I denne forstand kan man sige, at organismen skaber sine omgivelser. Stimulus tjener som den anledning, der bringer impulsen til udtryk. Stimuli er midler; handletendensen er det igangsættende udgangspunkt (*tendency is the real thing*). Intelligens må forstås som kapaciteten til at udvælge stimuli, der er i stand til at aktivere livsproccessen, og som tjener til livets opretholdelse og genopbyggelse. Handlingens formål fremtræder ikke nødvendigvis (*need not be 'in view'*), men en (videnskabelig) redegørelse for handlingen må inkludere en beskrivelse af det mål, handlingen bevæger sig hen imod. Dette er en naturalistisk teleologi og forenlig med et mekanisk grundsynspunkt".

En naturalistisk teleologi forenlig med et mekanisk grundsynspunkt. I en vis forstand skaber ("konstruerer") mennesket altså sine omgivelser, som Mead siger, men ikke på grundlag af tilfældige kausale påvirkninger eller stimuli, derimod i den forstand at det er finalt dømt til at skabe de omgivelser, som dets evolutionært tilblevne artskaraktéristiske egen-

skaber er anlagt på at frembringe. Derfor genkender vi den menneskelige socialitet overalt på jord (Katzenelson, 1994; 2004, 239-240), skønt den iklæder sig forskellige kulturelle gevandter.

Denne tankefigur benævnes undertiden argumentet om “den svage stimulus” (*the poverty of the stimulus*). Willert (2004) henleder opmærksomheden på dens slægtskab med ovennævnte ide hos Maturana & Varela om organisme-omverdensforholdet som kendetegnet ved informatorisk lukkethed, såvel som de øvrige forfattere anført samme sted tilføjer vi.

Hvad angår Brentanos distinktion mellem det fysiske og det psykiske og i fortsættelse af et tidligere afsnit kan vi da sige, at det fysiske er virkende bevægelse, det forløber i årsag-virkning kategorier. Årsag A (Stimulus) medfører virkning B (Response), med andre ord den klassisk behavioristiske S – R formel. Senere medgav behaviorismen at formlen næppe burde opfattes som en ren refleksbue direkte fra S til R, men at organismens indre mekanismer medvirkede i den skikkelse R fik som virkning af årsag S, hvilket udvidede formlen til S – O – R, hvor O står for Organisme. Det psykiske er derimod virksom bevægelse, det forløber som et subjekts selvaktive udfoldelse i forhold eller relationelt til et objekt. Objektets påvirkning kommer ikke nødvendigvis før subjektets aktivitet, fordi subjektets selvaktivitet som nævnt kan inddrage objektet, således at subjekt og objekt udgør momenter ved hinanden i en dialektisk relation (Engelsted, 1989 II, 6, 70). Denne selvvirksomhed er det psykisk levendes dynamik. Den uorganiske materies dynamik er passiv virkning af en kausal årsag, medens det med psyke udstyrede organiske stofs dynamik er selvvirksomt.

Det vil sige at det psykisk levende vil sig selv i form af foregribende opsøgen eller fremkalden af visse dele af omgivelserne snarere end andre. Jævnligt rammer omgivelserne ganske vist organismen uforberedt, uopsoget og uanmeldt, især hvis objektet eller begivenheden er ny i biotopen, således at organismens indre egenskaber ikke er artsspecifikt forberedt på det. Men i hovedparten af organismens dagligdag er dens omgivelser delvis beredt af den selv, fordi de er foregribet af organismens selvaktive rækken ud mod dem. Organismen møder oftest, hvad den har opsoget og anticiperet. Psykisk liv er således indrettet, at omgivelserne ikke bare indtræffer, skønt det også sker, men også kaldes frem af de medfødte biopsykiske programmer. Denne selektive selvaktive rettethed mod omgivelserne er adfærdens fundamentale træk langt ned i dyrerækken (Woodworth, 1958, 125).

Således går det til, at ofte foreligger organismens objekt i første omgang ikke håndgribeligt tilstedeværende. Det optræder anticiperet eller har “in-eksistens” med Brentanos ord. Objektet er et ikke-objekt eller et ikke sanseligt objekt, der i og ved organismens selvaktivitet bliver til et sanset objekt. Dvs. at objektet er “repræsenteret” i organismens aktivitet forud for dets håndgribelige eksistens, hvilket kan udtrykkes på den måde at responsen ligger forud for stimulus (Engelsted, 1989 II, 87). Dimen-

sionen fra ikke-objekt til objekt er dermed et centralt træk ved psykens væremåde, hvilket giver formlen $O - R - S - O$: Den virksomme Organisme O affyrer af egen drift en final Respons R , der opsøger en Stimulus S , som derpå virker kausalt ind på Organismen O .

Psykens væsentligste væsen

Den selvirsomme intentionalitet er, hvad psyken primært er til for, dens væsentligste væsen ned til de laveste organismer udstyret med en psyke. Den er til for at få objektet til at foreligge ved at blive foregribet, få verden til at komme organismen i møde på dens egne præmisser og ikke blot dukke op på må og få. Psyken bringer virkelighedens objekter til veje, således som den givne organisme nu engang formår det alt efter sin arts evolutionshistorisk tilblivne motivationelle mål og kognitive såvel som emotionelle objektrealiserende apparatur (Katzenelson, 2004).

Hvad angår de distale eller yderste ting, er denne objektrealiserende virksomhed arts-karakteristisk. Organismer opsøger objekter, der er adækvate for deres finalt styrende mål, hvilket vil sige at de forholder sig på bestemte artsgivne måder til de for deres biopsykiske design egnede objekter i biotopen, herunder og ikke mindst deres artsfæller. Enhver arts virksomhed søger at virkeliggøre en virkelighed, som er artsadækvat for den, og som den evolutionært er designet til at realisere. Med H. Lipps kan man sige, at levende væsener må forstås som deres væren-til deres "element" ud fra deres "habitus" (Pahuus, 1988, 178), hvilket kan oversættes til at organismens virksomhed griber efter bestemte udsnit af verden ud fra dens art.

Selvorganisering

Aristoteles' teleologiske eller finale årsagsforklaring blev angrebet. I 1600-tallet blev den af Galileo m.fl. afskediget til fordel for den kausale, og siden har der stået strid om teleologiens status, for tilsyneladende kan den ikke undværes i livsvidenskaberne. Teleologien fortsatte derfor med at hemsøge naturvidenskabsmændene i deres drømme, som Diderot sagde halvandet hundrede år senere (Prigogine & Stengers, 1984, 121). Det gjorde den bl.a. i form af "vitalistiske" betragtninger (hos fx Driesch og Bergson), hvilket vil sige antagelser om at der er blevet tilført den døde materie en skabende livskraft (*elan vital*), som ikke kan forklares videnskabeligt. Der er imidlertid ingen grund til at gribe til vitalisme, for det teleologiske årsagsprincip har lige så naturlovsmæssig en status som det kausale, hvilket formentlig var hvad Mead mente med "en naturalistisk teleologi forenlig med et mekanisk grundsynspunkt".

Autopoiese

Imod Galileo og den ham efterfølgende naturvidenskab mente Kant (1790, § 66), at levende organismer har en enheds- og helhedsdannende karakter, som ikke kan forklares ved fysikkens og kemiens kausalprincipper. Det fordrer dialektiske teleologiske forklaringer på, at hver komponent i helheden er hinandens forudsætninger, og man må forstå en organismes organer ved de mål eller funktioner, de opfylder set i lyset af den helhed, som de danner tilsammen, idet enhver komponent i helheden tjener en anden komponents mål. Det vil sige at enhver del af en levende organisme frembringer de andre dele, samtidig med at den selv frembringes ved de øvrige dele i deres fælles organiske helhed.

Dette princip kan formuleres på den måde, at levende stof er selvorganiserende systemer eller med Maturana & Varela (1980) besidder "autopoiese", selvskabene. Levende stof organiseres ikke passivt sammensat af dets forud givne enkelte komponenter, men organiserer disse aktivt gennem sin egen selvregulerende helhedsdannende virksomhed. Levende stof opretholder sig selv ved til stadighed at forny sig selv.

Selvorganisering kan ikke eller i hvert fald kun vanskeligt forklares kausalt, fordi kausaliteten jo kun har én retning fra årsag A til virkning B, men ikke tilbage igen fra B til A. Tilstedeværelsen af den omvendte bevægelse fra B til A er imidlertid forudsætningen for, at et system kan være indre selvorganiserende og ikke kun kausalt ydre organiseret. Ifølge kausalitetsprincippet kan som nævnt en virkning ikke være sin egen årsag eller kan ikke bevirke sig selv, hvorimod selvorganisering kan forklares teleologisk ved systemets finale mål. Selvorganisering og selvregulering er et livet iboende princip, der både vil til orde i egen ret og tjener opnåelsen af organismens iboende mål. Organismen anticiperer hvad den har brug for, siger Maturana & Varela (1987, 53), et selvorganiseret system igangsætter spontant virksomhed der er rettet mod omgivelserne og egnet til at udnytte dem til at holde sit system gående ("opretholde livsprocessen" med Mead) alt efter dets fuldt udviklede mønstre (*in its definitive patterns*); alt efter sin art kalder jeg det.

Dissipation

Trods forsigtighed i formuleringerne lægger Prigogine ikke skjul på, at hans teori om såkaldt dissipative strukturer, der er en anden betegnelse for selvorganiserende systemer (og som han hentede en nobelpris hjem på), står den aristoteliske metafysik om en "organiserende immanent intelligens" nærmere end det meste af den gængse fysik (Prigogine & Stengers, 1984, 220).

Erkendelseskategorien

Da den teleologiske årsagsforklaring af mange ikke blev betragtet som videnskabelig kurant, kunne den selvsagt ikke levere sand kundskab. Desværre kan vi ikke forstå levende væsener uden.

Det indså Kant som nævnt, men også han lå under for sin samtids opfattelse af, at kun kausalprincipperne i den dominerende Newtonske fysik var acceptable lovmæssigheder for naturens foreteelser. Derfor konkluderede han, at forståelsen af og forklaringen på livets finale processer må ligge uden for menneskets erkendelsesmuligheder (Mayr, 1976, 402). Det dilemma søgte Kant at slippe ud af på sin sædvanlige måde ved at gøre teleologien og selvorganiseringen til en erkendelsesform. Finaliteten er hos ham epistemisk i den forstand, at den er en af den menneskelige erkendelsens betingelser for at begribe liv. Dermed er den blevet en erkendelseskategori snarere end et ontologisk faktum, noget faktisk eksisterende derude i virkeligheden, skønt han ikke drog denne fulde konklusion muligvis af frygt for at tilskrive lavere organismer samme erkendelsesformåen som mennesker (Pahuus, 1988, 169). Denne frygt er imidlertid ubegrundet. Den Aristoteliske finalitet som her udlagt er ikke en erkendelseskategori, men en antagelse om biopsykiske værensmåder i livets orden, uanset om mennesker eller andre væsener erkender dem eller ej. Hvis man som tidligere nævnt sidestiller eller erstatter finalitet med funktionalitet, ser biologer som Somerhoff eller filosoffer som Nagel da heller ingen problemer i at forene teleologiske forklaringer på liv med kausale (Boden, 1972, 51).

Kants anerkender altså teleologiens eksistens, men fratager den i samme åndedrag ontologisk status. Kants teleologi er derfor subjektiv i tidligere omtalte Hegelsk forstand, ikke objektiv, idet finaliteten ikke er indlagt i naturen selv, men kun har status som en af menneskets erkendelsesbetingelser (Hartnack, 1990, 101).

Styringens styring

Gennem udforskningen af de kemiske mekanismer, der ligger til grund for genomet, mener moderne molekylarbiologi muligvis at kunne nøjes med kausalforklaringer på organismers finalt selvaktive og selvdannende egenskaber. Men uanset om man for argumentets skyld medgiver, at fysisk-kemisk set kan alle biopsykens *enkelte operationer* muligvis forklares kausalt, så er den fysisk-kemiske kausalitet i stoffet ikke des mindre teleologisk i *sin samlede udfoldelse*.

Lad os altså for argumentets skyld sige, at kausaliteten styrer organismen. Men hvad styrer denne styring både forud for og efter, at organismen fungerer kausalt? Det gør de medfødte programmer, som ud fra deres evolutionært udviklede artskaraktéristisk iboende mål og funktionsmåder tillader forekomsten af nogle kausale årsagsforbindelser i organismen,

men ikke andre, og tillader de kausale årsager at få de og de virkninger, men ikke nogle andre. De medfødte mekanismer eller programmer afgrænser og indrammer eller rammesætter som nævnt, hvilke kausale årsag-virkningsforbindelser der er muligt, mindre muligt og umuligt forekommende for medlemmer af denne art.

Virkninger bliver til årsager

Det er blevet pointeret, at både kausalt og finalt betragtet kan en virkning ikke være sin egen årsag. Det må imidlertid præciseres, at det kan den ikke i momentet, i et givet øjeblik, men det kan den over tid. I et evolutionært tidsperspektiv bliver virkninger faktisk til deres egne årsager.

Det sker på den måde, at resultatet af den naturlige selektion blandt de utallige kausale årsags-virknings forekomster i evolutionens forløb sætter sig genetisk som artens egenskabslager af biopsykiske dispositioner og kapaciteter. Dermed bliver kausalårsagernes virkninger (som er processer) over tid til mekanismer (som er strukturer) med iboende mål, hvilket vil sige at de er blevet deres egne teleologiske årsager. Sådan bliver kausalitet til finalitet, virkens virkninger bliver til virksomme årsager.

Teleologiske fejltagelser

Evolutionen som sådan har imidlertid ingen iboende mål, det være sig mennesket, den store hvide mand, verdensånden, universel harmoni etc.

Ganske vist har Prigogine & Stengers (1984) muligvis ret i, at der i alt stof findes en iboende udviklingsretning mod orden forstået som hierarkisk differentieret kompleksitetsstigning, hvilket ikke skal forklares nærmere (se Katzenelson, 1989), men det er meningsløst at kalde det for målrettet frem- eller tilbageskridt. Det er også meningsløst at lægge værdibetragtninger ind i gængse udtryk som en højt udviklet art og mennesket som den højeste. Sådanne betegnelser kan kun henvise til mere eller mindre kompleksitet. Arterne er, hvad de er, og så længe det fungerer, fortsætter de med at være sådan eller ændres mod stadig bedre tilpasning til deres biotop, mere er der ikke at sige om det. De mest primitive arter i betydningen mindst komplekse organismer, bakterierne, som alt andet liv formentlig nedstammer fra, er også de længstlevende, op mod fire milliarder år. Det må kaldes en enestående evolutionær succes, til trods for at de ikke har udviklet sig en disse i dette umådelige tidsrum. Ej heller opstår biologiske og psykiske organer, fordi de er tjenlige til fremtidige mål eller bare fremtidig fordelagtig tilpasning, for det kan lige så godt gå galt. Det har det gjort for 99,9% af alt levende.

Slige opfattelser er teleologiske fejlslutninger og faldgruber. Finale eller målrettede processer optræder hos det enkelte individ af en art, fordi individets mekanismer er programmeret sådan, men artens evolutionære udvikling er udelukkende resultatet af den naturlige selektion

mellem egenskabsvariationer. Teleologisk programstyring er evolutionens resultat, når virkninger er blevet til årsager, men ikke dens begyndelse, proces og forudsætning.

Tilfældighed og nødvendighed

Men derfor kan man ikke omvendt hævde, at evolutionen er tilfældig, skønt den også er det. Den har nødvendigheder, den er et samspil af tilfældigheder og nødvendigheder.

Evolutionen til- og fravælger tilfældigt varierende egenskaber mellem artens individer alt efter egenskaberne tilpasningsdygtighed. Disse tilfældige individuelt varierende egenskaber har kausal oprindelse som genernes rekombination gennem enten kønnet forplantning eller mutation. Men skønt selektionen af variationerne er en kausalt forårsaget tilfældighed i momentet, bliver selektionens resultater som anført til teleologiske nødvendigheder over tid, når virkninger er blevet til deres egne årsager. Over tid, lang tid, fører selektionen af individuelt tilfældigt varierende egenskaber til dannelsen af stabile egenskaber eller mekanismer eller programmer hos arten, og de udgør herefter artens nødvendigheder i form af finalt foregribende styringer som de rammesatte betingelser for, hvilke kausalt tilfældige individuelle variationer der fremtidig får lov eller ikke får lov at optræde inden for programmernes grænser. Derfor ender vi med at måtte forstå evolutionen såvel kausalt tilfældigt som teleologisk nødvendigt.

Finalitet som kontekstforståelse

Med opfattelsen af psyken som sin egen udfoldelses årsag (årsag som indre væsen) leverer Aristoteles et afgørende supplement til Darwin. Darwin forklarer evolutionen som gradvis selektion af tilfældige individuelle variationer. Dermed kan Darwin forstå hvordan evolutionen er kommet i stand, men i mindre grad hvorfor evolutionen eksisterer (Engelsted, 1980).

Man kan spørge kausalt efter livets "hvad er det" og "hvordan er det blevet til". Men man kan med Aristoteles også spørge teleologisk efter livets "hvad er det til for" og "hvorfor er det blevet til" (Mayr, 1976, 360). Om man stiller den ene eller den anden type spørgsmål, afhænger af hvilket slags svar man eftersøger, men begge spørgsmåls- og svartyper er gyldige, og finaliteten eller den målrettede årsag er ligeså "mekanisk" som den kausale årsag (ibid., 365). En naturalistisk teleologi er forenlig med et mekanisk grundsynspunkt, blev Mead citeret for tidligere.

Videre kan man godt stille et teleologisk spørgsmål uden et kausalt, men ikke omvendt, for når som helst man stiller et kausalt spørgsmål, er det teleologiske altid til stede om end oftest indirekte som baggrundsfaktor. Man kan nemlig kun forstå livets kausale "hvad er det" og "hvordan er det blevet til" i en teleologisk kontekst eller horisont. Al specifik erk-

endelse forudsætter en generel erkendelseskontekst, og for levende væseners vedkommende har denne horisont altid at gøre med det funktionelle eller teleologiske eller finale “hvad er det til for” og “hvorfors er det blevet til” (Mayr, 1976, 361). Som Leibniz sagde er *causa finalis* det erkendelsesbegreb, som *causa efficiens* leder op til (Hartnack, 1990, 59). For at nå fuld forståelse i livsvidenskaberne kan ingen årsagsanalyse af en egenskab afsluttes uden at der spørges finalt til, hvorfor det eksisterer og hvad dets funktion er (Mayr, 1976, 398). Som Aristoteles pointerede skal alle årsagsforklaringer i sving for at opnå fuld redegørelsen for tingenes væsen.

Det gælder også selv om det kausale spørgsmåls indlejring i den finale horisont er tavst forudsat eller forvist og tabuiseret. Vi kan først levere en kausal forklaring på forekomsten af en given adfærdsdisposition og kapacitet, efter at vi har identificeret og karakteriseret denne, og det indbefatter en teleologisk omvej. De basale eller yderste (distale) livsfænomener kan vi kun kategorisere i det mindste indledningsvist ved at spørge efter, hvad deres formål er eller hvad de er til for. Indgangsforståelsen til liv er altid final, først derpå kan der i nogle tilfælde stilles kausalt meningsfulde spørgsmål. Ellers er der slet og ret ingen forståelse, fordi der ingen overgribende forståelseskontekst er. Uden forståelse af se-kunnen, ingen forstand på øje.

Optimalitet

I dagligdagens opfattelse af evolutionen spiller begrebet overlevelse ofte en central forklaringsrolle såsom “det er de bedst egnede der overlever”. Det er selvfølgelig rigtigt, men ligesom at øjet skulle være svar på spørgsmålet om, hvorfor vi kan se, er overlevelse kun den halve evolutionshistorie og risikerer endda at køre historien af sporet.

Levelse og overlevelse

Spørger man kausalt til livets og sjælens “hvordan er det blevet til”, lyder svaret altså ofte “individernes overlevelse gennem konkurrence eller Kooperation og artens bevarelse gennem reproduktion”. Spørger man derimod finalt til livets “hvad er det til for”, bliver svaret “individernes optimale levelse eller funktionsmåde alt efter artens iboende mål og mekanismer”, for da undgår man ikke at måtte inddrage betragtninger over egenskaberne artsadækvate og dermed optimale udfoldelse fra øjeblik til øjeblik hos de enkelte individer (Bjerre, 1988, 106).

Optimal levelse er noget andet end overlevelse. Udsagnet “de bedst egnede overlever” som resumé af evolutionsprincippet er korrekt nok, men også en trivielt tautologisk version af evolutionserkendelsens cirkularitet (“hvem overlever?”; “det gør de bedst egnede”; “hvordan kan man vide, at de er de bedst egnede?”; “fordi de overlever”). Trivialiteten fremkommer, fordi man begår den fejl at opfatte overlevelse som belønningen for

bedst mulig tilpasning til habitatet. Det er ikke tilfældet og endnu et eksempel på en evolutionær faldgrube: Evolutionen uddeler ikke belønninger. Overlevelse er ikke *belønningen* for den tilpasning, der er gået for sig hen gennem evolutionshistorien, men *kriteriet* på den (Flew, 1967, 32). Optimal levelse vil derimod sige, at mange og måske de fleste (men næppe alle) egenskaber hos en organisme er bygget eller programdesignet til at bidrage til tilpasningen i den for arten naturlige habitat her og nu på dette givne stadium af evolutionen (Staddon, 1987). En sådan opfattelse af egenskaberens artsadækvate livsudfoldelse er den korteste vej til forståelsen af evolutionens "hvad er det til for" (Etter, 1978; Smith, 1987), idet man anmærker at som tilfældet er med hele den teleologiske programstyring er levelsens artsadækvate udfoldelse og dermed optimalitet evolutionens *resultat*, ikke dens begyndelse, proces og forudsætning (Smith, 1987).

Optimal levelse består i, at individerne her og nu gør det, som de efter deres art er programbyggede til at gøre på den efter programmet mest adækvate måde under de givne omstændigheder (der bortses her fra moralske bedømmelser af adfærden, det hører en anden fortælling til). Det er ikke individernes finale stræben at overleve – bortset fra ekstremssituationer (som der ganske vist kan være mange af til visse tider og steder) – ej heller at bidrage til artens bevarelse gennem reproduktion. Det er individernes finale mål at leve som de er skabt til at gøre der, hvor de nu engang er sat på jord, og herunder at lade deres seksualitet og yngelpleje komme til orde, hvilket for det meste *resulterer* i individuel overlevelse og artens bevarelse. Men at gøre artens bevarelse og (bortset fra ekstremssituationer) individets overlevelse til mål for levelsen fra øjeblik til øjeblik, er at begå den teleologiske fejltagelse at sammenblende egenskaberens slutprodukt med de faktorer, der evolutionært frembragte dem. Det er at lægge abstrakte mål ind i dagligdagen, men dér findes kun konkrete mål i individernes konkrete forhold alt efter hvad de er anlagt på som art, og det er disse motivationelle mål såvel som deres kognitive og emotionelle realiseringsredskaber og ledsageomstændigheder vi skal identificere, beskrive og forklare for at forstå artens individer (Katzenelson, 2004), hvad enten det drejer sig om amøber eller mennesker eller menneskers kulturelle omgang med hinanden (Bjerre, 1968, 100).

Amøber forstår næppe hvad overlevelse betyder. Men de ved hvad levelse er, for vi kan iagttage at de af egen drift opsøger nogle objekter og ignorerer, afviser eller fjerner sig fra andre, at de registrerer visse informationer, men ikke andre, og udfører visse bevægelser snarere end andre.

Domænespecificitet

Det er altså ikke forkert at karakterisere livets væsen som overlevelse eller lidt mere præcist som artsbevarende koordination af delfunktioner, men

dels risikerer begreberne overlevelse og artsbevarelse som just gennemgået at sammenblande proces og resultat, endda på en triviel måde, og de dels er for generelle og flertydige. Komplekse levende væsener rummer mangfoldige mål, der udmøntes i en lang række konkrete eller domænespecifikke funktioner, medens overlevelse og artsbevarelse er domænegenerelle antagelser af mindre brugbarhed.

Begrebet domænespecificitet kan ikke udredes nøjere her (se Katzenelson, 2007), blot følgende bemærkninger. Antagelsen om domænespecificitet imødegår den ellers udbredte opfattelse, at evolutionens naturlige udvælgelse kun har aflejret få generelle medfødte egenskaber som fx evnen til at lære og tilegne sig kulturens komponenter. Tværtimod mener teorien at evolutionen har frembragt en mosaik af specifikke dispositioner og kapaciteter, der er opstået hver for sig i løbet af fylogenesen, og som hver for sig er egnet til at løse bestemte specifikke livsopgaver og tilpasningsproblemer på hver deres måde "efter Princippet af en nødvendig trinvis Erhvervelse af hver sjælelig Evne og Færdighed" som Darwin sagde (1859/1981, 526). Under de skiftende habitaters forskelligartede selektionspres er egenskaberne udviklet tidsforskudt stykke for stykke over lange evolutionsstræk og mere eller mindre uafhængigt af hinanden til imødekomme af hver sin nytilkomne livsopgave, som de er specialiserede til at problembearbejde. Der findes ingen generelle problemløser, for der findes intet generelt adaptivt problem (Symons, 1992). Der er tusind måder at overleve og artsbevare på, hvortil tusind egenskaber bidrager hver især og med hver deres delmål.

I løbet af evolutionen erstatter nytilkomne egenskaber endvidere ikke de ældre, men bygges oven på og ind i dem efter princippet om en nødvendig trinvis erhvervelse osv. De ældre egenskaber påvirkes af de nye, men bevarer større eller mindre dele af deres oprindelige væsen som komponent i en nu mere omfattende og kompleks helhed (Maclean, 1990). I udviklingens løb skifter egenskaberne snarere orden, omstruktureres igen og igen og syes atter og atter sjusket sammen til det kludetæppe, som det endelige resultat fremviser. Egenskaberne tandhjul går ikke altid i hak, de kan være i konflikt med hinanden og blandes på mangfoldige måder. Det menneskelige sjæleliv er en konföderation, en broget "forsamling" (Minsky, 1985, 17), en hob af dispositioner og kapaciteter der både koopererer og konkurrerer.

Perfektion

Fordi psyken er et sådant kludetæppe af trinvist erhvervede egenskaber, må optimalitet ikke forveksles med perfektion. Tilpasningens omkostninger umuliggør at sjælens arkitektur kan udgøre et perfekt design.

Fx er den frygt for slanger, som snerter de fleste af os, ikke så stærk som den kunne have været. Hvis frygten var stærkere, havde den været så dominerende, at vore forfædre aldrig havde turdet bevæge sig uden

for hjemmebasen. Det ville reducere risikoen for slangebid betragteligt og udgøre en perfekt tilpasning til slangemiljøet, men på bekostning af en tilsvarende betragtelig reduktion af al anden virksomhed end slangeundgåelse, hvilket ville være en u hensigtsmæssig og sandsynligvis dødbringende tilpasning til hele habitatet, der forlanger tusind andre adfærds kapaciteter end slangeundgåelse. Dispositionen til slangeundgåelse samstemmer således ikke nødvendigvis eller kan være i direkte konflikt med andre dispositioner og kan derfor kun gøre sig gældende til en vis grad. Det er på ingen måde en perfekt mekanisme, men den er adækvat eller optimal givet de omkostninger, der ville have været prisen for en perfekt funktion. Selektionspresset fremmer mekanismer, hvis fordele overgår deres ulemper relativt til andre mekanismer, men ikke absolut. Mekanismerne udgør blot det bedst mulige resultat (Buss, 1999, 21; Dawkins 1982, 46). Evolutionen kender mestendels kun til kompromiser.

Det bliver ikke bedre

Så Sokrates havde ret i at tingene er, som de er, fordi det er bedst sådan. De yderste ting, stadigvæk. Det er selvorganiseringens livline at det adækvate gøres på det rigtige sted til den rigtige tid, og i så fald er det nødvendigvis også optimalt. Evolutionært tilblivne egenskaber kan ikke være anderledes end de er, hvilket definatorisk er det bedste, uanset at det ikke er perfekt, og uanset at talrige omstændigheder kan lægge sig hindrende i vejen for deres optimale udfoldelse.

Ekskurs om identiteten

Det kan ikke overraske, at et sjusket syet kludetæppe har problemer med at hænge sammen. Det opfører sig snart sådan og snart på anden vis, der dementerer det første, det mener både det ene og det modsatte samtidig, som ville forbløffe blot en beskeden smule logik, og er skæv i identiteten. Det forholder sig sådan med psyken at dens ene egenskab oftere, end vi holder af at tro, ikke har meget med den anden at gøre. Stalligere sagt med Sigrid Unset i sin Nobeltale 1944 er mennesket stort nok til, at både Gud og Djævelen kan kæmpe om dets sjæl.

“Han fandt det mærkeligt, at så dybe følelser kunne eksistere side om side med så udbredt en fantasiløshed”. Et tilfældigt hukommelsescitat fra en ældre amerikansk roman (forfatter og titel glemt). Det er imidlertid ikke mærkeligt, det er tværtimod det almindelige at der kører mentaltale parallelspor uden påfaldende indbyrdes sammenhæng. Bødlen kan sagtens være en kærlig familiefar, både Gud og Djævelen har fat. Nogle af psykologiens pionerer som James, Binet, McDougall og sådan set også Freud samt psykiatere som Charcot og Janet var da også af den opfattelse at det, der i datidens sprog gik under navnet multiple personligheder og i vore dages diagnosesystem benævnes dissociativ identitetsforstyrrelse

(*dissociative identity disorder*), kunne være et egnet afsæt for en model af sindet. I nutidig psykologi kaldes det legionteori (Manning & Manning, 2007) af det latinske *legio* = en skare; i dansk dagligsprog betyder det utallige.

Vil det sige at menneskesjælen er kaos i idel splittelse, at identiteten forlods er ruineret af vor evolutionære tilblivelseshistorie og i moderniteten ovenud fragmenteret, atomiseret, opløst i almindelig konfusion, er subjektet død osv. som der ofte nok lamenteres i dens litteratur og filosofi afskrivende al substans og varighed?

Selvsagt skal mennesket i dag navigere mellem flere vilkår og bestride mere forvirrende opgaver end den feudale bonde eller stenalderjægeren, men at psyken skulle være et klinisk tilfælde strider mod al intuition og kan ikke passe, så kunne vi ikke kende forskel på op og ned. Det ville være en så eklatant evolutionær fiasko, at arten måtte uddø.

Det gør den måske også. Men det kommer ikke denne fortælling ved, den lader sig blot inspirere af planteriget. Tænk på psyken som en drueklase: Her har vi egenskab nr. 1 (fx se-kunnen), det er dén drue, og der er drue nr.2 og 3 osv., hundrede, tusinder. Druerne triller ikke hen ad bordet i hver sin tilfældige retning, de hænger sammen i en klase. Hver drue er både sig selv adskilt fra de andre og forbundet, integreret, med dem i en klase.

Det ville sprænge rammerne for en ekskurs at redegøre for, hvordan det nøjere forholder sig hermed, men det kan man læse om hos de ovenstående og den tidligere nævnte Minsky.

Levelse og mættelse

Det hævdes ofte at reklamebranchen skaber behov eller at et liberalistisk økonomisk system til stadighed formerer nye behov. Forudsætningen herfor antages at være, at vore behov er grænseløse og umættelige. Det er en udbredt opfattelse i sociologi, samfundsteori og filosofisk antropologi at mennesket er at forstå som et sådant kronisk mangelvæsen, og samfundets institutioner er til for at inddæmme behovenes uendelige ekspansion og opøve vor selvadministration af deres uopfyldelige kar.

Hvis der i disse antagelser og talemåder med behov menes naturlige egenskaber med naturlige mål som her forstået, passer det ikke. Hverken reklame eller andre kulturelle foreteelser kan skabe natur, den er forlods skabt. Kultur kan skabe ønsker, og de er muligvis umættelige, men vore naturale programmer (her illustreret med behovene) er ikke. Behov hører til de fjerne distale eller yderste ting, ønsker til de nære proksimale. Behov og ønsker er således ikke det samme. Vi kan behøve noget, vi ikke ønsker (kærlighed, en opsang, vitaminer), og vi kan ønske noget, vi ikke behøver (magt, smiger, en lyserød cadillac), og behovene er ikke umættelige. De er mættelige, ligesom andre dyrs er det. Deres væsen er at leve sig selv ud efter deres art i fortrøstning eller uro, og under passende vilkår gør de

det som naturlig selvreguleret levelse. Sådan er det med natur, den ikke blot passer sig selv, men passer på sig selv, hvis den får lov til det under passende vilkår. Mennesker vil ikke have det godt, siger Karen Blixen, de vil gøre det gode. Og onde, tilføjer vi, gud og djævel.

Naturlig mættelse betragter vi som en selvfølge for vore legemsorganernes vedkommende og foruroliges, hvis det ikke er tilfældet. Hvorfor skulle det ikke gælde for psykeorganerne, der efter denne forfatters opfattelse udgør sjælens massive fundament? I så fald er den selvregulerende mættelse målets målestok, appelinstanten for vor artstypiske normalfunktion, og den er som nævnt god nok, definatorisk, bliver i hvert fald ikke bedre. Menneskets sind er ikke et umætteligt begær (hvilket den empiriske lykkeforskning i øvrigt synes at bekræfte). Det umættelige begær har modernitetens litteratur og filosofi opfundet ligesom identitetssløsheden uden at bemærke, at det er uvirkeligt for mennesket som det artsvesen, der er ældre end en kort epoke i en enkelt enklave på jord, men som man her ikke har haft øje for blændet af et par hundrede års dyrkelse af forestillingen om det instinktfri over naturen hævede endeløst formbare plasticmenneske (lige udbredt i liberalistiske og socialistiske ideologier).

Alarmsignal og appelinstant

Som enhver anden art er mennesket tværtimod "vidunderlig hensigtsmæssigt indrettet" (Bjerre, 1968, 153), dvs. optimalt om end ikke perfekt. Hertil bemærker Villy Sørensen i forordet til Bjerres bog, at ansigt til ansigt med de herskende tilstande virker sådan en påstand let afsindig. Men han tilføjer, at man jo kunne vende det om og sige, at ansigt til ansigt med en sådan anskuelse virker de herskende tilstande let afsindige. Påstanden står i modstrid med de herskende anskuelser, men bekræftes af følelsen af, at det ikke er normalt at have det rædselsfuldt, uanset at man måske normalt har det rædselsfuldt, turnerer han med vanlig spids pen.

Menneskets problem er, at det siden civilisationens indtræden for godt 10.000 år siden resulterende i de mange fortrinligheder, vi nødvendig ville være foruden, har skabt kulturer der er drastisk anderledes end dem, vi blev til i og ved, og derfor ikke nødvendigvis stemmer godt overens med vore evolutionært opståede egenskaber eller ligefrem påfører dem lidelse. Menneskene er gode nok, som også Prædikereren siger, de har bare så meget for, og dét. Desårsag kan de givne omstændigheder være lidet fremmede vilkår for optimaliteten. Den indsigt fastholdt Freud (1930) standhaftigt under navn af kulturens byrde i dansk oversættelse; den tyske titel *Das Unbehagen in der Kultur* er nok så prægnant.

Dette ubehag for den artsadækvate levelse og mættelse er alarmsignalet for fejlfunktion (Bjerre, 1968, 105) og det eneste holdbare afsæt for kulturkritik (også forstået på tysk som "en nøje undersøgelse af", ikke som brok), såvel som for terapi og anden helsegering. Alarmsignalet

fortæller os, at noget burde være anderledes. Gennem registreringen af en mangel eller overdosis, af noget skævvredet eller forkrøblet skimter vi omridset af dette nogets normale og optimale skikkelse. Det gør vi, fordi vi som en selvfølge betragter målenes genkendelige form som det sædvanlige i menneskelivet, normen for at være menneske, appelinstanten for vor artstypiske normalfunktion, ja som selve humaniteten (Katzenelson, 2004, 239-242). Målenes genkendelige form er målestok for menneskets væsen forstået "som 'art', som en indre, stum almenhed, der forener de mange individer med *naturlige* bånd" som Marx & Engels (1846/1974, 11) fortrydeligt tilskriver Feuerbach i deres såkaldt sjette tese vendt mod samme, hvor de plæderer for at menneskets væsen tværtimod er "ensemblet af de samfundsmæssige relationer" (ikke "indbegrebet af de sociale relationer" som den danske oversættelse lyder). Ærgerligt at de mener det modsatte af arten som den indre stumme almenhed, der forener individerne med naturlige bånd, for meget bedre kan dette korrekte standpunkt ellers næppe formuleres.

Bemærk de ironiske citationstegn omkring ordet "art". Hvor dum kan man være, siger de, artsvesen, *naturlige* bånd, *my foot!* Naturens ekskommunikation i frygt eller foragt, den indre som den ydre natur, er fortforende et tungt arvelod i Vestens åndsliv.

Den negative metode

Optimalfunktionen kan lide. Den kan være læderet fra begyndelsen af eller gå i stykker hen gennem livshistorien, men det kalder vi da også ødelagthed, særhed, uforløsthed, hæmmethed, fortvivelse, afvig, sygdom m.m. Målenes manglende adækvate indløsning henpeger på funktionsmådens eller de kulturelle vilkårs eller begge deles fraværende normalitet og optimalitet. Normalfunktionen er endda forudsætningen for vor forståelse af programmernes mål og facon, fordi vi kun begriber noget, hvad som helst, ved eksplicit eller implicit at henvise til og sammenligne det med en standard eller appelinstant for dets formodede normalitet, prototypiskhed, kategorialitet eller sædvanlige udgave. Vi forstår foreteelser ud fra deres standardfunktion og ikke ud fra deres fejlfunktion, selv om undersøgelse af fejlfunktionen hjælper til at bestemme standardfunktionen, ja kan være afgørende herfor.

Det kan kaldes den negative metode. Den søger at bestemme det normale og optimale ved skridt for skridt at identificere det unormale og mindre mønstergyldige. På den måde når man aldrig frem til en absolut bestemmelse af det normale og optimale, heldigvis, for det rummer farer for stempling af afvig eller ligefrem eugenik. Den negative metode er mere beskeden, den indkredser standardfunktioner gennem diagnose af fejlfunktioner. Sådan går kroppens doktorer frem, hvis de ellers skænker det en tanke, og det gør sjælens og kulturens også. Det lader sig lige så lidt gøre at udføre kulturkritik som helsegering uden en eksplicit eller

implicit antagelse om standardfunktioner opfattet som naturlige optimale mål og funktionsmåder såvel som naturlige omstændigheder og vilkår herfor (Maslow, 1954, 125).

Litteratur

Aristoteles. *Fysikken*.

Bjerre, P. (1968) *Menneskets natur: Erkendelsespsykologi*. København.

Bjerre, P. (1988) *Opbrud*. København.

Boden, M. A. (1972) *Purposive explanation in psychology*. Sussex.

Brentano, F. (1874) *Psychologie vom empirischen Standpunkte* 1-2. Leipzig.

Buss, D. M. (1999) *Evolutionary psychology: The new science of mind*. Boston.

Darwin, G. (1859/1981) *Arternes oprindelse*, 5. udg. København.

Dawkins, R. (1976) *The selfish gene*. New York.

Engelsted, N. (1980) Psykogenesen. *Psyke & Logos*, 1, 91-112.

Engelsted, N. (1989) *Personlighedens almene grundlag* 1-2. Århus.

Etter, M. A. (1978) Sahlin's and sociobiology. *American Ethnologist*, 5, 160-169.

Flew, A. G. N. (1967) *Evolutionary ethics*. New York.

Freud, S. (1930) *Das Unbehagen in der Kultur*. G. W. XIV, 421-506. London.

Gazzaniga, M. S. (1992) *Nature's mind: The biological roots of thinking, emotions, sexuality, language, and intelligence*. London.

Gibson, J. J. (1979) *The ecological approach to visual perception*. Boston.

Hartnack, J. (1990) *Hegels logik*. København.

Hegel, G. W. F. (1817) *Enzyklopädie der philosophischen Wissenschaften im Grundrisse*. Heidelberg.

- Heidegger, M. (1927/1977) *Sein und Zeit*. Tübingen.
- Johansen, K. F. (1988) *Platons filosofi: Udvikling og enhed*, 2. udg. København.
- Kant, I. (1790) *Kritik der Urteilkraft*. Berlin.
- Karpatschof, B. (2006) *Udforskning i psykologien: De kvantitative metoder*. København.
- Katzenelson, B. (1989) *Psykens verden, i verden: Et naturevangelium*. Århus.
- Katzenelson, B. (1994a) *Homo socius: Grundlaget for menneskeligt samkvem*. København.
- Katzenelson, B. (1994b) Mennesketilblivelsens koreografi og kundskabens grundlag. I: A. Neumann (red.) *Det særligt menneskelige: Udviklingspsykologiske billeder*, 59-105. København. Også i: B. Katzenelson (2000) *Psykologiske indblik*, 235-268. København.
- Katzenelson, B. (1996) Handling og adfærd. *Bulletin fra Forum for Antropologisk Psykologi*, 1, 2-27, 70-84. Også i: B. Katzenelson (2000) *Psykologiske indblik*, 73-98. København.
- Katzenelson, B. (2004) *Drivkræfter, følelser og erkendelse*. København.
- Katzenelson, B. (2007) Evolutionspsykologi. I: B. Karpatschof & B. Katzenelson (red.) *Klassisk og moderne psykologisk teori*, 301-323. København.
- Køppe, S. (1990) *Videnskabens niveauer*. København.
- Lewin, K. (1926) Vorsatz, Wille und Bedürfniss. *Psychologische Forschung*, 7, 294-385.
- MacLean, P. D. (1990) *The triune brain in evolution: Role in paleocerebral functions*. New York.
- Manning, M. L. & Manning, R. L. (2007) Legion theory. *Theory and Psychology*, 17, 839-862.
- Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco.
- Marx, K. & Engels, F. (1846/1974) *Den tyske ideologi*. København.

- Maslow, A. H. (1954) *Motivation and personality*. New York.
- Maturana, H. R. & Varela, F. J. (1980) *Autopoiesis and cognition: The realization of the living*. Boston.
- Maturana, H. R. & Varela, F. J. (1987) *The tree of knowledge: The biological roots of human Understanding*. Boston/London.
- Mayr, E. (1976) *Evolution and the diversity of life: Selected essays*. Cambridge MA.
- McDougall, W. (1933/1960) *An introduction to social psychology*, 33rd ed. London.
- Mead, G. H. (1934/1967) *Mind, self, and society*. Chicago.
- Minsky, M. (1985) *The society of mind*. New York.
- Neisser, P. (1976) *Cognition and reality: Principles and implications of cognitive psychology*. New York.
- Nørretranders, T. (1993) *Mærk verden: En beretning om bevidsthed*. København.
- Pahuus, M. (1988) *Naturen og den menneskelige natur: En filosofisk og litterær analyse af mennesket i naturen og naturen i mennesket*. Århus.
- Polanyi, M. (1969) *Knowing and being*. London.
- Prigogine, I. & Stenger, I. (1984/1985) *Den ny pagt mellem mennesket og universet: Nye veje i naturvidenskaberne*. København.
- Rubin, E. (1949) *Experimenta psychologica*. København.
- Smith, E. A. (1987) Optimization theory in anthropology: Applications and critiques. In: J. Dupré (ed.) *The latest of the best: Essays on evolution and optimality*, 201-249. Cambridge MA.
- Spencer, H. (1881) *The principles of psychology* 1-2, 3rd ed. London.
- Staddon, J. E. R. (1987) Optimality theory and behavior. In: J. Dupré (ed.) *The latest of the best: Essays on evolution and optimality*, 179-198. Cambridge MA.

Symons, D. (1987) On the use and misuse of Darwinism in the study of human behavior. In: J. Barkow, L. Cosmides & J. Tooby (eds.) *The adapted mind*. New York.

Willert, S. (2004) Brobygning til George Herbert Mead. *Bulletin fra Forum for Antropologisk Psykologi*, 14, 83-88.

Woodworth, R. S. (1958) *Dynamics of psychology*. New York.

Wright, G. H. v. (1971) *Explanation and understanding*. London.

Østerberg, D. (1966) *Forståelsesformer*. Oslo.

Det psykofysiske problem – supervenienens og emergens

Af *Simo Køppe*

Det psykofysiske problem er mildest talt et af de filosofihistoriske klassikere som der er produceret biblioteker fulde af tekster om. Dette skal selvsagt ikke afholde en fra at diskutere relationen. Spørgsmålet om psykens emergens – altså dens opståen og tidlige dannelse – er tæt knyttet til en forsøgsvis afklaring eller blot stillingtagen til det psykofysiske problem. Et begreb som superveniens er velegnet til at føre denne diskussion idet begrebet forener det psykofysiske problem og psykens emergens.

Supervenienens

Begrebet superveniens har det i de sidste 20-25 år være relativt uomgængeligt i forbindelse med forskellige tilgange og diskussioner af det psykofysiske problem og spørgsmålet om psykisk kausalitet. Superveniens er således først og fremmest en filosofisk teori om kausalitet, herunder kausalitet mellem forskellige niveauer. Det er imidlertid også sådan, at superveniensbegrebet meget ofte diskuterer den psyko-fysiske relation, både indirekte og direkte. Således bruger en af de vigtigste superveniensteoretikere J. Kim meget ofte den psykofysiske relation som diskussionseksempel i sine superveniensartikler. Superveniensteorien er således også en teori om psykens relation til nervesystemet.

Henimod superveniensteorien

Som overgang til en relativt tæt gennemgang af J. Kims teori om superveniens skal der kort nævnes nogle kernebegreber som fra slutningen af 1970erne og frem bliver centrale i diskussionerne af reduktion og emergens. I de fleste tilfælde er der tale om klassiske definitioner som vækkes til live igen. Termerne bruges i andre sammenhænge end de her behandlede, f.eks. diskussioner af logik, filosofisk semantik m.m.

| | | | | |
|-----------------------------|--------------------------|--------------------------|--------------|-----------------|
| Type egenskaber 1, 2, 3 ... | | | Intensionel | Naturlig klasse |
| Token 1 egns. 1,2,3,4 | Token 2 egns. 1,2,3,5 | Token 3 egns. 1,2,3,6 | Ekstensionel | |

Umiddelbart er type/token distinktionen meget enkel. En række forekomster i verden (tokens) er alle realiseringer af en bestemt type. Distinktionen dækker således også almen begrebsdannelse, hvor der ud fra et vist antal eksempler generaliseres et begreb, en kategori som dækker forekomsten. Den ældre skelnen mellem intensionel og ekstensionel begrebsdefinition er en parallel skelnen, som ikke er identisk, men alligevel indeholder samme element af abstraktion som type/token. Det er endvidere almindelig antaget, at både type og token kan defineres i kraft af et sæt af egenskaber og at typeegenskaberne skal være til stede ved en tilhørende token.

Når diskussionerne af type/token m.m. er relevant i denne sammenhæng, så er det fordi en stor del af den amerikanske diskussion af reduktionisme - specielt reduktion af det psykiske til det fysiske (neurofysiologiske) udfoldes som en diskussion af forholdet mellem type og token. Det er også sådan, at nogle af de kendteste positioner inden for denne debat, f.eks. funktionalisme og fysikalisme relativt enkelt kan defineres ud fra type/token distinktionen.

I forskellige sammenhænge har den amerikanske filosof Fodor diskuteret reduktionismeproblemet og har i den sammenhæng anvendt type/token diskussionen. Fodor er formentlig bedst kendt for sin udvikling af den funktionalistiske teori om forholdet mellem nervesystem og psyke og for sin sprogteori. Den sidste vil ikke blive diskuteret her. I en af Fodors kendteste artikler om reduktionisme argumenteres der også imod enhedsvidenskaben og for de specifikke videnskabers egenart.

En svag form for reduktionisme er hvad Fodor kalder en token fysikalisme. Token fysikalisme indebærer, at alle begivenheder som videnskaberne beskæftiger sig med i sidste ende er fysiske. Token fysikalisme er svagere end type fysikalisme, som hævder at alle egenskaber der beskrives i videnskab som sådan er fysiske egenskaber. Token fysikalisme er svagere end den gængse reduktionisme. fordi reduktionisme ud over token fysikalisme hævder, at der til enhver naturlig klasse i alle videnskaber svarer en naturlig klasse i fysikken. Fodors første argument mod reduktionismen i denne form er, at det er ganske usandsynligt, at enhver naturlig klasse i andre videnskaber skulle svare til en naturlig klasse inden for fysikken idet de naturlige klasser inden for andre videnskaber ofte er formuleret via generaliseringer af fænomener som intet fysisk har til fælles.

Der er især to yderligere argumenter som er relevant i denne sammenhæng. For det første diskussionen af Nagels brolove og for det andet hans

væsentligste argument for af specialvidenskaberne har en ontologisk funderet begrundelse. Fodors andet argument er en kritik af Nagels brolove. Ifølge disse (Nagel, 1961) gælder at for alle A som findes i den sekundære (den der skal reduces), men ikke i den primære (den der skal reduceres til) findes et B i den primære, således, at når A eksisterer så eksisterer B også. Så selv om A ikke eksisterer i den primære og B ikke eksisterer i den sekundære, så vil A ikke forekomme uden B. Hvis dette er opnået, hævder udledningsbetingelsen at det nu er logisk muligt, at udlede den sekundære teori fra den primære, samt alle brolove.

Man må gå ud fra at de fysiske love er betingelsesløse eller rettere undtagelsesfrie. Man må ligeledes antage, at jo mere kompleks videnskabens genstandsområde bliver, des flere undtagelser vil der gælde. Hvad der i psykologien måtte fremtræde som lovmæssigheder er enten klassisk behavioristiske (men den kan ikke dække hele genstandsområdet) eller er kun tilnærmede lovmæssigheder i den udstrækning, der altid er undtagelser. Brolovene skal altså forbinde (og hermed reducere) lovmæssigheder uden undtagelser (fysikkens) og lovmæssigheder med undtagelser, og det kan kun lade sig gøre, hvis brolovene også har undtagelser. Men hvis brolovene har undtagelser, så er reduktionismen ikke længere sikret.

Det tredje argument er knyttet til de egenskaber der definerer de fænomener der undersøges i specialvidenskaberne. En måde at formulere reduktionisme på er: entiteter der tilhører forskellige fysiske klasser kan ikke have ens beskrivelser, og hvis to elementer beskrives forskelligt fysisk, så må de beskrives forskelligt i enhver lov overhovedet.:

But why should we believe that this is so? Any pair of entities, however different their physical structure, must nevertheless converge in indefinitely many of their properties. Why should there not be, among those convergent properties, some whose lawful interrelations support the generalizations of the special sciences? Why in short, should not the kind predicates of the special science *cross-classify* the physical natural kinds? (Fodor, 1974, s. 513)

Hvis der foregår en krydsklassificering så kan kravet om at reduktion opnås ved at specialvidenskabernes natural kinds svarer en-til-en til fysiske natural kinds ikke længere opretholdes.

J. Kim og superveniens

Iflg. Kim kan man ret præcist tidsfæstne indførelsen af de ontologiske diskussioner i forbindelse med det psyko-fysiske problem inden for den analytiske filosofi. Det drejer sig om to artikler af henholdsvis H. Feigl (1958) og J. Smart (1959). Det nye som iflg. Kim sker her er indførelsen af identitetsteorien (type fysikalisme, hjernetilstands teorien) - eller reduktiv materialisme. Nu er det ikke fordi teorien egentlig er særlig udbredt - den fik allerede i slutningen af 1960erne og begyndelsen af 1970erne al-

ternativer som var mere udbredte, men pointen er, at selv om de færreste er materialister som Feigl og Smart - og hævder, at psykiske tilstande er identisk med hjernetilstande - så er det ikke sådan, at man som reaktion genså dualismer af cartesiansk oprindelse. Langt de fleste inden for den analytiske tradition fastholdt fysikalismen, men der blev hurtigt udviklet andre former - funktionalismen af Putnam i 1968 og den "uregelmæssige" monisme af Davidson i 1970. Putnams funktionalisme hævder at psykiske fænomener er af en anden og abstrakt karakter end de fysiologiske processer og at en given psykisk tilstand kan realiseres af forskellige materielle. Det er altså ikke muligt, at identificere en psykisk tilstand med én og kun én bestemt fysiologisk. Davidson mener heller ikke at de psykiske fænomener kan reduceres til de fysiske, men yderligere, at der ikke kan formuleres lovmæssigheder mellem de fysiske og de psykiske - og at de psykiske fænomener ikke engang i sig selv følger strikte love. Davidson formulerer iflg. Kim nærmest en pasant principperne for superveniensteorien som i løbet af 1980'erne bliver den dominerende teori på feltet. Superveniens formuleres hos Davidson sådan, at der kan ikke eksistere to psykiske hændelser som er fysisk identiske, men adskiller sig på det psykiske plan og der kan ikke indtræde en ændring på det psykiske plan uden at der også sker en ændring på det fysiske. Kim afslutter den meget korte historiske gennemgang med formuleringen af tre grundantagelser, som dels udspringer af bruddet med den fysikalistiske identitetsteori (og grundlægger en ikke-reduktiv fysikalisme eller ikke reduktiv materialisme), dels har været dominerende de sidste 20 år af 1900'tallet:

1. Superveniensteorien - det psykiske supervenierer på det fysiske.
2. Antagelsen af, at den psykiske entitet realiseres eller implementeres af det fysiske
3. Antagelsen af, at det psykiske emergerer fra det fysiske

I sin mest almene og almindelig definition betyder superveniens følgende:

Mental properties supervene on physical properties, in that necessarily, for any mental property M, if anything has M at time t, there exists a physical base (or subvenient) property P such that it has P at t, and necessarily anything that has P at a time has M at that time (Kim, 1998, s. 9).

Psykiske egenskaber superveniere på fysiske egenskaber på en sådan måde at der til enhver psykiske egenskab altid eksisterer en fysisk egenskab og således at hver gang en tilsvarende og identisk fysisk egenskab findes, så eksisterer den psykiske egenskab. Teorien tillader, at der findes forskellige fysiske egenskaber som kan realisere samme psykiske egenskab, men ikke omvendt.

Her anvendes det psykofysiske problem som udgangspunkt, men superveniensteorien kan også generaliseres til en anvendelse inden for et hierarki af ontologiske distinkte niveauer, således at der er tale om egenskaber på et niveau i forhold til egenskaber på de øvrige niveauer:

For any x and y belonging to level L (other than the lowest level) if x and y are indiscernible in relation to properties at all levels lower than L (or, as we may say, x and y are *microindiscernible*) then x and y are indiscernible with respect to all properties at level L (Kim, 1998, s. 17).

Hvis alle egenskaber på niveauerne under L er identiske med egenskaber på L , så er de uskelnelige m.h.t. alle egenskaber på L . Igen siger udsagnet ikke, at x og y kun kan realiseres af én tilstand - men blot, at hvis egenskaberne på niveauerne under L er identiske, så er fænomenerne på L det også. Det udelukker altså ikke, at de kan være forskellige samtidig med at x og y stadig er uskelnelige. Dette kaldes *mereologisk superveniens*. Den kan principielt bruges på alle planer i den gængse lagmodel af verden - elementarpartikler, atomer, molekyler . . . samfund. Den mereologiske superveniens beskriver det forhold, at hvis to entiteter på et givet niveau ikke kan skelnes fra alle underliggende niveaues entitetsegenskaber, så er de to entiteter identiske. Denne formulering udsiger ikke engang om det er muligt at reducere fra et niveau til et andet - den beskriver blot en relation mellem to sæt af egenskaber på to forskellige niveauer.

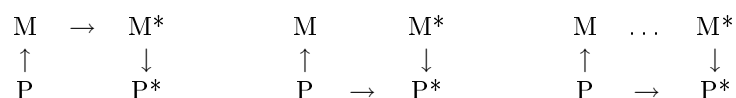
I sine mest defensive øjeblikke definerer Kim superveniens som sådan, som en beskrivelse af to mængder af egenskaber som varierer sammen:

“We must conclude then that mind-body supervenience itself is not an *explanatory theory*; it merely states a pattern of property covariation between the mental and the physical and points to the existence of a dependence relation between the two. Yet the supervenience is silent on the nature of the dependence relation that might explain why the mental supervenes on the physical. Another way of putting the point would be this: supervenience is not a *type* of dependence relation - it is not a relation that can be placed alongside causal dependence, reductive dependence, mereological dependence, . . . Rather, any of these dependence relations can generate the required covariation of properties and thereby qualify as a supervenience relation. Supervenience therefore is not a metaphysically ‘deep’ relation; it is only a ‘phenomenological’ relation about patterns of property covariation, patterns that possibly are manifestations of some deeper dependence relation . . .” (Kim, 1998, s. 14)

Kim indrømmer at han engang mente, at superveniens var mere end den blotte beskrivelse af mønstre i samvariation. På denne måde bliver superveniens imidlertid en overordnet fællesbetegnelse for mange forskellige teorier, også typer af funktionalisme. Den afgrænses på den ene side over for den identitetsteoretiske fysikalisme - iflg. superveniensteorien er de psykiske fænomener ikke identisk med de fysiske, men supervenierer på dem. Den afgrænses på den anden side over for emergente teorierne,

som i Kims udgave nærmest er synonymt med en dualisme, idet der er tale om fænomener der opstår på tværs af to ontologisk adskilte domæner uden nævneværdig relation til hinanden.

Når Kim tilsyneladende er relativt defensiv i sin formuleringer af hvad superveniensteorien egentlig kan løse, skyldes det muligvis vanskelighederne ved at håndtere spørgsmålet om kausalitet. For en identitetsteoretisk fysikalisme er den tilsyneladende sammenhæng mellem en række psykiske egenskaber i adfærdsmæssigt handlingsforløb ikke noget problem de er kun tilsyneladende, reelt er der tale om en række fysiske kausaliteter. I den udstrækning superveniensteorien indeholder antagelsen af ikke-reducerbarhed, så er der ikke så mange modeller at vælge imellem. Nedenfor betegner M og M^* to psykiske egenskaber i en sekventiel række, mens P og P^* betegner realiseringerne eller instantieringer af dem. Den



første model udsiger, at en given fysisk egenskab er instantiering af en psykisk egenskab. Den psykiske egenskab er udgangspunkt for den næste egenskab i en handlingssekvens, som instantieres af en ny fysisk egenskab. Denne opfattelse indebærer en nedadgående kausalitet, hvor et psykisk (og fysiologisk baseret) element skaber en andet psykisk element, som først derefter giver anledning til et nyt fysisk element. Der er to ting som er forkert ved denne opfattelse (målt med bl.a. superveniensteorien). For det første er der tilsyneladende ingen sammenhæng mellem P og P^* , hvilket forekommer uholdbart. For det andet, og værre, er der en direkte nedadgående kausalitet, hvor den fysiske egenskab P^* dannes tidsmæssigt efter den psykiske. Den mellemste model er umiddelbart den mest sandsynlige, nemlig at de to psykiske egenskaber er instantieringer af et kausalt sammenhængende fysisk forløb. Det er imidlertid heller ikke helt tilfredsstillende, for hvad adskiller så egentlig identitets fysikalismen fra superveniensteorien. Selv om de lodrette pile skal forstås som superveniensrelationer, så er der alligevel sådan, at der forekommer at mangle en sammenhæng mellem M og M^* . I den tredje model er der tilføjet en stiplede linje mellem M og M^* . Linjen skal blot antyde, at der på en eller anden måde må være en tidsmæssig sammenhæng mellem M og M^* . Hvis der ikke er en sammenhæng er det tilsyneladende vanskeligt præcist at angive hvor superveniensteorien adskiller sig fra identitetsteorien.

I en diskussion af bl.a. Fodor og Searls antagelser præciseres nogle bestemte typer af funktionalisme. Superveniensteorien deler umiddelbart "multirealiseringshypotesen" med funktionalismen - det at en given psykisk størrelse har mange realiseringer. Superveniensteorien hævder en markant indskrænkning af realiseringmulighederne, at der er tale om

fysiske egenskaber, hvad funktionalismen principielt ikke gør. Når de psykiske egenskaber opfattes som funktionelle, så er de ikke kompositorisk bundet til de kompositionelle eller strukturelle detaljer hos deres realisatorer. Om en bestemt fysisk egenskaber realiserer en psykisk egenskab eller ej er et spørgsmål om mønsteret i de fysiske egenskaber. Enhver baseegenskab med et bestemt sæt af relationer til andre base-egenskaber kan tjene som instantiation af den psykiske egenskab.

Spørgsmålet om psykiske fænomeners realisation er vigtigt i denne sammenhæng. Til denne diskussion indføres en skelnen mellem sekundære og primære egenskaber (secondary-order properties), hvor de sekundære er de supervenient opståede og de primære den fysiske basis. Det er endvidere sådan, at relationen mellem de to slags egenskaber er indlejret i et større system. Sådanne systemer er kausalt nomologiske. Superveniens indeholder således et element af kausalitet. De sekundære psykiske egenskaber relaterer sig til de primære fysiske egenskaber kausalt nomologisk. Hver gang nervesystemet specificerer en specifik konfiguration, har en specifik fysisk egenskab P, da eksisterer den psykiske egenskab M. Men hvorvidt den psykiske entitet skabes eller ej, er et produkt af hele systemets tilstand. Man kan også sige det på en anden måde. Hvis et givent system er kausalt-nomologisk, så er der tale om funktionelle egenskaber. Men man skal her være opmærksom på, at der med kausalt-nomologisk ikke henvises til en specifik årsagssammenhæng mellem de primære og de sekundære egenskaber. De betingelser som systemet befinder sig i er kausal-nomologisk - det er et kausalt-nomologisk system. Dette indebærer *ikke* at så snart der foreligger en fysisk tilstand P som er grundlaget for det psykiske fænomen M, så eksisterer M hver gang P eksisterer. Om en given sekundær egenskab opstår eller ej er afhængig af hele systemets samlede funktion. I nogle tilfælde vil en given kemisk konstellation fremkalde et bestemt psykisk fænomen, i andre fremkaldes det ikke. P er en nødvendig betingelse, men ikke en tilstrækkelig. Der skal altså være nogle betingelser i systemet som skal være opfyldt for at den psykiske egenskab eksisterer. Dette kan ekspliciteres på tre måder:

1. De specifikke M'er og P'er kan ikke i sig selv formuleres som nomotetiske regler, idet de varierer fra organisme til organisme. Man kan således ikke beskrive en P - kemisk, fysiologisk el. lign. - og dernæst hævde at når den forekommer, så forekommer M også. Hele systemet skal tages i betragtning.
2. Det samme kan siges om en enkelt organisme. Fordi nociocptive celler fyrer - hvilket sædvanligvis giver anledning til en smerteoplevelse - så er denne fysiologiske egenskab ikke tilstrækkelig til altid at fremkalde en smerteoplevelse.
3. Syntaktikalisme er den opfattelse, at det er de syntaktiske egenskaber ved psyken som er det afgørende i beskrivelsen af psykiske

kausalteter - og ikke det semantiske (indholdsmæssige eller repræsentationelle). Synspunktet er især fremhævet i forbindelse med diverse diskussioner af computation. Kims pointe er, at man ikke behøver at godkende distinktionen mellem syntaks og semantik i denne forbindelse for at nå til at indholdsbestemme af psykiske fænomener ofte indebærer eksterne relationer. I intentioner, beliefs osv. er organismen rettet ud ad mod omverdenen, termene indebærer en relation mellem organismen og omverdenen, hvilket indfører "omverdenen" som parameter som skal medtages.

"Thus, that a given intentional state of an organism instantiates a certain semantic property is a *relational* fact, a fact that involves the organisms relationship to various external environmental and historical factors. This makes semantic properties relational, or extrinsic, whereas we expect causative properties involved in behaviour production to be nonrelational, or intrinsic, properties of the organism." (Kim, 1998, s. 36).

Andre steder plejer Kim i disse sammenhænge at skelne mellem svag og stærk superveniens. I en historisk fremstilling af superveniensteorien (Kim, 1990) argumenteres der for at skelne mellem tre elementer i teorien: Kovarians, afhængighed og ikke-reduktibilitet, hvor det især er kovariansen der detaljeres. (Det var også den karakteristik som Kim ovenfor beskrev superveniensteorien som helhed ud fra). Kovarians er bestemmelsen af superveniens ud fra en beskrivelse af to sæt af egenskaber som varierer synkront. Der skelnes mellem to svage varianter og to stærke varianter - her er det tilstrækkeligt med svag-I og stærk-II (A er en mængde af egenskaber der supervenierer på B). Den svage variant af kovariansen hævder, at der ikke eksisterer nogen mulig verden, som indeholder x og y, hvor x og y er uskelnelig m.h.t. egenskaber i B og samtidig skelnelige m.h.t. egenskaber i A. Sagt på en anden måde: det vil aldrig være muligt, at én bestemt neurofysiologisk enhed giver anledning til to forskellige psykiske egenskaber. Den stærke variant hævder: Det er med nødvendighed sådan, at hvis noget har egenskaben F i A, så eksisterer der en egenskab G i B sådan at G med nødvendighed har F. Sagt på en anden måde: givet en neurofysiologisk konstellation så vil der med nødvendighed altid opstå en bestemt psykisk egenskab. Den vigtigste forskel mellem den svage og den stærke er termen "med nødvendighed", for den indebærer at det må gælde for alle tænkelige verdener.

Kim mener ikke at den svage superveniens er tilstrækkelig til at specificere *hvilken* form for kovarianser der kan være tale om. Begrebet om sammenhæng eller afhængighed kan ikke præciseres inden for den svage form. Det skyldes selvsagt at den svage form blot hævder at der er en en-til-mange relation, men hvordan relationen mere præcist er, siges ikke. Tesen er konsistent med, at der findes en verden nøjagtig ligesom vores (i fysisk forstand), men uden nogen psykiske egenskaber overhovedet; eller

at der findes en verden nøjagtig ligesom vores, hvor det kun er encellede størrelser der har bevidsthed eller at der findes en verden nøjagtig ligesom vores hvor alt har samme grad af psykiske egenskaber. Hvilket må siges at forekomme kontraintuitivt.

Kim konkluderer, at det kun er den strenge form for superveniens der er gyldig. Det er den eneste form der kan sandsynliggøre specifikke relationer af kausal karakter. På den anden side skal de heller ikke være entydigt kausale, og der kan aldrig være tale om en en-til-en relation mellem fysik og psyke - så var man tilbage ved identitetsteorien.

Udviklingen i superveniensteorien (set ud fra Kims vinkel) er som de fleste andre teoriers. En sprudlende start med masser af løsningsforslag, de fleste kun som foreløbige, men med fremtidig indløsning. Med tiden bliver flere og flere spørgsmål ladt åbne tilbage. Ligesom i Putnams reduktionismeartikel afsluttes argumentationerne med at det "i sidste ende" er et empirisk problem - hvilket forekommer noget ufilosofisk. Der er imidlertid én ting der er afgørende ved superveniensteorien, nemlig at den har bidraget til, at ikke-identitetsteorier er de dominerende og at den har insisteret på at ontologien er nødvendig at medtage. Måske skyldes Kims problemer, at det i en analytisk-filosofisk ramme er vanskeligt at håndtere ontologi. Det er nærmest præmissen for hele diskussionen. Et sted skriver Kim at hvis emergente teorier har ret på noget punkt overhovedet, så er det i forbindelse med qualia (1998, s. 103) hvilket heller ikke er et specielt opsigtsvækkende synspunkt.

I sin diskussion af den analytiske filosofis genfødsel af emergente teorier (en anden pointe er at den i mange hovedtræk gentager emergente teorierne fra 1920-30'erne) anvender Stephan (1999) en skelnen som er udmærket. Han skelner mellem en punktuelt reduktion, en lokal reduktion og en global reduktion. En punktuelt reduktion er en reduktion af en entitet til ét bestemt tidspunkt. Den lokale reduktion beskriver reduktion gældende for f.eks. en bestemt art eller et bestemt individ, mens den globale reduktion omfatter en opnået reduktion som er gyldig til enhver tid i enhver form. Disse tre former er for så vidt alle strukturelle, idet de ikke behandler emergens af nye niveauer, så det er ikke en alt dækkende distinktion. Det er anvendelig i den analytiske tilgang idet en stor del af diskussionen går på forskellen på en lokal og en global reduktion.

Konklusion

Det var Putnam som i sin tid indførte funktionalismen (i den betydning den har her). Det gjorde han i en artikel fra 1967, hvor han fremstiller kernehypotesen bl.a. på denne måde: Givet en række forskellige tilstande af en organisme, som producerer ét bestemt psykisk indhold, er der ingen tvingende grund til at tro, at der er afgørende ligheder i de fysiske tilstande af organismen der instantierer det psykiske indhold. Altså en en-til-mange relation hvor "mange" ikke har noget relevant fælles. Antagelse

af "multirealisation" er i første omgang egentlig rettet mod identitetsteorien - det er et argument mod en-til-en overensstemmelsen mellem det psykiske og det fysiske. Det bliver så også grundlaget for funktionalismen. I beskrivelsen af funktionalismen, er det også en afgørende pointe, at psykiske tilstande kan betragtes som computationelle, og at de principielt kan udføres af en Turing-maskine. Putnams senere kritik af funktionalismen (bl.a. Putnam, 1988) har flere led, hvor et af dem er kritikken af Turing-maskine sammenligningen. Det næste led i kritikken er at overføre funktionalismens en-til-mange på de psykiske tilstande selv. Der er ikke noget tvingende argument for, at x antal ens psykiske tilstande forstået som tilstande med samme indhold, skulle have en computationel enshed. Har de ikke det er der ikke noget tilbage af funktionalismen.

Putnam diskuterer en indvending (mod hans kritik) af funktionalismen, som ekspliciterer en principiel antagelse vedr. reductibilitet. Hvis man som Putnam argumenterer for (Putnam, 1988, kap. 5 - 7), at det er yderst vanskeligt at begrunde, at to tilsyneladende ens psykiske størrelse - f.eks. at "sne er hvidt" - vitterligt er ens, og hermed åbner for en referenceteori som inddrager diverse lag af kontekster, så kræver sammenligningen en beskrivelse af disse lag af kontekster. Funktionalisten kunne her hævde, at godt nok er det måske kritisabelt at betragte en psykisk tilstand som lukket inde i hovedet, og man kan muligvis argumentere for, at en given psykisk tilstand er bundet til og hægtet sammen med omgivelserne, men så kan man principielt medtage disse omgivelser, såvel den fysiske kontekst, den biologiske kontekst, den interaktionistiske kontekst og den samfundsmæssige/kulturelle horisont.

En psykisk entitet bestående af et begrænset antal egenskaber, f.eks. tanken, bevidstheden om, "sne er hvid". Hvordan relaterer denne størrelse sig til de andre niveauer man kan operere med.

1. Den materialistiske og reduktionistiske identitetsteori vil hævde, at der til enhver mulig psykisk entitet eksisterer en og kun en fysisk (via neurofysiologien) entitet. Der er en strikt en-til-en relation, og i realiteten er der kun én entitet, nemlig den fysiske.
2. Der er i det ovenstående skelnet mellem forskellige en (psykisk entitet) -til-mange (neurofysiologiske entiteter) teorier. Der er funktionalismen i sin klassiske form, der er forskellige justerede funktionalismer og der er superveniensteorien. Superveniensteorien kan i denne forbindelse ses som en funktionalisme med begrænsninger. Til én psykisk entitet kan der svare en mængde af forskellige fysiske tilstande, men der svarer altid én, og den samme fysiske tilstand kan ikke give anledning til to forskellige psykiske entiteter.
3. Samme neurofysiologiske tilstand kan give anledning til to forskellige psykiske entiteter - hvilket samtidig er et definerende træk ved emergens.

For at præcisere forskellene mellem disse antagelser er det nødvendigt at stille spørgsmålet: "hvad er identiske entiteter". Det er på sin vis eklatant banalt at gøre opmærksom på, at der ikke er noget faktisk eksisterende som er identisk. Når person A til tiden t tænker "sne er hvid" og vi mener at denne psykiske entitet er baseret på en neurofysiologisk tilstand i A til tiden t , så er det dels sådan, at den neurofysiologiske tilstand betragtet som et samlet tilstandsbillede af nervesystemet, aldrig nogen sinde vil gentages i identisk form noget sted, hverken hos person A til t^* eller hos nogen anden person idet nervesystemet som helhed ændres konstant og der er så mange parametre involveret, at en identisk gentagelse er urealisabel. Det samme kan imidlertid siges om tanken "sne er hvid". Der er forskel den psykiske og sociale kontekst, som tanken "sne er hvid" realiseres i, den kan ledsages af et utal af forskellige konnotationer, affektive undertoner osv.osv. Og det er blot hos samme person. Hvis der er forskellige personer af forskellige aldre fra forskellige kulturer og historiske epoker, så er der selvsagt ikke ret meget der er ens ved tanken "sne er hvid".

MEN, det er der nu alligevel. Eller rettere, det er et fælles grundtræk ved både videnskab og ontologisk teori, at man må hævde den reale eksistens af kategorier, typer, begreber, naturlige klasser m.m. Hvis vi hævder, at verden kun eksisterer som tokens - altså konkrete ekstensionelle instantieringer, og at kategorisering blot er et mere eller mindre tilfældigt net der lægges ned over verden, så vil det uden tvivl ende i relativisme. Menneskets psyke er opstået i kraft af en begrebsdannelse som vi via sproget kan reflektere over og med, og denne begrebsdannelse svarer for så vidt udmærket til videnskaberens idealisering af deres genstandsområde og fjernelse af "støj" i form af uønskede egenskaber. Begreber, typer osv. dannes ved at aggregere det fælles og set bort fra det forskellige. Psykisk fungerer typer (og i mindre grad videnskabelige begreber) dynamisk i den forstand det aggregerede - det kernesæt af egenskaber der definerer typen - varierer afhængig af konteksten. Til nogle tidspunkter er det nogen kerne-egenskaber der er definerende for hele typen, til andre tider andre (jvf. E. Rosch).

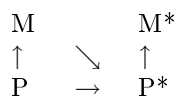
I forbindelse med superveniensteorien forekommer der en tendens til at se nervesystemets tilstande som tokens og de psykiske tilstande som typer. Dette forekommer imidlertid uholdbart og ender for så vidt i en uendelig regres. Hvis den psykiske entitet er en type, hvad er så realiseringen af den psykiske entitet - en token af typeform? Og hvis man har beskrevet to neuro-tokens som identiske til to forskellige tidspunkter, så har man for så vidt konstitueret en type. Bortset fra at der ikke er grund til at antage, at der ikke også er reelt eksisterende typer i nervesystemet. Så vidt vi kan se

giver anvendelsen af type-token distinktionen anledning til mere forvirring end godt er.

Hvis man hævder, at en given tilstand i nervesystemet kan give anledning til flere forskellige psykiske entiteter - altså version (3) ovenfor - så kan det begrundes på følgende måde:

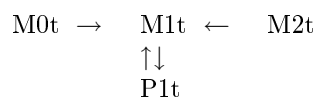
Til ethvert givet tidspunkt manifesterer nervesystemet en tilstand, som kan give anledning til mange forskellige psykiske entiteter. Forstået som initialbetingelser for psyken er nervesystemet i en tilstand, hvor der ikke entydigt og kausalt bestemmes en og kun en psykisk entitet. Mulighedsrummet indsnævres i og med den konkrete dannelse af en konkret psykisk entitet. Dette indebærer en selektiv udvælgelse, som kan forstås på flere måder.

- a) Det psykiske fungerer som begrænsende betingelser på det neurofysiologiske i den forstand at det er det psykiskes historicitet, der er baggrund for selektionen. Hvad der selekteres er bestemt af relationer til andre psykiske entiteter, og når en given psykisk entitet er dannet, virker den i sig selv som begrænsende betingelse på den følgende neurofysiologiske tilstand: Denne model er uklar i mælet omkring tidsligheden.



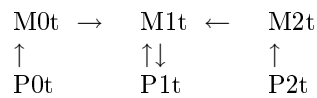
Hvis MP er til tiden t og M^*P^* er til tiden t_1 i hvilken tid eksisterer så den diagonale pil?

- b) Hvis problemet med tidsligheden skal afklares, må man hævde, at den givne psykiske entitet dannes i og med selektionen, og at der altså ikke er nogen forskel i tid på M og M^* . Dette kan fremstilles således: Alle fire entiteter eksisterer til samme tid.



- c) Her ville en indvending være, at $M0t$ og $M2t$ ikke kan hænge og svæve i luften, men at de i sig selv er bestemt af entiteter i nervesystemet, altså:

Her er der to problemer. Dels er det vigtigt, måske afgørende, at der ikke er nogen pile mellem P 'erne, svarende til pilene mellem M 'erne, eller at de ihvertfald af en anden slags (bemærk at det ikke er en tidsfølge men



en synkron model). Vi er her tæt på en anden/ny version af den gamle gestaltpsykologiske pointe, at samme fysiske størrelse kan perciperes på flere forskellige måder. Det der er betydende relationer (et mønster) på det psykiske niveau er ikke betydende, relevante, bestemmende på det neurofysiologiske niveau. Denne antagelse kan formentlig forsvares. Dels er der problemet med dobbelt-pilen. Den indeholder det element af nedadgående kausalitet, som er nødvendig for version (3) som helhed. En løsning ville være, at se logikken i modellen som parallel til bølgefunktions kollaps - blandt et antal mulige/sandsynlige tilstande realiseres én, som igen giver anledning til en ny sandsynlighedstilstand osv. En anden mulighed er at undersøge om man kunne finde eksempler på en selektiv kausalitet mellem andre niveauer - eller rettere, hvis det skal gælde som emergens-kriterier for overgangen biologi-psyke må det også gælde for andre overgange. En tredje mulighed er at se selektion ud fra logikken i Rosch's dynamiske begrebsdannelse. Samme pulje af egenskaber kan danne baggrund for mange forskellige selektioner, dog indenfor samme ekstensionelle omfang. Det intensionelle kan variere, afhængig af konteksten. Bestemte træk fremhæves som det essentielle ved en kategori - i andre situationer fremhæves andre egenskaber osv. Men det er ikke sådan at der slet ikke er nogen grænse mellem begreber. De begrebsdannelser, som involveres i psyken er så komplekse, dvs. indeholder så mange egenskaber, at dette system af egenskaber selvsagt ikke er stabilt og fasttømret. Det der afgør hvilke egenskaber der aktuelt opprioriteres er konteksten.

Til noget af dette vil version (2) indvende at ligegyldig hvilken konstellation af mønstre de psykiske entiteter realiseres som vil der altid svare en specifik fysisk tilstand. Dette svarer til superveniens-teorien. Hvis man endvidere forestillede sig at man til samme, dvs. identiske tilstand i nervesystemet, vitterligt havde to forskellige psykiske entiteter, så ville (2) argumentere, at så var den fysiske forskel blot et andet sted, f.eks. i konteksten, hvilket vil sige, at der altid vil være en fysisk forskel (jvf. Putnams argument ovenfor og Kims skelnen mellem interne og eksterne relationer). Spørgsmålet er hvordan man med denne antagelse undgår at ende i en regres: det er næppe relevant for den psykiske entitet, at der på Sirius lige nu flyttes adskillige atomer. Og mens det er umiddelbart tilforladeligt at diskutere den fysiske tilstand for en psykisk entitet, hvor det altid implicit forudsættes af det er via et organiseret nervesystem, da er det mindre indlysende at det overhovedet er relevant at tale om en fysisk baggrund for samfund.

Noget andet er, at hvis man insisterer på denne glidende kontekst og dermed på denne slags forskellighed, er det vanskeligt at se hvordan den psykiske identitet i supervenienstesens kan opretholdes - altså en-til-mange relationen. For hvorfor er ikke også den psykiske entitet da altid forskellig. Og problemet er da, at hvis der ikke nogensinde er identiske entiteter, så vil der ikke være den store forskel til identitets-fysikalismen - at der til enhver psykisk entitet svarer én og kun én fysisk tilstand. Der vil ikke være noget epistemologisk argument og niveauerne vil være udelukkende ontologisk begrundet.

Litteratur

Feigl, H. (1957). The "Mental" and the "Physical". *Minnesota Studies in the Philosophy of Science*, vol II, 370-498.

Fodor, J. (1974). Special sciences. In Kim, J. og Sosa, E. (ed.): *Metaphysics*, ss. 504-514. London: Blackwell, 1999.

Kim, J. (1978). Supervenience and nomological incommensurables. *American philosophical quarterly*, 15, 149-156.

Kim, J. (1984a). Concepts of supervenience *Philosophy and phenomenological research*, vol. XLV.

Kim, J. (1984b). Epiphenomenal and supervenient causation. *Midwest studies in philosophy*, vol IX, 257-270.

Kim, J. (1992). "Downward Causation" in Emergentism and non-reductive physicalism. In A.Beckermann m.fl. (ed.): *Emergence or Reduction?* ss. 119-139. New York: Walter de Gruyter.

Kim, J. (1998). *Mind in a physical world*. Cambridge: MIT Press.

Nagel, T. (1961). *The structure of science*. London: Routledge and Kegan.

Putnam, J. (1967). The nature of mental states. In *Philosophical papers*, 2, 429-441. Cambridge: Cambridge University Press, 1975.

Putnam, J. (1969). On properties. In J. Kim og E. Sosa (ed.): *Metaphysics*, 243-252. London: Blackwell, 1999.

Putnam, J. (1988). *Representation and reality*. Cambridge: MIT Press.

Stephan, Achim (1999). *Emergenz*. Dresden: Dresden University Press.

Smart, J.J.C. (1959). Sensation and brain processes. *Philosophical Review* 68, 141-156.

Om betingelserne for objektiv erkendelse af kvantitative forhold. En praktisk-realistisk teori

Af *Jens Mammen*¹

Indledning

I filosofiens og psykologiens tæt forbundne historier har forståelsen af den menneskelige erkendelse været udspændt mellem to poler, en empiristisk og en rationalistisk.

Det fælles udgangspunkt for de to forståelser har været, at mennesket modtager elementær erfaring om verden gennem sine sanser. Sanserne er menneskets umiddelbare kontaktflade til den verden, som skal erkendes. Sanserfaringen udgør erkendelsens elementære stof.

Forskellen mellem de to forståelser har angået forholdet mellem dette simple erfaringsstof og den menneskelige erkendelses færdige form, som den ytrer sig i vor begrebsmæssige og teoretiske erkendelse, en erkendelse som retter sig imod dybere og mere komplicerede forhold i verden end dem, som vi umiddelbart møder med sanserne.

Empiristerne har i forlængelse af en videnskabelig tradition for teoretisk økonomi forsøgt at forstå transformationen fra sanseerkendelse til begrebserkendelse ved hjælp af så enkle associative eller logiske principper som muligt. Et moderne eksempel er forsøgene på ved hjælp af modeller for induktive neurale netværk at simulere menneskelig erkendelse og begrebsdannelse. Denne såkaldte 'konnektionisme' er den moderne afløser for den klassiske associationisme.

Rationalisterne har i forlængelse af en lige så prisværdig tradition for antireduktionisme forsøgt at fastholde rigdommen i de menneske-

¹ Dette bidrag er en lettere omarbejdet udgave af Mammen (1994).

lige erkendelsesformer ved at forudsætte disse som givet i deres færdige form forud for det stof, som sanseerkendelsen tilvejebringer. Et moderne eksempel er forsøgene på at analogisere menneskelige tankeformer med komplicerede algoritmer defineret som computerprogrammer og tilsvarende analogisere sanseerfaringerne med inputsekvenser.

Jeg vil ikke her gøre meget mere ud af disse traditioner. Blot påpege, at de står over for store vanskeligheder, som jeg anser for principielle og uovervindelige. Empiristerne vil ikke kunne nå frem til en gyldig psykologisk forståelse af de menneskelige begreber ud fra deres simple mekaniske principper. Og rationalisterne undgår ikke at måtte forlade sig på aprioriske antagelser om erfaringsuafhængige begreber eller semantiske strukturer, hvis artshistoriske oprindelse samt livshistoriske og kulturhistoriske udvikling ikke kan forklares videnskabeligt.

Både empirister og rationalister står også over for et mere filosofisk eller erkendelsesteoretisk problem vedrørende erkendelsens objektivitet. Begge parter har svært ved at begrunde den begrebsmæssige erkendelsesobjektivitet, da ingen af parterne formår at forankre den begrebsmæssige erkendelsesform i virkelighedens former.

Jeg har i anden sammenhæng lidt mere indgående diskuteret disse traditioner og deres blandingsformer (f.eks. Mammen, 1986b), hvilket må tjene som undskyldning for den summariske behandling her. Jeg vil her koncentrere mig om et forsøg på at etablere en tredje forståelsesmåde, som afviger fra empirismens og rationalismens fælles antagelse om den simple sansning som den umiddelbare kontakt mellem verden og erkender. En sansning, som i sin simpelhed endnu ikke har den færdige erkendelses form. I stedet vil jeg se på forskellige forsøg på at forstå erkendelsens form ud fra et aktivt subjekts udvælgelse og erkendelse af omverdensobjekter gennem en formgivende praksis eller virksomhed.

Det drejer sig om en tradition med rødder i tysk idealisme efter Kant og dialektisk materialisme (Fichte, Hegel, Marx), som på nogle punkter er beslægtet med den amerikanske pragmatisme (Peirce, Dewey, Mead). I psykologien har de væsentligste repræsentanter nok været den russiske psykolog A.N. Leontjev og den amerikanske psykolog J.J. Gibson. (Se f.eks. Mammen, 1983; 1986b). Jeg vil her fremdrage, hvilket lys denne tradition, som jeg vil kalde *praktisk realisme*, kan kaste over et specielt emne, nemlig erkendelsen af kvantiteter. Jeg vil mene, at der her er et frugtbart berøringspunkt mellem psykologien og andre fag, der beskæftiger sig med kvantitative vurderinger og deres udvikling. Her er den økonomiske videnskab vel en af de mere nærliggende. Den samme forbindelse blev fremhævet af den tidligere professor i Teoretisk Statistik som Redskab inden for Samfundsvidenskaberne ved Københavns Universitets Statistiske Institut, Georg Rasch. Dette skete bl.a. i hans afskedsforelæsning den 9. marts 1972, som siden i udvidet form blev publiceret i Nationaløkonomisk Tidsskrift (Rasch, 1972). Men naturvidenskab er nok eksemplet frem for alle andre.

Jeg vender tilbage til Rasch's begreb om 'specifikt objektive målinger'.

G.T. Fechner

Den moderne empirisk funderede psykologi skilte sig ud fra filosofien i det 19. århundrede. Startskuddet bliver ofte anset for at være Gustav Theodor Fechners arbejde fra 1860, som havde det formål at finde en lovmæssig sammenhæng mellem fysiske sansepåvirkningers styrke og deres oplevede eller psykiske intensitet, deraf navnet 'psykofysik' (Fechner, 1860).

Fechner funderede bl.a. sit arbejde på empiriske resultater opnået i 1830'erne af fysiologen Ernst Heinrich Weber. Weber havde fundet, at for forskellige sansekvaliteter, som f.eks. lydstyrke og lysstyrke, var forskelstærsklen ΔI (med god tilnærmelse) proportional med den fysiske intensitet I , en sammenhæng, som siden er blevet kendt som *Webers lov*, altså

$$\frac{\Delta I}{I} = k \quad (1)$$

Forskelstærsklen ΔI er her den forskel i sansepåvirkningens fysiske intensitet, som en forsøgsperson vil opdage i netop halvdelen af de præsenterede tilfælde.² Proportionalitetskonstanten k er forskellig for forskellige påvirkningskvaliteter, men forholdsvis ens for forskellige personer. F.eks. er k for lydstyrke ca. 0.1, for lysstyrke ca. 0.02.³

Fechner gjorde nu den dristige antagelse, at der til en given fysisk intensitet I svarer en psykisk kvantitet P (for psyke), og at P er en funktion af I , dvs. $P = f(I)$. Desuden forudsatte han, at forskelstærskler altid opleves kvantitativt ens og med en størrelse, som udgør en måleenhed for oplevet forskel. Med andre ord, hvis to intensiteter afviger med netop en forskelstærskel, opleves forskellen som en elementær forskel. Hvis to intensiteter afviger med n forskelstærskler, opleves forskellen som n elementære forskelle og dermed som n gange så stor som den første forskel. Endelig er forskellen mellem to oplevede styrker lig med oplevelsen af forskellen. Den oplevede styrke har altså ifølge Fechner en *additiv struktur*, eller er struktureret som en differensskala, med den oplevede forskelstærskel som *enhed*.

Ifølge Fechner har vi altså, at hvis det for to intensiteter I_1 og I_2

² Der ses her bort fra de metodiske problemer ved denne definition, som først er løst i den moderne signaldetektionsteori, der bygger på en matematisk model for tvungne beslutninger.

³ Der ses her bort fra tekniske detaljer vedr. definitionerne af påvirkningskvaliteterne, dvs. frekvens, bølgelængde, areal etc.

gælder, at⁴

$$I_2 - I_1 = n\Delta I, \quad (2)$$

så vil det for de tilsvarende oplevede styrker $P_1 = f(I_1)$ og $P_2 = f(I_2)$ gælde, at

$$P_2 - P_1 = n. \quad (3)$$

Af ligningerne (1), (2) og (3) får vi nu, at

$$\frac{P_2 - P_1}{I_2 - I_1} = \frac{1}{kI} \quad (4)$$

hvor I er en værdi i intervallet $I_1 < I < I_2$. Vi kan derfor tillade os at erstatte differenskvotienten i (4) med differentialkvotienten dP/dI og får nu

$$\frac{dP}{dI} = \frac{1}{kI} \quad (5)$$

eller ved integration

$$P = \frac{1}{k} \ln\left(\frac{I}{I_0}\right), \quad (6)$$

hvor P sættes til 0 for $I = I_0$, og hvor I_0 kan være en vilkårlig valgt udgangsværdi. Eller på den etablerede form, hvori *Fechners lov* er blevet kendt

$$P = c \log(F), \text{ eller } P = KF\text{dB}, \quad (7)$$

hvor $c \approx 2.3/k$, $K \approx 0.23/k$, og hvor $F = I/I_0$ (F for fysisk). $F\text{dB}$ betyder, at F måles i dB (decibel) i forhold til I_0 .⁵

Fechners lov udgør en væsentlig del af rationalet for at måle intensiteten af sansepåvirkninger, f.eks. støjniveau, i logaritmiske enheder, nemlig i decibel.

Som det ses, er Fechners metode til bestemmelse af den psykiske størrelse P indirekte og, via hypotetiske antagelser om enhed og additivitet, afledt af personers empirisk konstaterede skelneevne i forhold til de fysiske intensiteter F (eller I). Selvom Fechner behandler variabelen P , som havde den en selvstændig realitet, er den imidlertid en konstruktion, hvis realitet udelukkende består i de implikationer, der kan afledes af Fechners lov. Dvs. empiriske forudsigelser, som ikke selv rummer henvisning til P , men udelukkende til vor evne til at skelne og sammenligne sansepåvirkningers intensiteter F indbyrdes.

⁴Det forudsættes, at ΔI er relativt konstant i intervallet mellem I_1 og I_2 for små n og ellers, at vi i henhold til middelværdisætningen kan erstatte summen af de enkelte ΔI med $n\Delta I$ for et I i intervallet.

⁵Her betegner 'ln' den naturlige logaritme og 'log' titallogaritmen. Det gælder, at $\ln(x) = \ln(10)\log(x)$, hvor $\ln(10) \approx 2.3$. Heraf $c \approx 2.3/k$. Forholdet mellem to positive tal, x og y , er n bel (B), hvis $\log(x/y) = n$. Forholdet er tilsvarende n decibel (dB, tiendedele bel), hvis $\log(x/y) = n/10$. Heraf $K = c/10 \approx 0.23/k$.

En implikation af Fechners lov (7) er, at for to fysiske intensiteter F_1 og F_2 af en given sansekvalitet, f.eks. lysstyrke, vil vi for de tilsvarende oplevede intensiteter P_1 og P_2 have, at

$$P_2 - P_1 = \text{clog}(F_2) - \text{clog}(F_1) = \text{clog}\left(\frac{F_2}{F_1}\right). \quad (8)$$

Den oplevede forskel afhænger altså alene af forholdet mellem de fysiske intensiteter. Hvis begge fysiske intensiteter multipliceres med den samme faktor, skulle det altså ikke påvirke den oplevede forskel.

Denne implikation af Fechners lov modsvarer faktisk en objektiv egenskab ved dyrs og menneskers livsrelevante omverden eller 'økologi'. Hvis en overflade med farver i forskellig lyshed (eller albedo) belyses med indfaldende lys i varierende styrke, vil forholdet mellem det reflekterede lys fra de forskellige dele af overfladen være konstant eller invariant. Hvis en lydkilde, f.eks. et talende menneske, høres i varierende afstand, vil forholdet mellem lydintensiteterne i de forskellige frekvenser og i de forskellige dele af forløbet med god tilnærmelse være invariant. Disse såkaldte økologiske *invarianser* modsvarer altså ifølge Fechners lov invariante oplevede forskelle.

Enhver vil kunne bekræfte denne lovmæssighed i dagligdagens oplevelser. Ting ser stort set ens ud i forskellig belysning, vi kender den samme stemme på tæt og lang afstand, etc.

Den tanke er nærliggende, at dyrs og menneskers erkendelse er optimal, hvis den genspejler objektive omverdensforhold relativt uafhængigt af erkendelsespositionen, og at den biologiske udvikling har fremmet en sådan lovmæssig sammenhæng mellem påvirkninger og oplevelser, der lader objektive egenskaber fremtræde som konstante. Fechners og Webers love er formentlig afledte og implicite udtryk for vor økologiske invarianser.

Endnu en implikation bør nævnes af Fechners lov (7). Lad os tænke os, at vi har f.eks. en variabel lyskilde og en variabel lydkilde. Vi udvælger nu en given lydstyrke og en given lysstyrke og fortæller en forsøgsperson, at de modsvarer hinanden. Der skrues derefter op for lyset, og man beder forsøgspersonen regulere lyden, så lys og lyd igen svarer til hinanden. Sådanne såkaldte tværmodale sammenligninger opleves af de fleste mennesker som meningsfulde og kan udføres med temmelig stor præcision.

Hvis vi sætter den oplevede styrke af de to første påvirkninger til nul, drejer det sig altså, i henhold til Fechner, om at afgøre, for hvilke andre par af lys- og lydstyrker (F_{lys}, F_{lyd}) det gælder, at

$$P_{lys} = P_{lyd}, \quad (9)$$

eller ifølge (7), at

$$c_{lys} \log(F_{lys}) = c_{lyd} \log(F_{lyd}), \text{ eller} \quad (10)$$

$$F_{lys} = F_{lyd}^q \quad (11)$$

hvor eksponenten $q = c_{lyd}/c_{lys} = k_{lys}/k_{lyd}$, og hvor k 'erne er de empirisk fundne Weber-brøker (1).

En lang række sådanne forsøg med mange forskellige sansekvaliteter er udført af den amerikanske psykofysiker S.S. Stevens (1960; 1961), der har bekræftet den forudsagte *potenslov* (11), ganske vist med eksponenter q , der i visse tilfælde afviger fra de værdier, der via Fechners lov kan forudsiges ud fra Webers k -værdier.

Stevens har dog først og fremmest udført forsøg, hvor den ene af påvirkningerne er erstattet med talværdier, hvad han kalder *direkte skalering*. Også her findes en potenslov, selvom det er meningsløst, at tale om en k -værdi for tal. At vi imidlertid også oplever tal logaritmisk, forekommer nærliggende, al den stund nutidens mennesker lever i en økologi, hvor tal ofte er udtryk for målestørrelser i forskellige måleenheder. Det er netop forholdet mellem to tal (forskellen mellem deres logaritmer), der siger noget om et objektivt invariant forhold, uafhængigt af, om enheden er cm, meter eller tommer, hvorimod tallenes differenser kun afbilder noget invariant, så længe den arbitrære måleenhed holdes fast.

Det er imidlertid påfaldende, at Stevens selv insisterer på at forstå tal som direkte mål for den oplevede indre psykiske intensitet P , som han reificerer fuldt så meget som Fechner, i stedet for at anskue tallene som en del af vor kulturelle økologi, som vi forholder os praktisk til. Derfor mener Stevens også, at hans potenslov er i direkte modstrid med Fechners logaritme-lov (Stevens, 1961). Som en strid om, hvilket af de to fiktive P 'er, der er virkeligt, er spørgsmålet uløseligt og illustrerer ganske godt det ufrugtbare i den traditionelle, mekaniske empirisme (Poulsen, 1994).

Allerede i 1874 havde Franz Brentano hævdet den logiske implikation mellem de to lovmæssigheder (logaritme-loven og potens-loven), under forudsætning af, at Fechners lov også beskriver vort forhold til tal. Måske afspejler Stevens' uenighed med Brentano en principiel forskel på den traditionelle empirisme og Brentanos læggen vægt på den praktiske virksomhed i menneskenes livsverden (Engelsted, 1989, p. 68f).

De sammenhænge, som de klassiske psykofysikere studerede, modsvarede en yderst begrænset del af vor sansemæssige kontakt med verden, og var i øvrigt som regel undersøgt empirisk under meget restriktive laboratoriebetingelser. Ikke desto mindre udgør den klassiske psykofysik en af psykologiens største succeser, idet dens afkast uden for den egentlige psykologi har været betydeligt. Hverken Kodak eller Philips ville sikkert have kunnet undvære den. Psykofysikkens oprindelige ambition om at udgøre et fundament for den almene psykologi har imidlertid været en tilsvarende fiasko, bundet som den har været til en empiristisk forståelse af erfaringens stof udelukkende som sansninger af simple fysiske påvirkninger.

J.J. Gibson

Et væsentligt brud med den klassiske psykofysik sker med den amerikanske psykolog J.J. Gibson (1966; 1979). Gibsons erklærede hensigt er at erstatte empirisme-rationalisme dikotomien med en realisme, funderet i dyrs og menneskers praktiske livsvirksomhed.

Gibsons studieområde er først og fremmest vor visuelle erkendelse af omverdensgenstandenes overflader. Han gennemfører en analyse af, hvilke invarianser der optræder i det samlede visuelle påvirkningsmønster, når vi færdes i dagligdagens landskaber og rum. Han konstaterer f.eks., at påvirkningsmønsteret, der hidrører fra faste flader, undergår matematisk beskrivbare kontinuerte perspektiviske transformationer, når vi bevæger os i forhold til dem. I disse transformationer optræder der *invarianser*, som *specificerer objektive geometriske relationer* i de faste flader.⁶

I en række forsøg påviser Gibson, at disse invarianser faktisk registreres og udnyttes af dyr og mennesker i erkendelsen af deres omverden, uanset at invarianserne er abstrakte matematiske relationer i højdimensionale variabel-rum ('higher-order variables'). Sanserne medvirker i denne erkendelsesproces. Men de tilvejebringer kun information om verden ved at være *indlejret i en praktisk virksomhed*, der forbinder individet med dets omverden.

Gibson udvider sin analyse til de tilfælde, som har særlig relevans for menneskenes omverdenserkendelse, og hvor vi erkender verden igennem *redskaber*. F.eks. kan vi føle en overflades form og elastiske beskaffenhed, ruhed m.m. ved hjælp af en pind, vi holder i hånden. Vor opmærksomhed er ikke fokuseret på pindens kontakt med hånden, men på dens kontakt med overfladen. Vi føler så at sige med spidsen af pinden, der fungerer som en slags forlængelse af vor krop.

Gibson har dog ikke udviklet nogen detaljeret teori for denne redskabserkendelse og heller ikke udført nogen nævneværdig empirisk udforskning af den. Tilsyneladende står vi over for en praktisk erkendelseform, som er meget kompliceret og vanskelig at analysere.

⁶ Et par særligt simple eksempler (der findes andre mere funktionelt effektive) viser sig i den central-projektion, som kaster reflekterede lysstråler fra omgivelserne ind i hvert af en betragters øjne gennem øjets optiske midtpunkt. Projektionen (strålebundtet eller 'the optic array') kan repræsenteres entydigt ved strålernes skæring af et vilkårligt projektionsplan i betragterens øje. Det gælder nu, at hvilke som helst tre punkter i omgivelserne, der i forhold til hinanden ligger på en ret linie, også altid på projektionsplanet vil ligge på en ret linie. Det samme gælder naturligvis for fire punkter A, B, C og D , der ligger fast på en ret linie. Her vil det endvidere gælde for projektionerne A', B', C' og D' på planet, at $(A'C'/B'C')/(A'D'/B'D')$ er invariant og lig med $(AC/BC)/(AD/BD)$, det såkaldte dobbeltforhold, 'cross-ratio'. For at disse invarianser skal specificere konstante forhold i omgivelserne, er det nødvendigt, at de optræder på baggrund af en variation i strålebundtet produceret af, at individet og omgivelserne bevæger sig i forhold til hinanden. For at være effektive skal de også gøre sig gældende som globale relationer i et helt felt og ikke blot mellem få punkter.

Jeg vil imidlertid her gøre et forsøg på at analysere et meget simpelt tilfælde af redskabsformidlet erkendelse, ganske vist med fare for at foretage en indsnævring af feltet, der ikke er mindre end den, som de klassiske psykofysikere foretog, hvad angår den mere direkte sansformidlede erkendelse. Min undskyldning er imidlertid, at jeg på trods af indsnævringen indfanger *et specifikt træk ved den menneskelige erkendelse*, som i sin almene form gennemsyrrer hele vort omverdensforhold, nemlig vor evne til *praktisk og erkendelsesmæssig fastholdelse af enkeltting* (Mammen, 1983).

G. Rasch

Analysen af de særlige invarianser, som afdækkes i verden ved den redskabsformidlede erkendelse, og som jeg vil beskrive i det følgende, er foretaget af den danske statistiker Georg Rasch, hvis indsats her i flere henseender er parallel med Gibsons, som jeg i øvrigt ikke ved, om han var bekendt med.

Georg Rasch arbejdede i mange år som statistisk konsulent inden for flere forskellige fagområder, bl.a. medicin, økonomi og psykologi. I sin lærestol som professor var han knyttet til den nationaløkonomiske faggruppe ved Københavns Universitet. Selv kom jeg i berøring med ham i forbindelse med en dyrepsykologisk undersøgelse, hvor jeg fungerede som assistent (Reventlow, 1970). Jeg har fortalt om Rasch's bidrag til denne undersøgelse i en tidligere publikation (Mammen, 1986a). Desuden fulgte jeg som studerende i en årrække Rasch's forelæsninger, studiekredse og seminarer. En god introduktion til Rasch's betydningsfulde arbejde kan findes i Karpatschof (2007).

Rasch's særlige bidrag til statistikkens teori vedrører beskrivelsen af interaktionen mellem en udvalgt mængde undersøgelsesobjekter (objekter) og et inventar af redskaber (af Rasch kaldet agentia). Objekterne kan være skoleelever, der underkastes prøver, eller personer, der udspørges med spørgeskemaer. Redskaberne kan tilsvarende være psykologiske prøver, standpunktsprøver eller spørgeskema-items. Interaktionen resulterer i data (reaktioner), som typisk kan beskrives på ordinale eller nominale skalaer. F.eks. antal stavfejl i en læseprøve eller 'ja'/'nej' til et spørgsmål. Som hovedregel antages resultatet af interaktionen at være stokastisk (ikke-deterministisk) og betinget af konstante egenskaber hos objekter og agentia, samtidig med at de enkelte interaktioner er indbyrdes stokastisk uafhængige, dvs. at de ikke gensidigt påvirker hinanden.

Det ideal, som Rasch opstillede for de statistiske modeller til at beskrive denne situation, hentede han ikke fra samfunds- og adfærdsvidenskaberne, men fra naturvidenskaben, især fysikken. Han henviste her til den analyse, som den tyske fysiker og psykolog Kurt Lewin (1931) havde foretaget af naturvidenskaben efter Galilei. Det karakteristiske for den mod-

erne fysiks anvendelse af modeller, herunder sandsynlighedsmodeller, er at modellerne, ligesom fysikkens love i det hele taget, er 'population-suafhængige', dvs. at interaktionerne beskrives uden henvisning til fænomenernes absolutte hyppigheder, men alene til deres betingede hyppigheder, som anses for universelle og gældende for ethvert enkelttilfælde. Fysikken før Galilei, f.eks. den Aristoteliske fysik, havde derimod ifølge Lewin fejlagtigt identificeret det lovmæssige med det absolut hyppige, dvs. det typiske eller almindeligt forekommende i den faktiske fænomenerverden (population), som det ytrede sig i det empirisk gennemsnitlige.

Fysikken før Galilei havde derfor også været 'botaniserende', passivt og ekstensivt iagttagende, hvorimod den efter Galilei koncentrerede sig om den intensive og aktivt-eksperimentelle undersøgelse af enkelttilfældets dynamik, gerne under ekstreme (og sjældne) betingelser.⁷

Rasch mente, at der ofte skete en sådan 'før-galileisk' sammenblanding af den deskriptive og teoretiske statistik, der tildelte normalfordelte værdier i populationer en fejlagtig teoretisk status i de statistiske modeller.⁸ I stedet udviklede Rasch et sæt *populationsuafhængige målingsmodeller*, som han mente levede op til de lewinske og galileiske krav om universel beskrivelse af enkelttilfælde. Jeg mener, at det faktisk lykkedes, og at modellerne desuden havde stor matematisk elegance, f.eks. Rasch (1960), et tegn på, at man er på rette vej! Derudover viste modellerne sig faktisk anvendelige på meget forskellige materialer.

Det var karakteristisk for den generalitet, hvormed kravene til målingsmodellerne blev formuleret, at disse krav også kunne anvendes på tilfælde, hvor interaktionerne var deterministiske. Disse tilfælde er matematisk simple end de stokastiske tilfælde, og jeg skal derfor af hensyn til fremstillingens længde og læserens tålmodighed holde mig til dem. Desuden kan de behandles mere udtømmende, hvad angår deres matematiske form.

Jeg vil yderligere indskrænke mig til det simple endimensionale deterministiske tilfælde, som Rasch har analyseret flere steder, f.eks. Rasch (1960, p. 110-114; 1972, p. 161-172).

Rasch vælger som illustrativt eksempel interaktionen mellem en række legemer (objekter) og en række kraftpåvirkninger (agentia)⁹ og de resulterende accelerationer, således som den kan beskrives med Newtons 2. lov. Hvis legemerne kan karakteriseres ved masseparametrene $M_i (i = 1, \dots, m)$, og kraftpåvirkningerne med kraftparametrene $K_j (j=1, \dots,$

⁷ Det hører naturligvis med til historien, at Lewin vender sin brod imod samtidens psykologi, der med få undtagelser (f.eks. Freud) karakteriseres som før-galileisk. En tilsvarende kritik af senere psykologiske teorier findes hos Poulsen (1982).

⁸ Jævnfør f.eks. Rasch's skeptiske holdning til populationsbegrebet i Rasch (1955).

⁹Strengt taget er det hos Rasch ikke selve kraftpåvirkningerne, der udgør agentia, men derimod det konkrete sæt identificerbare og stabile instrumenter eller redskaber, der udøver kraftpåvirkningerne. Rasch bygger her videre på en analyse fra 1876 foretaget af den skotske fysiker James Clerk Maxwell (1831-79), der redegør for, hvorledes det er muligt på én gang at indføre de gensidigt afhængige begreber masse og kraft og deres måltal ud fra mønstret i iagttagne accelerationer uden at have i en cirkularitet.

n), vil accelerationerne have værdierne

$$A_{ij} = \frac{K_j}{M_i}, \quad (12)$$

hvis der måles i passende enheder.

Hvis alle legemer efter tur påvirkes med alle kræfter, vil vi få $m \times n$ ligninger af formen (12). Hvis vi har registreret alle $m \times n$ accelerationer, har vi altså $m \times n$ ligninger med $m + n$ ubekendte, idet vi eksempelvis antager, at vi ikke på forhånd kender kraft- og masseparametrene.¹⁰

Hvis f.eks. $m = 3$ og $n = 3$, har vi 9 ligninger til bestemmelse af 6 parametre. Ligningssystemet er overbestemt, og strukturen i matricen A_{ij} vil kunne bruges til at teste vor formel eller model (12), uanset at ingen af parametrene kendes. I nærværende eksempel skal matrixens rækker og søjler således være proportionale. Selvom vi alene har empirisk kendskab til A_{ij} matricen og ikke kender nogen af parameter værdierne, kan vi altså alligevel ved en avanceret form for induktiv erkendelse danne os et begreb om formen af ligning (12), blot vort empiriske grundlag svarer til en overbestemt matrix eller endda blot et overbestemt udvalg af matrixens værdier i forhold til antallet af parametre.¹¹

Hvis vi betragter den simple situation, hvor $m = 2$ og $n = 2$, har vi netop 4 ligninger og 4 ubekendte. I det almindelige tilfælde vil vi derfor kunne beregne alle 4 ubekendte, forudsat at vi kender ligningernes (fælles) form. Men vi vil ikke have nogen mulighed for alene ud fra A_{ij} matrixens 4 værdier at teste, om selve ligningernes form er som antaget.

Men det særlige tilfælde, som er beskrevet i ligning (12), afviger fra dette almindelige tilfælde. Her kan vi faktisk alene ud fra A_{ij} matrixens 4 værdier, dvs. uden at kende nogen af de 4 ubekendte, teste, om ligningernes form er som beskrevet i (12). Søjlerne og rækkerne i A_{ij} matricen skal som sagt være proportionale. Til gengæld for denne erkendelsesgevinst, kan vi i dette tilfælde ikke beregne de 4 parametre, selvom vi kender ligningens form og hele A_{ij} matricen. Uanset hvor stor en $m \times n$ matrix vi har, vil vi ikke kunne beregne parametrene. Ligningssystemet vil i alle tilfælde have ingen eller uendeligt mange løsninger, aldrig kun én.

Denne særlige situation viser sig dog at indeholde endnu et afgørende erkendelsesmæssigt træk, som Rasch har kaldt 'specifik objektivitet'. Lad os betragte et udvalg af matricen alene bestående af 2 værdier i samme

¹⁰ Sættet af værdier (A_{ij}), hvor $i = 1, \dots, m$ og $j = 1, \dots, n$, udgør altså en $(m \times n)$ -matrix A eller i det følgende blot A_{ij} . Den matematisk kyndige læser advares mod at læse matricen A_{ij} som en tensor eller lineær afbildning. A_{ij} repræsenterer alene højre- eller venstresiden i et ligningssystem $f(M_i, K_j) = A_{ij}$. Matrix-algebraiske begreber som rang (A) etc. er derfor ikke relevante i nærværende sammenhæng.

¹¹ Denne sidste situation forelå f.eks. i Rasch's analyse af anvendelsen af læseprøver på skoleelever, hvor ikke alle prøver med rimelighed kunne anvendes på alle klassetrin, det såkaldte 'brobygningsproblem'. Tilfældet er dog mere kompliceret end det ovenfor diskuterede ved at være stokastisk (se Rasch, 1960)

søjle A_{1j} og A_{2j} svarende til, at vi lader den samme (ukendte) kraft K_j virke på to legemer med (ukendte) masser M_1 og M_2 . Hvis vi indsætter værdierne i ligning (12) og dividerer, får vi nu, at

$$\frac{A_{1j}}{A_{2j}} = \frac{K_j/M_1}{K_j/M_2} = \frac{M_2}{M_1} \quad (13)$$

Forholdet mellem accelerationerne er altså uafhængigt af den anvendte kraft og afhænger alene af de to masser. Eller med andre ord, forholdet mellem accelerationerne er en *invarians*, der direkte specificerer eller informerer om forholdet mellem masserne, uafhængigt af kraften. Vi kan altså ikke ud fra 2×2 matricen beregne nogen af de 4 ubekendte. Til gengæld kan vi i denne særlige situation alene ud fra 2 af matrixens værdier (i samme søjle, eller række) beregne en relation mellem 2 af de ubekendte, hvilket ikke er muligt i det almindelige tilfælde.

Generelt formuleret eksisterer der altså i vort eksempel en funktion α og en funktion¹² β således at

$$\alpha(A_{ij}, A_{kj}) = \beta(M_i, M_k) \text{ for alle } j. \quad (14)$$

Funktionen α er en *invarians*, som entydigt *specificerer* og dermed informerer om en *relation* β i *objekterne* uafhængigt af valget af agentium (K_j). Det var dette træk, Rasch kaldte for 'specifik objektivitet'. Gibson ville nok have kaldt det for 'specificeret objektivitet', men ellers have ment det samme.

Lad os nu i stedet for legemer, kræfter og accelerationer helt alment tale om objekter (med parameter o), agentia (med parameter a) og reaktioner (med parameter r), og i stedet for den særlige ligning (12) helt alment udtrykke r som en funktion¹³ af o og a

$$r = f(o, a) \text{ eller } r_{ij} = f(o_i, a_j) \quad (15)$$

Kravet om *specifik objektivitet* kan nu i mere almen form end i (14) formuleres som et krav om, at der skal eksistere funktioner α og β , så at

$$\alpha[f(o_i, a_j), f(o_k, a_j)] = \beta(o_i, o_k) \quad (16)$$

for alle o_i, o_k og a_j .

Det kan nu bevises (Rasch, 1972, p. 170-171), at klassen af de funktioner f , der opfylder kravet om specifik objektivitet, netop er mængden af *latent additive* funktioner, dvs. funktioner, der kan omskrives på en

¹² Der skal være tale om reelle, monotone (og dermed én-éntydige) funktioner i begge variable.

¹³ Funktionen antages at være to gange differentiabel og strengt monoton i begge reelle variable. Dvs. de partielle differentialkvotienter skal være forskellige fra nul for alle værdier.

additiv form. Mere præcist formuleret er funktionen $r = f(o, a)$ latent additiv, hvis og kun hvis der eksisterer funktioner¹⁴ ϕ , g og h , således at

$$\phi[f(o, a)] = g(o) + h(a) \quad (17)$$

for alle o og a ¹⁵.

Da ligning (17) er symmetrisk mht. o og a , ses det endvidere, at hvis funktionen f opfylder kravet om specifik objektivitet med hensyn til objekter, gør den det også med hensyn til agentia. Hvis der altså eksisterer en invarians i reaktionsmatricen, der specificerer relationer i objekterne uafhængigt af agentia, eksisterer der også altid invarianser i reaktionsmatricen, der specificerer relationer i agentia uafhængigt af objekterne. *De objektive betingelser for erkendelse af relationer i objekterne er de samme som for erkendelse af relationer i redskaberne (agentia)* (jf. Rasch, 1972, p. 172).

I vort eksempel med legemer og kræfter vil forholdet mellem to accelerationer med det samme legeme således fortælle om forholdet mellem de påvirkende kræfter uafhængigt af legemets masse.

I eksemplet vil funktionerne ϕ , g og h i (17) være logaritmefunktioner. Således vil den latent additive ligning (12) kunne omskrives til

$$\log(A_{ij}) = [-\log(M_i)] + \log(K_j), \quad (18)$$

som har en eksplicit additiv form. Der findes matematiske midler til for en given funktion $f(o, a)$ at afgøre, om den er latent additiv, og til at finde, hvilke funktioner ϕ , g og h , der i givet fald gør den eksplicit additiv.¹⁶

¹⁴ Dvs. strengt monotone, reelle funktioner af én reel variabel.

¹⁵ Det ses let, at latent additivitet (17) er en tilstrækkelig betingelse for specifik objektivitet som defineret i (16). Lad nemlig $\alpha(x, y) = \phi(x) - \phi(y)$ og $\beta(x, y) = g(x) - g(y)$. Vi har nu, at $\alpha[f(o_i, a_j), f(o_k, a_j)] = \phi[f(o_i, a_j)] - \phi[f(o_k, a_j)] = [g(o_i) + h(a_j)] - [g(o_k) + h(a_j)] = g(o_i) - g(o_k) = \beta(o_i, o_k)$. Beviset for, at (17) også er en nødvendig betingelse for (16), er mere teknisk.

¹⁶ Ligning (17) kan omskrives til

$$f(o, a) = F[g(o) + h(a)], \quad (17a)$$

hvor F er den inverse til ϕ . Vi har nu, at

$$\frac{\partial f}{\partial o} = F'[g(o) + h(a)]g'(o) \text{ og} \quad (17b)$$

$$\frac{\partial f}{\partial a} = F'[g(o) + h(a)]h'(a), \text{ hvoraf fås} \quad (17c)$$

$$\frac{\partial f / \partial o}{\partial f / \partial a} = \frac{g'(o)}{h'(a)} \text{ eller} \quad (17d)$$

$$\log \left| \frac{\partial f / \partial o}{\partial f / \partial a} \right| = \log |g'(o)| - \log |h'(a)|, \quad (17e)$$

hvoraf igen

$$\frac{\partial^2 \log \left| \frac{\partial f / \partial o}{\partial f / \partial a} \right|}{\partial o \partial a} = 0. \quad (17f)$$

Det må antages, at *fysiske interaktioner* mellem redskaber og genstande, for så vidt de følger fysikkens love, er latent additive¹⁷ og således byder os særligt gunstige *objektive erkendelsesvilkår*. Hvis vi ud fra relativt stabile objekter og redskaber har erfaret interaktionens latent additive form, vil vi derefter også kunne erkende relationer mellem variable objekters egenskaber, blot redskaberne er relativt stabile (og vice versa).

For den umiddelbare tanke er det ellers nærliggende at mene, at den redskabsformidlede erkendelse indeholder det paradoks, at vi ikke kan erkende verden uden at kende redskabet, og at vi ikke kan erkende redskabets egenskaber uden at kende verden. Et tilsvarende synspunkt bliver da også ofte fremsat som argumentation for erkendelsens almindelige subjektivitet. Men som det er vist, er vor verden så heldigt indrettet, at paradokset kun optræder i vor refleksion, ikke i den praktiske virkelighed.¹⁸

Vekselvirkningen mellem sten og nødder, mellem knive og træstykker, osv. har en sådan form, at vi ud fra resultaterne induktivt kan erfare og danne begreber om stenenes og nøddernes relative hårdhed, knivenes skarphed og træets sejhed, etc., erfaringer, som ikke ville være tilgængelige for os, hvis vi ikke lod redskaber og objekter interagere.

De afgørende *subjektive og praktiske betingelser* for denne erkendelse er, ud over at vi kan erkende og *sammenligne* reaktionerne mellem redskaber og objekter, at vi kan *identificere* disse. Ikke i den forstand, at vi kan erkende deres parameterværdier, hvilket jo først er et resultat af vor praktiske erkendevirksomhed, men i den forstand at vi genkender objekter og agentia fra gang til gang og ikke forveksler dem indbyrdes. Vi skal blot kunne skelne deres *numeriske identitet* og ikke deres kvalitative identitet, for at udtrykke det filosofisk. Selve det at *fastholde* et redskab fra gang til gang i stedet for ad hoc at benytte et nyt hver gang indeholder altså en kilde til erkendelse af objektive invarianser. Redskabet bliver et sammenligningsgrundlag og dermed medvirkende i en begrebsliggørelse

Ligning (17f) er en *nødvendig og tilstrækkelig betingelse for latent additivitet* af funktionen $f(o, a)$, som er to gange differentiable og strengt monoton i begge reelle variable. Af ligning (17d) kan g og h bestemmes, hvorefter ϕ kan bestemmes af (17). En simpel funktion, som ikke er latent additiv, vil f.eks. være $f(o, a) = oa^2 + ao^2$.

¹⁷ Jeg kan ikke bevise denne påstand. Men jeg har ikke kunnet finde eksempler på love for fysisk interaktion, der ikke er latent additive. Selv en så kompliceret funktion som den, der indgår i loven for relativistisk hastighedsaddition $w = f(u, v) = (u + v)/(1 + uv/c^2)$ er latent additiv med $\phi(v) = g(v) = h(v) = c \operatorname{Artg}(v/c)$.

¹⁸ Et godt eksempel på, at filosofien ikke, som mange tror, kan lade sig nøje med dagligsprogets og snusfornuftens logik, men (via praksis i vid forstand) må forudsætte kendskab til verdens faktiske indretning, som den bl.a. er reflekteret i realvidenskabernes.

I øvrigt vil jeg vare mig for ontologiske spekulationer over, hvorvidt universet og menneskene, eller blot matematikken og fysikken, overhovedet kunne eksistere uden latent additive fysiske lovmæssigheder eller uden relativt stabile legemer m.v. Jeg har altså ikke noget veldefineret sammenligningsgrundlag for mine påstande om "gunstige" eller "heldige" erkendelsesvilkår. Der kan strengt taget kun sammenlignes med den manglende refleksion af disse vilkår i erkendelsesteorien og i psykologien.

af objektet (jf. Mammen, 1983, p. 260).

I praksis kan vi sikre redskabernes identitet og stabilitet ved at opbevare og beskytte dem. Netop dette særlige tilknytningsforhold til redskaber ser ud til at være afgørende for den *særligt menneskelige erkendelse* i modsætning til dyrenes, der også kan være redskabsmæssig, men ikke udviser den samme 'solidaritet' med redskabet som hos mennesker (Mammen, 1985; 1993; 2002; Poulsen, 1985; 1989).

En generalisering til ikke-parametrisk interaktion

Hvis vi for en stund vender tilbage til de objektive betingelser for specifik objektivitet, vil det være nærliggende at forsøge at generalisere begrebet ud over det tilfælde, hvor funktionen f i (17) er mellem reelle variable, men stadig inden for det deterministiske tilfældes rammer. Hvad hvis interaktionen er 'kvalitativ', dvs. mellem instanser, som vi beskriver i forskellige adskilte (diskrete) og evt. ikke-ordnede (nominale) kategorier i stedet for som en kontinuert variation? Eller hvis sammenligningerne ikke er talmæssige (metriske), men blot beror på ordningsrelationer, 'større end', 'mindre end'. Jeg diskuterede i sin tid muligheden med Rasch, og vi havde en korrespondance om emnet, som desværre blev afbrudt af Rasch's sygdom og død.

Det viser sig faktisk, at selvom den eneste mulighed for sammenligning af reaktioner beror på *ækvivalensrelationer*, altså hvorvidt reaktionerne tilhører samme kategori (som specialtilfælde er ens eller ej), er det muligt at formulere en nødvendig og under visse omstændigheder tilstrækkelig betingelse for specifik objektivitet.

Allerede i det kontinuerte, parametriske tilfælde, som vi har diskuteret ovenfor, kan der formuleres en nødvendig betingelse for specifik objektivitet, og dermed latent additivitet, som kun retter sig imod $f(o,a)$ som *relation* og ikke som funktion. Ved indsættelse i (17) ses det, at

$$\begin{aligned}(\phi[f(o_1, a_2)] = \phi[f(o_2, a_1)] \wedge \phi[f(o_1, a_3)] = \phi[f(o_3, a_1)]) & \quad (19) \\ \Rightarrow \phi[f(o_2, a_3)] = \phi[f(o_3, a_2)].\end{aligned}$$

Da ϕ er monoton og dermed én-éntydig, følger heraf, at det for alle sæt $(o_1, o_2, o_3, a_1, a_2, a_3)$ skal gælde, at

$$\begin{aligned}(f(o_1, a_2) = f(o_2, a_1) \wedge f(o_1, a_3) = f(o_3, a_1)) & \quad (20) \\ \Rightarrow f(o_2, a_3) = f(o_3, a_2).\end{aligned}$$

Formentlig er denne nødvendige relationelle betingelse for specifik objektivitet også tilstrækkelig i det kontinuerte tilfælde. Efter forslag fra Benny Karpatschof vil jeg kalde betingelsen i ligning (20) for *kvasi-transitivitet*.

Lad os nu se på det tilfælde, hvor objekter og agentia blot beskrives ved *ordningsrelationer* og ikke parametrisk. Specifik objektivitet vil i

dette tilfælde betyde, at vi i stedet for funktionerne α og β i (16) blot indsætter de ikke-refleksive ordningsrelationer ' $>$ ' og/eller ' $<$ ', at vi altså i stedet for at måle reaktionerne blot sammenligner dem efter deres orden i en eller anden henseende. Et eksempel vil være den fra skolens geologiundervisning kendte ordinale skala for mineralers hårdhed, hvor mineraler ordnes efter, om de kan ridses med et synligt spor af en negl, en kniv, en diamant etc. Uanset, hvad vi ridser med, kan vi være sikker på, at ridser bliver dybest i det blødeste af to mineraler. Og uanset, hvad vi ridser i, vil vi kunne ridse dybere med kniven end med neglen. Altså er der specifik objektivitet som defineret i (16).

I dette tilfælde vil det altid kunne lade sig gøre at tildele en endelig mængde objekter en heltallig *rang*. Og tilsvarende for redskaberne (agentia). Hvis vi lader g og h i (17) være disse rangtal, vil ϕf være summen af rangtallene¹⁹, som hvis vi er heldige, vil inddele reaktionerne i ordnede kategorier.²⁰ Hvis lighedstegnet i (20) tolkes som ækvivalens, altså fællesskab i forhold til disse ordnede kategorier, vil også (20) være opfyldt. Altså er *kvasi-transitivitet*, dvs. tilfredsstillelse af (20), også i dette særlige *ordinale* tilfælde en *nødvendig betingelse for specifik objektivitet*.

Generelt ser det ud til at gælde, at hvis reaktionerne mellem en mængde objekter og agentia lader sig ordne i ækvivalens-kategorier, der tilfredsstiller (20), vil der eksistere en ordinal struktur i interaktionerne, som beskrevet ovenfor, således at *kvasi-transitivitet* altså i dette tilfælde også er en *tilstrækkelig* betingelse for specifik objektivitet. Jeg vil dog undlade argumentationen her.

Endelig skal det tilføjes, at i de probabilistiske tilfælde, som Rasch har beskrevet, er der tale om parametriske modeller for sandsynligheden af diskrete reaktioner, f.eks. svar på spørgsmål eller antal fejl i prøver. Angående den matematiske formalisering af den specifikke objektivitet og det tilknyttede krav om latent additivitet i disse tilfælde henvises f.eks. til Rasch (1960).

Afslutning

Forudsætningen for at overvinde den begrænsning, som empirismen og rationalismen med deres fælles forudsætninger har sat for forståelsen af den menneskelige erkendelse, er ikke blot en almen henvisning til menneskelig praksis eller virksomhed. Dvs. en virksomhed, som producerer et formmæssigt rigere og mere objektivt genspejlende stof for erfaringen

¹⁹ Hvis vi i vort eksempel har tildelt mineralerne rang efter stigende hårdhed og redskaberne efter stigende skarphed, skal en af rangordnerne vendes.

²⁰ Hvis reaktionskategorierne svarende til de forskellige rang-summer er ordnede, dvs. ikke-overlappende, vil det i vort geologiske eksempel være et tegn på, at vort udvalg af mineraler og redskaber er særligt velegnet som grundlag for en ordinal skala, altså en ordinal reference-ramme for mineralers hårdhed. Vi har i dette tilfælde en primitiv heltallig differens-skala for hårdheder.

end den simple og passive sansning. Henvisningen til denne praksis må konkretiseres og eksemplificeres for at kunne frembyde et reelt alternativ til empirismen og rationalismen.

Jeg har ovenfor fremdraget to sådanne eksempler, repræsenteret ved J.J. Gibson og G. Rasch.²¹ Begge har analyseret vore muligheder for erkendelse af kvantitative forhold i vor omverden. Den særlige rolle, som de to teorier tilskriver invarianser i vor interaktion med verden, giver desuden en forståelsesramme for den traditionelle psykofysik, her repræsenteret ved G. T. Fechner og hans langt senere efterfølger og opponent S. S. Stevens.

I Gibsons teori er forudsætningen for vor erkendelse af geometriske relationer i vor omverden, at der er bevægelse i forholdet mellem individ og verden, bevægelse, som producerer invarianser i interaktionen, der igen specificerer konstante relationer i verden. Som hovedtilfælde produceres disse bevægelser ved individernes aktive eksploration, lokomotion og manipulation i forhold til deres omverden.

I Rasch's teori er forudsætningen for erkendelse af objektive relationer mellem objekter, at det erkendende subjekt udfører en aktiv sammenligningsvirksomhed, som indebærer, at objekter og agentia kombineres, at reaktionerne sammenlignes, og frem for alt, at der holdes rede på objekters og redskabers (agentias) numeriske identitet og tages vare på redskabernes stabilitet. Kun inden for rammerne af en sådan identificerende og fastholdende praksis kan sammenligningerne specificere objektive relationer mellem objekterne. Samtidig specificeres objektive relationer mellem redskaberne. Og endelig specificeres selve interaktionens form. *De praktiske forudsætninger for erkendelse af objekter, redskaber og deres lovmæssige interaktion er de samme, og alle indebærende en identificerende og fastholdende praksis i forhold til enkeltting.*

En sådan praksisform er efter alt at dømmes et menneskeligt privilegium, og desuden en forudsætning for vort generelle erkendelsesmæssige og følelsesmæssige forhold til medmennesker, ejendele, kulturens objekter

²¹ Karl Marx har som bekendt foretaget en detaljeret analyse af etableringen og erkendelsen af varers kvantitative værdi i en praktisk interaktion defineret ved produktion og udveksling af varer i et marked. På mange måder er den parallel med de fremdragne eksempler, blot mere kompleks. Forudsætningen for introduktion af en "almen værdiækvivalent" eller et universelt betalingsmiddel er et bestemt mønster i mængden af alle de bilaterale bytteforhold, som matematisk kan beskrives som *transitivitet*. Populært sagt vil det sige, at det skal give samme resultat at bytte direkte som via en vilkårlig tredje vare. Marx' analyse har ved sin logisk-induktive metode ligheder med Maxwell's tidligere omtalte. Interessant i denne forbindelse er i øvrigt Marx's henvisning til Aristoteles' manglende mulighed for at erkende det reelle grundlag for etableringen af værdierne, fordi han kun så på selve markedet og ikke på varenes frembringelse. I slavesamfundet var produktionen af varer så at sige skjult for markedet, fordi arbejdskraftens tidsforbrug som grundlag for værdierne tilhørte en anden verden uden forbindelse til markedet. Først med indførelsen af arbejds løn bliver grundlaget for bytteværdierne synlig ifølge Marx (1973, p. 73-74), og den transitive struktur opståen, som var en gåde for Aristoteles, bliver forståelig.

og naturens. Men det er en anden historie (Mammen, 1986b; 1993).

Litteratur

Engelsted, N. (1989). *Personlighedens almene grundlag I*. Århus: Aarhus Universitetsforlag.

Fechner, G.T. (1860). *Elemente der Psychophysik*. 2 Bd. Leipzig: Breitkopf & Härtel.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J.J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Lewin, K. (1931). The Conflict between Aristotelian and Galilean Modes of Thought in Contemporary Psychology. *Journal of General Psychology*, vol. 5, pp. 141-177. Optrykt i K. Lewin, *A Dynamic Theory of Personality*, pp. 1-42. New York: McGraw-Hill, 1935.

Karpatschof, B. (2007). Psykisk målbarhed. I: B. Karpatschof & B. Katzenelson (red.). *Klassisk og moderne psykologisk teori*. København: Hans Reitzel, s. 534-565.

Mammen, J. (1983). *Den menneskelige sans. Et essay om psykologiens genstandsområde*. København: Dansk psykologisk Forlag. (2. udg. 1989; 3. udg. 1996).

Mammen, J. (1985). Menneskets bevidsthed. I O. Fenger & S. Jørgensen (red.), *Skabelse, udvikling og samfund. En forelæsningsrække. Acta Jutlandica LX, Samfundsvidenskabelig serie 16*, pp.73-81 og 271. Århus: Arkona.

Mammen, J. (1986a). Om matematik og psykologi. Eller noter om matematik, intuition og virkeligheden (bl.a. hundestejler). I I.D. Petersen & A.F. Petersen (red.), *Delhed og helhed. Teoretiske og metodiske studier over komplicerede psykobiologiske fænomener. Festskrift til Iven Reventlow*. 2.6.1986, pp. 25-37. København: Forlaget Politiske Studier.

Mammen, J. (1986b). Erkendelsen som objektrelation. *Psyke & Logos*, vol. 7, nr. 1, pp. 178-202.

Mammen, J. (1993). The Elements of Psychology. I N. Engelsted, M.

- Hedegaard, B. Karpatschof & A. Mortensen (eds.), *The Societal Subject*, pp. 29-44. Århus: Aarhus University Press.
- Mammen, J. (1994). En praktisk-realistisk teori for erkendelse af kvantiteter. I: N. G. Bolvig, H. Brøns, S. Jørgensen & E. Yndgaard (red.). *Perspektiver i samfundsvidenskaben*. Århus: Aarhus Universitetsforlag, s. 165-185 (resumé s. 12).
- Mammen, J. (2002). Mapping the subject: The renewal of scientific psychology. *Bulletin fra Forum for Antropologisk Psykologi* (nu: *Journal of Anthropological Psychology*), No. 11, 77-89.
- Marx, K. (1973). Das Kapital. Bd. I. I Marx/Engels: *Werke*. Bd. 23. Berlin: Dietz Verlag.
- Poulsen, H. (1982). Leontjev, genspejlingsbegrebet og den almene psykologi. *Psyke & Logos*, vol. 3, nr. 1, pp. 161-75.
- Poulsen, H. (1985). Om redskabskonstruktion og menneskelig psyke. I O. Fenger & S. Jørgensen (red.), *Skabelse, udvikling og samfund. En forelæsningsrække. Acta Jutlandica LX. Samfundsvidenskabelig serie 16*, pp. 49-55 og 270. Århus: Arkona.
- Poulsen, H. (1989). Forholdet mellem subjekt og objekt i dyrs og menneskers erkendelse. I S. Brock & P. Petersen (red.), *Dømmekraft. Objektivitet, subjektivitet og videnskab*, pp. 155-162. Århus: Aarhus Universitetsforlag. (Se også Poulsen, H. (1991). The Relationship between Subject and Object in Human and Animal Cognition. I H. Poulsen: *Conations*, pp. 24-30. Århus: Aarhus University Press).
- Poulsen, H. (1994). Om mekanistiske forståelsesformer i psykologien. I: N. G. Bolvig, H. Brøns, S. Jørgensen & E. Yndgaard (red.). *Perspektiver i samfundsvidenskaben*. Århus: Aarhus Universitetsforlag, s. 71-80 (resumé s. 11).
- Rasch, G. (1955). Om populationer, standarder og profiler. 6. nordiske militærpsykologkonference, København, 9.-11. sept. 1954, pp. 47-57.
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. København: Danmarks Pædagogiske Institut. Også Chicago: Chicago University Press, 1980.
- Rasch, G. (1972). Objektivitet i samfundsvidenskaberne. Et metodeproblem. *Nationaløkonomisk Tidsskrift*, bind 110, nr. 3-4, pp. 161-196.

Reventlow, I. (1970). *Studier af komplicerede psykobiologiske fænomener*. København: Munksgaard.

Stevens, S.S. (1960). On the New Psychophysics. *Scandinavian Journal of Psychology*, vol. 1, pp. 27-35.

Stevens, S.S. (1961). To Honor Fechner and Repeal his Law. *Science*, vol. 133, pp. 80-86.

Statistiske niveauer og Structural Equation Modeling i psykologisk forskning og til beskrivelse af behandlingseffekt

Af *Jan Ivanouw*

Indledning

I denne tekst vil jeg fremstille nogle centrale nyere måder at bearbejde data, herunder nogle metoder til at undersøge behandlingseffekt. Det drejer sig om en familie af meget kraftfulde og nyttige metoder, som imidlertid ind til videre ikke har været så meget anvendt i psykologien, og som derfor her introduceres for et dansk publikum.

Fremstillingen består i en gennemgang af en række niveauer af statistiske metoder. Niveauerne er en slags platforme som hver for sig består af en række analysemetoder. Niveauerne udgør også en slags ramme for forskeres og praktikers progressivt øgende forståelse af statistiske metoder¹.

De nederste niveauer udgøres af de for de fleste kendte metoder, dem som hyppigt ses anvendt i psykologien. De øverste niveauer er af en kompleksitetsgrad som er fremmed for mange psykologer. Det er imidlertid ærgerligt fordi psykologer i deres centrale fagområder som regel beskæftiger sig med emner af ganske høj kompleksitet. Det er ikke usædvanligt at høre psykologer afvise kvantitativ forskning ud fra en opfattelse af at man med statistiske metoder alligevel ikke rigtigt får fat i det der

¹Begrebet "statistiske niveauer" skal ikke opfattes som et statistisk fagudtryk. Ej heller indeholder det et postulat om at statistikken historisk har udviklet sig gennem disse niveauer. Der er derimod tale om et forsøg på at karakterisere grupper af statistiske metoder som udgør samlede sæt af analyseredskaber som hver for sig typisk står til rådighed for den enkelte psykolog i vedkommendes arbejde. Niveauerne udgør efter min erfaring med at undervise psykologer og psykologistuderende en udviklingsrække i den enkeltes uddannelse i forståelsen af statistiske metoder.

er vigtigt i psykologien. Denne opfattelse baserer sig imidlertid nok på manglende kendskab til metoderne på de øvre niveauer, som netop kan beskrive og analysere meget komplekse fænomener på en præcis måde. Hvor metoder af denne kompleksitetsgrad tidligere også teknisk var så vanskelige at gennemføre at de var en specialitet for særligt statistisk kyn-dige, findes der i dag PC-programmer der gør det teknisk ukompliceret at anvende dem.

Som det *første statistiske niveau* beskriver jeg metoder til simple ana-lyser af enkelte variable og sammenligninger mellem to og flere variable, den såkaldte deskriptive statistik.

På *andet statistiske niveau* suppleres den deskriptive analyse med statistisk hypotesetestning som en hjælp til at vurdere hvornår de em-piriske resultater blot må opfattes som tilfældig variation uden relation til den problemstilling der undersøges, og hvornår sandsynligheden snarere taler for at der netop kan ses systematiske tendenser som peger på fagligt relevante lovmæssigheder.

Tredie statistiske niveau introducerer statistiske metoder, med regres-sionsanalysen som grundform, der kan anvendes til at undersøge psyko-logiske modeller. På dette niveau søger man at forklare en enkelt variabel (den afhængige variabel, eller effektvariablen) ud fra et system af andre variable (de uafhængige variable, eller prædiktorer).

På *fjerde statistiske niveau* udvides disse modeller til de såkaldte *path-modeller*. Med disse kan man undersøge flere variables indflydelse på flere effektvariable – og således at man kan inddrage relationerne mellem de uafhængige variable i en mere kompleks struktur hvor nogle variable kan mediere andre.

Disse indtil nu omtalte metoder lider af en række mangler som kan overskrides ved anvendelse af de nyere metoder på *femte statistiske niveau*, som har flere forskellige navne. Ofte sammenfattes de under betegnelsen *Structural Equation Modeling* (SEM), på dansk *strukturelle ligninger*. Feltet omtales også med begreber som *covariansstrukturanalyse*, og mere generelt som *modellering med latente variable*. Dele af denne type model-ler benævnes som *konfirmatorisk faktoranalyse* (CFA). De i dansk psyko-logi ret kendte *itemresponsmodeller* (IRT) – herunder *Rasch-modellen* – kan opfattes som en særlig form for CFA med kategoriale data.

Modellerne på dette niveau udmærker sig blandt andet ved at arbej-de med latente begreber på linie med psykologiens begreber. Endvidere ved at det er muligt i disse modeller – på samme måde som i psyko-logisk praksis – at anvende flere forskellige undersøgelsesmetoder til at indkredse et bestemt psykologisk begreb. Endvidere giver disse modeller et kvalificeret skøn over de usikkerheder der indgår i modellen, både med hensyn til målemetoderne og med hensyn til en række af de indgående (latente) variable.

Jeg vil efter denne gennemgang give en introduktion til statistiske metoder på forskellige niveauer til at undersøge behandlingseffekt.

Datatyper og skalaer

Som basis for alle former for statistisk bearbejdelse må informationerne identificeres med hensyn til type: Vi regner med *kategoriale data*, som er *uordnede* (eks. patienter opdelt efter diagnose), *ordnede kategoriale data* (eks. grad af enighed opdelt i “helt uenig”, “lidt uenig”, “lidt enig” og “helt enig”), som ofte kaldes for ordinale data, samt *metriske data* der måles på en ækvidistant skala, hvor det er meningsfuldt at regne med at afstandene er lige store i hele skalaens forløb (f.eks. reaktionstid). Metriske data opdeles af og til efter Stevens i intervallskala- og forholdsskaladata, afhængigt af om der kan findes et absolut nulpunkt. For praktiske statistiske formål er denne skelnen imidlertid ofte ikke nødvendig. Derimod kan det i nogle tilfælde være hensigtsmæssigt at skelne mellem *tælledata* som kun kan være hele tal og *kontinuerte data* som også kan antage de mellemliggende værdier.

Der kan også være særlige egenskaber ved data som kan få betydning ved statistiske opgørelser. I nogle tilfælde har man at gøre med såkaldt *zero-inflated count data*, altså tælledata med en ophobning omkring nul (f.eks. antal angstanfald inden for et bestemt tidsrum, opgjort for et udvalg fra normalbefolkningen). I en række tilfælde har man data hvoraf nogle er komplette og andre ukomplette, nemlig de såkaldte *overlevelsesdata*. Det morbide navn stammer fra den medicinske forskning i forbindelse med opgørelser af levetid f.eks. efter udbrud af en bestemt sygdom. Nogle af patienterne er døde på undersøgelsestidspunktet, mens andre stadig lever, og det karakteristiske ved den dataform er at man jo ikke ved hvor længe de sidste stadig vil leve. Et mindre morbide eksempel på denne datatype er personer i psykoterapi hvoraf nogle har afsluttet terapien mens andre stadig er i behandling.

En særlig datastruktur udgøres også af *hierarkiske data*. Her er der tale om data som naturligt falder i grupper som indbyrdes har større sammenhæng end uden for gruppen, og som derfor ikke tilfredsstillende ofte stillede krav i statistisk databehandling om at data skal være indbyrdes uafhængige. Skoleelever er f.eks. inddelt i skoleklasser, som også er samlet i skoler. Da der er en del fælles for elever i samme klasse – og for klasserne på samme skole – må man regne med en anderledes variation blandt eleverne inden for klassen, end blandt elever på tværs af klasserne. Det samme er tilfældet hvis man undersøger patienter fra forskellige psykiatriske afdelinger. Det er vigtigt at denne datastruktur også findes når man gentager målinger, f.eks. ved undersøgelser af behandlingseffekt. Målingerne fra den samme person på forskellige tidspunkter i behandlingsforløbet har selvfølgelig en indbyrdes sammenhæng som ikke findes på tværs af personerne, og det må der tages hensyn til i databehandlingen.

Når man måler eller tæller data for en variabel, samler man data sammen så der fremkommer et tal på en skala. Et problem kan være at skalaværdierne er bundet til den metode der har været anvendt til

at fremskaffe data. Eksempelvis vil man ved en måling på depressionsspørgeskemaet BDI få en skala der går fra 0 til 63. Hvis man i stedet anvender Zungs Depressionsskema (ZDS), får man en anden skala der går fra 20 til 80 (Beck, 2006). For at anvende disse tests skal man altså kende til deres skala.

Specielt skal man gerne vide hvor i skalaen man typisk ser resultater fra depressive og hvor man ser resultater fra ikke-patienter. F.eks. henviser på BDI resultater i skalaområdet 30-63 til svær depression, mens det tilsvarende område på ZDS er 70-80.

Der findes imidlertid en måde hvor man kan udtrykke resultater på en skala på en måde der er lettere at forstå og hvor det er lettere at sammenligne resultater på tværs af målemetoder. Det drejer sig om *standardisering* af skalaerne, således at gennemsnittet i den undersøgte gruppe² får tildelt værdien 0 og spredningen sættes til 1³. Se figur 2 hvor data på x-aksen er standardiseret udgave af x-aksen på figur 1. I en række statistiske metoder anvendes standardiserede skalaværdier.

1. Statistiske niveau: Deskriptiv statistik

Den simpleste form for kvantitativ dataanalyse er den såkaldt *deskriptive statistik*. Denne består i at sammenfatte og strukturere de informationer der er indsamlet på en måde som gør dem mere overskuelige.

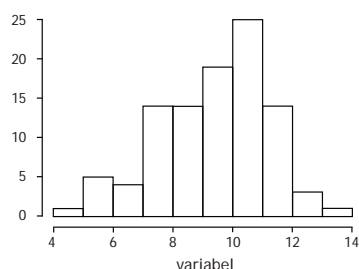


Figure 1. Fremstilling af en enkelt variabel

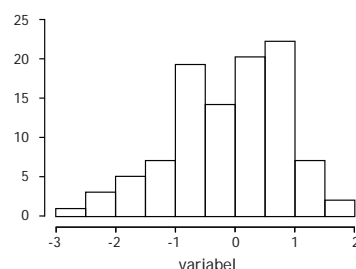


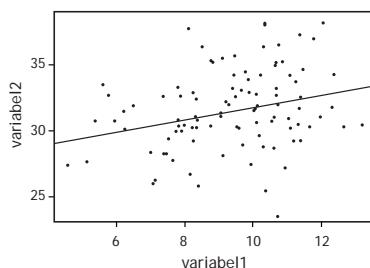
Figure 2. En variabel på standardiseret skala

Den simpleste form er opgørelser inden for en enkelt variabel. Hvis man har uordnede kategoriale data, vil man kunne foretage en simpel optælling af frekvenser, eksempelvis hvor mange patienter der har hver diagnose. Med ordnede kategoriale og med metriske data kan man fremstille data i én variabel i et diagram (histogram) (se figur 1). Fordelingens egenskaber kan sammenfattes ved en række *fordelingsindikatorer*. Den mest

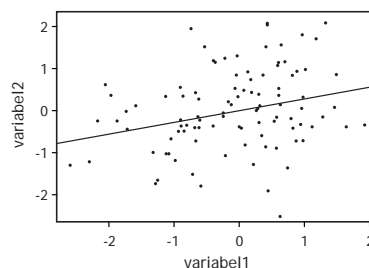
²eller en bestemt referencegruppe.

³Det foregår ved at værdierne på den oprindelige skala fratrækkes gennemsnittet i gruppen, hvorefter resultatet divideres med standardafvigelsen i gruppen.

kendte er centralskønnet, der altså fortæller hvor tyngden, “midten” i data ligger. Dette bør suppleres med et spredningsskøn der beskriver hvor stor variation der er omkring dette centralskøn. Med metriske data vil man typisk bruge gennemsnit som centralskøn og standardafvigelse som spredningsskøn, mens man med ordnede kategoriale data for eksempel kan anvende median og kvartilafvigelse. Fordelinger kan også beskrives med hensyn til om de er symmetriske og hvor spidse de er. Dette fremstilles ved metriske data som en talværdi for skævhed og en for kurtose⁴.



Figur 3. Sammenhængsdiagram med to variable



Figur 4. Samme variable med standardiserede skalaer

Ofte er man imidlertid interesseret i at sammenholde forskellige variable. To uordnede kategoriale variable sammenholdes typisk i en krydstabel. To metriske variable kan afbildes i et sammenhængsdiagram. Sammenhængen mellem sådan to variable kan beskrives som kovarians hvis variablene er i deres oprindelige skala, og som korrelation hvis variablene er standardiserede⁵

En meget stor del af den tidlige psykologiske forskning er foretaget som forsøg på forstå sammenhænge mellem to variable. Af og til tales der direkte om *sammenhæng*, eksempelvis når graden af konflikt i opvæksthjemmet sammenholdes med aggressiv adfærd hos et barn i familien. I andre tilfælde taler man i stedet for om *forskel*, eksempelvis hvis man vil undersøge om der er forskel på fædre og mødre med hensyn til hvor meget de anvender straf i deres børneopdragelse. Hvis man omformulerer denne problemstilling til at der er tale om to variable, forældrenes køn og anvendelsen af straf, kan man også her tale om sammenhæng mellem

⁴ En fordeling der er symmetrisk, har en skævhed på 0, mens en højreskæv - altså en fordeling med en lang hale til højre - har en positiv skævhed, og en venstreskæv har en negativ skævhed. En fordeling der er mere spids end en normalfordeling, har en kurtose der er positiv, mens en normalfordeling har en kurtose på 0, og en mere flad fordeling har en negativ kurtose.

⁵ Når man i et PC-program vil beregne en sammenhæng mellem to variable, får man som regel automatisk en korrelationskoefficient, og variablene bliver dermed automatisk standardiserede uden at man bemærker det, men man skal være klar over at der faktisk er tale om standardiserede variable. Kovarians omdannes til korrelation ved at kovariansen divideres med spredningen fra hver af de to skalaer ganget med hinanden.

variable (det kan beskrives således at når forældrenes køn “varieres” fra far til mor, vil man se en bestemt ændring i straffanvendelsen).

2. Statistiske niveau: Signifikansundersøgelser og effektstørrelser

Den deskriptive forarbejdning resulterer altså i en række statistiske størrelser og opstillinger som har til formål at gøre data mere overskuelige. Krydstabeller, sammenhængsdiagrammer, beregning af gennemsnit, standardafvigelser, kovarianser og korrelationskoefficienter giver fingerpeg om strukturen i datamaterialet.

Imidlertid er der i praktisk forskning så mange forhold der spiller ind, at disse sammenfatninger sjældent er fuldstændigt entydige. Selvom man i et forsøg prøver at tage højde for indflydelsen fra uvedkommende faktorer, vil sådanne altid alligevel spille ind. Ved forsøgsdesignet prøver man især at undgå at uvedkommende faktorer medfører en systematisk skævvridning af resultaterne (bias). Selvom der ikke forekommer en systematisk skævvridning, vil sådanne faktorer dog som regel sløre resultaterne ved at påvirke hver enkelt observation lidt den ene vej eller lidt den anden. Samtidig vil der altid være en vis måleusikkerhed i de anvendte målemetoder (tests). Tilsammen betyder dette at man skal vurdere de statistiske sammenfatninger med hensyn til om resultaterne blot er udtryk for den tilfældige variation der er et resultat af uvedkommende faktorer og måleusikkerhed, eller om variationen er udtryk for de lovmæssigheder man undersøger. Det er i denne sammenhæng man taler om *statistisk signifikans*. Signifikans kan oversættes ved “betydende”. Det spørgsmål man søger at besvare, er derfor: Kan man i de bearbejdede data se tendenser som ud fra en statistisk vurdering overskrider de forventede tilfældige variationer (fra uvedkommende faktorer og fra måleusikkerhed) og dermed bliver statistisk betydende, eller statistisk signifikante?

Løsningen på dette problem er de *statistiske signifikanstests*. Fremgangsmåden ved disse er at foretage nogle yderligere omformninger af de bearbejdede forsøgsresultater (gennemsnit, krydstabeller m.m.) således at man får nogle størrelser som er statistisk kendte i den forstand at man ved hvor hyppigt de forekommer (=at man kender deres fordeling). Man kan eksempelvis omforme en krydstabel til et tal som kaldes χ^2 . Afhængigt af størrelsen af tabellen (antal celler), kender man den statistiske fordeling af χ^2 -størrelsen. Det betyder at man kan vurdere om den konkrete krydstabel hyppigt vil kunne forekomme udelukkende på basis af tilfældigheder, eller om den kun meget sjældent vil kunne fremkomme som tilfældig variation når de to variable ikke har nogen systematisk sammenhæng.

På samme måde kan man omdanne resultater fra to forskellige grupper, udtrykt ved hjælp af de to gennemsnit og de to standardafvigelser,

til en enkelt statistisk størrelse, kaldet t . Man ved hvorledes fordelingen af t vil opføre sig hvis gennemsnittene kun adskiller sig som følge af tilfældigheder, og der ikke er nogen systematisk forskel⁶.

I begge disse tilfælde kan man ud fra en bestemt værdi af χ^2 og af t finde den tilsvarende p -værdi, som er sandsynligheden for at forsøgsresultatet kunne være opstået ved en tilfældighed uden at der var nogen faktisk systematisk sammenhæng i datamaterialet (for så vidt angår de undersøgte variable). Denne sandsynlighed vil man som regel gerne have til at være meget lille, således at hypotesen om at resultatet kun er en tilfældighed kan afkræftes som usandsynlig, og man i stedet må slutte at man faktisk har en sammenhæng mellem variablene.⁷

Der er her nævnt to forskellige statistiske såkaldte *teststørrelser* (χ^2 og t). Der findes en række forskellige statistiske størrelser som beregnes ud fra forskellige typer af data i forskellige forsøgssituationer, men gangen i den statistiske hypoteseprøvning er den samme i alle tilfælde, og man ender i alle tilfælde med en p -værdi.

Konfidensgrænser En anden måde at vurdere usikkerheden i et resultat er at beregne sikkerhedsgrænser omkring de fundne talstørrelser, f.eks. det interval som "den sande værdi" af størrelsen med 95 % sandsynlighed ligger inden for. Sikkerheds- eller konfidensgrænserne findes ved til den fundne værdi at lægge til og trække fra et produkt af to tal. Det ene er karakteristisk for det empiriske fund, nemlig *standardfejlen*, SE . Det andet tal afhænger af hvilken sandsynlighed man ønsker. For at opnå 95 % sikkerhedsgrænser skal man anvende den værdi som anvendes til at afgrænse 2.5 % i en normalfordeling, hvilket altid er tallet 1.96⁸. Formlen for 95 % sikkerhedsgrænserne bliver altså $estimat \pm 1.96 * SE$. *Estimatet* er det tal vi vil finde grænserne for. Hvordan SE beregnes, afhænger af situationen⁹.

Problemer med statistisk hypotesetestning Den psykologiske forskning har trukket meget store veksler på deskriptiv statistik kombineret med statistisk hypoteseprøvning. Der findes imidlertid en række problemer i dette. Det ene er at man ikke ustraffet kan gennemføre en lang række statistiske tests på det samme datamateriale. Jo flere gange man bruger et bestemt sandsynlighedsniveau (5% eller 1%) til at bestemme at noget er "usandsynligt", jo mere sandsynligt bliver det at de "usandsyn-

⁶ Også vurderet i lyset af størrelsen af standardafvigelse.

⁷ Her har man brug for en målestok for hvornår sandsynligheden for et tilfældigt resultat er så lille at man kan afkræfte tilfældighed som forklaring. Der findes forskellige 'signifikansniveauer' hvoraf 5% og 1% er de mest anvendte.

⁸ Hvis man ønsker 99 % sikkerhedsgrænser, er det tilsvarende tal, 2.58.

⁹ Hvis estimatet for eksempel er gennemsnitet i en fordeling for et bestemt sample, findes SE som standardafvigelsen i fordelingen divideret med kvadratroden af antal personer som indgår i samplet.

lige" ting faktisk optræder (sml. jo flere gange man spiller kort, jo højere er sandsynligheden for at man på et eller andet tidspunkt får en bestemt god hånd, selv om den i sig selv er sjælden)¹⁰.

Effektstørrelser Det andet problem er at signifikansundersøgelse kun fortæller om det er sandsynligt at der er en systematisk sammenhæng mellem variable, eller ej, men ikke hvor stor en sådan sammenhæng eventuelt er. Dette er i sig selv et problem. Man kan opdage signifikante sammenhænge som imidlertid ikke har nogen praktisk betydning fordi effektstørrelsen er for lille. Omvendt kan man have situationer hvor der faktisk er en ganske stor effekt, men hvor en statistisk hypotesetestning ikke viser noget signifikant resultat fordi den *statiske power*, det vil sige sandsynligheden for at finde en faktisk lovmæssighed med en statistisk test, er for lav. Ved at indsamle store samples får man en højere sandsynlighed for at den statistiske hypoteseprøvning ender med at vise statistisk signifikans. Det har vist sig at en stor del af forskningen inden for psykologien har lidt af for lav statistisk power på grund af for små samplestørrelser¹¹.

Disse overvejelser har ført til at man i forskningssammenhænge er begyndt ikke kun at interessere sig for statistisk signifikans, men også for *effektstørrelser*. Når man i en forskningsartikel f.eks. sammenligner de gennemsnitlige resultater i to grupper, er det i mange tidsskrifter nu et krav at man ikke kun angiver størrelsen af t og den dertil svarende p -værdi, men at man også angiver et mål for størrelsen af forskellen mellem gennemsnittene der ikke afhænger af samplestørrelsen.

Når det drejer sig om at vurdere størrelsen af gennemsnit, anvendes ofte Cohens d ¹² (Cohen, 1988). Når det drejer sig om sammenhænge mellem to metriske variable, kan korrelationskoefficienten i sig selv opfattes som et mål for styrken af sammenhængen. Også med hensyn til effektstørrelser findes nogle konventioner som kan bruges når man skal vurdere dem¹³.

¹⁰Man kan korigere for dette, f.eks. gennem *Bonferronis korrektion* hvor signifikansniveauet divideres med antal statistiske tests af samme slags (*familywise*) der udføres. For at acceptere f.eks. én test ud af 10 som signifikant på 5%-niveauet skal p -værdien være mindre end $0.05/10 = 0.005$.

¹¹Der er udviklet en statistisk metodik kalder *metaanalyse* som råder bod på de små samples der ofte anvendes i enkelte forskningsundersøgelser. Ved metaanalyse beregner man effektstørrelser for hver undersøgelse inden for et forskningsområde og sammenfatter dette ved en statistisk viderebehandling af disse effektstørrelser (Hunter og Schmidt, 2004).

¹²Differencen mellem de to gennemsnit divideret med en beregnet gennemsnitlig standardafvigelse i de to fordelinger.

¹³Baseret på resultater inden for psykologien generelt anses en d på 0.2 for at være en lille effektstørrelse, en d på 0.50 for en mellemstor, mens 0.80 anses for at være stor. Tilsvarende anses en korrelationskoefficient, r på 0.10 for at være en lille sammenhæng, en r på 0.30 for en mellemstor, mens 0.50 anses for en stor effekt.

Ovenfor har jeg hele tiden taget udgangspunkt i forskning der sammenholder kun to variable. Der findes imidlertid metoder hvor man kan inddrage flere end to variable, f.eks. sammenligne gennemsnit i tre eller flere grupper¹⁴.

Krydstabeller mellem flere end to variable (f.eks. en variabel med ja/nej til et spørgsmål, en anden med køn, en tredje med socialgruppe, en fjerde der beskriver to indsamlingstidspunkter og en femte der opdeler mellem en behandlings- og en kontrolgruppe) kan opstilles i store tabeller, eller som et system af tabeller med to variable. De er imidlertid vanskelige at overskue, og der findes metoder til at foretage en samlet vurdering af sådan et tabelsystem¹⁵.

3. Statistiske niveau: Forklaring af sammenhænge mellem variable, regressionsmodeller

Skal vi imidlertid gå videre og kunne forklare hvorledes en række forskellige (uafhængige) variable kan influere en bestemt variabel (den afhængige), må vi anvende en ny type statistiske metoder. Grundtypen blandt disse er (multivariat) regressionsanalyse.

For enkeltheds skyld ser vi først på regressionsanalyse med kun to variable, en uafhængig og en afhængig. Situationen er den samme som ved et sammenhængsdiagram mellem to variable. I dette tilfælde har vi blot opdelt variablene i de to kategorier, uafhængig, eller prædiktor, og afhængig, eller effekt. Vi forsøger f.eks. at forklare aggressiv adfærd hos børn med klima i hjemmet. Vi operationaliserer aggressiv adfærd ved hjælp af en målemetode, f.eks. optælling af nogle nærmere bestemte typer af konfliktadfærd hos barnet i et bestemt tidsrum i en nærmere bestemt sammenhæng. Dette bliver den afhængige variabel. Vi operationaliserer tilsvarende klimaet i hjemmet til en bestemt målemetode, f.eks. et spørgeskema (dette blive den uafhængige variabel). Resultaterne fremstilles som en punktsværm i et sammenhængsdiagram, og vi lægger den linie der passer bedst med punktsværmen således at summen af et mål for afstanden mellem punkterne og linien bliver mindst mulige (jfr. figur 3 og 4 på side 121). Bl. a for at undgå negative værdier kvadrerer vi afstandene før vi lægger dem sammen. Når vi har fundet den mindst mulige sum af de kvadrerede afstand har vi den bedst passende linie. Metoden hedder derfor også de mindste kvadraters metode (Least Squares).

Kommet så langt, kan vi beskrive linien med en formel

$$y = a + b * x \tag{1}$$

¹⁴Her kan anvendes en række metoder under overskriften *variansanalyse*, ANOVA (Stevens, 1996).

¹⁵kaldet log-lineære modeller (Agresti, 1996).

Dette har den fordel at vi i stedet for alle datapunkterne kan nøjes med at kende de to faste værdier som karakteriserer linien (a og b). Disse kaldes for *parametre* til linien, og den ene (a) kaldes for *interceptet* (det stykke som afskæres på y-aksen) mens den anden (b) beskriver hældningen af linien og kaldes *hældningskoefficienten*, eller *regressionskoefficienten*. Denne beskriver hvor stor ændring der bliver på y-aksen når man forøger x-værdien med 1. Pointen er altså at man i stedet for at kigge på alle punkterne, kan nøjes med at kende denne linie. Derved har man beskrevet lovmæssigheden. I figur 3 på side 121 er interceptet 27.5 og hældningskoefficienten 0.46. Efter standardisering af de to variable (figur 4) er interceptet 0 (nulpunktet er så at sige “flyttet midt ind i punktsværmen” så linien går lige igennem det). Hældningskoefficienten bliver nu 0.27, hvilket samtidig er *korrelationskoefficienten*, *r*.

Imidlertid ligger punkterne i praksis på grund af måleusikkerhed og indflydelse fra uvedkommende faktorer netop rundt omkring linien og ikke direkte på den. Hvis vi måler afstandene fra hvert punkt til linien – målt på y-aksen – og ser på disse positive og negative afvigelser, får vi en fordeling med en enkelt variabel.

Dette kaldes for *residualerne* idet de består i den variation på y-aksen der er “til overs” (residual), altså den variation som ikke “forklares af” x-variablen. Residualerne betegnes ofte med bogstavet *e* for “error”, og regressionsligningen kommer altså til at se således ud:

$$y = a + b * x + e \tag{2}$$

Vi regner med at gennemsnittet af disse residualer er 0 – hvis linien er lagt rigtigt, skal de positive og negative afvigelser nemlig vægte lige meget.

Det interessante er hvor stor *standardafvigelsen* er i fordelingen af residualerne. Hvis standardafvigelsen er stor, betyder det at selvom linien ligger rigtigt, så er der store afstande mellem punkterne og linien, og sammenhængen mellem x- og y-variablen er dermed ikke særligt stærk. Der er i dette tilfælde med andre ord store påvirkninger på y-variablen ud over den som kommer fra x-variablen og/eller variablene er målt på en måde der indebærer stor måleusikkerhed.

Pointen i denne måde at tænke på sammenhænge er dels at man har en måde at sammenfatte al informationen i modellen¹⁶, nemlig i de to parametre, interceptet og hældningen, dels at man gennem variationen (standardafvigelsen) i residualerne har en måde at vurdere hvor god modellen er. Hvis variationen i fordelingen af residualerne er stor, kan det være tegn på at der er en eller flere vigtige faktorer som ikke er inddraget i modellen.

Styrken i denne måde at tænke på ses først rigtigt når der er flere x-variable, altså en række mulige påvirkninger som i fællesskab spiller ind

¹⁶Altså i eksemplet: teorien om at børns aggressivitet har sammenhæng med klimaet i hjemmet.

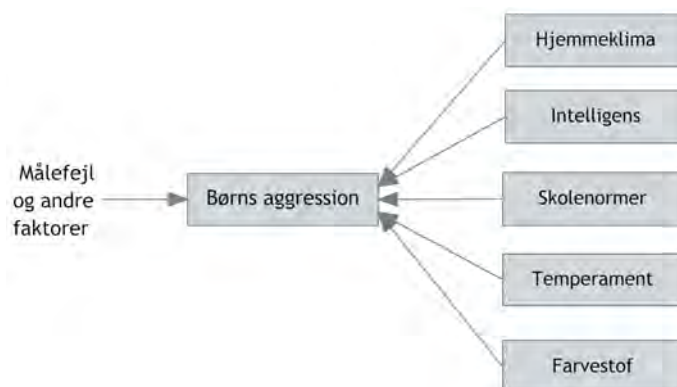


Figure 5. Regressionsmodel for børns aggression

på y-variablen. I eksemplet ovenfor kan man ud over klimaet i hjemmet forestille sig en række andre mulige faktorer der kan spille ind på barnets aggressive adfærd, eksempelvis barnets intelligens, normer i skolen, temperamentsforskelle, indtagelse af farvestoffer i slik m.m. Man kan tænke sig at man afsætter et punkt for hvert barn med hensyn til barnets aggression og alle de nævnte informationer for derefter at lægge en linie der skal passe til punktsværmen. Da der i eksemplet er fem uafhængige variable, kan man ikke forestille sig hvordan det vil se ud i et koordinatsystem (som ville blive 6-dimensionalt), men vi kan alligevel bruge samme metode. Formlen for sammenhængen mellem variablene bliver (med navne på variablene):

$$\begin{aligned}
 \text{Barnsaggression} = & \text{basalaggression} + \text{vægtning}_1 * \text{hjemmeklima} \\
 & + \text{vægtning}_2 * \text{intelligens} + \text{vægtning}_3 * \text{skolenormer} \\
 & + \text{vægtning}_4 * \text{temperament} + \text{vægtning}_5 * \text{farvestof} \\
 & + \text{fejl}
 \end{aligned} \tag{3}$$

eller med de noget kortere matematiske betegnelser:

$$y = a + b_1 * x_1 + b_2 * x_2 + b_3 * x_3 + b_4 * x_4 + b_5 * x_5 + e \tag{4}$$

idet vi kalder den afhængige variabel (børns aggression) for y , og de fem uafhængige variable for x_1, x_2, x_3, x_4 og x_5 . Vi har nu som parametre stadig et intercept (a), og desuden en *hældningskoefficient* for hver x-variabel (b_1, b_2, b_3, b_4 og b_5). Nu kalder man dem dog som regel i stedet for *regressionskoefficienter*. Hver regressionskoefficient fortæller hvor meget ændring der vil komme i y-variablen hvis man forøger den pågældende uafhængige variabel med 1 (forudsat at alle de andre uafhængige variable ikke ændrer sig samtidig). Interceptet (a , som ofte i stedet betegnes som b_0) angiver den værdi y-variablen vil have i den situation hvor

alle x-variablene, og/eller regressionskoefficienterne, til x'erne er 0, og indflydelsen fra dem derfor ophører (derfor her benævnt "basalaggression"). Interceptet bliver en slags udgangspunkt, startpunkt – hvad det nærmere betyder, afhænger af situationen. Vi vil se at det i forbindelse med udviklings- eller behandlingsprocesser ofte kommer til at betegne situationen ved starten af undersøgelsen.

Regressionskoefficienten, "hældningen", angiver hvor meget den bestemte uafhængige variabel influerer på y-variablen. Hvis den er lille, betyder det altså at en forøgelse af den pågældende uafhængige variabel med 1 kun vil give en meget lille forøgelse på y-variablen. Størrelsen af regressionskoefficienten er således et mål for hvor vigtig den pågældende variabel er for det fænomen vi undersøger (den afhængige variabel).

For at spare plads, udelader man ofte gangetegnet (*) i ligninger, så ovenstående regressionsligning vil typisk se sådan ud:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + e \quad (5)$$

Der kan udføres en statistisk signifikanstest for hver enkelt parameter i regressionsanalysen som fortæller om den er signifikant forskellig fra 0. Hvis den ikke er det, er det det samme som at sige at man ikke kan være sikker på at den pågældende uafhængige variabel overhovedet har nogen indflydelse på y-variablen – hvilket ofte er det samme som at sige at den variabel med rimelighed kan udgå af modellen. Man har altså ikke med sikkerhed fundet at denne variabel har nogen betydning for y-variablen. Da hver regressionskoefficient er beregnet set i lyset af de øvrige, kan der i praksis være problemer med at fortolke indflydelsen fra de forskellige uafhængige variable. Specielt hvis disse er indbyrdes afhængige, kan det være svært at få en god entydig model, ligesom der er andre forudsætninger for anvendelse af regressionsanalysen, som vi ikke vil komme ind på her¹⁷.

De værdier som regressionskoefficienterne antager, er imidlertid bundet til de skalaer som variablene er målt på. Det skal forstås således at når en bestemt uafhængig variabel forøges med 1 på *sin egen skala*, så forøges y-variablen med størrelsen af den pågældende uafhængige variabels regressionskoefficient *på y-aksens skala*. Hvis skalaerne for de forskellige uafhængige variable er ret forskelligt indrettet (med hensyn til gennemsnit og standardafvigelse), er det svært at få et godt overblik så man kan sammenligne betydningen af de uafhængige variable med hensyn til hvor stærkt de influerer på y-variablen. Man omregner derfor ofte alle skalaerne til standardiserede skalaer. Dermed bliver alle regressionskoefficienterne udtrykt på samme skala (med gennemsnit

¹⁷Forudsætningerne er: Normalfordelte uafhængige variable og samme varians i residualerne (som måles på y-aksen) i de forskellige skalaafsnit for de uafhængige variable (= at datapunkternes spredning omkring den tilpassede linie skal være den samme for hele linien, og f.eks. ikke være større for højere værdier af de uafhængige variable).

0 og standardafvigelse 1) og de kan derfor sammenlignes med hensyn til deres indflydelse på y-variablen. De standardiserede regressionskoefficienter betegnes ofte med det græske bogstav, β , og koefficienterne bliver derfor $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ og β_5

Afhængigt af arten af måleskala for y-variablen anvendes forskellige variationer af regressionsmodeller. Ved en kategorial y-variabel med to muligheder (patient vs. ikke-patient) kan man anvende en *logistisk regressionsanalyse*, ved tælledata *Poissonregression* osv.

Man skal altså bruge forskellige typer af regressionsmodeller afhængigt af hvilken type data vi har som y-variabel. Dette betyder imidlertid mindre for de uafhængige variable. Man kan godt i en almindelig regressionsanalyse anvende kategoriale uafhængige variable. En variabel med to værdier (ja/nej) kan kodes som 0 og 1, og indgå som uafhængig variabel. En kategorial variabel med tre muligheder (skizofren, depressiv, ikke-patient) kan omkodes til to *dummy-variable* hver med værdien 0 og 1. Den første kunne være skizofren (ja/nej), den anden depressiv (ja/nej), og hvis begge er nej, har vi den tredje mulighed (ikke-patient).

Den regressionsanalysemetode der er gennemgået, forudsætter *lineær sammenhæng* mellem variablene. Det betyder at en forøgelse på 1 i en uafhængig variabel giver en bestemt forøgelse i den afhængige variabel, og at dette er det samme lige meget hvor på skalaen vi befinder os. I nogle tilfælde er dette imidlertid ikke rigtigt. Man kunne f.eks. tænke sig at en forøgelse i konfliktniveauet i hjemmet ikke har nogen særlig betydning for barnets aggressivitet når vi taler om meget lave konfliktniveauer, mens en forøgelse i et hjem der i forvejen er præget af konflikt, vil have større effekt på barnet. Dette kunne afbildes som en kurvet linie, og man vil tale om en ikke-lineær sammenhæng. Der findes metoder til håndtere sådanne sammenhænge, enten ved at ændre på skalaerne så man får en lineær sammenhæng mens man gennemfører analysen¹⁸ - og så ændre tilbage når man er færdig - eller også ved at bruge en anden type ligning med flere parametre som kan beskrive kurvede forløb.

Begrænsninger i regressionsmodellen Man skal bemærke at denne klasse af modeller beskriver relationen mellem en enkelt (afhængig) variabel og en eller flere uafhængige variable. Man kan altså ikke samtidigt operere med flere afhængige variable. Man skal endvidere bemærke at mens man ved målingen af y-variablen kan overveje hvorledes måleusikkerhed spiller ind (via residualerne i modellen), forudsættes alle de uafhængige variable i princippet målt helt præcist. Der er ikke nogen parameter for svingninger i hver af de uafhængige variable som følge af måleusikkerhed. Det indgår endvidere at hver uafhængig variabel direkte tænkes at påvirke y-variablen. Det forudsætter at der ikke er nogen forbindelse

¹⁸se Singer og Willett (2003, s. 211) for en oversigt over forskellige måder at transforme kurvede linier så de bliver lineære.

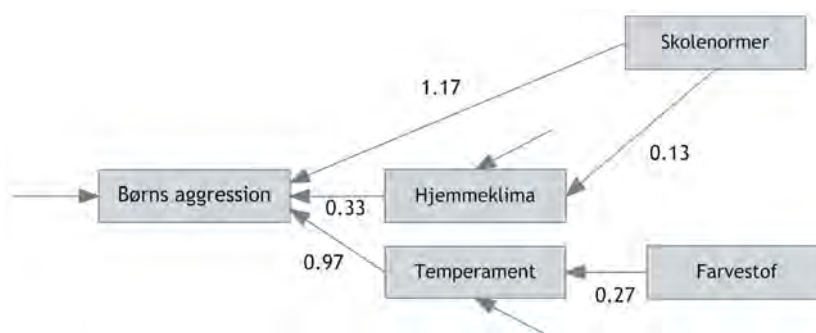
(korrelation) mellem de uafhængige variable, og der regnes ikke med indirekte, medierede, påvirkninger mellem en uafhængig og en afhængig variabel.

Dette vil imidlertid ofte være urealistisk, psykologisk set. Man vil have brug for statistiske modeller hvor man kan medtænke sammenhænge mellem variable, herunder medierende variable. I eksemplet kunne man f.eks. tænke sig at indflydelsen på barnets aggressivitet fra normerne i skolen modificeres af klimaet i hjemmet således at skolens indflydelse er stærk i hjem med højt konfliktniveau, men ikke særlig stærk i hjem med lavt konfliktniveau, at virkningen af farvestoffer modificeres af barnets temperament osv.

4. Statistiske niveau: Pathanalyse

Man kan udvide regressionsanalyseprincippet til at opstille en model med medierende variable.

Et sådant system kaldes en *pathmodel*, og det er muligt at indregne et helt system af variable der inddrager forholdet mellem de enkelte variable. Man har altså her en metode som ligger tættere på den måde vi normalt tænker på i psykologien. Vi arbejder jo som regel i psykologisk praksis (og forskning) med at der er en række forskellige forhold som har en række virkninger i et mere eller mindre kompliceret samspil. Ud over at beskrive hvilke sammenhænge der på denne måde er mellem en række variable, kan pathmodellen vise hvor stærkt de forskellige forhold influerer på hinanden. Hvis vi i vores eksempel for enkelthedsens skyld forudsætter at



Figur 6. Pathmodel for børns aggression

intelligens viser sig ikke at have nogen indflydelse på børns aggressive adfærd, og at indflydelsen fra skolenormer og farvestoffer er medieret af andre variable, kan vi eksempelvis få modellen i figur 6.

Ved estimation af modellen ud fra vores empiriske data – og ved hjælp af en estimationemetode som f.eks. en variant af maximum likelihood – finder vi *pathkoefficienterne*. I modellen skal disse pathkoeffi-

cienter forstås på samme måde som regressionskoefficienter i regressionsmodellen, altså som en beskrivelse af hvor store virkninger en bestemt variabel (eller system af variable) kan have på en bestemt anden variabel. Mere teknisk viser pathkoefficienterne hvor stor forandring der vil komme i den variabel pilen peger på hvis man øger (eller formindsker) variabelen med 1 (samtidig med at andre variable holdes konstante). Pathkoefficienter kan angives både i oprindelig skala og med standardiserede værdier. Pilene der ikke starter i en boks, repræsenterer målefejl, samt indflydelse fra faktorer der ikke er med i modellen¹⁹.

Når man i en pathmodel ser på en kæde af påvirkende variable, kan man finde påvirkningen fra den første til den sidste ved at gange de mellemliggende pathkoefficienter med hinanden. I eksemplet kan man med standardiserede pathkoefficienter beregne at skolens normer både influerer direkte og indirekte modificeret af klimaet i hjemmet. Pathkoefficienten for den direkte vej er 1.17, mens den for den indirekte vej findes ved at gange de to koefficienter 0.13 og 0.33, hvilket samlet viser at den indirekte påvirkning i eksemplet kun er meget lille ($0.13 * 0.33 = 0.04$). Når man ser på en bestemt variabel og vil beregne den samlede effekt af de forskellige påvirkninger, skal man summere disse. I eksemplet er den samlede påvirkning fra skolens normer fra både den direkte og den indirekte vej summen af koefficienterne: $1.17 + 0.04 = 1.21$.

Som man kan se, er det muligt at udtrykke ganske komplekse sammenhænge ved hjælp af pathmodeller. Imidlertid forudsættes det stadig at hvert enkelt begreb skal "operationaliseres" til at måles med én enkelt metode og at målefejl og indflydelse fra andre faktorer tilknyttes de samlede operationaliserede mål i modellen, men ikke uddifferentieres f.eks. på de enkelte items i en målemetode. Disse begrænsninger kan ophæves ved at anvende latente variable.

5. Statistiske niveau: Målingsmodeller og latente variable

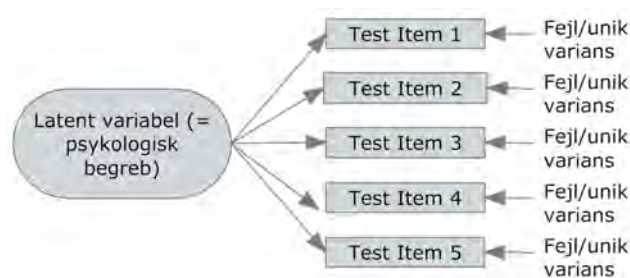
En *latent variabel* er en variabel som ikke observeres, altså som ikke måles direkte. I psykologien anvendes latente variable i helt overvældende grad, eksempelvis til at beskrive personlighedstræk, intelligens og holdninger²⁰. Det er i psykologien som regel erkendt at begreberne kan måles på forskellige måder, og at hver metode er forbundet med en række fejlkilder, som dels består i egentlig måleusikkerhed, dels i at andre psykologiske forhold

¹⁹ pilene der peger på de variable der står "midt i" modellen repræsenterer fejl og indflydelse der ikke er med i modellen, og kaldes ofte for *disturbances*.

²⁰ Et begreb som "reaktionstid", der anvendes i psykologien, kan måles ret direkte. Ganske vist er tid også et begreb, og det skal måles via en eller anden form for ur, men tidsmåling er veludarbejdet og veldokumenteret, og de mulige fejlkilder er velkendte. De fleste af psykologiens begreber har derimod ikke på samme måde etablerede og veludarbejdede målemetoder.

let indvirker på måleresultaterne²¹. Den almindelige løsning på problemet med at måle en psykologiske egenskab som nødvendigvis er uobserverbar, er at *operationalisere* begrebet, hvilket betyder at vælge en målemetode som herefter “står for” begrebet. I operationaliseringen af begreber indgår målefejl og andre indflydelser imidlertid ikke. Arbejdet med at afklare målemetodens funktionsmåde foretages typisk i selvstændige undersøgelser, typisk under udviklingen af målemetoden (testinstrumentet). Når man i forbindelse med et psykologiske forskningsprojekt vælger det instrument som skal måle de egenskaber man er interesserede i, vælger man selvfølgelig instrumenter med de bedst mulige psykometriske egenskaber, men når forskningsresultaterne bearbejdes, behandles målingerne som om de var fejlfrie udtryk for de egenskaber der skal belyses. Denne fremgangsmåde er blevet kritiseret, specielt fordi der findes bedre fremgangsmåder (Borsboom, 2006), som gennemgås nedenfor.

Ved også statistisk at arbejde med *latente variable* kan man imidlertid adskille de aspekter af måleresultaterne som direkte afspejler de undersøgte begreber, fra de aspekter der er udtryk for fejl og anden indflydelse på målingerne.



Figur 7. Målemodel med latent variabel

Latente variable indføres altså dels fordi man derved kan inkludere måleusikkerhed og indflydelse fra ukendte faktorer i sin model, dels fordi man har mulighed for at inkludere mere forskelligartet information i målingen. Man vil f.eks. på samme tid kunne anvende resultater fra en selv vurderingstest og fra observationer. Om det fungerer, er i hvert enkelt tilfælde et empirisk spørgsmål som kan undersøges ved afprøvning af målemodellen.

²¹Det kan dreje sig om andre psykologiske tendenser som overlapper det begreb man interesserer sig for (eksempelvis er dominans og aggression adskilte psykologiske begreber, men en målemetode som måler det ene, vil let blive påvirket af det andet), det kan dreje sig om situationsafhængige forhold i målesituationen (træthed, humør), og målemetoden kan også influeres af forhold vedrørende selve metoden (reaktion på at skulle svare på et spørgeskema eller at skulle løse en række opgaver), eller af personens reaktioner på undersøgelsen (vil gerne præsentere sig som velfungerende, eller måske det modsatte).

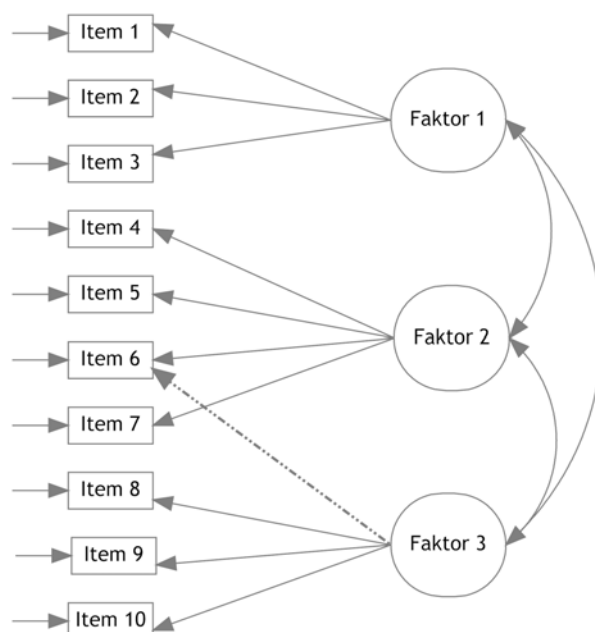
Modeller med latente variable kan have to dele, en målingsmodel, og (eventuelt) en strukturmodel. *Målemodellen* beskriver relationen mellem et begreb og de måder dette måles på. *Strukturmodellen* beskriver relationen mellem de latente variable.

Konfirmatorisk faktoranalyse (CFA)

Konfirmatorisk faktoranalyse anvendes typisk som *målemodel* i en statistisk model med latente variable. Det er i denne model muligt at anvende en række forskellige mål - også kaldet indikatorer - for et latent begreb. Det simpleste er at tænke sig en test med en række spørgsmål hvor disse spørgsmål udgør indikatorerne som tilsammen måler det pågældende begreb. Det er imidlertid muligt at anvende flere forskellige tests og andre assessmentmetoder i kombination for at måle et begreb. I sin typiske form anses de latente begreber for at være de uafhængige, mens observationerne er de afhængige variable, man siger at indikatorerne er *reflektive*. De symboliseres som regel med bogstavet *y*. Dette kan forekomme omvendt i forhold til hvad man normalt tænker sig. Imidlertid giver det god mening hvis man forestiller sig at der eksisterer en tendens, f.eks. et depressivt træk, hos personen, og dette træk bestemmer hvorledes personen besvarer de enkelte spørgsmål i en test. Jo mere deprimeret personen er, jo mere tilbøjelig vil han være til at svare ja til spørgsmål der vedrører depressive symptomer²². Den måde personen besvarer spørgsmålene, vil dog som regel også være påvirket af andre forhold end det latente træk, f.eks. andre uvedkommende træk, personens indstilling til testen (bagatelliserende versus overdrivende, f.eks.), samt tilfældig målevariation. Dette afbildes, som tidligere nævnt, med en pil uden "afsender" som ender på indikatoren og kaldes *fejlvariation*, eller *unik variation*. Målingsmodellen udformes ved hjælp af en beregningsmetode (estimationsmetode)²³ således at der fremkommer nogle koefficienter (her kaldet *loadings*) som beskriver hvor stærkt det latente træk influerer på de enkelte indikatorer, det vil sige itemscores, f.eks. svar på spørgsmål. Den bedste målingsmodel fås selvfølgelig med høje koefficienter og samtidigt lave andre påvirkninger (unik varians). Når hver indikator kan forklares som resultat af en indflydelse fra en latent variabel af en eller anden styrke og en resterende indflydelse

²²Der findes imidlertid situationer hvor det faktisk må være indikatorerne der bestemmer den latente variabel. Ofte nævnes som et eksempel socio-økonomisk status. Det er mere rimeligt at beskrive det således at *indikatorerne* løn, uddannelse, jobtype bestemmer den samlede socio-økonomiske status end at tænke sig at løn, m.m blot er et udtryk for en uafhængigt eksisterende socio-økonomisk status. Indikatorer af denne art kaldes *formative* i modsætning til de hidtil beskrevne *reflektive* indikatorer, også kaldet *effektindikatorer* (Brown, 2006).

²³Der er udviklet en række estimationsmetoder til forskellige situationer og forskellige datatyper. Nogle af disse bygger på de mindste kvadraters metode (least squares) og andre på det såkaldte maximum likelihood-princip. Detaljerne i dette er statistisk specialviden (se. f.eks. Bollen, 1989).



Figur 7. Konfirmatorisk faktoranalyse (CFA)

fra fejl og andre faktorer, så er vi tilbage i tankegangen fra regressionsanalysen. Indikatoren betegnes med y . Nu er det ganske vist ikke en observeret variabel der påvirker den afhængige variabel, men en latent variabel, som vi typisk kalder for eta (η). Til regressionskoefficienterne svarer nu loadings, som betegnes med det græske bogstav lambda (λ), mens den resterende indflydelse betegnes med epsilon (ϵ). λ_0 repræsenterer interceptet, eller indikatorens skalaværdi når den latente variabel er 0. Ligningen for hvad der influerer besvarelsen af et bestemt item (indikator) bliver altså (sammenlign med ligning (2)):

$$y = \lambda_0 + \lambda_1 \eta_1 + \epsilon \quad (6)$$

Der bliver sådan en ligning for hver af indikatorerne, og den samlede målingsmodel beregnes ud fra systemet af disse ligninger.

Den beskrevne målingsmodel kaldes også for konfirmatorisk faktoranalyse (CFA). Til forskel fra den mere almindeligt kendte eksploratoriske faktoranalyse (EFA), så tænker man sig typisk at hver indikator kun måler én latent variabel og at de latente variable kun påvirker deres 'egne' indikatorer, at påvirkningen (loading) på andre indikatorer således er 0. På figur 7 er med en stiplede linie angivet at et item (nr 6) loader på to forskellige latente variable. Det er en situation man for det meste søger at undgå, men som accepteres i nogle tilfælde.

Den samlede model vurderes med såkaldte globale *fit indexes* hvoraf der findes en række forskellige som alle på en eller anden måde sammenligner de faktiske resultater med dem som man med modellen vil kunne regne sig frem til. Det bedste er selvfølgelig hvis modellen kan bruges til at regne sig til de værdier som faktisk er observeret i undersøgelsen, men der vil altid være en forskel. Det er størrelsen af denne forskel som vurderes ved fit-indeksene²⁴.

Ud over de globale fit-indeks er undersøges også om de enkelte koefficienter, herunder loadings, (pilene i modellen) er store nok, og om den resterende varians (residualvariens repræsenteret ved pilene uden afsender) er tilstrækkelig lav. Styrken af tendenserne i modellen afbildes ved de koefficienter der svarer til hver pil. Modellen vurderes også på om de unikke varianser (pile uden afsender) er uafhængige af hinanden eller om der er korrelationer mellem nogle af dem. Korrelationer kan godt tillades, men der bør være en reel forklaring på korrelationerne, som i princippet bør kendes på forhånd. Et eksempel er *metodevariens*, dvs. at nogle observationer deler metode (selvrating, versus observation fra personale, eller samme type af spørgsmål i en test) og derfor har en vis indbyrdes korrelation²⁵.

Der er ikke nogen principiel begrænsning for hvor kompleks en model kan gøres. Imidlertid tilføjer hver ny faktor eller indflydelse i modellen *parametre* der skal bestemmes. Hvor mange parametre det er muligt at bestemme, afhænger primært af hvor mange forskellige (relevante) informationer, f.eks. antal items i en test, der indgår i modellen. Størrelsen af samplet har især betydning for hvor præcist det er muligt at bestemme de forskellige koefficienter (med mindre samples øges standardfejlen, *SE*).

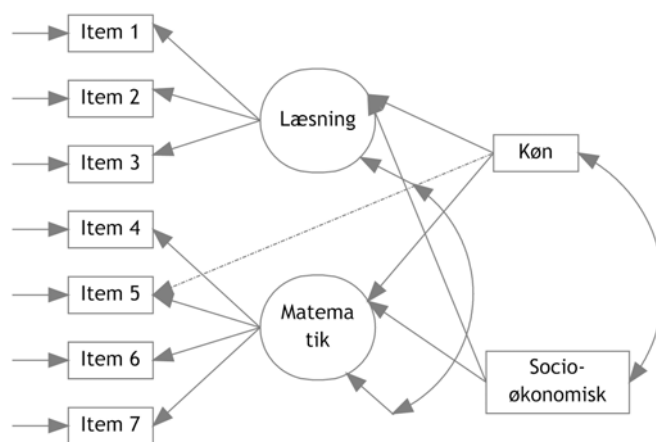
Når man tilføjer flere faktorer og/eller flere sammenhænge mellem faktorer i modellen, øger man antal parametre, og modellen bliver mere kompliceret. Der bør være teoretiske grunde til at gøre dette. Hvis man blot gør det for at få modellen til at passe bedre, risikerer man at den blot passer bedre til det aktuelle sample, og at man ikke vil kunne bekræfte modellen på andre samples²⁶.

En vigtig pointe ved en målingsmodel er at den gerne skal være identisk for forskellige grupper af personer, f.eks. mænd og kvinder, og personer med forskellig socio-økonomisk status (se fig 8). For at teste dette

²⁴De fleste fit-indeks baserer sig på χ^2 -testen. Lige som ved undersøgelse af en sammenhængstabel beregnes χ^2 som et mål for forskellen mellem den observerede og den teoretiske fordeling. I modsætning til hvad der er normalt ved sammenhængstabeller, "håber" man ved vurdering af en model at fit-indekset *ikke* vil vise en signifikant forskel. Man håber nemlig selvfølgelig af forudsigelsen, i form af den model man har opstillet, ikke afviger signifikant fra de observerede resultater.

²⁵Metodevariens er oprindeligt beskrevet af Campbell og Fiske (1959) som også angiver en metode til at vurdere denne. Denne hedder multitræk-multimetode-matricen, og bygger på visuel inspektion af en kompleks korrelationstabel. Ved at anvende korreleret unik varians kan man i CFA vurdere metodevariens på en mere præcis måde.

²⁶Dette kaldes *overfitting* af modellen.



Figur 8. CFA-model med baggrundsvARIABLE og DIF

kan man foretage en CFA-analyse af modellen i figur 8, eller man kan undersøge modellen for hver af disse grupper for sig.

Det er endvidere vigtigt at indflydelse fra baggrundsvARIABLE, eksempelvis køn, foregår via de latente variable, og ikke også ved direkte påvirkning af enkelte indikatorer (jfr. den punkterede linie fra køn til item 5 i figur 8), fordi dette demonstrerer at det pågældende item opfører sig forskelligt over for forskellige persongrupper²⁷.

I den mest almindelige udgave af CFA undersøges variationen, dvs. hvad der indvirker på hvad, hvad der får hvad til at variere. Derimod undersøger man ikke niveauet for skalaerne, f.eks. gennemsnit på de forskellige variable. Der er dog nogle situationer hvor gennemsnitsforskelle netop er pointen. Det er typisk situationer hvor man vil undersøge om der er forskel på persongrupperns niveau på en variabel. Det beskrives som at man inkluderer en *mean structure* i analysen.

IRT-modellen i CFA Itemresponsemodeller (IRT) kan opfattes som indeholdt i CFA. Når man anvender kategoriale indikatorer, kan en loading (f.eks. λ_1) i CFA-modellen opfattes som *diskriminationsparameter* i en 2-parameter IRT-model, mens interceptet (λ_0) for hvert item i CFA-modellen kan opfattes som *sværhedsparameter* i IRT-modellen. Man får her brug for at estimere mean structure for at få et mål for de enkelte intercepter (gennemsnit) for indikatorerne. Man kan opstille en 1-parameter eller Rasch-model ved at indlægge den begrænsning i CFA-modellen at loadings for alle indikatorerne på den samme faktor skal være ens (og teste om denne model passer).

²⁷der er i så fald såkaldt *differentiel itemfunktion*, DIF.

Eksploratorisk faktoranalyse (EFA)

I eksploratorisk faktoranalyse starter man med det udgangspunkt at alle indikatorer (items) har sammenhæng med alle faktorer – omend formentlig i forskellig styrkegrad. Ligningen for hver indikator bliver derfor som udgangspunkt, alt efter hvor mange latente variable man regner med:

$$y = \lambda_0 + \lambda_1\eta_1 + \lambda_2\eta_2 + \lambda_3\eta_3 + \lambda_4\eta_4 \dots + \epsilon \quad (7)$$

Formålet med EFA er dels af afklare hvor mange latente variable, eller *faktorer*, der med rimelighed skal anvendes for at forklare systemet af indikatorer²⁸, dels at afklare hvilke indikatorer der har sammenhæng med hvilke af faktorerne. Man håber at kunne udelade en række af led'ene i ligningen fordi loadingen (λ) fra den pågældende faktor (η) er så lav, at den i praksis kan regnes for at være 0.

For at opnå at så mange led som muligt forsvinder, er det almindeligt at omregne hele modellen. Idealet er at hver faktor kun kommer til at have høje loadings på en enkelt faktor. Denne omregning kaldes en *rotation*, fordi man når man afbilder to faktorer mod hinanden²⁹ kan vise omregningen ved at rotere koordinatsystemet så punkterne kommer til at ligge tæt på akserne. Der findes forskellige metoder til denne omregning. I nogle tilfælde foretages den så faktorerne ikke har nogen korrelation med hinanden, i andre tillades dette³⁰ Det er mest almindeligt at anvende vinkelrette akser, hvilket imidlertid kan undre fordi psykologiske begreber i mange tilfælde har en vis overlapning.

Rent praktisk tager EFA udgangspunkt i en komplet oversigt (matrix) over korrelationerne³¹ mellem alle indikatorerne (f.eks. testitems). Der findes en række forskellige metoder til at udtrække informationer om hvilke grupper af indikatorer der har høje indbyrdes korrelationer. Ved egentlig faktoranalyse regnes der med at faktorerne repræsenterer latente variable, som man giver (ofte lidt tentative) navne. Ved en af udtrækningsmetoderne, Principal Components, foretager man imidlertid udelukkende en datareduktion uden at der impliceres nogen latente variable³².

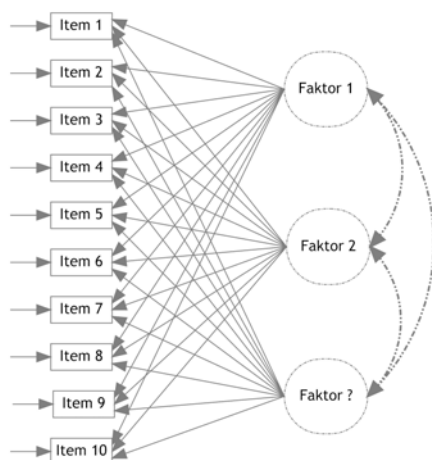
²⁸Her findes en række kriterier, eksempelvis bevare faktorer med eigenvalue større end 1, anvende scree plots og anvende Horns parallelanalyse (Stevens, 1996; Zwick og Velicer, 1986).

²⁹Man afbilder loadings fra hver enkelt item (indikator) på hver af de to faktorer som et punkt. Der bliver altså et punkt for hvert item, eller indikator, og de samler sig typisk i nogle forskellige grupper. Ved rotation forsøger man at få faktorernes akser til at ligge tæt på punktsværmen.

³⁰grafisk kan det sidste afbildes ved at man for bedre at tilpasse akserne til punkterne tillader akserne at stå skråt i stedet for vinkelret på hinanden. Den vinkelrette rotation kaldes *orthogonal*, den skrå kaldes *oblique*.

³¹Det kan også være kovarianserne.

³²Borsboom (2006) gør opmærksom på at Principal Componentsmetoden i nogle tilfælde bliver anvendt *som om* der impliceres latente variable, men at disse efterfølgende, pga. den fejlagtige anvendelse af PC, viser sig ikke at kunne bekræftes ved



Figur 9. Eksplorativ faktoranalyse (EFA)

Det eksploratoriske ved EFA-modellen understreges netop af at der ikke er givet nogen struktur på forhånd, men at man gennem analysen forsøger at få et fingerpeg om strukturen.

Dette indebærer også at man ikke kan beregne noget modelfit – der er jo ikke nogen bestemt model at vurdere. Man har heller ikke mulighed for at beregne sikkerhedsintervaller for hvor meget de forskellige items loader på hver faktor.

I de situationer hvor man i forvejen har en teori om hvilke faktorer der er tale om, og hvilke indikatorer der har sammenhæng med hvilken faktor, bør man inkludere denne forhåndshypotese i den statistiske undersøgelse og opstille en CFA, hvorved hypotesen testes.

Som svar på problemet med “misbrug” af CFA til gentagne modelafprøvninger og -tilpasninger er udviklet en særlig form for EFA hvor det er muligt at estimere loadings fra alle indikatorer på alle faktorer (latente variable), men hvor en på forhånd opstillet model også specificeres, og det både er muligt at foretage rotation af faktorløsningen, som i EFA, og at beregne standardfejl (*SE*) og vurdere modelfit af den på forhånd specificerede model, som i CFA. Metoden kaldes *Exploratory Structural Equation Modeling* og er beskrevet af Asparouhov og Muthén (2008).

Struktur af latente variable (SEM)

Den anden del af en model med latente variable kan være en *strukturdel* hvis man nemlig - lige som i en path-model - tænker sig at latente variable påvirker andre latente variable.

en CFA-analyse. Dette anfører han blandt andet som en alvorlig kritik mod big-five-modellen i personlighedspsykologien.

I den tidligere opstillede pathmodel over børns aggression var det nødvendigt at vælge en operationalisering af hvert begreb således at det kun måles af en enkelt målemetode. Man skulle for eksempel vælge om man ville måle barnets aggression ud fra observationer i skolen, observationer i hjemmet, observation under legekontakt, observation i undervisningssituationen, eller ud fra vurderinger fra enten lærer eller forælder.

Dette er ikke nødvendigt med en SEM-model. Barnets aggression kan opfattes som en latent variabel som måles med forskellige indikatorer. Alle de nævnte målemetoder vil potentielt kunne bruges til samlet at måle den latente variabel. Relationen mellem indikatorerne og den latente variabel udtrykkes i målingsmodellen som kan undersøges med CFA.

Et udviklingsarbejde kan bestå i at finde frem til den bedste kombination af indikatorer. Ved at anvende flere indikatorer (målemetoder) får man mulighed for en bedre tilnærmelse til det begreb man er ude efter at undersøge. Det bliver på den måde muligt at inddrage de forskellige aspekter som definerer begrebet, det som også kaldes *det nomologiske netværk*. Resultaterne bliver ikke så afhængige af lige netop hvilket mål man har anvendt. Når man "indkredser" den latente variabel med flere slags målemetoder får man endvidere mulighed for at vurdere hvor stor måleusikkerhed³³ der er impliceret i de forskellige metoder. Det er muligt at trække noget af denne måleusikkerhed ud som særlige parametre således at man får et bedre mål for hver af de latente variable. Det giver et bedre mål for sammenhænge med andre latente variable.

På samme måde kan man lade de forskellige influerende faktorer (de uafhængige variable i regressions- og pathmodellerne) måle med flere forskellige metoder (indikatorer) og opnå de samme fordele for disse variable. Man kan f.eks. måle klimaet i familien med flere typer af spørgeskemaer, og man kan supplere med observation af interaktion i familien.

I modsætning til pathmodellen vil koefficienterne der beskriver sammenhængen mellem de latente variable (i nogen grad) være befriet for måleusikkerhed og indflydelse fra ukendte faktorer. Det betyder at man vil få et mere rigtigt billede af hvor stærk påvirkning forskellige faktorer udøver på hinanden i en psykologisk model. Det er almindeligt i empirisk forskning at finde korrelationer mellem psykologiske faktorer på omkring .30. Disse korrelationer er imidlertid påvirket af attenuation på baggrund af måleusikkerhed, og de viser derfor ofte en lavere sammenhæng end der faktisk må tænkes at være.

En fordel ved anvendelse af SEM-modellen er altså at man så at sige 'trækker målefejl ud af modellen' i deres egne parametre, og at man på en måde renser målene for sammenhæng mellem de latente variable og på den måde kan korrigere for attenuation.

³³ Måleusikkerhed skal forstås som både egentlig måleusikkerhed og indflydelse på en indikator fra andet end den latente variabel, f.eks. kan forældrens vurdering af barnets aggressive adfærd også være påvirket af hvordan vedkommende gerne vil fremstille familien over for undersøgeren.

Latent klasseanalyse Indtil nu har det været underforstået at latente variable er kontinuerte. Imidlertid kan man også arbejde med latente kategoriale variable, de såkaldt *latente klasser*. Situationen er den at man ud fra en række målinger gerne vil opdele i klasser, eller typer, der er indbyrdes forskellige. For eksempel vil man undersøge om det er meningsfuldt at afgrænse grupper af patienter fra hinanden og opstille nye, ikke på forhånd kendte diagnostiske grupper eller undergrupper. Et andet anvendelsesområde er typer af forandringsprocesser under behandlingsforløb (se senere). Sådanne modeller blev først indført med den restriktion at der ikke må være variation inden for hver klasse på de variable der definerer klassen. Al variation skulle foregå *mellem* klasserne. Imidlertid er der i de seneste år udviklet metoden *mixed models* der kombinerer latente klasser og latente kontinuerte variable (Muthén, 2004).

Et eksempel på anvendelse af disse modeller er undersøgelse af om en model med afgrænsede diagnoser, eller en model med dimensionel beskrivelse af psykopatologi bedst beskriver en bestemt patientgruppe. (Muthén, 2006). Generelt kan man anvende disse modeller til at undersøge inhomogenitet i en population (Lubke og Muthén, 2005).

Statistiske metoder til at undersøge behandlingseffekt

Ovenfor er beskrevet niveauer i statistisk forståelse i almindelighed. Herefter vil de statistiske analysemetoder til brug for behandlingseffektundersøgelser blive gennemgået. Det drejer sig om undersøgelser af ændringsprocesser, og analysemetoderne kan derfor lige så vel anvendes på normale udviklingsprocesser, f.eks. børns kognitive udvikling.

Analyse af to undersøgelsestidspunkter: 1. og 2.

Statistiske niveau:

Behandlingseffektundersøgelser har ofte taget form af to målinger af de samme personer, før og efter en behandlingsintervention. Dette design udbygges ofte med sammenligninger mellem forskellige grupper (behandlet gruppe versus kontrolgruppe, eller forskellige grupper der udsættes for forskellige typer af behandling). Ofte inddrages flere forskellige mål ved de to undersøgelser (f.eks. forskellige tests). Endvidere udbygges det ofte med en eller flere efterundersøgelser hvor personerne undersøges i forskellige intervaller efter behandlingens afslutning. Det generelle i denne type design er imidlertid at selve behandlingsinterventionen tit er repræsenteret ved kun to måletidspunkter, før og efter. I en klassisk tekst har Campbell og Stanley gennemgået dette og andre forskningsdesigns (Den seneste, opdaterede version er skrevet af Shadish, Cook og Campbell (2002)).

Som jeg vil påvise, er der så store mangler med dette grundlæggende design, at det stort set bør forlades. Først vil jeg imidlertid gennemgå de statistiske metoder der typisk anvendes ved dette to-tidspunktsdesign.

Analyse af data som to uafhængige fordelinger. Den simpleste, og mest utilstrækkelige, statistiske behandling af undersøgelsen med to dataindsamlinger er at opfatte hver dataindsamling som en selvstændig fordeling og sammenligne fordelingerne på de to indsamlingstidspunkter.

Hvis man arbejder med intervallskaladata, er det normale at beregne gennemsnit og standardafvigelse på de to forskellige tidspunkter, og derefter udføre en statistisk test (typisk en *t-test for uparrede data*) for forskel mellem gennemsnit.

Traditionelt vil man se efter om den statistiske test viser et signifikant resultat, og hvis det er tilfældet, konkluderer man at behandlingen har haft effekt. Som beskrevet side 124 er der vanskeligheder med denne undersøgelsesmetode, og mange behandlingsundersøgelser er gennemført med for lav statistisk power.

Man kan beregne en effektstørrelse der viser hvor stor indflydelse behandlingen i følge de indsamlede data har haft. Når der er tale om forskel mellem to grupper, anvender man ofte Cohens *d*. Denne gør det muligt at få en fornemmelse af effekten uafhængigt af samplestørrelsen, og man har med metaanalyser af denne form for undersøgelse fået mulighed for bedre at udnytte resultaterne fra de mange små behandlingsundersøgelser der typisk er foretaget.

Undersøgelse af individforskelle mellem de to testninger: Differensscoren Den oven for skitserede metode er imidlertid decideret dårlig, idet den ikke udnytter den information man faktisk har indsamlet. Når de samme personer er undersøgt to gange, før og efter en behandlingsintervention, ved man mere om gruppen end gruppernes gennemsnit før og efter. Man kender nemlig også den enkelte persons testresultat før og efter. Det er altså muligt på individplan at koble testresultaterne mellem de to testninger. Hvis denne viden inddrages, vil det f.eks. være muligt at skelne mellem en situation hvor ingen person ændrer sig fra før til efter (og dermed et uændret gennemsnit for hele gruppen), fra en anden situation hvor halvdelen af personerne bliver bedre og halvdelen bliver dårligere efter behandlingen, og det samlede gennemsnit ganske vist ikke ændres, men hvor det måske vil være muligt at skelne mellem to undergrupper af personer som reagerer forskelligt på behandlingen.

Statistisk set, kan informationen om de enkelte personers resultater ved de to testninger udnyttes ved at man for hver enkelt person trækker f.eks. det andet testresultat fra det første, og dermed opnår en række differenser. Hvis der måles med en symptomtest, vil en positiv differens vise en forbedring, mens en negativ vil vise en forværring, og størrelsen

vil vise graden af forbedring hhv. forværring. Man kan nu arbejde med fordelingen af differenser. Et gennemsnit i denne fordeling større end nul vil vise en gennemsnitlig forbedring fra før til efter behandlingen. Man kan med en statistisk test (*t-test for parrede data*) undersøge om gennemsnittet i denne differensfordeling er signifikant større end nul. Man kan vise at denne test har større statistisk power for de samme data, end testen for uparrede data.

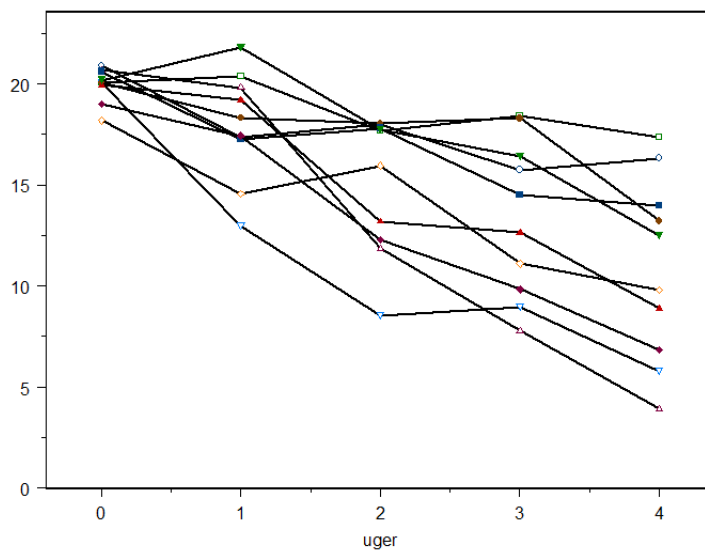
Kritik af metoden med differensscore Der har imidlertid været forskellige kritikpunkter af den skitserede fremgangsmåde, og det har været diskuteret på hvilken måde man statistisk set bedst kan sammenfatte resultaterne fra to testninger. Det har i denne forbindelse været opfattet som et problem at der som regel, netop fordi man måler forandringer for enkeltpersoner, er korrelation mellem før- og eftertest. En metode som *residual gain* har været anvendt for at kompensere for denne korrelation, men resultatet af denne diskussion er at den simple differens mellem scores fra de to testninger er den bedste indikator for ændringen.

Kritik af designs med kun to dataindsamlinger Imidlertid er denne form for forskningsdesign blevet kritiseret for i det hele taget at mangle mulighed for at belyse ændringsprocessen på en relevant måde. Det er endvidere ikke muligt at inddrage måleusikkerhed i differensmålingen. Det har også været opfattet som en vanskelighed at det ikke er muligt at være sikker på at flere grupper der skal sammenlignes ved behandlingsundersøgelsen, faktisk ligger på samme niveau på de relevante variable ved starten af undersøgelsen. En randomiseret fordeling er ikke i sig selv en garanti for at man opnår sammenlignelighed mellem grupperne (selv med randomiseret tildeling af personer til de forskellige behandlingsgrupper er det muligt at få ganske store tilfældige udsving i startniveauet). Disse problemer viser sig imidlertid delvis at forvinde ved en anden tilgang til behandlingseffektundersøgelsen.

Growth modeling med observerede data: 3. og 4. Statistiske niveau

Ved growth modeling tages et andet udgangspunkt for behandlingsundersøgelsen, og der anvendes andre metoder. At der er en korrelation mellem scores før behandlingen og senere, er baggrunden for at opfatte behandlingsdata som hierarkiske data, med enkeltpersonen som den grupperende faktor som medfører større sammenhæng internt i "gruppen", personens resultater på forskellige tidspunkter. Endvidere er udgangspunktet at se på behandlingen som et forløb der skal dokumenteres ved en række målinger, snarere end en simpel undersøgelse af en forandring fra før til efter.

På figur 10 kan vi se forløbet af behandling over 4 uger for 10 personer. På figuren er afsat en symptomscore ved 5 målinger, ved start og efter hver uge. Vi kan se en generel nedgang, men med ret forskellige forløb³⁴.



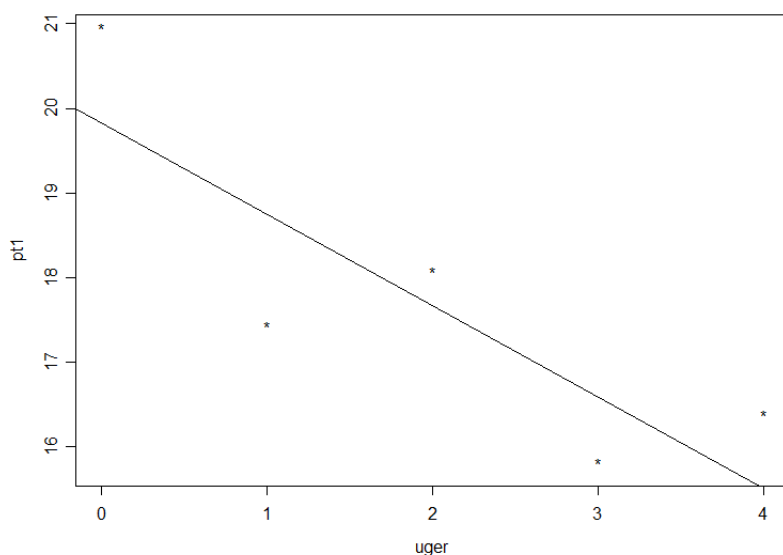
Figur 10. Reduktion i symptomer for 10 personer ved 5 målinger under terapiforløb

Det bliver parametrene ved de individuelle ændringsforløb der karakteriserer den enkelte forløb og som kan anvendes ved sammenligninger på gruppeplan. Ved lineært udviklingsforløb bliver det altså intercept og hældning der er de to parametre der arbejdes med³⁵. Ved ikke-lineære forløb anvendes evt. mere komplicerede parametre (f.eks. hældningskoefficienten til en kvadreret tidsvariabel).

Man kan på figur 11 se udviklingen for en enkelt person i løbet af de fem testninger. Der er indlagt en linie som bedst muligt passer med datapunkterne. Man kan se at linien ikke kommer til at passe nøjagtigt med punkterne. Det kan være fordi den faktiske ændring for personen ikke sker lineært. Imidlertid kan det også være at det er måleusikkerhed der gør at punkterne ikke ligger helt præcist på linien. Med kun to testninger

³⁴Det ser ud til at udgangspunktet for personerne ikke er så forskelligt, men at forløbet er ret forskelligt.

³⁵På figur 10 er det interceptet der ikke er så forskelligt, mens hældningen er ret forskellig for de forskellige personer.



Figur 11. Reduktion i symptomer for en enkelt person med 5 målinger

ville vi ikke kunne se tegn på måleusikkerhed, fordi en linie altid kan passe nøjagtigt med to punkter.

Lad os først tænke os at forbedringsprocessen i undersøgelsesperioden sker lineært, altså at personen støt og roligt ændrer sig i et bestemt tempo, som kan være forskelligt fra person til person, men altså er ensartet for den enkelte (nedenfor vil vi gennemgå hvorledes man kan undersøge om det også er tilfældet). Dette svarer til at man skal kunne vise personens udvikling med en ret linie. Hvis dette er tilfældet, er afstanden mellem linien og punkterne udtryk for måleusikkerhed. Selve linien kan beskrives ved hjælp af to tal: dens hældning, og dens placering (højt eller lavt). Placeringen kan bedst måles ved at anbringe y-aksen ved det første måletidspunkt³⁶ og så se hvor personens linie skærer denne y-akse, der som bekendt kaldes for *interceptet*. Hældningen viser hvor hurtigt ændringen foregår. Interceptet viser testresultatet ved første testning. Hvis testen viser personens symptombelastning, er interceptet altså et mål for personens tilstand ved starten af behandlingen, og hældningen udtryk for hvor hurtigt personen er blevet bedre i løbet af behandlingen.

En fordel ved denne metode er også at man ikke er bundet af at testningerne skal foretages med nøjagtigt samme tidsintervaller for alle personer. Testresultaterne afsættes blot som punkter i koordinatsystemet

³⁶ved at sætte starttidspunktet til 0.

ud for de tidspunkter hvor de er målt, og man kan stadig finde startniveau (intercept) og hældning (ændringshastighed) således at personerne kan sammenlignes.

Den rette linie kan, som nævnt, karakteriseres ved en ligning der forbinder to variable og indeholder to *parametre*, nemlig hældning og intercept (jfr. ligningen $y = a + b * x + e$).

$$symptom = a + b * behandlingsuge + e \quad (8)$$

Det interessante ved denne måde at afbilde resultaterne er at hver person nu ikke kun er karakteriseret ved en række testresultater som man kan forsøge at sammenligne direkte, men at testresultaterne for hver person kan sammenfattes ved personens intercept og hældning. Ved behandlingens start, hvor behandlingsugen = 0 og b derfor udgår, bliver effektmålet = a. Interceptet er et mål for startniveauet (f.eks. mængden af symptomer ved behandlingens start). Hældningskoefficienten angiver hvor stor ændringen bliver i effektmålet når behandlingsmånedene øges med 1. Den er altså med andre ord udtryk for ændringshastigheden i behandlingen.

Når man har en række personer der er blevet behandlet, får man et intercept og en hældningskoefficient fra hver person. Alle intercepterne kan beskrives med en fordeling, som kan analyseres, og det samme kan gøres med alle personernes forskellige hældningskoefficienter. Man kan nu interessere sig for hvad der influerer på disse parametre.

Med hensyn til ændringshastigheden, *b*, kan man f.eks. undersøge om den er påvirket af behandlingsmetoden. Hvis vi tænker os et design med en behandlingsgruppe og en kontrolgruppe, kan vi give variabelen *behandlingsmetode* værdierne 0 (= ingen behandling) og 1 (= behandling). Vi opstiller en ligning til beskrivelse af ændringshastigheden:

$$b = c + d * behandlingsmetode + e_1 \quad (9)$$

c er den ændringshastighed som kan observeres når behandlingsmetoden er 0, dvs i kontrolgruppen, og leddet med *d* derfor udgår. *d* er øgning (eller formindskning) af ændringshastigheden ud over *c* som vi kan se i den behandlede gruppe (behandlingsmetode = 1).

$$a = f + g * behandlingsmetode + e_2 \quad (10)$$

Med hensyn til startsituationen, så kan vi opstille en ligning til undersøgelse af om der er forskel på symptomniveauet i behandlingsgruppen og kontrolgruppen. *f* er effektmålet (symptombelastningen) i kontrolgruppen, mens *g* er den højere eller lavere symptombelastning der var ved starten i behandlingsgruppen.

En sådan model kaldes på engelsk for en *multilevel random-effects model* (Singer og Willett, 2003), "multilevel", fordi ligningen for enkeltpersonernes udvikling kaldes for niveau 1, mens ligningerne til forklaring

af parametrene a og b kaldes for niveau 2, "random-effects", fordi de to parametre a og b tænkes at variere tilfældigt inden for gruppen af klienter. Man kan fremstille den samme model ved at sætte ligningerne for niveau 2 ind i ligningen for niveau 1 og få en enkelt samlet ligning:

$$\begin{aligned} symptom &= [f + g * behandlingsmetode + e_2] \\ &+ [c + d * behandlingsmetode + e_1] \\ &* behandlingsmdr + e \end{aligned} \quad (11)$$

I denne form kaldes modellen på engelsk for en *mixed-effects linear model* (Pinheiro og Bates, 2000).

Det er muligt at udbygge modellen med forskellige andre faktorer, både nogle som er konstant over tid (f.eks. køn), og nogle som varierer over tid (f.eks. samtidig medicinindtagelse).

Ikke-lineære forløb Ved en konkret undersøgelse må man selvfølgelig starte med at undersøge om de individuelle forløb med rimelighed kan beskrives som rette linier, og først derefter arbejde med fordelingerne af intercepter og hældninger.

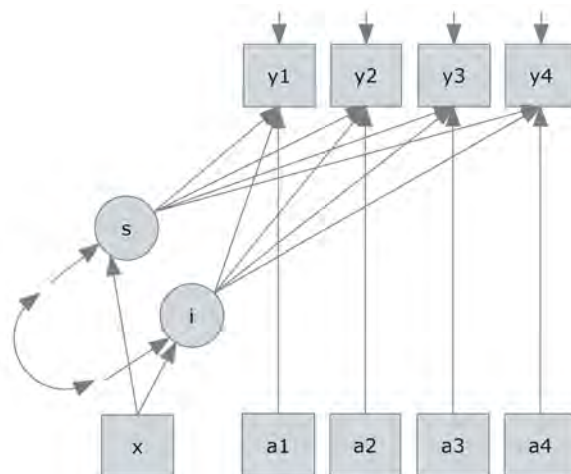
I de tilfælde hvor den rette linie ikke kan anvendes, kan man, som tidligere beskrevet side 129, undersøge om man kan ændre på skalaerne for måling af tid og testresultater således at man ender med alligevel at kunne beskrive resultaterne på rette linier. Dette kan godt tillades hvis man blot regner tilbage til de oprindelige skalaer når man ser på de endelige resultater, mens den statistiske analyse kan foretages på de transformerede (ændrede) skalaer på en enkel måde ved hjælp af den rette linies ligning, og altså kun to parametre (intercept og hældning).

I nogle tilfælde kan man imidlertid ikke ændre på skalaerne på en måde som resulterer i rette linier, og man må derfor finde andre måder at beskrive kurveforløbet. Der findes en række forskellige kurver som kan beskrives med simple formler - lige som den rette linies ligning. Et simpelt kurvet forløb kan beskrives ved at indføre et kvadratisk led i ligningen, som får formen $y = a + b * x + c * x^2 + e$.

Growth modeling med latente variable: 5. Statistiske niveau

Også med hensyn til longitudinelle undersøgelser, f.eks. undersøgelse af behandlingseffekt, findes der modeller der kan rumme flere forskellige måder at måle de psykologiske egenskaber man ønsker undersøgt for eventuel forandring. Man er således ikke bundet til at operationalisere et begreb med en enkelt målemetode, men kan kombinere flere og derved øge validiteten af målingerne. Modellerne kan endvidere tage højde for måleusikkerhed, og dermed (i nogen grad) kompensere for ufuldkommenhederne i målingerne således at man får mere sande mål for de latente

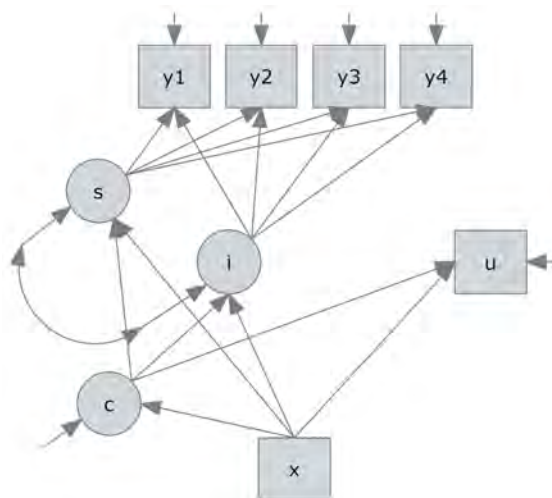
variable. Det skulle tendentielt medføre sikrere resultater, og vil ofte også sige at effekterne vil vise sig at være større (sml. problemstillingen med korrektion for attenuation).



Figur 12. SEM-model for behandlingsforløb

I modellen på figur 12 er y_1, y_2, y_3 og y_4 observerede mål, f.eks. testitens. Parametrene til de lineære ændringsforløb er i , intercept og s , slope, eller hældning. Dette er den grundlæggende growthmodel med latente variable. Der kan endvidere tilføjes faktorer som påvirker tidsafhængigt (x), f.eks. opdeling i personer der er i forskellige slags psykoterapeutisk behandling. Der kan endvidere være faktorer som kan variere på de forskellige tidspunkter (a), f.eks. medicinering med psykofarmaka på de forskellige behandlingstidspunkter. Med den buede dobbeltpil mellem i og s , angives muligheden af at disse er korrelerede. Det vil f.eks. ikke være mærkeligt hvis patienter med højere grad af symptomer har et andet behandlingsforløb end patienter med færre symptomer. I modellen er beskrevet et retliniet forløb. Imidlertid er behandlingsforløb ofte kurvede, eksempelvis med hurtig bedring af nogle symptomer og en senere langsommere ændring (se f.eks. Monsen et al., 2002). Dette kan, som tidligere nævnt, beskrives ved transformation af skalaerne under analysen. Det er imidlertid også muligt at indføre et kvadratisk led i ligningen, som får formen $y = a + b * x + c * x^2 + e$. Det kvadratiske led (som kunne kaldes q for "kvadratisk") ville så skulle indgå som en tredje latent variabel ved siden af i og s .

Modellen i figur 13 er en *growth mixture model*, eller *GMM* (Muthén, 2004). Modellen indeholder, som den foregående, y 'er der repræsenterer observationer på 4 forskellige tidspunkter. Disse observationer anses for bestemt af alvorlighedsgraden af tilstanden for personerne ved behandlingsstart (intercept, i), og ændringshastigheden under behandlingsfor-



Figur 13. SEM-model for behandlingsforløb med latent klassevariabel

løbet (hældningen, s). Observationerne er også bestemt af andre forhold (unik varians, repræsenteret ved pilene uden afsender). De to kontinuerte latente variable, i og s måles ved hjælp af y -observationerne. Som i forrige model, kan det også tænkes at de er påvirket af baggrundsvariable, x . Det kunne være behandlingsform, eller det kunne være køn, således forstået at det ene køn har et dårligere udgangspunkt ved starten af behandlingen (i) og/eller har en langsommere eller hurtigere bedring under behandlingen (s). En baggrundsvariabel som køn, opdeler personerne i klasser som er kendt på forhånd. Imidlertid er det muligt at behandlingsforløbet kan vise at klienterne falder i flere klasser med forskellige forløb. Dette afbildes som en latent klasse (c). Det kan f.eks. være der viser sig to undergrupper, hvoraf den ene reagerer hurtigt på behandlingen, mens den anden kun reagerer langsomt, eller slet ikke. Denne forskel er vigtig ved undersøgelser af behandlingsformer for at finde frem til hvilke behandlingsformer der er velegnede over for hvilke patienttyper. Pointen er at opdelingen i patienttyper ikke er kendt i forvejen, og at man må finde klasseopdelingen ad empirisk vej ved at undersøge om kurverne over bedringsprocessen (jfr figur 10) er tilfældigt fordelt, eller de kan opdeles i to klasser, evt. med en vis intern variation. Her er tale om en klasseopdeling som ikke findes ved at analysere de direkte observationer (som ved traditionel clusteranalyse), men om en latent klasseopdeling som findes ud fra fordelingerne af s og i , som igen er fundet igennem y -observationerne. Endvidere betyder det at der ved analysen tages højde for fejlvariation og unik variation (pilene uden afsender) når den kategorielle variabel, c , bestemmes. Af hensyn til overskueligheden er de tidsvarierende baggrundsvariable (a -erne) fra figur 12 udeladt, men der

er ikke noget i vejen for også at inkludere sådanne variable i modellen. På tegningen er endvidere indtegnet en kategorial observeret effekt, u , som kan være bestemt af opdelingen i latente klasser (c). Et eksempel kunne være drop-outpatienter, altså personer der ikke fuldfører behandlingen, som holder op fordi deres forløb er af typen der reagerer langsomt eller slet ikke på behandlingen. Det er også antydnet at baggrundsvariablen x kan have betydning for drop-out-tendensen, eksempelvis kan det være at mænd lettere slutter behandlingen end kvinder. Som tidligere beskrevet skal modellen undersøges med empiriske data.

Det er således muligt med disse statistiske modeller – som for tiden hastigt udvides med flere analysemuligheder – at håndtere en lang række forskelligartede forhold der er vigtige for undersøgelse af behandlingseffekt.

Manglende eller censorerede data Et særligt problem ved longitudinelle undersøgelser er ukomplette data. Der vil ofte mangle data fra nogle undersøgelsestidspunkter fra bestemte personer. Tidligere blev det anset for mest renfærdigt at udelukke personer med ukomplette data, men dels er det spild af information, dels vil der ofte være forskel på personer med komplette data og personer med forskellige manglende data. Der er imidlertid i den senere tid udviklet en lang række metoder til dels at analysere indflydelsen fra manglende data, dels til at kompensere for manglerne. Hvis man analyserer data som forløb og beskriver disse med relevante parametre, vil man typisk for hver person have indsamlet en række målinger på forskellige tidspunkter, og man vil kunne beregne modellens parametre selvom enkelte datapunkter mangler.

En anden situation kan forekomme hvor man undersøger samples med løbende indtagning af klienter. Her vil det ofte forekomme at nogle af de personer der indgår i undersøgelsen har afsluttet behandlingen, mens andre stadig er under behandling. Den almindelige måde at tænke på er vel at man må vente til alle personer er færdige med behandlingsforløbet. Det er ikke en relevant metode kun at analysere færdige forløb, fordi der kan være inhomogeniteter, som f.eks. at flere af de klienter der stadig er i behandling kan have en langsommere bedringsproces eller en generelt dårligere prognose. Det er imidlertid også misvisende at analysere alle forløb som om de var afsluttede – data fra slutningen af de endnu ikke afsluttede forløb mangler og kan derved give bias i analysen. Man kan imidlertid løse problemet ved at anvende de nævnte statistiske modeller og specificere at datatypen er *overlevelsesdata*.

Praktisk anvendelse af de beskrevne metoder

Statistiske metoder til og med det der her kaldes 3. statistiske niveau beskrives normalt med almindelig omnibusstatistikpakker. Ofte anvendes

inden for psykologiens område programmet *SPSS*. Imidlertid er metoderne på de højere niveauer ikke lette, og i nogle tilfælde ikke mulige, at implementere i *SPSS* og tilsvarende programmer. Det har ført til udvikling af en række specialprogrammer. Eksempelvis er *RUMM 2030* udviklet til analyse af Raschmodeller, og *MLwin* til analyse af longitudinelle data med hierarkiske modeller. Statistikprogrammet *S-plus*, med gratisvarianten *R*, kan anvendes til modellering på højt niveau, men de højeste niveauer kræver en hel del teknisk arbejde med programmet.

Af programmer til analyse af modeller med latente variable er *LISREL* det tidligste og det mest kendte. Det er imidlertid samtidigt et program som er teknisk komplekst og derfor kan være vanskeligt at anvende. Nogle lettere programmer er *EQS*, *AMOS*, mens først *M-plus* kombinerer på den ene side mulighed for at modellere en lang række forskelligartede modeller med stort set alle kombinationer af datatyper med på den anden en meget simpel betjening, og desuden med omfattende undervisningsaktiviteter der er gratis tilgængelige på internettet (Muthén og Muthén, 1998-2007). Resultaterne af analyserne er dog en lang række tal til beskrivelse af de forskellige parametre som det kræver en nærmere viden om SEM-modeller at fortolke.

Litteratur

Agresti, A. (1996) *An Introduction to Categorical Data Analysis*. New York: Wiley.

Asparouhov, T. & Muthén, B. (2008). *Exploratory Structural Equation Modeling*. Upubliceret manuskript, Version 3C, 23. maj, 2008.

Bech, P. (2006). Depressionsspørgeskemaer med særligt henblik på Major Depression Inventory (MDI). I Elsass, P., Ivanouw, J., Mortensen, E.L. og Poulsen, S. *Assessmentmetoder. Håndbog for psykologer og psykiatere*. København: Dansk Psykologisk Forlag.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71 (3), 425-440.

Brown, T.A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York: Guilford.

Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*,

56(2), 81-105.

Cohen, J. (1988). *Statistical Power analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.

Hunter, J.E & Schmidt, F.L (2004). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Thousand Oaks, CA: Sage Publications.

Lubke, G.H & Muthén, B. (2005). Investigating population heterogeneity with Factor Mixture Models. *Psychological Methods*, 10(1), 21-39.

Monsen, K., Monsen, J.T, Svartberg, M. & Havik, O.E. (2002). Chronic pain patients: Patterns of change in interpersonal problems, pain intensity and depression-anxiety. *Psychotherapy Research*, 12(3), 339-354.

Muthén, B. (2004). Latent Variable Analysis. Growth mixture modeling and Related Techniques for Longitudinal Data. In Kaplan, D. (Ed.) *Handbook of quantitative methodology for the social sciences*. Newbury Park, CA: Sage.

Muthén, B. (2006). Should substance use disorders be considered as categorical or dimensional? *Addiction*, 101(suppl. 1), 6-16.

Muthén, L.K. & Muthén, B.O. (1998-2007). *Mplus User's Guide. Fifth Edition*. Los Angeles, CA: Muthén & Muthén.

Pinheiro, J.C. & Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.

Singer, J.D. & Willett, J.B. (2003). *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

Stevens, J. (1996). *Applied Multivariate Statistics for the Social Sciences*. Mahwah, NJ: Erlbaum.

Zwick, W.R. og Velicer, W.F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

Statistiske modeller – forenklinger, forvanskninger og/eller fortrængninger?

Af Svend Kreiner

“Everything should be made as simple as possible, but not simpler.”
(Einstein)

“A model is an intentionally simplified representation of an object.”
(Karpatchof, 2000)

“Of course, all postulated models have at best limited and approximate validity. The point is here that the use of a model, the nature of whose limitations can be foreseen, is not wise, except for very limited purposes.”
(Cox, 1969, p. 18).

“Essentially, all models are wrong, but some are useful.” (Box & Draper, 1987, p. 424)

Forord

Anledningen til, at jeg har skrevet denne artikel, er af en sådan karakter, at jeg har behov for at supplere indledningen med et forord af lidt mere personlig karakter, end man normalt bliver præsenteret for i videnskabelige artikler. Årsagen er den enkle, at min forståelse af hvad statistiske modeller er, først og fremmest bygger på det, som Benny Karpatchof for næsten 40 år siden (da han hyrede mig som studentermedhjælper på Danmarks pædagogiske Institut) fortalte mig om sin forståelse af modeller i al almindelighed og statistiske modeller i særdeleshed. Jeg var på det tidspunkt en meget

uerfaren og forvirret statistikstuderende med en særdeles uklar fornemmelse for, hvad statistik var for noget, og jeg kan stadig huske lyden af de faldende 10-ører, da Benny tog sig tid til at forklare mig om statistisk *metode* og om forskellen på statistisk metode og statistisk teori. De ti måneder, hvor jeg arbejdede for Benny, og de efterfølgende mange år, hvor vi arbejdede sammen, gav mig en forståelse af og en glæde ved mit eget fag som stadig holder, og som jeg her 40 år efter stadig først og fremmest tilskriver Bennys indflydelse.

Så derfor kære Benny: Jeg oplevede at du for 40 år siden drejede mig 180^o og udstak en hel anden kurs, end den jeg havde i forvejen. Det har jeg altid været både særdeles bevidst om og meget taknemmelig for, og jeg håber bare, at du ikke synes, at jeg siden er kommet alt for meget ud af kurs. Der er meget af det, som jeg har lært af dig, som har været en stor inspiration for mig også længe efter at vores formelle samarbejde på Psykologisk Laboratorium stoppede. Bl.a. – men ikke kun – dit modelbegreb, som stadig optager mig, og som jeg har valgt at skrive om, fordi jeg synes, at der er behov for nogle supplerende betragtninger. Jeg håber du er enig – om ikke i selve betragtningerne, så i hvert fald i behovet for dem.

Indledning

Bag enhver statistisk analyse ligger der en statistisk model. Den kan være så enkel, at de fleste ville have svært ved overhovedet at se den, eller den kan være så kompliceret, at det for mange vil være vanskeligt at se andet end modellen, men uanset hvad, så er modellen der altid.

Der er mange forskellige opfattelser af, hvad statistiske modeller er, og derfor også forskellige holdninger til, hvordan man skal forholde sig til dem. De fire forskellige citater, som jeg har valgt at indlede denne artikel med, opsummerer tilsammen min forståelse af det statistiske modelbegreb er: En model er et resultat af et forsøg på at beskrive noget kompliceret på en *forenklet* måde. Efterspørgslen efter enkelhed er som bekendt udtryk for et generelt hæderkronet videnskabeligt princip, og i den forstand er statistisk modelsnedkeri altså i godt selskab, men læg mærke til formuleringen. Det er ikke et tilfælde, at jeg skriver “forenklet” og ikke blot “enkel”.

Udover at forståelsen af modellerne som konsekvenser af forenkling giver en vis frihed i forbindelse med konstruktionen og anvendelsen af modellerne, har forenkling også en pris, som der på forskellig måde refereres til og advares mod i citaterne. Det er klart, at man ikke kan vurdere modeller ud fra et sandhedskrav på samme måde, som man ville gøre det

i forhold til teorier, fordi ideen om en forenklet beskrivelse betyder, at beskrivelsen i sagens natur være ufuldstændig. Der må være et eller andet, der enten er udeladt af modellen eller beskrevet på en forkert måde. Synspunktet, at alle modeller er forkerte (og specielt citatet fra Box), bliver ofte opfattet som *carte blanche* til anvendelse af dårlige modeller, men alle tre citater inklusive Box'es advarer mod denne holdning. Einstein advarer generelt mod *overforenkling*, Cox advarer mod at bruge urealistiske modeller, hvor manglerne er åbenlyse, mens Box understreger, at ikke alle modeller er *nyttige*.

Og her kommer vi så til Karpachofscitatet. Det, der er specielt ved Karpachofs modelbegreb, er ikke ideen om forenkling. Den idé er rigtig, men den deler han med de andre forfattere. Det, der er centralt i Karpachofs definition af modeller, er kravet om, at der skal være tale om en *bevidst* forenklet beskrivelse, et krav som er en nødvendig konsekvens af erkendelsen af, at en modelbeskrivelse kan være overforenklet, urealistisk og/eller unyttig. Hvis man vil undgå disse faldgruber er man i sagens natur nødt til at gøre sig nogle overvejelser om, hvordan modellen skal konstrueres, om hvad den skal indeholde, og om hvad den ikke skal indeholde. Der skal træffes en række valg, som man bør *tænke* over. Og hvis man tænker over det, er det jo pr. definition en bevidst beskrivelse.

Dette er naturligvis nemmere sagt end gjort. Der er mange eksempler på rutineprægede (bevidstløse) lineære regressionsanalyser, hvor det er åbenlyst, at den, der har foretaget den statistiske analyse, har benyttet netop denne metode, fordi den er let tilgængelig i statistiske standardprogrammer, uden at tænke nærmere over den model, der ligger bag netop denne form for analyse. Regressionsmodellen kan i disse tilfælde bedst beskrives som implicit givet, og det er vel et spørgsmål, om der overhovedet er tale om en model i Karpachofs forstand.

Men lad det ligge her. Sagen er, at det lyder så enkelt at sige, at en model er det, der er tilbage, når vi har elimineret alt det, som vi betragter som irrelevant, men at det dels kan være ganske vanskeligt at forstå hvordan forenklingen er foregået og dels er sådan, at en model udover at indeholde det, der er relevant, også indeholder meget, der ikke har noget at gøre med det, som modellen beskriver. Det hele hænger på de antagelser vi gør, når vi konstruerer modellen, men den rolle, som alle modellens forskellige antagelser spiller, selv når der foretages et bevidst modelvalg, er meget vanskelig at gennemskue. Det er dette problem, som jeg vil forsøge at kaste lidt lys over i denne artikel. Jeg vil i løbet af denne diskussion tillade mig at referere til nogle af de modeltyper, som jeg selv har arbejdet med, men diskussionen er helt generel, fordi jeg mener, at de fleste af de problemer, som jeg vil diskutere er almene problemer, som (i større eller mindre grad) findes i alle statistiske modeller. Det er også min hensigt at diskussionen skal være metodisk og ikke teoretisk statistisk med en vægt, der i højere grad er lagt på spørgsmålene om, *hvad* det er man gør, når man arbejder statistisk, og *hvorfor* man gør

det på denne specielle måde, end på *hvordan* man gør det.

Af de mange ting, der kan diskuteres i forbindelse med statistiske modeller, har jeg valgt at fokusere på følgende emner i nogenlunde følgende rækkefølge:

- Modellens referenceramme
- Modelantagelserne
- Forholdet mellem realistiske og urealistiske modeller
- Referenceløse modelegenskaber
- Inferensrammer
- Utilgængelige problemer

Listen er på ingen måde udtømmende, og andre ville formodentlig have valgt at tage andre forhold op, hvis de var blevet bedt om at sige noget om statistiske modeller. Diskussionen er også mere skitseagtig end godt er – ikke fordi der ikke har været tid nok til at skrive den, men fordi emnet er både svært og uhåndgribeligt. Ovennævnte liste over emner, der kan betragtes som en disposition for artiklen, vil forhåbentlig gøre det lidt nemmere at følge argumenterne i artiklen, hvor ufærdige og utilstrækkelige de end måtte være.

Forenklingen og dennes konsekvenser

En statistisk model består dels af et endeligt antal variable og dels af et sæt regneregler (formler) ved hjælp af hvilke man kan beregne sandsynligheder for alle de forskellige hændelse, der kan defineres ud fra variablenes værdier. Set ud fra denne synsvinkel er spørgsmålet om, hvad det er for en forenkling der finder sted, når man konstruerer en statistisk model, i princippet let at besvare. Det foregår i tre forskellige trin:

1. For det første i udvælgelsen af variable, der beskriver det, man er interesseret i, og dermed også i et fravalg af variable, der ikke kommer med. Fravalget af det, der ikke tages med, er formodentlig den største kilde til de metodeproblemer, som statistikken må slås med.
2. For det andet i en beslutning om at observere og kode de udvalgte variable på en bestemt måde. Disse beslutninger er typisk også beslutninger om fravalg af nogle detaljer i iagttagelserne, som måske kunne være af interesse.

3. Valg af funktioner, der kan benyttes til beregning af sandsynlighederne for de hændelser, der kan defineres efter at kodningen har fundet sted, funktioner, der ofte omtales som statistiske fordelinger. Disse fordelinger afhænger som regel af et begrænset antal parametre, der referer tilbage til egenskaber ved det objekt, det genstandsfelt, som modellen beskriver. Parametrene er typisk ukendte, men hensigten med den efterfølgende statistiske analyse er som regel at kaste lys over, hvilke værdier parametrene rent faktisk har.

En diskussion af de problemer, som modelforenklingen medfører, fokuserer typisk på de metodeproblemer, som udeladelsen af oplysninger indebærer, og på de statistiske fejl, som valget af ”forkerte” fordelinger indebærer. Der er ikke meget jeg kan bidrage med til denne diskussion udover at påpege at terminologien ”*forkert* fordeling” måske ikke er særlig naturlig, hvis man ser modellerne som udtryk for bevidste forenklinger. Personligt vil jeg – som det vil fremgå af det efterfølgende – foretrække at man skelner mellem urealistiske og realistiske modeller. Der følger imidlertid andre og mere subtile problemer med som konsekvenser af forenklingen, bl.a. at modellen tilføres egenskaber, som intet har at gøre med det, som modellen beskriver, og at selve modelkonstruktionen udelukker mulighederne for at rejse en række relevante spørgsmål til det, der rent faktisk er inkluderet i modellen. Det er disse metodeproblemer, som jeg ønsker at diskutere i denne artikel. Diskussionen vil komme til at dreje sig om nogle af de standardmodeller, som de fleste på et eller andet tidspunkt har benyttet sig af, hvis man overhovedet har foretaget statistiske analyser og om nogle af de modeltyper – Rasch modeller og grafiske modeller – som jeg selv har beskæftiget mig meget med, men de problemer jeg kommer ind på er efter min mening helt generelle.

Modellens referenceramme

Før vi går videre er det nødvendigt at minde om, at en model først er en model, hvis den er en model af noget. Formlen

$$P(Y_v = 1|X_v = x) = \frac{\exp(\alpha + \beta x)}{1 + \exp(\alpha + \beta x)} \text{ for } v = 1, \dots, n \quad (1)$$

er altså ikke en logistisk regressionsmodel, før man har fortalt hvad Y og X er, og Rasch-modellen,

$$P(Y_{vi} = 1) = \frac{\exp(\theta_v - \alpha_i)}{1 + \exp(\theta_v - \alpha_i)} \text{ for } v = 1, \dots, n \text{ og } i = 1, \dots, k \quad (2)$$

er heller ikke en statistisk model, før der foreligger oplysninger om, hvilken population personerne tilhører, hvilke items der er tale om, og hvad θ er et

udtryk for. Formlerne er altså udtryk for prototype-modeller, som først bliver rigtige modeller, når de relateres til specifikke referenceramme, hvor det er helt klart, hvad modellen egenskaber henviser til.

Begrebet referenceramme har altid stået centralt i forbindelse med diskussionen af Rasch modellerne, fordi et af argumenterne for interessen for Rasch modellen er, at den garanterer specifikt objektive målinger af latente egenskaber eller færdigheder. Det, der for mig er det specielle ved Raschs definition af specifik objektivitet, er ikke objektiviteten, men derimod specificiteten. Det forhold, at et givet måleinstrument kun kan forventes at give objektive målinger inden for en helt konkret referenceramme. Når dette er sagt, skal det dog også bemærkes, at hverken Rasch eller nogen af de mange andre, der har skrevet om Raschmodellerne, har gjort særlig meget ud af at forklare, hvad en referenceramme egentlig er for noget. Referencerammebegrebet har derfor altid har stået for mig som en black box, hvis indhold jeg har haft en intuitiv fornemmelse af, uden at jeg egentlig følte mig sikker på, hvad Rasch i realiteten tænkte på, da han definerede sit objektivitetsbegreb.

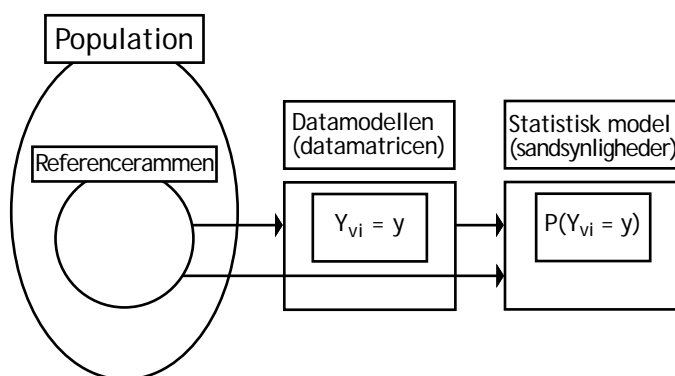
Sådan ser jeg ikke længere på det. Referencerammen hos Rasch er i realiteten hverken indviklet eller dybsindigt. Det svarer til det, som vi i forbindelse med statistiske modeller i al almindelighed kunne omtale som modellens genstandsområde. I de sammenhænge, hvor jeg som regel foretager statistiske analyser, består referencerammen af en kombination af en afgrænset population på den ene side og en teori, der dels definerer en række egenskaber ved personerne og dels udtaler sig om relationer mellem disse begreber på den anden. Man kan naturligvis forestille sig andre former for referencerammer, men hvorom alt er, så vil en model altid referere til et bestemt genstandsområde, uanset om der er tale om en Raschmodel eller om en mere almindelig statistisk model. Det er dette genstandsområde som er modellens referenceramme. Referencerammen kan bestå af en delmængde af individer fra en større population, eller der kan være tale om en hel population, og teorien kan være mere eller mindre færdig med både klart og uklart definerede begreber, men det er under alle omstændigheder afgørende, at det er entydigt klart både hvilke individer, der er omfattet af referencerammen, og hvilke teoretiske begreber, som den indeholder. Og selvfølgelig afgørende, at man ved præcis hvilke modelegenskaber, der referer til begreber inden for referencerammen, og hvilke der ikke gør det.

Modelkonstruktion

Det næste, der skal diskuteres, er spørgsmålet om, hvad der skal med i modellen og i hvilken forstand, der er tale om forenkling? Modellen omfatter, som nævnt tidligere, dels variable med et bestemt indhold, der er kodet på en bestemt måde for en række individer, der er omfattet af

referencerammen og dels formler – teoretisk statistiske fordelinger - der fortæller os, hvorledes sandsynligheder knyttet til variablenes udfald skal beregnes. Så vidt kan alle sikkert blive enige.

Det datamateriale, der skal analyseres ved hjælp af modellen, er typisk organiseret med rækker, der referer til individer, og søjler, der referer til generiske variable defineret ved observationsmetode og procedure til kodninger. Disse generiske variable omtales ofte som modellens variable, men det er vigtigt at forstå, at *modellens* variable findes inde i datamatricen og ikke som overskrift til søjlen. Jeg har i andre sammenhænge¹ beskrevet selve datamatricen som en datamodel, der ligger placeret mellem referencerammen og den statistiske model som vist i figur 1, og understreget det forhold, at datamatricen i sig selv kan betragtes som en model af genstandsfeltet (referencerammen). Den pointe, jeg gerne vil have stå fuldstændigt klart, er, at den statistiske models variable er variablene Y_{ij} inde i datamatricen. Det kan således være, at overskriften til den i 'te søjle er Socialklasse, men det er Y_{vi} , der angiver den v 'te persons socialklasse, der er en af modellens variable – ikke søjlen.



Figur 1. Sammenhæng mellem referenceramme, datamodel og statistisk model

Når sandsynlighederne i modellen skal fastlægges, sker det ved at man foretage visse antagelser om disse sandsynligheder. En af de almindeligste af disse er, at alle individerne dels er indbyrdes (stokastisk) uafhængige og dels er identisk fordelte. Hvorvidt denne antagelse er realistisk, og hvorvidt det er en antagelse, der henviser til noget i referencerammen, er et andet spørgsmål, som vi vil vende tilbage til. Konsekvensen af antagelsen er imidlertid bekvem. Hvis antagelsen holder, betyder det nemlig, at vi kun behøver at beskrive, hvorledes sandsynlighederne for udfald knyttet til et enkelt individ skal beregnes, fordi sandsynlighederne for alle de andre individer skal beregnes på samme måde, og fordi vi på grund

¹ Statistisk problemløsning (Kreiner, 2007)

af antagelsen om uafhængige individer kan beregne sandsynligheder for udfald defineret for flere individer ved at gange sandsynlighederne for de enkelte individer med hinanden.

Dette er så grundlæggende, at man måske kan undre sig over, at det overhovedet skal diskuteres her. Når jeg alligevel gør det, er det for at understrege, at det, den statistiske model beskriver, er individer og ikke populationer. Hvis læseren af denne artikel har været udtrykket til en epidemiologisk surveyundersøgelse af risikoen for at få en blodprop i løbet af de næste fem, år vil det, den statistiske model fortæller, være en forenklet vurdering af, hvorledes læserens risiko er. Forenklet, fordi den statistiske analyse i princippet kun kan indeholde oplysninger om et endeligt udvalg af de relevante risikofaktorer, og forenklet, fordi den statistiske model påstår, at alle personer med præcis de samme risikofaktorer har præcis den samme risiko, men alligevel en vurdering, der forhåbentlig er god nok til, at læseren og læserens læge sammen kan finde en måde at håndtere problemet på.

Det er en ofte formuleret forståelse, at modellen beskriver risikoen for en bestemt delpopulation, og at modellen derfor beskriver populationen og ikke individerne i populationen, men det er altså en misforståelse. Modellen indeholder estimater af $P(\text{Blodprop}_{vi} | \text{Risikofaktorer}_{vi})$, dvs. estimatet af risikoen for den v 'te person givet netop denne persons sæt af risikofaktorer. At modellen påstår, at risikoen er den samme for alle personer med præcis de samme risikofaktorer, er en del af modellens forenkede måde at beskrive forholdene på, som ikke ændrer ved det forhold, at det er individernes risiko, der beskrives.

Estimatet af risikoen for at få en blodprop vil formodentlig bliver foretaget ved hjælp af en eller anden form for regressionsanalyse, hvorefter resultaterne kan henføres til referencerammen, fordi der er en række naturlige referencer mellem analysen og det modellen indeholder på den ene side og det som modellen beskriver på den anden:

- Variablen Blodprop refererer til en veldefineret hændelse eller tilstand.
- Variablene inkluderet i Risikofaktorer refererer til tilstande eller hændelser, som er inkluderet fordi epidemiologiske teorier om årsager til blodpropper fortæller, at de kan være en del af årsagerne.
- Den epidemiologiske definition af risikoen for en sygdom er lig med sandsynligheden for at få sygdommen. Præcis hvilken sandsynlighed, der er tale om, er ikke altid entydigt defineret, men den betingede sandsynlighed, der kan defineres for forskellige udfald af variabelen Blodprop, dvs. $P(\text{Blodprop}_{vi} | \text{Risikofaktorer}_{vi})$, er altså modellens forslag til hvordan den skal defineres.

- Og til sidst: Den teori, der ligger bag en sådan epidemiologisk analyse, vil typisk være en kausal teori, der kan forstås på den måde, at risikofaktorerne kan være årsag til at sandsynlighederne – risikoen – kan være større eller mindre, hvorimod selve forekomsten af blodproppen er en tilfældig hændelse, der kun afhænger af sandsynligheden, men ikke direkte af risikofaktorerne. Set ud fra denne synsvinkel er det ikke kun estimatet af risikoen, men også selve analyseformen – en regressionsanalyse med Blodprop som afhængig og risikofaktorer som uafhængig variabel – der står i et naturligt forhold til den kausale referenceramme².

Problemet med tolkningen af regressionsanalysen som en kausal analyse er et problem i sig selv, som jeg vil vende tilbage til senere i forbindelse med en diskussion af såkaldte inferensrammer – reducerede statistiske modeller, der kan udledes fra en statistisk grundmodel.

Modelproblemer

Vi kan skelne mellem flere forskellige former for modelproblemer, der alle er konsekvenser af, at der er tale om foreklede beskrivelser af nogle forhold, og som alle kan få betydning både for den statistiske analyse og for tolkningen af resultaterne.

Problemer med manglende variable og for grove kodninger er velkendt og behøver ikke nogle supplerende kommentarer. Epidemiologernes interesse for confounding og interessen for klinisk kontrollerede forsøg er alle udtryk for forskellige måder at løse sådanne problemer på. Men der er andre problemer end confounding i forbindelse med statistiske analyser, og nogle af disse er forholdsvis subtile. De problemer, som jeg vil knytte nogle kommentarer til, er følgende:

- Problemer med urealistiske modeller
- Problemer fordi der er visse modelegenskaber, der mangler reference til noget i referencerammen.
- Problemer med tolkning af resultater fra forskellige inferensrammer.
- Problemer fordi modellerne lægger naturlige grænser for, hvilke spørgsmål, der kan stilles til data.

² Læseren skal være opmærksom på, at der er visse teorier for kausale modeller, der betragter kausalitet som deterministiske og ikke stokastiske fænomener, og hvor referencerne mellem den statistiske model og referencerammen er af en helt anden karakter end den jeg har skitseret her.

Alle disse problemer har noget at gøre med de antagelser, der ligger bag den forenkling, som de statistiske model er udtryk for. Jer er derfor nødt til at knytte nogle få kommentarer til disse antagelser før jeg kaster mig over problemerne.

Modellens antagelser

Modelegenskaber uden reference dukker op i forbindelse med de antagelser, man gør sig i forbindelse med modelkonstruktionen. Disse antagelser kan med fordel opdeles i fire forskellige typer som vist i efterfølgende skema, hvor teoribaserede antagelser i sagens natur typisk resulterer i modelegenskaber der henviser til forhold i referencerammen, mens designbaserede antagelser resulterer i referenceløse egenskaber. De to sidste typer, paradigmatisk antagelser og bekvemmelighedsantagelser, kan resultere i begge typer af modelegenskaber

| | Egenskaber med objektreference | Referenceløse egenskaber |
|---------------------------|--------------------------------|--------------------------|
| Teoribaserede antagelser | + | |
| Designbaserede antagelser | | + |
| Paradigmatisk antagelser | + | + |
| Bekvemmelighedsantagelser | + | + |

Teoribaserede antagelser vil som regel føre til modelegenskaber med reference til egenskaber ved referencerammen, fordi antagelserne drejer sig om egenskaber og relationer mellem egenskaber, der er defineret inden for referencerammen.

Designbaserede antagelser – som f.eks. den antagelse, at det er lykkedes at udtrække personerne ved tilfældig udtrækning fra referencerammen, og at eventuelt bortfald intet har at gøre med de variable, som modellen indeholder – er typiske antagelser om modelegenskaber uden referencer.

Paradigmatisk antagelser er udtryk for metodiske leveregler, som man retter sig efter, fordi “det er sådan man plejer at gøre” Det er ikke ualmindeligt at få artikler tilbage fra tidsskrifter med besked om at den statistiske metode, man har anvendt, ikke er den samme, som man normalt bruger i tidsskriftet. Hvis man skulle falde for fristelsen og ændre den statistiske analyse for at tækkes tidsskriftet, bliver analysen påvirket af paradigmatisk antagelser. I psykometriske sammenhænge stilles man ofte over for valget mellem faktoranalyser på den ene side og item responsanalyser på den anden. Der er dem, der altid vil vælge den ene form og

dem, der altid vil vælge den anden. Hvis man hører til en af disse fløje, er ens modeller derfor bl.a. præget af paradigmatisk antagelse. Inden for itemresponslejren har vi fløje med personer, der som princip vil vælge Rasch modellen som udgangspunkt for analysen, mens der er andre, der vil vælge den såkaldte 2-parameter model. Begge fløje bygger modeller, der er delvist baseret på paradigmatisk antagelse.

Bekvemmelighedsantagelser er antagelser, der tilføjes modellen med det formål, at gøre det muligt at gennemføre analyserne i praksis. En antagelse, om at sammenhængen mellem to variable er lineær, er som regel en bekvemmelighedsantagelse, der lægges til en referencerelateret antagelse om, at sammenhængen mellem variablene er monoton. Antagelsen om, at den betingede variabel i en lineær regressionsmodel har en (betinget) normalfordeling med en varians, der ikke afhænger af de uafhængige variable, er (næsten) altid en ren bekvemmelighedsantagelse.

Og – for nu at tale om noget af det, som jeg selv arbejder med – så er de psykometriske modellers krav om at items er *lokalt* uafhængige, også udtryk for et rent bekvemmelighedskrav. At definitioner af validitet og unidimensionalitet inkluderer kravet om lokal uafhængighed, kan måske få det til at se ud som om lokal uafhængighed er et udtryk for en fundamental item-egenskab, men det ændrer ikke – sådan som jeg ser det – ved det forhold, at der er tale om en ren bekvemmelighedsegenskab.

Urealistiske og realistiske modeller

Når der er tale om alt andet end meget elementære statistiske modeller, er modelkontrollen en vigtig del af de fleste statistiske analyser.

Overfladisk set har statistisk modelkontrol meget tilfælles med statistiske test, fordi modelkontrollen ofte omfatter beregning af en række teststørrelser for afledte hypoteser, der alle er sande, hvis modellen er korrekt. Analogien til hypotesetestningen kan derfor synes oplagt, fordi modelkontrollen kommer til at fremstå som et test af en nul-hypotese, der siger at modellen er sand, i forhold til et alternativ der siger, at den er forkert. Analogien til de statistiske test er så slående, at man er undskyldt, hvis man ikke rigtig kan se forskellen, men analogien er ikke desto mindre falsk. Hvis man tager definitionen af modeller for pålydende, kan der i sagens natur ikke være tale om nogen hypotese om at modellen er sand, fordi modeller er forenklinger. Det er tværtimod hele pointen med modellen, at den er falsk, men at den er det på en hensigtsmæssig måde.

Samtidig med at jeg fastholder, at dette er den rigtige måde at se på modellerne på, må jeg også erkende at synspunktet er problematisk, fordi det af mange bliver taget som et udtryk for, at man ikke behøver at bekymre sig så meget, når man får forkastet sine modeller. “Modeller er jo pr. definition *altid* forkerte” lyder det glade budskab.

Personlig kan jeg ikke dele glæden over dette budskab, selvom jeg er

enig i at det er misvisende at betragte modellen som en statistisk nulhypotese. Modelkontrollens opgave er efter min opfattelse ikke at prøve om modellen er sand, men kun at afprøve om en model er realistisk eller ikke realistisk. Ikke om det er den rigtige model, der er blevet konstrueret.

Spørgsmål om hvad det vil sige, at en model er urealistisk kan besvares fra to forskellige sider. En model er urealistisk, hvis den siger, at noget, der rent faktisk er forekommet, er (meget) *usandsynligt*. Og modellen er urealistisk, hvis den siger, at noget kan forekomme, som vi ved er *umuligt*.

Det er det første problem, som vi forsøger at vurdere i forbindelse med den statistiske modelkontrol. Modellen er urealistisk, hvis variationer i data tydeligt viser, at der er et eller andet, som modellen ikke tager højde for, enten fordi der er tale om systematiske tendenser, der ikke skulle være der ifølge modellen, eller fordi data fremtræder på en måde, der kun forekommer med meget lille sandsynlighed. Vi benytter bl.a. statistiske teststørrelser med beregninger af p-værdier for at vurdere graden af usandsynlighed, men logikken bag anvendelserne af de statistiske test til dette formål er en helt anden end i forbindelse med egentlige statistiske test, fordi der reelt set hverken er hypoteser eller alternativer på spil i forbindelse med modelkontrollen.

Den første betragtningsmåde ser altså på data fra modellens synsvinkel og stiller spørgsmål om, hvorvidt data indeholder noget der er usandsynligt ifølge modellen. Den anden betragtningsmåde ser på modellen ud fra datas synsvinkel og stiller det modsatte spørgsmål om, hvorvidt modellen påstår at noget er muligt, som vi ud fra vores viden om, hvordan data er indsamlet og kodet, på forhånd kan sige er umuligt. Hvis det er tilfældet, hvis der visse værdier, der har en positiv sandsynlighed for at forekomme ifølge modellen, selvom vi på forhånd med absolut sikkerhed kan sige, at sandsynligheden for disse værdier er lig med nul, så er det også et udtryk for, at modellen er urealistisk.

Det er det sidste problem, som Cox refererer til i det tredje af de fire citater, der indledte artiklen. Citatet kommer fra en introduktion til logistiske regressionsmodeller som stammer fra en tid, hvor lineære regressionsmodeller var det eneste praktisk tilgængelige redskab for mange mennesker. De logistiske regressionsmodeller er modeller, hvor den afhængige variabel kun kan antage to værdier, der typisk er kodet med tallene 0 og 1. Hvis man i stedet for en logistisk regressionsmodel bruger en lineær regressionsmodel, bruger man en model, der påstår at variabelen kan antage værdier fra hele talaksen fra $-\infty$ til $+\infty$. Nogle af disse værdier kan kun forekomme med meget små sandsynligheder, hvilket kunne få nogen til at argumentere for at modellen kunne fungere som en god approksimation til en realistisk model, men der er visse intervaller, mellem 0 og 1 vil ifølge modellen indeholde værdier med meget stor sandsynlighed. Cox's argument var, at det ikke var særlig smart at bruge lineære regressionsmodeller til denne type data, og at det f.eks. ville være bedre at bruge en logistisk regressionsmodel, fordi man på forhånd – uden overhovedet

at se på den empiriske tilpasning mellem data og model – kunne sige, at dette havde en mulighed for at være en realistisk model.

Problemet med de lineære regressionsmodeller for binære data var så indlysende, at Cox ikke behøvede at gentage argumenterne i den næste udgave af bogen om logistiske regressionsmodeller. Valget af logistisk regressionsanalyse til binære variable er i dag så indgroet, at det er at betragte som et paradigmatisk refleksbetinget og ikke et bevidst modelvalg, selvom der rent faktisk findes andre modeller, der ville passe til denne form for data.

Selvom forståelsen af betydningen af at vælge realistiske modeller til binære data er slået igennem i forbindelse med regressionsanalyserne, er der andre steder, hvor man tilsyneladende stadig befinder sig der, hvor regressionsanalytikerne befandt sig for 50 år siden. Betragt f.eks. en opgave i en pædagogisk test, som enten kan besvares rigtigt (kodet 1) og forkert (kodet 0). En psykometrisk model vil beskrive, hvorledes chancerne for at svare rigtigt på opgaven afhænger af en latent variabel, der er et udtryk for, hvor dygtig eleven er. Hvis Y er lig med svaret på opgaven, og T er lig med dygtigheden, skal den psykometriske model derfor være en regressionsmodel, der beskriver den betingede sandsynlighed, $P(Y=1|T)$.

I psykometriens barndom var den lineære regressionsmodel den eneste praktisk anvendelige model til dette formål. Det førte til udviklingen af faktoranalysemodellen, der beskriver svarene på et stort antal opgaver ved hjælp af lineære regressionsmodeller for forskellige afhængige variable (items), der alle antages at afhænge af de samme uafhængige latente variable. Faktoranalysemodellen er imidlertid præcis lige så urealistisk for dikotome items som den lineære regressionsmodel er det for en binær afhængig variabel og på præcis den samme måde. Det er denne erkendelse, der førte til udvikling af psykometriske itemresponsmodeller, hvor faktoranalysens lineære regressionsmodeller blev erstattet af logistiske regressionsmodeller. Dvs. modeller med den ene fordel i forhold til faktoranalysemodellerne, at man ikke på forhånd kunne sige, at modellerne var urealistiske.

Itemresponsmodellerne har en historie, der er næsten lige så lang som de logistiske regressionsmodeller, og argumenterne for at bruge dem frem for faktoranalysemodeller er præcis de samme som argumenterne for at bruge logistiske regressionsmodeller i stedet for lineære regressionsmodeller. Det er derfor ganske bemærkelsesværdigt, at det skift, som man har kunnet se mht. til præferencen for de logistiske regressionsmodeller frem for lineære regressionsmodeller, ikke har fundet sted i forbindelse med de psykometriske modeller. Der drejer naturligvis i lige så høj grad tale om paradigmatisk valg som inden for regressionsanalyserne, men der tilsyneladende er ikke noget dominerende paradigme inden for psykometri. Der er psykometrikere der altid vælger (bestemte typer af) itemresponsmodeller, og der er psykometrikere, der altid vil vælge faktoranalysemodeller, uanset at modellerne a priori må betragtes som fuldstændig

urealistiske.

For fuldstændighedens skyld skal det nævnes, at der eksisterer en tredje mulighed, der giver nyt liv til faktoranalyserne. Antag, at der for hver opgave er tale om en latent respons variabel, der angiver kvaliteten af besvarelsen af opgaven på en skala fra $-\infty$ til $+\infty$, men at det, der er blevet registreret i forbindelse med dataindsamlingen, kun er en vurdering om kvaliteten af besvarelsen ligger under eller over en mindste værdi for en tilstrækkelig god besvarelse. Antag desuden, at fordelingen af den latente respons svarer til en lineær regressionsmodel – at det, der er observeret er en kategoriseret normalfordeling. Sådanne modeller, der omtales som probitmodeller, kan for det første ikke på forhånd – før den statistiske modelkontrol - bortdømmes som urealistiske, og det, de i øvrigt siger om sandsynlighederne for korrekte svar på opgaverne, ligger så tæt på det som itemresponsmodellerne siger, at man ikke empirisk kan skelne mellem modellerne. Den ene modeltype altid kan altså altid fungere som approksimationer af den anden

Referenceløse modelegenskaber

Når den statistiske model skrues sammen, er det naturligvis hensigten, at gøre det på en sådan måde, at der er visse egenskaber ved objekterne i referencerammen, der modsvares af egenskaber ved modellerne, således at man efterfølgende kan tolke modelegenskaberne som udtryk for forhold, der er defineret i referencerammen. En væsentlig del af modelkonstruktionen drejer sig altså om at tage stilling til hvilke egenskaber ved individerne, som modellen skal indeholde referencer til, og det metodeproblem, som man naturligt bekymrer sig om, er om modellen er et udtryk for overforenkling, dvs. om der er noget, man har glemt at få med i modellen. Ved siden af dette problem er der imidlertid det andet og mere subtile problem, at modeller ofte indeholder meget andet og mere end forhold, der referer tilbage til referencerammen.

En epidemiologisk analyse af risikoen for blodprop vil bl.a. resultere i estimater af en række parametre, der er udtryk for kvantitative mål for hvor meget forskellige risikofaktorer bidrager til den samlede risiko. Det dominerende epidemiologiske metodeproblem er i denne forbindelse spørgsmålet om den confounding af risikoestimerne, der vil forekomme, hvis man enten har glemt at inkludere nogle vigtige risikovariabler i modellen eller hvis nogle af risikovariablerne er blevet kodet på en for grov og uhensigtsmæssig måde. Men der er også et andet problem, nemlig at estimaterne kun kan tolkes under den forudsætning, at det har god mening at effekten af risikofaktorerne kan kvantificeres, og at det har god mening, at sammenligninger mellem forskellige risikofaktorer altid giver samme resultat, uanset hvilke personer, der er tale om, og hvilke værdier de andre risikofaktorer har. De fleste, der foretager sådanne analyser, vil

næppe se dette som et stort problem, men det er værd at understrege, at den mening, som risikoparametrene måtte have, *skal* defineres i referencerammen. Det forhold, at parametrene findes i den statistiske model, betyder i hvert fald ikke i sig selv, at de har mening. Parametrene kunne jo være dukket op i forbindelse med forsøg på at gøre en ikke-realistisk model mere troværdig ved at tilpasse den effekt, som modellen i første omgang beskrev som en lineær effekt, til en mere kompliceret ikke-lineær funktion, der afhang af nogle flere ukendte parametre end dem, som man i første omgang havde forsøgt sig med, og som man syntes havde mening.

Og det fører os så til det problem der skal diskuteres i dette afsnit: at der er dele af den statistiske model, der intet har med virkeligheden at gøre. For modeller som helhed er dette ikke nogen overraskelse. Et landkort kan være en fremragende model af Europa, men der er ingen, der vil tage landkortet så bogstaveligt, at de tror at Jorden er flad, at den er lavet af papir, og at forskellige vejtyper i virkeligheden har den farve, som de har på landkortet. For statistiske modeller er situationen den samme, men her er det både mere uigennemskueligt, hvad der har og hvad der ikke har referencer til virkeligheden, og mere paradoksalt fordi det kan være, at der er nogle af modellens mest centrale egenskaber, dem der er hele forudsætningen for at kunne gennemføre statistiske analyser, der kan vise sig at være referenceløse.

Tænk f.eks. på selve sandsynlighedsbegrebet. Det er et fælles træk ved alle statistiske modeller, at de fænomener, som modellerne beskriver, beskrives som tilfældige og ikke som deterministiske fænomener. Hvis jeg havde deltaget i en surveyundersøgelse af vælgeradfærd efter sidste folketingsvalg, ville den statistiske model beskrive min vælgeradfærd som tilfældig, selvom jeg til det sidste vil argumentere for, at hvis der var noget, den ikke var, så var det tilfældig. Mit påstand vil være, at den tilfældighed, som modellen postulerer, er en konsekvens af den måde, som surveyinstituttet havde designet undersøgelsen på. At tilfældigheden med andre ord skyldes surveyforskernes synsvinkel, og at den ikke henviser til noget som helst i referencerammen som sådan. Det er i sig selv galt nok, men argumentet har den besværlige konsekvens, at det også må gælde for sandsynlighederne, der er det helt centrale aspekt ved de statistiske modeller. Hvis det ikke har mening at beskrive fænomener som tilfældige, så er det naturligvis lige så meningsløst at udtale sig om sandsynligheder som noget, der har med virkeligheden at gøre. De kan ikke henviser til noget i referencerammen, hvis fænomenerne ikke er tilfældige. Og hvis sandsynlighederne ikke betyder noget, hvad så med de parametre, som indgår i de formler, vi bruger til at beregne sandsynlighederne? De må i sagens natur også være referenceløse.

Der er flere måder at imødegå disse argumenter på. Frekventister³

³ Bayesianere, der bruger modeller, hvor det ikke bare er variablene, men også parametrene, der beskrives som tilfældige, vil formodentlig argumentere for at

vil med en vis ret argumentere for at sandsynlighederne i virkeligheden er udtryk for de relative hyppigheder hvormed fænomenet forekommer i en bestemt population, selvom argumentet får problemer, lige så snart man begynder at generalisere til andre (eventuelt endnu ikke eksisterende) populationer. Som fagstatistiker kan man også slippe uden om problemet ved – med rette – at sige at det egentlig er ligegyldigt om sandsynlighederne referer til et eller andet som forskerne har skabt ved at designe undersøgelsen på en bestemt måde eller om de (også) referer til fænomener, der virkelig er tilfældige. Den statistiske analyse bliver den samme uanset, om det er den ene eller anden situation, der er gældende.

Problemet bliver imidlertid ikke mindre af at blive ignoreret. Uanset om man synes at problemet er besværligt eller ej, så er det altså sådan, at sandsynlighederne i mange statistiske modeller ikke henviser til noget inden for modellens referenceramme. Og hvis dét er tilfældet – hvis det virkelig er sådan at de helt dominerende egenskaber ved de statistiske modeller ikke henviser til noget – så er det ikke vanskeligt at forestille sig, at der også kan være andre af modelegenskaberne som er referenceløse.

Og så er der faktisk situationen, hvor det er afgørende, at man er enige om hvorvidt modellens sandsynligheder betyder noget eller ej. I hvert fald, hvis man vil forstå, hvad man taler om, og vil forstå baggrunden for udviklingen af forskellige statistiske metoder.

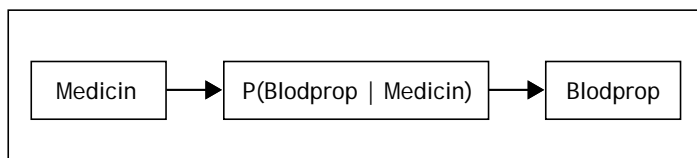
Som eksempel på dette kan vi f.eks. se på det epidemiologiske risikobegreb på den ene side og synsvinkler på og metoder til kausalanalyse på den anden side. Risikobegrebet kan ud fra en overfladisk betragtning synes enkel. Risikoen for at få en blodprop er ganske enkelt lig med *sandsynligheden* for at få en blodprop. Med hensyn til metoderne til kausal analyse er det desværre mindre enkelt. Der er i knap 20 år blevet arbejdet intensivt med statistiske metoder til analyse af kausalitet, men der er stadig ikke konsensus omkring metoderne, bl.a. fordi det ikke er ganske klart, hvad forskellige mennesker mener med sandsynligheder og dermed med selve risikobegrebet. Er forekomst af en blodprop noget, der sker tilfældigt, hvor hver af os har en vis sandsynlighed for at det vil ske, eller er der tale om deterministisk bestemte forhold, hvor det i princippet er forudbestemt, om vi får eller ikke får en blodprop. Hvis det er det første, der er tilfældet, referer den statistiske models sandsynlighed til noget i referencerammen – til den enkelte persons risiko. Hvis det er det andet refererer sandsynligheden kun til den forudbestemte hyppighed af blodpropper i *populationen*. Risikoen er populationsrelateret, og det har ikke mening at udtale sig om den enkelte persons risiko. At det kan se mere eller mindre tilfældigt ud skyldes udelukkende undersøgelsens design på fuldstændig samme måde, som det var tilfældet med undersøgelsen af

sandsynlighederne er udtryk for forskerens *subjektive* vurdering af chancerne for at variable og parametre har forskellige værdier, hvorved tilfældigheden tillægges en helt ny betydning, langt fra referencerammen.

vælgeradfærden.

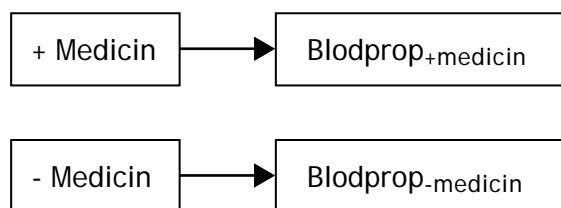
Disse overvejelser har haft en gennemgribende effekt på udviklingen af metoder til kausalanalyse og har ført til mange misforståelser, og megen ”snakken ved siden af hinanden”. Tænk igen på eksemplet med blodpropperne. Risikoen for at få en blodprop afhænger bl.a. af blodtrykket og kolesteroltallet. Hvis nogle af disse tal er for store, er der en forøget risiko for at få en blodprop, af hvilken årsag de fleste læger bl.a. vil anbefale at man tager medicin, der kan hjælpe til at få tallene ned på et uproblematisk niveau. Tænkningen bag disse anbefalinger er rendyrket kausal. Blodtryksmedicin er en direkte årsag til et reduceret blodtryk og dermed indirekte årsag til en reduceret *risiko* for at få en blodprop. Manglende blodtryksmedicin har den modsatte effekt i form af et fortsat højt blodtryk og derfor en fortsat forøget risiko for at få en blodprop. Det er i hvert fald teorien. For at afprøve den i forhold til en ny medicin vil man typisk foranstalte et kontrolleret forsøg, hvor et tilfældigt udtræk af patienterne får medicinen, mens resten får placebo. Forsøget tilfører altså noget tilfældighed, der kommer til at påvirke sandsynlighederne i den statistiske model, der skal bruges til analysen. Måden den påvirker sandsynlighederne på er imidlertid stærkt afhængig hvilken synsvinkel man anlægger i forhold til risikobegrebet. De to forskellige synsvinkler er illustreret i figurerne 2 og 3. Begge modellerne indeholder deterministiske aspekter, men på meget forskellig måde.

I den probabilistisk kausale model (figur 2), antages det, at medicinen har en deterministisk effekt på sandsynligheden – personrisikoen – for at personen får en blodprop. Hvorvidt der rent faktisk forekommer en blodprop afhænger derimod kun af sandsynligheden, og ikke af hvorfor risikoen skulle være større eller mindre. At forsøget er randomiseret har konsekvenser for sandsynligheden, men ikke for den effekt, som sandsynligheden har på forekomsten af blodproppen.



Figur 2. Probabilistisk kausalmodel.

For at håndtere den statistiske analyse indenfor en referenceramme, hvor kausaliteten betragtes som deterministisk, er man (jf. figur 3) nødt til at antage, at der eksisterer to counterfactuals med oplysninger om hvad der sker, hvis personen havde fået medicin ($\text{Blodprop}_{+\text{medicin}}$), og hvad der sker, hvis personen ikke tager medicin ($\text{Blodprop}_{-\text{medicin}}$). Ideen er, at begge variable eksisterer i hvert sit parallelle univers og at randomiseringen er udtryk for at der vælges tilfældigt, hvilken en af de to variable man har i sinde at iagttage.



Figur 3. Deterministisk kausalmodel med såkaldte counterfactuals

Når der kan være uklarhed om noget så fundamentalt som betydningen af tilfældigheden i de statistiske metoder, er det klart, at referenceløse modelegenskaber kan dukke op på mange forskellige måder. Risikoen er formodentlig størst i situationer, hvor den statistiske model er baseret på for mange bekvemmelighedsantagelser som tilpasses i lyset af de foreliggende data. Resultatet af en lineær regressionsanalyse er f.eks. en model, hvor den betingede fordeling af den afhængige variabel er normal, men hvor antagelsen om normalfordelingen ofte er en typisk bekvemmelighedsantagelse. Den er inkluderet fordi den statistiske analyse bliver særlig enkel, hvis antagelsen er realistisk, men der er sjældent nogen saglige argumenter for antagelsen. Hvis modelkontrollen afslører, at modellen er urealistisk, fordi residualerne har en tydelig skæv fordeling, vil man ofte forsøge med transformationer af den afhængige variabel (logaritme- eller kvadratrodstransformationer er særlig populære), men i sådanne situationer er der ingen, der ville rapportere dette som et værdifuldt resultat af undersøgelsen, fordi selve påstanden om en normal fordeling ikke har nogen selvstændig mening hverken for variable selv eller for den transformerede variabel.

Inferensrammer

En inferensramme er en delmængde af sandsynligheder, der kan udledes ud fra modellens grundsandsynligheder, og som kan danne udgangspunktet for en del af den statistiske analyse, fordi analysen indenfor inferensrammen kan afdække resultater om nogle af modellens parametre.

Inferensrammer kan defineres på mange forskellige måder. De mest almindelige er inferensrammer defineret ved

- marginale fordelinger, der kun indeholder et mindre antal af modellens samlede sæt variable,
- betingede fordelinger, hvor der ses på fordelingen af visse variable givet værdierne af andre variable,
- fordelinger af summariske størrelser.

En inferensramme er i sig selv en reduceret og dermed forenklet statistiske model, men referencerammen for en inferensramme-modellen udgøres af den statistiske model og kun indirekte (formidlet af grundmodellen) af den oprindelige referenceramme.

I teorien for Raschmodellerne optræder der således (mindst) tre forskellige betingede inferensrammer:

Den første inferensramme er defineret af den betingede fordeling af itemresponserne givet den samlede score for hver person. Denne modeltype omtales ofte som den *betingede* Raschmodel. Inferensrammen defineret af den betingede Raschmodel bruges til at estimere modellens itemparametre og til at kontrollere visse aspekter af Raschmodellen. I forhold til den oprindelige Raschmodel er den betingede Raschmodel en forenkling, fordi personparametrene er blevet elimineret, men i forhold til referencerammen for modellen er den *betingede* Raschmodel et udtryk for en ødelæggende overforenkling som model for personerne fordi det er personparametrene og ikke itemparametrene der er modellens primære interesse-parametre.

Den anden inferensramme er defineret af den betingede fordeling af itemresponserne, givet item-marginalerne – antallet af positive responser på de enkelte items. Denne inferensramme, hvor itemparametrene er blevet elimineret, er sjældent anvendt, men kunne bruges til at estimere personparametrene.

Og endelig er der den tredje inferensramme defineret ved den betingede fordeling af itemresponserne, givet både person scores og item-marginal. I denne inferensramme er alle ukendte parametre blevet elimineret. Denne inferensramme kan derfor ikke benyttes til estimation af nogen parametre men kun til kontrollere modellen at modellen empirisk set er realistisk.

Mulighederne for at skifte mellem forskellige inferensrammer har haft stor betydning for den måde vi forstår og arbejder med Raschmodellerne, men det forekommer også i mere simple rutineprægede statistiske analyser, hvor der vælges referenceramme alt efter hvilke redskaber, man har til rådighed.

Tænk f.eks. på en undersøgelse, hvor der foreligger sideordnede oplysninger om personens skattepligtige indtægt og personens selvrapporterede helbred registreret i et endeligt antal kategorier fra godt til dårlig, og antag, at det man er interesseret i, er en beskrivelse af sammenhængen mellem de to variable, og at man i den sammenhæng ønsker at kontrollere for køn.

En sådan analyse kan forekomme vanskelig, dels fordi der ikke er nogen veldefineret (kausal) retning mellem de to variable, og dels fordi der er tale om variable på meget forskellige skalaniveauer. Indtægten er en ratioskala, mens helbredet er en ordinal kategorial skala.

Den statistiske model, som skulle bruges til dette formål, skulle være en multivariate regressionsmodel, $P(\text{Indtægt, Helbred} \mid \text{Køn})$, hvor indtægt og helbred betragtes som afhængige af køn, og hvor relationen

mellem køn på den ene side og indtægt og helbred på den anden kunne tages som udtryk for en kausal relation. Interessen for sammenhængen mellem indtægt og helbred kan bedst beskrives ved en eller anden form for partiel korrelationskoefficient, men det vil desværre være ganske kompliceret for de fleste, både at finde den rigtige form for korrelationskoefficient fordi de to variable er på meget forskellige skalaniveauer og fordi en model, der i givet fald behandlede helbredet som en normalfordelt variabel, ville være en ekstremt urealistisk model. I stedet ville de fleste formodentlig foretrække at skifte inferensramme og se på den betingede fordeling af indtægten givet køn og helbred. Man ville altså erstatte modellen $P(\text{Indtægt, Helbred} \mid \text{Køn})$ med en regressionsanalyse baseret på $P(\text{Indtægt} \mid \text{Helbred, Køn})$, hvilket dels ville være helt tilladeligt og dels enkelt, fordi spørgsmålet om relationen mellem indtægt og helbred kunne klares med en tosidet variansanalyse. Det eneste man skal huske, når resultaterne tolkes, er, at de parametre der beskriver effekten af helbred på indtægt, ikke beskriver kausale effekter. At kausale analyser kræver en eller anden form for regressionsanalyse, men at regressionsanalyser ikke kræver kausalitet. Hvis man har fuldstændig styr på det, vil der ikke være problemer med analysen. Hvis man derimod forsøger at tolke resultaterne på en måde, der antyder kausale forhold, begynder man at tolke på forhold, der ikke henviser til noget i referencerammen. Man fejltolker resultaterne.

Utilgængelige problemer

Det bør ikke komme som en overraskelse, at en analyse af data inden for en bestemt inferensramme automatisk medfører, at der er visse problemer, der ikke kan tages op, fordi inferensrammen netop er defineret således at visse forhold er blevet udeladt, mens andre ikke er. En betinget Raschanalyse af itemresponsere, givet de samlede personscores, kan således fortælle noget om itemparametrene, men ikke om personparametrene. Det besværlige er imidlertid, at dette også er tilfældet for den statistiske model som helhed, også selvom modellen er realistisk med perfekt tilpasning til data og hverken inkluderer paradigmatisk antagelser eller bekvemmelighedsantagelser. Inden for en given grundmodel er der visse problemer, der kan rejses, og andre, der er utilgængelige.

Figur 4 og 5 viser to diagrammer – to netværk – der illustrerer resultaterne af to multivariate statistiske analyser⁴. Den første er en analyse af data fra et panelstudie, der inkluderede oplysninger fra 1968, hvor paneldeltagerne gik i 8. klasse, og oplysninger om livsforløbet frem til og med 1992. Den anden viser resultaterne af analysen af data fra et

⁴ Analyserne og en række metodiske problemer med anvendelsen af grafiske modeller er beskrevet i Kreiner (2005) og Kreiner og Keiding (2006).

longitudinelt studie med gentagne målinger af forbruget af det danske sundhedsvæsen i to forskellige amter.

Netværk af den type, der er vist i de to figurer, optræder i forbindelse med to forskellige typer af statistiske modeller. For det første i forbindelse med modeller for strukturelle ligninger og for det andet i forbindelse med grafiske modeller. Begge modeltyper er defineret ved graferne i den forstand, at graferne fortæller alt det om de statistiske fordelinger af variablene, som vi har brug for at vide for at kunne foretage en statistisk analyse. Tolkningen af graferne forekommer også intuitivt tilforladelige. En del af pilene kan tolkes i termer af kausale relationer, og vi kan spore både direkte og indirekte effekter i modellerne. Tolkninger som på overfladen er de samme, uanset om vi har brugt den ene eller den anden form for modellerne.

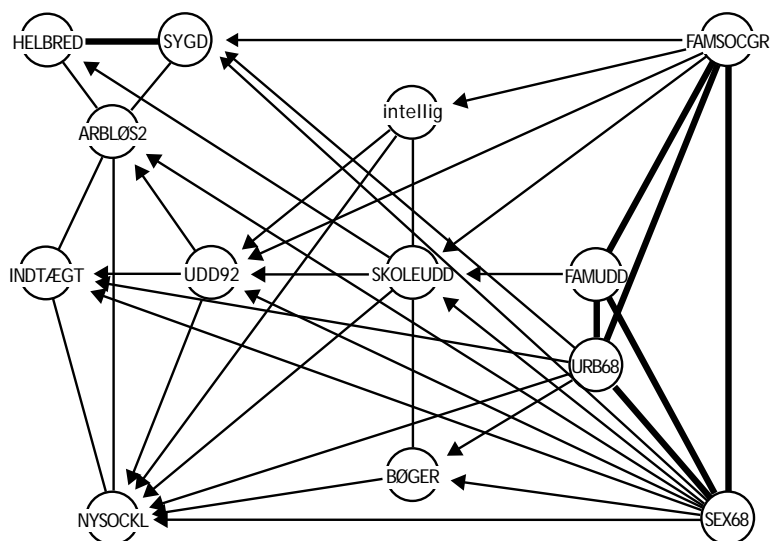
Hvis vi ser nærmere på forskellene mellem de to typer af modeller, viser det sig imidlertid, at de betyder helt forskellige ting, og at disse forskelle har at gøre med de problemer, som de forskellige former for modeller tillader os at tage op.

Begge modeltyper beskriver sammenhængen mellem variablene, men de gør det på meget forskellig måde. Modeller for strukturelle ligninger fokuserer på korrelationerne mellem variablene, mens de grafiske modeller interesserer sig for *partielle* korrelationer, dvs. korrelationer, hvor der er kontrolleret for den betydning, som andre variable måtte have.

I forbindelse med en analyse ved hjælp af strukturelle ligninger fokuserer man på kanterne og pilene i modellen og stiller det krav, at modellen skal kunne reproducere de marginale korrelationer på en sådan måde at modellens estimat af en bestemt sammenhæng skal svare så godt som muligt til den marginale (ukontrollerede) korrelation i data. En manglende forbindelse mellem to variable i en model for strukturelle ligninger betyder hverken, at de to variable er betinget uafhængige, eller at de er marginalt uafhængige. Spørgsmål om betinget uafhængighed er i særdeleshed ikke noget som disse modeller kan bruges til at afprøve. En analyse ved hjælp af modeller for strukturelle modeller er altså hverken en elaboreringsanalyse, sådan som den klassiske sociologiske metodologi definerer den, eller en analyse, der kontrollerer confounding, sådan som epidemiologerne forstår det.

Betinget uafhængighed står til gengæld i centrum i analyser ved hjælp af grafiske modeller. Mens modellerne for strukturelle ligninger definerer sig ved værdierne af korrelationerne mellem de variable, der ifølge grafen hænger direkte sammen, definerer de grafiske modeller sig ved at insistere på at den *partielle* korrelation mellem to variable *uden* en forbindelse i grafen er lig med nul – at de to variable er *betinget* uafhængige. Set fra denne synsvinkel ligger analyser ved hjælp af grafiske modeller tættere på sociologisk elaborering og epidemiologisk confounderkontrol end analyser ved hjælp af modeller for strukturelle ligninger. Dette kan man betragte som en fordel, men de grafiske modeller har til gengæld deres egne be-

grænsninger mht., hvad der kan stilles spørgsmål til, og hvad der ligger uden for det muliges rammer. Den grafiske model, der er defineret af figur 4, fortæller os således at intelligens og indtægt er betinget uafhængige – at intelligensen i 1968 ikke havde nogen direkte effekt på indtægten 25 år senere. Spørgsmålet om en påviselig marginal effekt, der kunne skyldes en indirekte effekt af intelligensen formidlet af uddannelse, arbejdsløshed og socialklasse, kan til gengæld ikke rejses inden for rammerne af den grafiske model. Dette spørgsmål besvares til gengæld af den model for strukturelle ligninger, der er defineret af figur 4. I denne model er det uklart, hvad relationen mellem variablene er fordi der ikke er nogen parametre, der beskriver sammenhængen: den partielle korrelation er forskellig fra 0, således at variablene ikke kan være betinget uafhængige hvilket igen betyder, at det er uklart i hvilket omfang der er tale om direkte og indirekte sammenhænge mellem variablene, men til gengæld er denne model mere fast i mælet omkring den marginale (ukontrollerede) korrelation mellem variablene. Den er lig med $r = 0.20$.

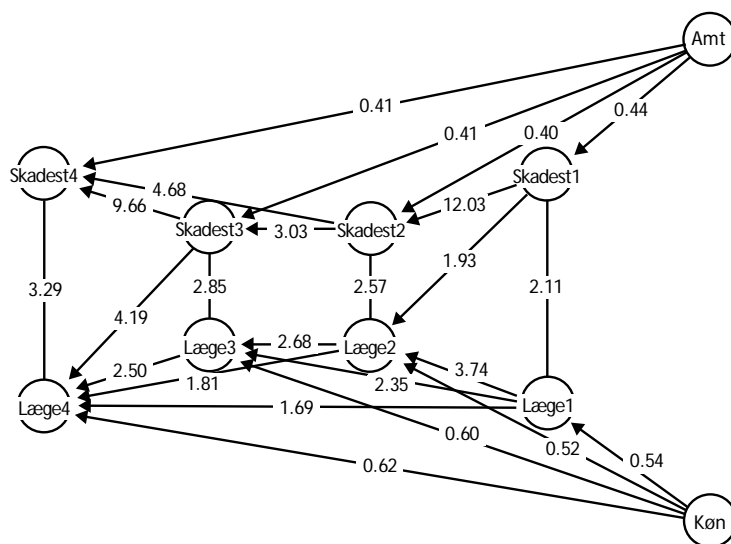


Figur 4. Multivariat statistisk model der beskriver sammenhængen mellem opvækst vilkår i 1968 og senere levevilkår i 1992. Figuren stammer fra Kreiner (2005), hvor analysen ved hjælp af grafiske modeller er diskuteret og beskrevet.

Foruden de nævnte forskelle mellem grafiske modeller og strukturelle modeller er der en yderligere forskel, der har interesse set fra både en klassisk sociologisk og en epidemiologisk synsvinkel. Begge metodeparadigmer interesserer sig for forekomsten af interaktioner i modellerne. Sociologerne i forbindelse med specifikationen af sammenhængene og epidemiologerne i forbindelse med analyse af effektmodifikation. Det inter-

essante problem er, at spørgsmål om interaktioner hverken kan tages op af grafiske modeller eller af modeller for strukturelle ligninger.

En grafisk model for kategoriske variable – den rå form uden andre antagelser end dem, der er defineret af grafen – vil tage det som givet, at der *er* tale om interaktioner – at der er tale om effektmodifikation, og vil end ikke være med til at formulere en nul-hypotese om at der ikke er interaktion. I modsætning hertil tager en model for strukturelle ligninger det præcis modsatte synspunkt, at interaktioner ikke kan forekomme. De eksisterer simpelthen ikke. De to modeltype er altså enige om at interaktioner er noget de ikke vil forholde sig til, men de er det på meget forskellig vis. De grafiske modeller vil nægte at formulere en nul-hypotese om at der ikke er interaktion, mens modeller for strukturelle ligninger vil nægte at være med til at formulere alternativet⁵.



Figur 5. Markovgraf for sundhedsadfærden beskrevet ved kvartalsvis opgørelse over besøg på skadestuen og kontakt med lægen. Tallene i figuren er Mantel-Haenszelestimer af styrken af sammenhæng (odds-ratio) beregnet under modellen. Figuren stammer fra Kreiner og Keiding (2006) hvor analysen ved hjælp af grafiske modeller er beskrevet og diskuteret.

⁵ Ekspertene i grafiske modeller og modeller for strukturelle ligninger vil – hvis de læser dette – formentlig belære mig om, at modeller for strukturelle ligninger kan indeholde baggrundsvARIABLE og interaktioner mellem effekten af disse baggrundsvARIABLE på modellens øvrige variable, samt om at parametriske grafiske modeller for normalfordelte variable heller ikke indeholder interaktioner. Dette er naturligvis korrekt, men det jeg diskuterer her er generelle grafiske modeller og generelle modeller for strukturelle ligninger uden supplerende antagelser.

Figur 5 viser en anden situation, hvor modellen kunne være både en grafisk model eller en model for strukturelle ligninger, og hvor der er præcis de samme problemer som i figur 4. Situationen i figur 5 er imidlertid endnu værre, fordi der er tale om gentagne målinger med naturlige og vigtige problemer som ikke kan tages op af hverken den ene eller den anden modeltype. Dette – samt løsningerne på problemet – er beskrevet andet steds (Kreiner og Keiding, 2006), så lad mig nøjes med at påpege, at ingen af de to modelapparater tillader at der stilles spørgsmål til om de sammenhænge, som modellerne påstår findes, er konstante eller varierende på tværs af tiden eller om hyppigheden af skadestuebesøg og lægebesøg er den samme eller forskellige i årets fire kvartaler.

Problemer med spørgsmål, som ikke kan tages op inden for rammerne af en bestemt model, er et generelt problem for alle typer af modeller og ikke kun de specielle modeller, som jeg har talt om her. Den kritiske læser vil måske sige, at så må man jo bare lave modellerne om, hvilket naturligvis er korrekt. Vi har i Kreiner og Keiding (2006) beskrevet flere forskellige måder at tilpasse de grafiske modeller således at vi kan svare de spørgsmål vi gerne vil stille. Men det er ikke alene korrekt, det er faktisk også pointen med hele argumentationen. Både de grafiske modeller og modeller for strukturelle ligninger nægter at forholde sig til en række spørgsmål. Hvis man insisterer på at få disse spørgsmål besvaret, må man altså konstruere nogle nye modeller. Det er nogle gange nemt – andre gange ikke så nemt. Og hvad værre er – det er ikke altid så let at gennemskue, at problemet overhovedet er der, fordi komplicerede statistiske modeller har en tendens til at overtage styringen af, hvad der foregår, således at det let kommer til at se ud som om formålet med analysen blot er at konstruere og fitte modellen.

Afsluttende kommentarer

Titlen på denne artikel indeholder et halvkvædet spørgsmål, som artiklen ligeså halvkvædet har forsøgt at besvare. Statistiske modeller ér både udtryk for forenklinger og for forvanskninger og fortrængninger. Jeg har forsøgt at argumentere med eksempler fra de modeltyper, som jeg selv har arbejdet med, og beklager at fremstillingen er blevet temmelig indforstået fordi pladsen ikke har tilladt, at jeg gjorde ret meget ud af at definere modellerne og forklare, hvordan de fungerer. Mit forsvar over for den kritik vil være, at læseren i stedet kan se på de statistiske modeller, som han eller hun selv har brugt. Jeg er sikker på, at de problemer, som jeg har fundet i mine modeller, også dukker op der.

Spørgsmålet om, hvor vigtige problemerne er, og om nogle modeltyper er bedre end andre, har jeg omhyggeligt undgået at sige ret meget om. Jeg deler det synspunkt, der er udtrykt i citatet af Cox, at det generelt ikke er særlig smart at bruge ikke-realistiske modeller, med mindre man

kan påvise, at den ikke-realistiske model er en god matematisk approksimation til en realistisk, men i praksis uanvendelig model. Bortset herfra kan modellens kvaliteter ikke vurderes uden reference til det, den er en model af. Hvis det virkelig er irrelevante forhold, der er udeladt, hvis en urealistisk model er en god approksimation til en realistisk model, hvis man ikke baserer sin tolkning af analysens resultater på referenceløse modelegenskaber, og hvis man ikke har interesse i de problemer, som modellerne fortrænger, så er det en god model. Men at diskutere om grafiske modeller generelt er bedre eller dårligere end modeller for strukturelle ligninger eller om faktoranalyse modeller er bedre eller dårligere modeller end itemresponsmodeller er efter min mening en diskussion af meningsløse spørgsmål.

Referencer

Box, GP & Draper NR (1987) *Empirical Model-Building and Response Surfaces*. Wiley

Cox, DR (1969) *Analysis of Binary Data*. Chapman & Hall

Karpatchof, B (2000) *Human Activity*. Psykologisk forlag

Kreiner S (2005) Intelligens og livsforløb. I Bryderup I (ed): *Social ulighed*. Side 155-176. Danmarks pædagogiske Universitets Forlag

Kreiner S & Keiding N (2006) Om analyse af gentagne kategoriserede målinger. I Vallgård og Koch (red) (2006) *Forskel og ulighed*. Side 123-143. Munksgaard.

Kreiner S (2007) *Statistisk problemløsning. Præmisses, teknik og analyse. 2. udgave*. Jurist og økonomforbundets Forlag.

On the choice between traditional item analysis and item response models in psychology

Af *Erik Lykke Mortensen*¹

Introduction

For several decades I have been involved in the analysis of many psychological tests, questionnaires, and rating scales. Typical tasks have been to analyze standardized tests that have been used for years in this country or to analyze questionnaires that are used for the first time — either because they have recently been translated into Danish or because they have been constructed from scratch in Danish for a specific research purpose. I have never been in an ideal situation for test construction and analysis — that is a situation with large sets of potential items that are being tried on large pilot samples. This fact has probably influenced my attitude to the practical use of item response models in psychology.

This situation has not changed substantially during the last 15 years, but in addition to small-sample clinical studies, self-report rating scales are now increasingly used in Danish population based studies with much larger samples.

¹ In October 1992, I presented a paper “on the choice between traditional item analysis and item response models in psychology” at an international meeting on Rating scales: Testing validity and reliability. It has been a pleasure for me to accept the editor, Jan Ivanouw’s invitation to contribute to the *Festschrift* to Benny Karpatschhof with an updated version of this paper. In updating the paper, I have kept the original structure and the original text to the extent that this was possible.

Advantages of item response models

Using item response models in test and scale construction offers a number of conceptual and practical advantages, and there are many good reasons why the sophisticated machinery of item response analysis should replace the comparatively primitive methods of analysis in traditional psychometrics. The advantages of item response models are in fact so obvious that we should be careful not to overlook the problems involved in routinely applying these models to existing instruments in psychology, and my main purpose is in fact to explain why item response analysis is not my first choice in many test analytic situations.

Basically, I will argue that item response models may be very useful in the construction of new tests and scales, if sufficient resources are available. However, the assumptions of item response models are generally unrealistic for the analysis of most existing psychological tests, because in both the cognitive and personality domains existing tests have rarely been developed on the basis of item response models.

The Danish attitude to item analysis

In Denmark the word “item analysis” almost exclusively describes analysis with the Rasch one-parameter item response model. This is perhaps understandable in the native country of Rasch, but I do believe that this has had two unfortunate consequences:

1. It is sometimes assumed that the one-parameter model is the only item response model.
2. It is also assumed that item analysis has not been part of traditional psychometric methods.

The description of classical item response models for binary items and the associated item characteristic curves (ICC) by Embretson & Reise (2000) shows that the first assumption is obviously wrong. The one-parameter Rasch model assumes that an item difficulty parameter is sufficient to describe ICCs, while the two- and three-parameter models assume that item discrimination and guessing parameters are also needed. Consequently, it may be argued that the defining characteristic of Rasch models is the assumption that the items in a scale can be described without including discrimination and guessing parameters (Lord & Novick, 1968). In 1992 a number of textbooks were available that compared the Rasch model with other item response models (Hambleton & Swaminathan, 1985; Hambleton et al, 1991).

In some areas of psychology it may indeed be possible to construct collections of items for which item discrimination can be assumed to be constant and for which guessing or random noise plays no role. However,

since most existing psychological tests were constructed without attention to item discrimination and guessing parameters, it is likely that items will differ with respect to these attributes, and consequently the one-parameter item response model seems unrealistic for these tests. In fact, in some areas of psychology — like the measurement of broad personality traits — it may be impossible to construct collections of items for which the one-parameter model is realistic.

Consequently, the conclusions of Jones & Appelbaum should be no surprise:

“Unfortunately, in most testing applications, the Rasch model provides a rather poor fit to the data. To achieve a better fit of model to data for multiple choice tests, evidence is mounting in favour of the 3-parameter item response model when sample size is sufficient for good parameter estimation” (Jones & Appelbaum, 1989).

Thus, research supports common sense: The one-parameter model is unrealistic for many instruments and the three-parameter model should be preferred. Apparently, the one-parameter model is often the only one used in Denmark, and in 1992 it was my impression that no three-parameter estimation program was generally available. The unfortunate consequence was that different item response models were rarely compared in Danish research — even when analyzing instruments for which one- or three-parameter item response models seemed much more realistic.

During several decades many item response models have been developed to handle different response formats and to handle different assumptions about the relationship between item responses and the latent trait of interest (Embretson & Reise, 2000; van der Linden & Hambleton, 1997). Many programs have also been developed to analyze these models (e.g. Du Toit, 2003), and researchers and data-analysts can choose among a large number of models and corresponding software, including non-parametric item response models (Sijtsma & Molenaar, 2002). It is, however, my impression that few Danish psychologists, researchers and students conducting item analyses are familiar with alternatives to Rasch models and the corresponding software

Actually, there appears to be a kind of paradox in Danish psychological statistics because in inferential statistics non-parametric - so called distribution free tests - have generally been preferred in spite of the fact that parametric statistical tests have been demonstrated to be relatively insensitive to violations of assumptions in many situations. In contrast, Danish item analytic parameter estimation has been based on strong and probably unrealistic assumptions — in spite of the fact that much less is known about the consequences of failure to meet assumptions in this area.

Since 1992, regression models are increasingly used in Danish psychological research (often replacing simple non-parametric methods), but in

relation to item analysis you still meet the attitude that either the Rasch model can be fitted to a set of items or you have to omit bad items or perhaps give up using the items as a scale. This attitude is in contrast to most areas of applied statistics where you look for an alternative statistical model if your data do not fit your initial model.

Testing whether the Rasch model can be fitted to a set of items is often called item analysis in Denmark, and perhaps because these analyses are often conducted by or in collaboration with a statistician the focus is on the item parameters and not the person parameters. This is remarkable because estimation of the item and person parameters is often done simultaneously, and most programs provide both sets of parameters.

The lack of focus on the theoretical person parameters may actually be a consequence of the exclusive focus on the Rasch model. Adherents of this model often point out that only if a one-parameter model holds, will the total score summarize all relevant information about test performance (statistical sufficiency, cf. Embretson & Reise, 2000, p. 57). For the two- and three-parameter models it is not sufficient to know the total number of correctly answered items, because the items differ with respect to both discrimination and guessing parameters. Statistical sufficiency is, however, only an advantage of the Rasch model as long as you assume that the total score will be used. If the theoretical person parameters are used instead, statistical sufficiency is not a problem, and since the theoretical person parameters have many advantages over total scores (e.g. in computer adaptive testing), I do not consider it progressive, but conservative to insist on using the total scores and the statistical sufficiency of the Rasch model (Mortensen, 2006).

Traditional item analysis

One way to look at traditional item analysis is to consider these techniques relatively free of the strong assumptions of item response models. It has sometimes been suggested that traditional item analysis is blind to the problem of the ICC and that the traditional “item difficulty index” implies a flat ICC — that the probability of a correct answer is independent of the ability or trait being measured (Suen, 1990, page 86-87). Considering the fact that the “item discrimination index” and the “item-total correlation” are parts of the machinery of traditional item analysis, it is simply not true that traditional psychometrics assume a flat ICC. However, comparing high and low scoring subjects by calculating item discrimination indexes and item-total correlations does not require exact knowledge of the ICC or choice of an explicit item response mathematical function or explicit description of item parameters. It does not require that the form of the ICC is the same for all items, and it is only based on the assumption that good items are items that distinguish respondents

who have more or less of the traits or abilities that a scale may measure. Thus, both traditional item analysis and item response analysis are based on interest in the item response function, but only *parametric* item response models make explicit assumptions about the ICC and the item response process.

Therefore, traditional item analytic techniques may be compared to non-parametric tests in inferential statistics: They are most appropriate in situations with a small sample and when there is reason to believe that the assumptions of parametric tests are not fulfilled. This comparison raises the question of sample size in relation to item analysis. The necessary sample size depends on the number of item parameters: A minimum of 1000 subjects has been recommended for estimation of three-item parameters. For the one-parameter model 200 subjects have sometimes been considered a sufficient sample size, but even this figure is unrealistically high in many clinical settings, and consequently traditional item analysis may be the only realistic possibility in many situations.

Since 1992, programs have been developed specifically for psychometric analyses of small samples (T-Rasch, Ponocny & Ponocny-Seliger, 1999), and the MSP program developed for non-parametric item response analysis is very flexible (Molenaar & Sijtsma, 2000).

While it must be admitted that clinical samples are often too small to be analyzed with parametric item response models, it should be observed that much too often coefficient alpha is the only psychometric statistic being reported – even for sample sizes that would permit detailed psychometric analyses, including application of parametric item response models. This may partly be explained by the fact that a standard statistical package like SPSS only includes few item analytic routines. However, item-total scale correlations can be calculated in most packages, and information about these correlations should be included as an absolute minimum.

Coefficient alpha has an ambiguous status being considered both an index of homogeneity and an index of reliability. It has now been 30 years since Green et al (1977) called attention to the fact that the alpha coefficient is dependent not only on inter-correlations among the items in a scale, but also on the number of items. This makes coefficient alpha a bad indicator of item homogeneity, but a good indicator of scale reliability. It seems that many authors still primarily think of coefficient alpha as an index of homogeneity, and a term like internal consistency contributes to the ambiguity of the term coefficient alpha.

The problem of uni-dimensionality

The traditional way to score tests has of course been to combine item scores to make up the total score. Traditional psychometrics do not spec-

ify item characteristic curves and assume that the relationship between responses to items and the latent trait of interest may vary substantially. It is, however, also assumed that if all items have a reasonably strong relation to the trait being measured, the noise associated with each item will not affect the total score substantially, and indeed the total score may have a strong, more or less linear relation to the trait being measured. It should, however, be observed that traditional psychometrics do not assume that items are strictly uni-dimensional: It is assumed that the individual item has considerable measurement error and may be influenced by specific factors, and combinations of sets of items are in fact used because individual items are not taken very seriously.

In contrast, most item response models assume uni-dimensional item collections, and uni-dimensionality may be considered a basic assumption in the practical application of item response models. Uni-dimensionality implies that the statistical dependence among items can be accounted for by a single latent trait, and this implies the assumption of local independence: Items should be statistically independent for subgroups of subjects who are homogenous with respect to the latent trait. Serial dependencies are, however, likely for many cognitive tests: An example is the recall of wordlists. It is indeed likely that the recall of individual words is not independent of the recall of surrounding words, and in this area it is difficult to see how uni-dimensional tests could be constructed. Another example of problematic items is tests of personality and emotions that ask several questions to the same stimulus material (e.g. vignettes or pictures describing or illustrating a certain situation).

Explorative factor analysis is sometimes described as a method of finding homogeneous sets of test items. However, historically, factor analysis was first used to test theories of intelligence, and early intelligence theorists did not factor analyze sets of single items, but sets of quantitative scores derived as sums of scores on cognitive items of the same type (Cooper, 1998). This makes sense since factor analysis is based on linear (Pearson) correlations among variables, and the question may be asked why item factor analysis became common in the construction of personality tests. The reason is perhaps that a collection of for example 400 personality items eminently illustrates what might be called the factor analytic problem: A situation with a large collection of variables and an unknown pattern of covariance(s) among the variables. In this situation, explorative factor analysis has been used to find sets of homogeneous items that may form a scale measuring a personality trait. However, discussions of the use of tetrachoric and polychoric correlations illustrate some of the problems in factor analyzing single items (Comrey & Lee, 1992), and in my opinion, explorative factor analysis can only be used to find items that may form a scale. Whether a set of items can actually be considered uni-dimensional needs further testing – for example by fitting item response models or confirmatory factor analysis.

A number of multidimensional item response models have been described (Embretson & Reise, 2000; van der Linden & Hambleton, 1997). These models may prove useful in specific contexts, but since the great majority of tests and scales have been developed to measure single traits, I find it unlikely that multi-dimensional models will be much used in the practical development of psychological tests (they could be used to analyze some of the existing personality tests). Consequently, I still see multi-dimensionality as a problem for practical applications of item response theory.

Multi-dimensional concepts in psychology I: Intelligence

Intelligence testing represents one of the few successes of applied psychology, and the WAIS is used in so many countries that it is setting a world-wide standard for individual intelligence tests. Since the publication of the first version of WAIS (Wechsler, 1955), several revised versions have been published (WAIS-R, WAIS-III and WAIS-IV). Naturally, item response theory played no role in the development of the original version, but perhaps it is more remarkable that it was not used in the development of later revisions. As a result, the test may be considered multi-dimensional at three levels which I will illustrate with examples from the original version of the WAIS:

1. IQ level: The test consists of 11 subtests, and the matrix of inter-correlations among the subtests was presented in the manual. The correlations had a range from 0.40 (between Digit Span and Comprehension) to 0.81 (between Information and Vocabulary). Thus, the size of the inter-correlations varies quite a bit, and it is quite unlikely that the collection of subtests can be considered uni-dimensional — even though the IQ is calculated by combining the 11 subtest scores. This is confirmed by factor analyses of the subtests which have typically found 2 or 3 factors and sometimes even 4 factors (Hill et al, 1985).
2. Subtest level: Several subtests consist of obviously multi-dimensional collections of items: Thus, the Information subtest consists of items covering a number of different areas of knowledge, and the Vocabulary subtest consists of words belonging to different areas of experience, and even some of the non-verbal tests appear to be multi-dimensional: Thus, some of the items in the Picture Completion tests seem to primarily test logical thinking (e.g. a missing shadow of a man), while other items seem to test attention to details (e.g. a missing finger) or geographical knowledge (e.g. a missing part of Jutland).

3. Item level: There is reason to believe that many individual WAIS items are influenced by several cognitive factors: Items in the Information and Vocabulary subtests may primarily be related to a verbal intelligence. However, any WAIS tester will be able to tell you that a person's knowledge of an item often is not primarily related to his or her verbal intelligence, but to specific life experiences, education, and interests. It is hard to see that this problem can be completely avoided in the construction of tests of general knowledge and vocabulary.

Clearly, the WAIS must be considered multi-dimensional at several levels, and consequently there is reason to consider possible test and item bias. Because the WAIS subtests are so obviously multi-dimensional, I decided to use chi-square techniques to analyse item bias or differential item functioning (DIF) in a sample of about 600 young Danes. For each subtest, the subjects were divided into total-score groups, and bias with respect to sex was evaluated within each score group. For the Information subtests 9 items showed significant DIF (6 items favouring males and 3 females), for Comprehension 6 items (4 favouring males, 2 females), for Similarities 2 items (both favouring males), for Vocabulary 15 items (11 favouring males and 4 females), and finally for Picture Completion 4 items (3 items favouring males and 1 females).

Thus, for each subtest relative item bias can be established for a number of items, and further analysis showed that a number of items are also biased with respect to educational level and tester. For most items sex bias is understandable as soon as the analysis has called attention to the item: Examples of items favouring males in the Information test are "length of Jutland", "explaining day and night" and "inhabitants in Scandinavia", while "constitution date", "author of Hamlet" and "author of Faust" are examples of items favouring females. In the Vocabulary subtest "movables", "ballast", and "well-box" are examples of words favouring males while "spangles", "colander", and "gelatine" favour females. Thus, it can be no surprise that - when the Information total score is controlled - males know the length of Jutland better than females, and more often are able to explain the shift between day and night. On the other hand, it can be no surprise that females know spangles and colander more often than males, and even for non-verbal tests item-bias seems understandable: In the Picture Completion test it was relatively easier for males to detect a missing part of a light-bulb, while it was relatively easier for females to detect a missing part of a female face.

We believe in the results of the statistical analysis because they make sense, and it should come as no surprise that in general the same items appear to be biased with respect to sex when the data of an almost fifty year older Danish WAIS sample are analyzed (Hess, 1974).

Some will argue that uni-dimensional and unbiased subtests could perhaps be constructed by using only a fraction of the items of each subtest of the WAIS. However, in that case we would not be working with the complete WAIS — that is the intelligence scale used all over the world and the test Danish clinicians have years of practical experience in using. In this context, it should be remembered that the psychometric properties of a test is only one of the factors that contribute to its clinical usefulness. In many areas the practical experience of psychologists using the test is much more important, and clinical psychologists are rightly conservative with respect to revising tests and introducing new tests in the clinic.

I still think that these considerations are important, but, that said, I find it problematic and surprising that item response theory apparently has not been used in the development of new versions of the WAIS.

Unfortunately, it seems to me that a new trend has developed since 1992: Test publishers regularly publish new versions of tests, and some clinical psychologists put high priority on using the most recent versions just as they prefer to use the most recent versions of computer programs. As a result they obtain little experience with each version and - what is perhaps worse - relatively few empirical studies are conducted with each version of the test.

How do we know that the WAIS is a good measure of intelligence — in short, how do we know that it is valid? We know that the WAIS is valid because the derived IQ fits into a theoretical and practical pattern of inter-correlations with other relevant tests and variables.

In other words, in general WAIS results correlate the way intelligence should correlate with other variables and constructs in psychology.

Intelligence tests that appear less multi-dimensional have been developed (e.g. Raven's progressive matrices), but since Binet constructed the first intelligence tests, obviously multi-dimensional tests have proved most useful in practical settings. This is primarily because test results may reflect specific abilities or talents rather than general mental ability if uni-dimensional tests are used.

Some item response theorists find it very important that a test score summarizes all available test performance information. However, the idea that the IQ should summarize all relevant information from the WAIS is almost ridiculous. Still, the IQ has proved immensely useful in many settings, and this is perhaps the reason that I do not find the statistical sufficiency of the one-parameter model all that important. Being a psychologist, I know that test scores can be useful even though they only summarize some aspects of test behaviour, and I would like to see somebody convince clinical psychologists that with the right psychometric model it is possible to summarize all relevant test behaviour in a single score.

Multi-dimensional concepts in psychology II: Personality

In 1990, Digman published a well-known review on the emergence of the five-factor model of personality. The five factors are of course extraversion, agreeableness (friendliness/hostility), conscientiousness, neuroticism and openness/intelligence. Digman (1990) pointed out that these five broad dimensions have been identified repeatedly for both rating scales and self-report questionnaires. It is important to notice that the five factors are abstract multi-dimensional concepts at the top of several hierarchical levels, and that consequently scales constructed to measure these dimensions must be multi-dimensional. This can be illustrated for extraversion which is a broad multi-dimensional concept including several concrete traits like sociability, activity, liveliness and sensation seeking (Eysenck & Eysenck, 1985).

The extraversion scale of Eysenck's well known personality questionnaire (Eysenck & Eysenck, 1975) is a mixture of items designed to catch a bit of several of the concrete traits or aspects of extraversion, even though the focus is on sociability. It is possible that uni-dimensional instruments can be constructed to measure these lower level personality traits, but it also possible that uni-dimensional instruments can only be obtained for even more concrete and specific behavioural habits. As instruments focus on more and more specific and concrete aspects of behaviour, the likelihood increases that they will have good psychometric properties from both a traditional item analytic and item response perspectives. However, the likelihood also increases that they measure so concrete and specific forms of behaviour that it becomes relatively uninteresting from both a theoretical and practical point of view. You risk that your set of items all ask the same question with slight differences in formulations – resulting in scales with high alpha coefficients and scales that are uni-dimensional as indicated by the fit of an item response model.

A reasonable compromise may be to construct uni-dimensional subscales corresponding to the concrete traits that make up extraversion and to derive the extraversion according to explicit formulas for combining these subscale scores. This is in fact the way the NEO-PI-R is constructed (Skovdahl Hansen, H. & Mortensen, E. L., 2004): The NEO assesses the five broad personality dimensions of the five-factor model and for each the five broad factors six concrete traits or facets are assessed. Thus, scores on the concrete traits of anxiety, angry hostility, depression, self-consciousness, impulsiveness, and vulnerability are combined to derive a composite score on the broad personality dimension of neuroticism.

The composite neuroticism score will obviously be multi-dimensional, but one may expect the concrete traits (or facets) to form uni-dimensional scales. However, explorative analyses of the Danish normative sample

with the programs Parscale (du Toit, 2003) and MSP (Sijtsma & Molenaar, 2000) showed that for most traits or facets neither parametric nor non-parametric item response models could be fitted. This suggests that the NEO-PI-R is multi-dimensional even at the concrete trait or facet level, and this is also indicated by DIF or item bias analyses.

I have observed DIF with respect to gender not only for the NEO-PI-R, but also for an instrument like the Schedule of Normal and Abnormal Personality (Eplov et al, 2006), and DIF with respect to gender may in fact be a major problem in many existing personality tests.

One might think that substantial DIF with respect to gender will result in different patterns of correlations between test scores and external variables for males and females. Apparently, this is not necessarily the case since we found similar patterns of correlations for males and females between Zuckerman's sensation seeking scales (1994) and external variables (Ripa et al, 2001) – in spite of the fact that preliminary analyses showed substantial DIF with respect to gender for most items in the inventory. However, if DIF is prevalent, sex differences on personality scores may of be a direct result of item selection.

Validity

Psychological tests are valid to the extent that they precisely measure the traits that they are supposed to measure. The word precisely indicates that a reasonable degree of precision of measurement is usually considered part of the concept of validity, but apart from that there is no direct relation between the internal structure of tests and scales and their validity.

People's real life behaviour and some tasks in experimental psychology are interesting and important without much interpretation (this is for example the case for some tests of perception, memory, and reaction time). In contrast, the concrete behaviour of people answering questions and items in psychological tests is relatively uninteresting. Test behaviour and test results usually only derive meaning from their intercorrelations with other more or less interesting behavioural variables. Therefore "construction validation" is not the same as "construct validation". The construction of a test may be validated by traditional item analyses or by fitting an item response model, but construct validation primarily involves establishing a meaningful network of intercorrelations for a given test score. The construct that the test is supposed to measure should fit into a meaningful pattern of relationships with other psychological constructs and external variables that are assumed to be associated with the measured construct (Nunnally, 1967).

Psychologists usually want to demonstrate that constructs like intelligence and personality traits are somehow "real" — inside people's heads.

Consequently, construct validation often involves relating psychological concepts to neuropsychology and experimental psychology. An example is extraversion: The concept has been given a neuropsychological interpretation, and it has been demonstrated that introverted and extroverted people behave differently in both experiments and real life (Eysenck & Eysenck, 1985). It should be added that since 1992 an increasing number of studies focus on the neurological basis for personality traits like neuroticism (Frokjaer et al, 2008).

It will be understood that construct validation can never be established by item analytic procedures, but only by relating test scores to other variables. The fact that a given item analytic model fits a set of test scores can at most be taken to indicate that it may be worthwhile to initiate construct validation studies. In that respect, item analytic results may be compared with reliability studies in traditional psychometrics: If a measure is reliable to a sufficient degree, it may be a good idea to proceed with validation studies.

It has sometimes been argued that you can evaluate the validity of existing psychological tests by observing whether a particular item response model fits a set of observed test data (in Denmark usually the one-parameter item response model). As discussed, this is problematic because many psychological constructs are multi-dimensional and existing tests were designed to measure these constructs (e.g. the five NEO factors). Furthermore, lack of fit for a given item response model may reflect an unrealistic model of the included items and the response process.

However, for the sake of the argument I presented data on Millon's MCMI in 1992 (Millon, 1982). This inventory had been analyzed by Svend Kreiner using the one-parameter item response model (Kreiner et al., 1990). The test consists of 8 personality scales and 12 clinical syndrome scales. Not surprisingly, the item response model could not be fitted to the personality scales, but only to the clinical symptom scales. For the latter scales, a number of items did not fit the item response model and had to be dropped. Consequently, it was possible to compare traditional scoring of these scales and the new item response model based scoring: The Pearson correlations ranged from 0.78 to 1.00, but for 10 out of 12 symptom scales the correlations were higher than 0.90 and the median correlation was 0.97. The number of items that did not fit the Rasch model varied from 0 to 19 and obviously the correlation between the traditional scoring and the scoring revised according to the Rasch analysis depended on the number of deleted items. Thus the correlation was 0.78 for the drug abuse symptom scale (19 items did not fit the Rasch model) and 1.00 for the borderline-cycloid scale (4 items did not fit the model). The data illustrate a basic problem in analysing existing tests with item response models: If too few items fit the model, you end up analysing another test than the one you started out analysing. On the other hand, if most items fit the model, the item response analysis does not really

make a difference.

These MCMI data appear to be typical of research comparing traditional test construction and construction based on item response models: Correlations are frequently in the high .90s, and this indicates a need to demonstrate that item response analysis leads to tests that are significantly more valid than traditional tests. Furthermore, the high correlations between traditional scores and the theoretically more valid item response derived scores also illustrate the ‘robustness’ of psychological measurement: Tests and inventories often consist of a substantial number of items, and because of the large number of items, scores are not necessarily invalidated by a few problematic items. Obviously, scores on short tests with few items will be sensitive to bias resulting from problematic items (e.g. very short scales used in population studies), but in general psychological measurement is perhaps more robust than people tend to think, and in many circumstances deleting a few items with sub-optimal psychometric properties may be more a question of aesthetics than a question of a significant difference in measurement precision.

To my knowledge, there have been few attempts in this country to demonstrate that use of various item analytic methods makes a difference in terms of correlations between test scores and other variables. Apparently, focus has been on what has been called “scale validation” – e.g. evaluating whether an item response model can be fitted to a set of items. For tests and rating scales comprising a fixed number of items, item response theorists still need to demonstrate that the choice of item analytic methods makes a significant difference — in terms of the practical usefulness and the theoretical validity of psychological tests.

It should, however, be acknowledged that for computer adaptive testing and other procedures adapting the items to the ability or trait level of the examinee, it is clearly an advantage to base tests and rating scales on an item response model (Wainer, 1990). In general, this use of item response models requires that the theoretical person parameters are used to describe ability and trait levels rather than the traditional raw score totals.

Conclusions and perspectives

In 1992, I summarized my major points:

1. Item response analysis has a number of advantages, and consequently these models should be considered in the construction of new tests.
2. The one-parameter item response model should be chosen if it can be considered realistic for a given set of problems
3. Many psychological constructs and tests are clearly multi-dimensional.

4. Consequently, there rarely seems to be much idea in using item response models to analyse existing psychological tests, questionnaires, and rating scales.

These were my conclusions in 1992. Although my views on the choice between traditional item analysis and item response models have not changed fundamentally since 1992, there has been new developments during the last 15 years and some of these developments should be discussed in a 2008 version of the presentation:

1. Test norms: It is sometimes argued that the use of item response models makes test norms unnecessary. This may be true for some criterion or domain referenced ability tests, but for most psychological tests, the theoretical person parameters do not give some magic meaning to a set of items that by itself has little meaning (e.g. items on a personality scale). To interpret an individual's test performance it is still necessary to be able to compare the individual with normative samples and demographic classes. The person parameters reflect the item responses, but - like simple proportions of item endorsements - by itself have no more meaning than the set of items making up the scale.
2. Reliability: For a given item response model, the test information function describes the amount of information provided by a set of test items at an ability or trait level and this function can be used to calculate standard errors of ability estimates (Hambleton & Swaminathan, 1985). Since there are good reasons to assume that measurement precision varies across ability or trait levels, this is a major improvement over the traditional standard error of measurement derived from the reliability coefficient and sample standard deviation. It should, however, be observed that there are other methods of estimating measurement error at different ability levels (Traub, 1994) and that classical reliability theory describes a much broader range of problems than measurement error at a given level of ability (Nunnally, 1967).
3. Confirmatory factor analysis: It is not unusual to see confirmatory factor analysis used to conduct item analyses. The popularity of confirmatory factor analysis may partly reflect the fact that most psychologists are trained in interpreting correlations and explorative factor analysis, partly the fact that a number of well-known standard programs are available for confirmatory factor analysis (Amos, EQS and Lisrel), and partly the fact that measurement models may be incorporated in broader structural equation causal models. Use of correlations based on categorical data is, however, problematic and confirmatory factor analysis should rather be

based on “item parcels” or mini quantitative scales. Item response models may be used to derive item parcels or mini scales as is well illustrated by a recent Danish Ph.D. thesis (Kristensen, 2008).

Considering the advantages of item response analyses and the easy access to relevant software, I find it remarkable – and problematic – that many test constructors and researchers apparently prefer traditional item analysis – even when item response analyses would obviously be more appropriate. Item response analyses should not replace traditional analyses, but a flexible use of both approaches seems more useful than the exclusive use of one of the two approaches. The terms “classical” and “modern” test theory are still being used, but item response models was first described about 50-years ago (e.g. Rasch, 1960). It seems to be time for a full integration of the two approaches to psychometrics.

References

- Comrey, A. L. & Lee, H. B. (1992). *A First course in Factor Analysis*. Hillsdale: Lawrence Erlbaum.
- Cooper, C. (1998). *Individual Differences*. London: Arnold.
- Digman, J.M. (1990). Personality Structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, *41*, 417-440.
- Embretson, S. E. & Reise, S. P. (2000). *Item Response Theory For Psychologists*. Hillsdale, New York: Lawrence Erlbaum.
- Eplov, L. F., Petersen, L. & Mortensen, E. L. (2006). Schedule for Non-adaptive and Adaptive Personality (SNAP). I: Elsass, P.; Ivanouw, J.; Mortensen, E. L.; Poulsen, S. & Rosenbaum, B. (red.) *Assessmentmetoder. Håndbog for psykologer og psykiatere*. (kapitel 15, side 353-389). København: Dansk Psykologisk Forlag.
- Eysenck, H. J. & Eysenck, M. W. (1985). *Personality and Individual Differences. A Natural Science Approach*. New York & London: Plenum Press.
- Eysenck, H. J. & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult)*. London: Hodder and Stoughton.
- Frokjaer, V. G., Mortensen, E. L., Adams, K. H., Haugbol, S., Pinborg, L. H., Nielsen, F. Aa., Svarer, C., Hasselbalch, S. G., Holm, S., Paulson, O. B., Knudsen, G. M. (2008). Frontolimbic serotonin 2A receptor binding in healthy subjects is associated with personality traits predisposing

to affective disorder. *Biological Psychiatry*, 63, 569-576,

Green, S. B., Lissitz, R. W. & Mulaik, S. A. (1977). Limitations of coefficient alpha as an index of test unidimensionality. *Educational and Psychological Measurement*, 37, 827-838.

Hambleton R. K. & Swaminathan H. (1985). *Item Response Theory. Principles and Applications*. Boston: Kluwer Nijhoff Publishing.

Hambleton R. K., Swaminathan H. & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park & London: Sage Publications.

Hess, G. (1974). *WAIS anvendt på 698 50-årige*. København: Akademisk Forlag.

Hill, T. D., Reddon, J. R. & Jackson, D. N. (1985). The factor structure of the Wechsler scales: a brief review. *Clinical Psychological Review*, 5, 287-306.

Jones L. V. & Appelbaum, M- I- (1989). Psychometric Methods. *Annual Review of Psychology*, 40, 23-43.

Kreiner, S., Simonsen, E. & Mogensen, J. (1990). Validation of a personality inventory scale: The MCMI P-scale (Paranoia). *Journal of Personality Disorders*, 4, 303-311.

Kristensen, A. S. (2008). *The structure of anxiety symptoms*. An investigation of the factor structure of anxiety symptoms in panic disorder and social phobia and of the relationship between symptom dimensions and personality. Aarhus: Faculty of Health Sciences, University of Aarhus.

van der Linden, W. J. & Hambleton (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York: Springer.

Lord, F. M. & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores*. Mass: Addison-Wesley.

Millon, T. (1982). *Millon Clinical Multiaxial Inventory Manual*. Minneapolis, MN: National Computer Systems.

Molenaar, I. W., Sijtsma, K. & van Schuur, W. H. (2000). *MSP5 for Windows. A program for Mokken Scale analysis for polytomous items*. Groningen: ProGamma.

Mortensen, E. L. (2006). Om testmanualer. I: Elsass, P.; Ivanouw, J.;

- Mortensen, E. L.; Poulsen, S. & Rosenbaum, B. (red.) *Assessmentmetoder. Håndbog for psykologer og psykiatere*. (kapitel 5, side 159-178). København: Dansk Psykologisk Forlag.
- Nunnally, J. (1967). *Psychometric Theory*. New York: McGraw-Hill.
- Ponocny, I & Ponocny-Seliger, E. (1999). *T-Rasch*. Groningen: ProGamma
- Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. København: Danmarks Pædagogiske Institut.
- Ripa, C. P. L., Hansen, H. S., Mortensen, E. L., Sanders, S. A. & Reinisch, J. M. (2001). A Danish version of the Sensation Seeking Scale and its relation to a broad spectrum of behavioral and psychological characteristics. *Personality and Individual Differences*, 30, 1371-1386.
- Sijtsma, K. & Molenaar, I. W. (2002). *Introduction to Nonparametric Item Response Theory*. Thousand Oaks & London: Sage Publications.
- Skovdahl Hansen, H. & Mortensen, E. L. (2004). Dokumentation for den danske udgave af NEO PI-R og NEO PI-R Kort Version – I: Costa & McCrae (2004). *NEO PI-R: Manual – klinisk*. (side 53-85). København: Dansk psykologisk Forlag.
- Suen, H. K. (1990): *Principles of test theories*. Hillsdale: Lawrence Erlbaum
- Du Toit, M. (2003). *IRT from SSI*. Lincoln Wood, IL Scientific Software.
- Traub, R. E. (1994). *Reliability for the Social Sciences: Theory and Applications*. Thousand Oaks & London: Sage Publications.
- Wainer, H. (ed.) (1990). *Computerized Adaptive Testing. A Primer*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychological Corporation.
- Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. USA: Cambridge University Press.

Computer Adaptive Testing: An introduction to the practical considerations related to developing a computer adaptive test for occupational testing

Af Guido Makransky

The past 15 years have seen a steady increase in the use of computerized testing applications in a variety of testing settings. The potential advantages of computerized assessment include immediate score reporting, decreased administrative burdens, increased security of testing materials, and more flexible test scheduling (Way, 2005). One of the largest advantages of using computerized testing is the increased statistical accuracy that can be achieved with computerized adaptive testing (CAT).

Instead of giving each examinee the same fixed test, a computer adaptive test (CAT) is designed to adapt to the ability level of the individual examinee. This tailored item selection can result in increased reliability with only a limited number of properly selected items. CAT algorithms successively select items so as to maximize the precision of the exam based on what is known about the examinee from previous items. For example, if an examinee performs well on an item, he/she would then be presented with a more difficult item. Or, if he/she performs poorly, he/she would be presented with a simpler one. CAT item selection algorithms do this through the use of item response theory (IRT). Item response theory is a set of mathematical models whereby the probability that a test taker will answer an item correctly is related to the ability level of the test taker and the characteristics of the item. After each response the examinee's ability estimate is updated and the CAT algorithm selects the subsequent item to have optimal properties at the new estimate. The implementation of CAT is attractive because research indicates that CATs are more precise (Runder, 1998a; Van der Linden & Glas, 2000a), take a shorter

period of time to complete (Runder, 1998a; Wainer, 2000), can be more motivating for the test taker (Daville, 1993), and are easier to maintain and improve (Linacre, 2000; Wainer, 2000) than traditional tests.

CATs have traditionally been used in large scale testing programs because of practical limitations such as the added costs associated with developing CATs, and the need for computer access. However recent research developments and the broad availability of personal computers and the internet have made CATs more attractive and accessible for a wide variety of testing contexts. The current paper provides an introduction to the most important issues related to CAT, and is organized according to the practical considerations that are necessary when designing a CAT for occupational testing. These include the IRT model that will be used, item bank development and calibration, the CAT algorithm, current research topics related to CAT, advantages and disadvantages of CAT for occupational settings, and informing and educating stakeholders about CAT. The discussion in this paper is meant for practitioners who have an understanding of basic psychometric theory and are interested in gaining a deeper understanding of CAT, for more extensive introductions to CAT, the reader is referred to Wainer (2000), and Van der Linden & Glas (2000a).

Item Response Theory

The challenge with CAT is to compare examinees on a similar scale even though they have responded to different items. This is accomplished with the help of IRT models which are used to score examinees and items on a common scale. IRT models apply mathematical functions that specify the probability of a response to an item, in terms of person and item parameters. Person parameters may, for example, represent the ability of a respondent or the strength of his/her attitude. Item parameters can include difficulty, discrimination (slope), and guessing (lower asymptote). IRT models have the following assumptions:

1. Measurement of a uni-dimensional trait.
2. Local independence (once a trait is accounted for item responses are independent of each other).
3. A specific form of relationship between trait level and probability of a response.
4. A stable trait.

For a more detailed description of the fundamentals of IRT see Hambleton et. al. (1991).

The most common IRT models differ in terms of two factors: The number of response options available in the items, and the number of item parameters that are included. IRT models can be divided into dichotomous and polytomous models based on the number of response options in the items. Dichotomous models are used when there are two response categories such as correct/incorrect, or agree/disagree options in the items. Polytomous models include items with more than two response options and are often statements that allow respondents to indicate a level of agreement on a given scale, or ability tests with partial credit items. Most of the operational CATs in use today are ability tests that use dichotomous models, however there has been a significant amount of literature that has focused on polytomous models and many new operational CATs are being developed with these models. Several polytomous models include: the Graded Response Model (Samejima, 1969), the Nominal Response Model (Bock, 1972), the Rating Scale Model (Andrich, 1978), and the Partial Credit Model (Masters, 1982). Since most operational CATs use dichotomous item responses, these will be the focus in this paper.

There are three commonly used dichotomous IRT models that differ based on the number of item parameters that are used. These include: the 1 Parameter Logistic or Rasch model, the 2 Parameter Logistic model, and the 3 Parameter Logistic model.

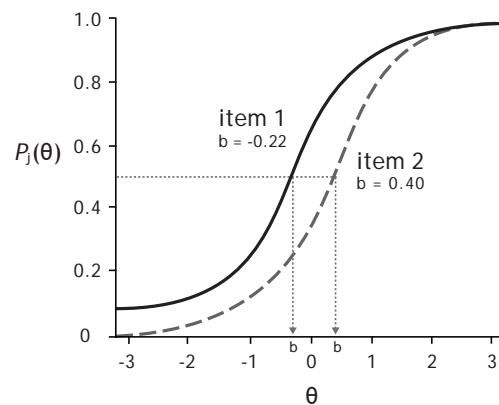


Figure 1. A graph illustrating the item characteristic curves of two items with the Rasch model.

The 1 Parameter Logistic model (1-PL) or Rasch model (Rasch, 1960) includes only a difficulty parameter (b) which is the point on the ability scale where the probability of a correct response is 0.5. As the difficulty parameter increases so does the ability that is required for an examinee to have a 50% chance of getting the item right. Therefore items with larger

difficulty parameters are harder. Figure 1 illustrates the item characteristic curves of two items with the Rasch model. The horizontal axis is the ability scale, ranging from very low (-3.0) to very high (+3.0). Given that ability follows the normal curve, 68% of the examinees will have an ability score between -1 and +1, and 95% will be between -2.0 and +2.0. The vertical axis is $P_i(\theta)$, the probability of responding correctly to item i based on the IRT model. In the figure Item 2 is more difficult than Item 1; this can be seen because the difficulty parameter (b) is higher for Item 2 than Item 1. In the example an examinee with an average ability (0 on the horizontal axis) will have approximately a 60% chance of answering Item 1 right and about a 40% chance of answering Item 2 right.

Since the Rasch model only includes a difficulty parameter, the assumption is made that all of the items have the same discrimination parameter, which means that the slopes in the curves are parallel at the point where the probability of a correct response is 0.5.

The 2 Parameter Logistic (2-PL) model (Birnbaum, 1968) includes a discrimination parameter (a) for each item in addition to the difficulty parameter. The discrimination parameter (a) defines the slope of the curve at its inflection point. The curve would be flatter with a lower value of a ; steeper with a higher value. Thus a , denotes how well the item is able to discriminate between examinees of slightly different ability. Figure 2 illustrates the item characteristic curves of two items. Item 2 has a steeper slope than Item 1, therefore Item 2 has a higher discrimination parameter. Another way to think of the discrimination parameter is as a loading. Some items load on to a construct more than others. In the 2-PL model the assumption is made that there is no lower asymptote or guessing parameter. Therefore there will be a probability of 0 for getting an item correct, if the ability level is low enough.

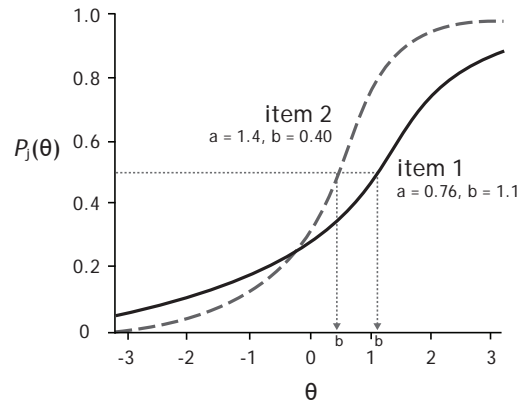


Figure 2. A graph illustrating the item characteristic curves of two items with the 2-PL model.

The 3 Parameter Logistic (3-PL) model (Birnbaum, 1968) includes a guessing parameter or lower asymptote (c) in addition to the other two parameters. The guessing parameter provides the possibility of a nonzero lower asymptote for the item characteristic curve and represents the probability of answering the item correctly even though the examinee has a low level of ability. This model is specifically useful when there is a chance for guessing, such as in multiple choice items, therefore the probability of a correct response would not reach 0. Figure 3 illustrates the item characteristic curves of two items that differ in the guessing parameter (c). It is clear from the graph that the lower asymptote is higher for Item 1 compared to Item 2.

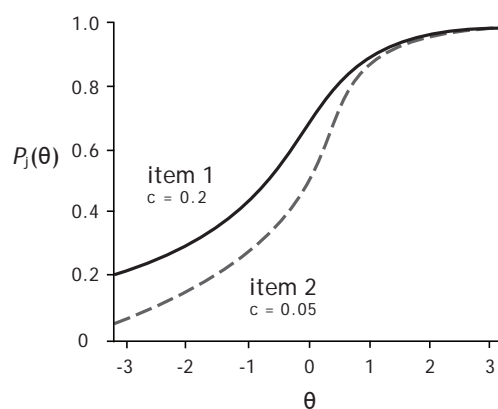


Figure 3. A graph illustrating the item characteristic curves of two items with the 3-PL model.

A discussion of the practical advantages and disadvantages of the three models is presented consequently. The 2-PL model will be used in this section to provide a better understanding of a dichotomous IRT model.

Under the 2-PL model, the probability of a correct response to a given item i is a function of an examinee's true ability and two item parameters

$$P_i(\theta) = \frac{1}{1 + \exp(-a_i(\theta - b_i))} \quad (1)$$

(Birnbaum, 1968). Here $P_i(\theta)$ is the probability of a correct response to item i , θ is the test-taker's ability, a_i is the item discrimination parameter, and b_i is the item difficulty parameter. Each item i has a different set of these two parameters, that are usually calculated based on prior administrations of the item. The model states that the probability of a correct response to item i for a specific examinee is a function of the two item parameters and the examinee's true ability.

There is debate in the literature about the advantages and disadvantages of the different dichotomous IRT models. From a pragmatic perspective the considerations regarding the different models could be summarized as follows. The item parameters used in the model describe the probabilistic function of a correct response on a particular item for test takers with a certain level of ability. The probabilistic function attempts to model the empirical data from a trial (calibration) sample as closely as possible based on the item parameters in the model. Figure 4 represents the item response functions for a 2-PL and a 3-PL IRT model. The goal of these probabilistic functions is to describe empirical data in a generalizable way in order to describe what the function would be in the testing population. The model will fit the empirical data better as you go from a simple to a more complex IRT model because of the addition of extra parameters. There are statistical tests such as the Lagrange multiplier (LM) test (Glas & Suarez, 2003) that assess if the additional fit observed with the more complex model is statistically significant.

However the goal of testing is to predict the testing population that will subsequently take the test, and not to fit the empirical data that is available. Therefore, as the model gets more complex the chances for uncertainty increases. This is because there are more parameters to calculate, and because the discrimination parameter and specifically the guessing parameter are not as easy to estimate as the difficulty parameter. Therefore a larger calibration sample is needed, in order to have certainty in the item parameters that are estimated the complex models.

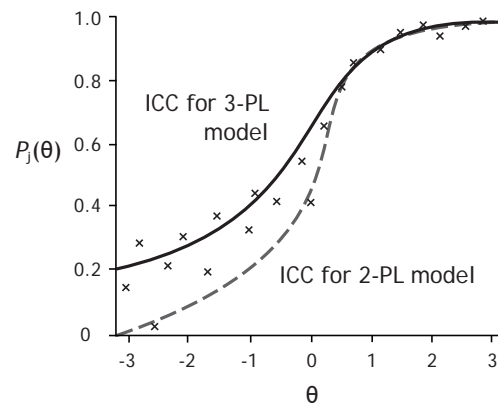


Figure 4. A graph consisting of empirical data fit to a 2-PL and a 3-PL model.

In developing occupational tests it is difficult to gather data in the development phases, and the time and resources required for collecting a large calibration sample are extensive. Therefore a possible solution is

to start with a simple model and when enough data becomes available statistical tests can be conducted to investigate if the added fit of the complex model is statistically significant.

In general the 1-PL model is appropriate when a small calibration size is available, and when there is empirical evidence for assuming that there are not significant differences between the discrimination parameters of the different items in the test. The 2-PL model is appropriate for situations where the items differ in terms of their discrimination parameters. The 3-PL is appropriate when there is a good chance to guess on the items such as in a multiple choice test with few answer categories. The 3-PL model has the disadvantage that large calibration sample sizes consisting of several thousand examinees are necessary.

Item Bank Development and Calibration

An item bank is a database of developed, reviewed and calibrated test items. Since the CAT algorithm selects the best items for each examinee from the item bank, the breadth and quality of items in the bank determine the quality, precision, security, life span and flexibility of a CAT. Therefore a well developed item bank is often considered the most important component of CAT. Item bank development can either start based on an existing test (usually a pre-existing conventional test), or is developed from scratch. The same considerations have to be taken when developing items for a CAT as would be taken for developing a conventional test (e.g. Irvine & Kyllonen, 2002). However many more items are necessary in a CAT item bank. The number of necessary items depends on several factors, including the number of items in the test, the level of security required, and the constraints to be used in the test etc.

There are two main types of CATs; equiprecise, and classification CATs. Equiprecise CATs are designed to measure all examinees with a similar level of precision regardless of their ability level. These tests include most ability or achievement tests used in occupational or educational settings. Classification CATs are designed to classifying examinees relative to a cutoff score and are usually used in certification exams or health outcome measures. These two types of CATs require different item bank compositions. An equiprecise CAT requires a large number of items that span the entire ability range, whereas a classification CAT requires a large number of items at the cutoff score and fewer items at the extremes. Since the tests within occupational settings are usually equiprecise CATs, these will be the focus of this paper. The six major steps involved in developing an item bank for CAT are covered next.

Steps in developing an item bank for CAT.

1. Simulation studies
2. Development and review of items
3. Item calibration
4. Item parameter estimation and the assessment of item quality
5. Calculation of an item bank information function
6. The addition of items as needed to approximate the desired target bank information function

1. Simulation studies

The first step in developing a CAT is to conduct simulation studies, which are used to provide previews of how a CAT system will function in practice. They give the test developer the opportunity to control certain variables in order to obtain rough estimates regarding important applied questions before resources are used in test development. Two types of simulations are used in CAT research: Monte Carlo and post hoc or real simulation studies.

A Monte Carlo simulation is usually used before any real data is available, to evaluate the potential performance of various approaches to CAT. This approach to simulation provides information about practical considerations. Some of these are the number of items needed in a test and item bank, appropriate start values, appropriate termination criteria, and can also provide information about the most effective estimation and calibration procedures.

A post hoc or real data simulation is used when there is data available from a previously existing conventional test, and the CAT is to be used to reduce the length, or to increase the precision of the test. The item bank used in this case would be the items in the conventional test. The objective of applying simulation procedures is to determine how much reduction in test length can be achieved, or to estimate the possible increase in precision in the test by simulating the re-administration of the items adaptively.

2. Development and review of items

Once there is a clear idea of the number of items necessary for a CAT, item development can begin. Here item development can take place in the same way as in a conventional test; however a greater number of items are necessary. Care should be taking in ensuring that the items are well developed and revised before item calibration can begin. This is particularly important in CAT, where each examinee's total score is determined by a relatively few number of items. A poor test item could have an inordinately large effect on some scores, and not others because different items may be used with different examinees (Wainer, 2000).

3. Item calibration

Once a large number of items have been developed and reviewed for a CAT, the following step is to calibrate the items. However the calibration of CAT items is usually more complicated than for conventional tests, because the large quantity of items in a CAT makes it impossible to administer the entire test to a single examinee. Even if it was possible, security issues regarding the confidentiality of the items would make this a bad idea. Therefore a calibration design is necessary that allows items to be calibrated on a similar scale even though they are administered to different test takers. Traditional methods for calibrating items include booklet designs where items are divided into several booklets. Booklets are made up of a certain number of unique items and a certain number of linking items that are included in all of the booklets. Although commonly used, booklet calibration designs lack efficiency which makes them sub-optimal for certain settings.

Occupational testing is one setting where calibration designs must be as efficient as possible because it is very difficult to obtain data in the development phases of a test. Alternative online calibration strategies have been developed for settings where there are a limited number of test takers (e.g. Makransky and Glas, 2008). In addition to being more efficient these designs provide more flexibility which allows for changes and additions while in the calibration process, improving the item bank development process, and decreasing development costs.

An additional challenge when developing an item bank for an occupational test is the sample that is used. It is difficult to obtain data from a calibration sample that is similar to the intended test population. This is specifically the case in certain high stake circumstances such as employee selection. To circumvent this problem test developers usually access respondents from a context other than the one in which the test is to be used. The use of a low stakes calibration sample comes with several limitations. First, there is evidence that large motivational differences exist between low stakes calibration samples and the intended population

of respondents (Wise & DeMars, 2006). These motivational differences introduce bias in the estimation of parameters in the calibration and can result in biased test scores. Research has found models (e.g. Wise & Kong, 2005, Wise & DeMars, 2006) that improve the quality of the calibration of items through filtering low motivation respondents from the data, based on the time it takes examinees to respond to the items.

4. Item parameter estimation and the assessment of item quality

There are many IRT approaches to item parameter estimation including; Joint Maximum Likelihood (JML), Conditional Maximum Likelihood (CML), Marginal Maximum Likelihood (MML), Markov Chain Monte Carlo (MCMC), and Bayesian estimation methods that are similar for CAT and conventional tests. Most current operational tests use MML estimation because it is simple and produces relatively unbiased parameter estimates.

There are countless data analysis programs available for estimating item fit, and calculating item parameters (e.g. QUEST, WINSTEPS, RUMM, DIGRAM, BILOG, MULTILOG, XCALIBRE, and OPLM). Different programs use different methods to assess the quality of the items in the test, depending on the IRT model, and the statistical test that is used to assess fit. Regardless of the method that is used, items are either identified as fitting or miss-fitting the IRT model. Once items have been identified that do not fit the model these are either eliminated from the test, or are reviewed. Using a revision process to identify the specific content problems in the test has several advantages over simply eliminating the items that do not fit the model. The qualitative judgment of items is important because it gives a theoretical reason for eliminating items instead of simply relying on statistical evidence. Revising items can also save resources because some items may be recycled, and can improve the chances that future items will be of higher quality.

5. Calculation of an item bank information function

The amount of information provided by an item bank is related to the precision with which ability can be estimated at a specific difficulty level. Unlike in classical test theory where precision is incorrectly presumed to be uniform across the ability range, precision in IRT is calculated at each point across the entire range of test scores. Scores at the extremes of the test generally have more error associated with them than scores closer to the mean because there are usually more items that assess ability accurately close to the mean.

In IRT Fisher's information function can be used to calculate the amount of information an item supplies at each level of ability. For the

Rasch model precision is simply the probability of a correct response multiplied by the probability of an incorrect response at each level of ability, or,

$$I(\theta) = p_i(\theta)q_i(\theta). \tag{2}$$

Where $I(\theta)$ is the information of an item for a given level of ability, $p_i(\theta)$ is the probability of a correct response and $q_i(\theta)$ is the probability of an incorrect response at the given level of ability. In IRT the standard error of estimation is the reciprocal of the test information at a given ability level,

$$SE(\theta) = 1/\sqrt{I(\theta)}. \tag{3}$$

Thus more information implies less error of measurement. For the 2-PL model Fisher's information is calculated by taking the square of the item discrimination parameter (a) and multiplying it by the product of the probability of a correct and incorrect response

$$I(\theta) = a_i^2 p_i(\theta)q_i(\theta). \tag{4}$$

Therefore the discrimination parameter has a large effect on the information for the 2-PL model. In general information is higher when the difficulty of the item (b) is close to θ , and information is generally higher when the discrimination parameter (a) is high.

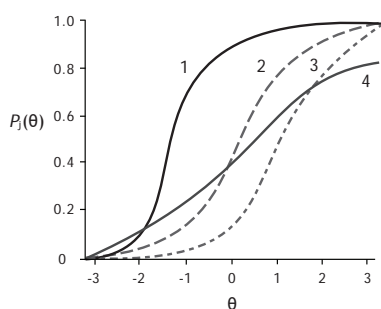


Figure 5. Item characteristic curves for four items.

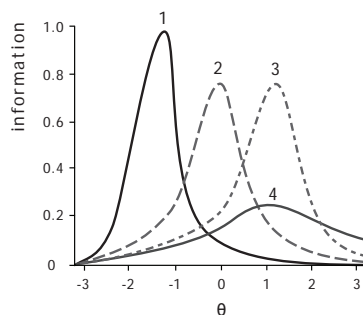


Figure 6. Information functions for the same four items.

Because of local independence, item information functions are additive. Thus, the item bank information function is simply the sum of the information functions of the items in the item bank. The item bank information function is affected by 1) the number of items in the item bank 2) the quality of the items, and 3) the match between item location and examinee ability level (Baker & Kim, 2004).

In general, item information functions tend to look bell-shaped. Figure 5 illustrates the item characteristic curves for four items in an item bank. Figure 6 represents the information function for the same four

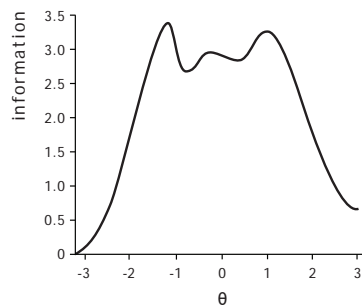


Figure 7. The test bank information function, adding the four items from figure 6.

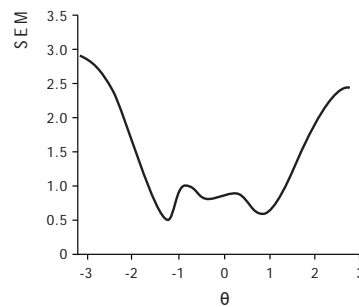


Figure 8. The standard error associated with the item bank from figure 7.

items. The item bank quality can be assessed by plotting the item bank information function which is simply the addition of the item information functions as shown in figure 7. Figure 8 depicts the standard error associated with this item bank which is the inverse of the information function (figure 7).

6. The addition of items as needed to approximate the desired target bank information function

The example illustrated above is unrealistic because the item bank only consists of four items and most operational item banks consist of hundreds of items. However the bank information function and standard error provide evidence that many more items are needed in the item bank. It is also clear that most of the new items should be more difficult than the current items because the current items provide most information for lower levels of ability. Finally the items differ greatly in terms of their discrimination parameters and thus differ in terms of the amount of information that they provide. Therefore items can be analyzed to see the characteristics that may make some items better than others. In an operational test Item 4 would be eliminated from the test because it does not discriminate well between test takers and provides almost no information. These are typical considerations that are taken when analyzing the item bank and deciding on the possible addition of new items. At this point the decision is usually made between launching the test, or developing more items before the test is launched.

The CAT Algorithm

The testing algorithm is a set of rules specifying the questions to be answered by the examinee and the order of their presentation (Thissen

& Mislevy, 2000). The CAT algorithm consists of four major components:

1. A starting rule for selecting the first item.
2. A procedure for scoring item responses and estimating trait level.
3. A method of selecting the next item.
4. A rule for ending the test.

The CAT algorithm uses these components in an iterative process as follows (Rudner, 1998b):

1. All of the items that have not yet been administered are evaluated to determine which will be the best one to administer next given the current estimated ability level.
2. The best available item is administered and the examinee responds.
3. A new ability estimate is computed based on the responses to all of the administered items. 4) Steps 1 through 3 are repeated until a stopping criterion is met.

Starting rule for selecting the first item

A starting rule for a CAT can either be similar for all test takers or differ based on previously available information. Previous information can be a result from the same test at a previous point in time, the result from a correlated test, or other information that is hypothesized to increase the accuracy of the test. There are often issues of fairness involved in starting test takers at different starting points. Therefore most CATs assume that all examinees have an average ability level at the beginning of the test.

Most operational CAT algorithms start a test by assign a random item from a group of starter items that are at some level right below the mean. Starting with a fairly easy item has the advantages that it provides examinees with a chance to get an impression of the test, it improves the test takers' general level of motivation, and lives up to most examinees' expectations of beginning a test with a fairly easy item. Finally allowing examinees to begin a test with fairly easy items gives examinees that have no previous experience with the item types a chance to get use

to them, which limits the advantages of having had previous experience with similar tests.

The disadvantage of starting with a fairly easy item instead of an item of average difficulty is that easy items do not provide a lot of information about most of the examinees in the population. This means that more items would be necessary to attain the same level of precision. Therefore a decision regarding how the items should be selected at the beginning of the test will depend on the practical considerations of the particular test. Providing good instructions and an explanation of the functionality of CAT, as well as a number of practice items before the test begins, can improve the examinees' understanding of the test and the CAT system, and can limit the need for starting the test with many easy items. Due to security issues it is recommendable to start test takers with a random item from a group of starter items so that individual items do not receive too much exposure.

Scoring Procedure

After an item is administered, the CAT algorithm updates its estimate of the examinee's ability. If the examinee answers the item correctly, the CAT algorithm will likely estimate their ability to be somewhat higher, and vice versa. This is done by using the item response function from IRT to obtain a likelihood function of the examinee's ability. IRT estimation procedures weigh each item by its item parameters based on whether the examinee answered each item correctly. There are three commonly used methods for calculating ability in CAT: Maximum Likelihood Estimation (MLE), Bayesian estimation, and Weighted Maximum Likelihood estimation.

MLE is the most commonly used method for estimating ability in operational CATs. With MLE, item parameters are treated as true parameters, and the ability estimate is the one that makes the item response pattern most likely. For a more detailed description see Baker & Kim (2004). A disadvantage regarding MLE is that an ability estimate cannot be calculated for response patterns where all of the items are answered correctly or incorrectly. Some possible solutions to this problem are to administering items of extreme difficulty levels until a mixed pattern is obtained, or to use Bayesian estimation methods at the beginning of a test until a mixed pattern of responses is observed.

Bayesian estimation includes an a priori distribution of examinee ability (usually a normal distribution) in the estimation formula, and has two commonly used estimators: expectation a posteriori and maximum a posteriori. Bayesian estimation has the advantage that an ability estimate can be obtained with all answer patterns. In addition Bayesian methods can be used to incorporate possible error in the item parameters which is an advantage when test results are given for a CAT that is still in

development. The disadvantage with using Bayesian estimation methods is that the inclusion of a prior means that ability scores tend to shrink toward the mean, which can be of practical importance for tests that consist of few items, as is often the case in occupational tests.

An alternative to these methods is the use of a Weighted Maximum Likelihood estimator (Warm, 1989). Warm found that Weighted Maximum Likelihood estimation was less biased than both MLE and Bayesian methods and can be calculated with most response patterns.

Item selection algorithm

What differentiates CAT from traditional tests is that items are selected optimally based on an item selection algorithm. Here a function is used to find the next available item from the item bank that provides the most information about the examinee. Since IRT places examinees and items on the same scale, the information provided by each item at each level of ability can be calculated fairly easily. A function that is often used in item selection for CAT is Fisher's information function. For dichotomously scored items, the function has the following form:

$$I_i(\theta) = \frac{(\theta)P'_i(\theta)^2}{P_i(\theta)Q_i(\theta)}, \quad (5)$$

where $P_i(\theta)$ is the response function for item i , $P'_i(\theta)$ its first derivative with respect to θ , and $Q_i(\theta) = 1 - P_i(\theta)$. In CAT, the function is used to define the item in the pool that yields the largest value at $\theta = \theta^*$ where θ^* is the current estimate of the ability of the examinee (Van der Linden & Glas, 2000b).

An example of the information function for an item bank consisting of 10 items is presented in Figure 9. The highest point on the item information function on the ability scale for each item reflects the difficulty of the item, the height of the item information function reflects the item discrimination. Therefore Item 1 is the easiest item because it is farthest left on the ability continuum, and Item 10 is the most difficult. Because Item 9 is the most discriminating it has the highest item information function, and Item 7 is the least discriminating. In this example Item 6 would be the next item that is selected because it provides the most information at the current ability estimate of $\theta^* = 0$ (Weiss, 2007).

The previous example illustrates a CAT algorithm with a selection criterion based on the probability of a correct response 50% ($P_i(\theta) = 0.5$), meaning that examinees would on average answer half of the items in the test correct. However there are certain circumstances where it may be appealing to have examinees respond correctly to 60% or 70% of the items in the test. This could increase their level of motivation but would decrease the efficiency of the test. In fact optimal item selection where the best item is always selected for each examinee is almost never possible

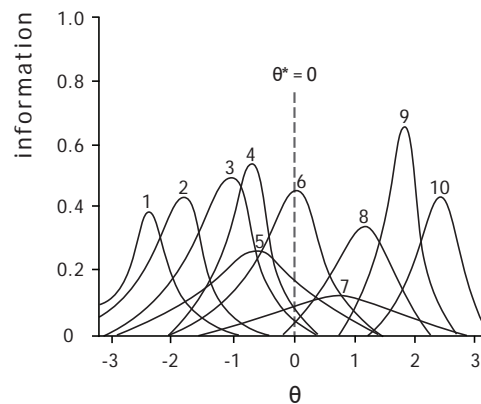


Figure 9. A graph consisting of the information functions for a test with 10 items.

in an operational CAT. There are usually constraints such as content balancing or item exposure that decrease the efficiency of the algorithm in favor of other practical considerations. These constraints are discussed in more detail later in this paper.

Termination criterion

The CAT algorithm is designed to repeatedly administer items and update the estimate of examinee ability until the termination criterion is met. The termination criteria usually include either one or a mixture of the following: Fixed standard error of the ability estimate, minimum information, a fixed number of items, or a fixed amount of time.

In terms of psychometric fairness the best termination criterion is minimum standard error where the test is concluded when the standard error reaches a specified lower bound. All examinees are measured with roughly the same level of precision with this method. However in practice a termination criterion consisting of a fixed number of items is often chosen because of the perceived lack of fairness with different test lengths. A disadvantage with a fixed number of items is that examinees are assessed with different levels of precision. However stakeholders who do not understand the principles of IRT usually have difficulty accepting the idea that fair decisions can be made between two test takers who are not administered the same number of items. Operational tests may also use fixed amount of time as a stopping rule when it is important to have a limit for the maximum amount of time an examinee can spend on a test. The minimum information criterion where the test is concluded when there are no remaining items that provide information over a certain value can be used when the efficiency of precision vs. the amount of time and effort

in testing is of concern. Many operational CATs use a combination of these methods.

Current Research Topics Related to CAT

There are many research topics that dominate the literature related to CAT that are of interest for occupational testing. Most of these topics are methods for taking practical considerations that are important for stakeholders into consideration in the testing algorithm. These usually cause constraints that limit the efficiency of the maximum information selection algorithm used in CAT. Some of these include:

- Content balancing
- Multiple scales
- Testlet based adaptive testing
- Item exposure
- Item parameter drift
- The CAT delivery system
- Increasing the feasibility of CAT development
- Examinee considerations

Content balancing

Content balancing is used in situations where there is the necessity to ensure that all examinees receive items from several different content areas, even though a construct may be statistically unidimensional. A good example of a test where content balancing is used is an arithmetic test where items consist of addition, subtraction, multiplication, and division. Although these different item types may load on a single scale, the addition items would tend to be easiest and the division items would be the most difficult. As a consequence high ability students would tend to get mostly division items and low ability students would get mostly addition items. Thus the test would be measuring the construct according to the model; however the examinees would not get similar content.

The procedures used for content balancing in IRT (e.g., Kingsbury & Zara, 1991) usually modify the maximum information item selection procedure by including the content category of the items in the item selection algorithm. Once an item is selected by the maximum information algorithm, its content classification is examined relative to target values set up in advance for each examinee. If the selected item represents a

content area that is underrepresented at that stage in the examinee's test, the item is administered. If not, the next best item is selected and examined (Weiss, in press).

Although content balancing is most common in educational achievement testing, it can also be used in occupational testing. One example could be a logical reasoning test that is statistically uni-dimensional but measures three sub-scales: verbal, numeric and spatial reasoning. Personality tests may also include scales that measure broad constructs with several sub-scales. If the purpose of the test is to measure a broad domain such as logical reasoning or conscientiousness and not the specific sub-scales within this domain, content balancing can be used to ensure that examinees receive similar content in the test.

Multiple scales

An alternative to content balancing is the use of multiple scales. Broad constructs are usually dealt with by treating each sub-scale as a separate uni-dimensional test. Thus scores for each sub-scale are obtained separately and these are combined to attain a score for the broad domain. The advantage of using multiple scales is that each examinee is measured as accurately as possible on each scale. The result is a profile of scores for each examinee that can be used for applied purposes. A disadvantage of using multiple scales is that it is more time consuming than using content balancing because accuracy in each of the sub-scales must be established, which is not necessary if only the score for the broad construct is of interest.

A method commonly used when a test consists of multiple scales is to establish starting values for the scales based on the previously administered scales. Since most test batteries are made up of scales that have relatively high correlations, the CAT algorithm can use this information to start examinees with a more accurate starting value that can result in a more precise result with fewer items, because examinees start at a value that is closer to their ability. Brown & Weiss (1977) proposed a method using multiple regressions to order subtests by how well they can be predicted from the other subtests. The multiple regression equations can be used to predict an examinee's initial ability on each new test in the battery as a starting value for that test. Using this method Brown and Weiss (1977; Gialluca & Weiss, 1979; Maurelli & Weiss, 1981) demonstrated that test lengths for later tests in a battery could be reduced by 80% or more of their full test length with no reduction in measurement precision.

An alternative method is administering multiple scales simultaneously. Here the CAT algorithm can simultaneously keep track of the different ability estimates for several different scales for each examinee. This method does not take advantage of using the other scales to improve

starting values, however it can be an advantage for certain domains, such as preference, where the administration of multiple scales simultaneously can ensure that examinees consider their preferences broadly.

Testlet based adaptive testing

There are many tests that include items that must be administered as a group and not individually. These items typically have a stimulus followed by a group of questions. Some examples are reading assessments where the examinee reads a passage and responds to several questions concerning that passage, or a chart that is followed by several questions regarding the information from the chart. Here each item group can be treated as a testlet and the CAT item selection algorithm can use a polytomous IRT model to calculate ability, after the completion of each testlet, and select the next testlet based on that estimate.

Item exposure

A CAT algorithm selects the item that is statistically best for each examinee which results in some items being presented more than others. In high stakes testing situations it is important to ensure that items are not overly used, because overexposure can compromise the confidentiality of the items. Therefore limiting the exposure of items on the test enhances test security and increases the life-span of the items in the test. However, item exposure also degrades the quality of the adaptive test, requiring a longer test. Several methods for controlling item exposure while maintaining the efficiency of the CAT test have been suggested. One approach to control exposure is to randomly select the item to be administered from a small group of best fitting items. A common method suggested by Simpson and Hetter (1985) uses a probability model to control item exposure.

Implementing an item exposure control in the CAT algorithm is important if the test is high stakes and there are many test takers, however item exposure control is not necessary in low stakes tests, and tests with a limited number of test takers.

Item parameter drift

The stability of the item parameters is one of the assumptions for IRT based adaptive testing. However, item parameters may shift over time, compromising the validity of the CAT. Some possible reasons for item parameter drift could be that the answers to the items become available to a segment of the testing population. Changes may be due to the mode of administration. Courses might become available that train the content

of the test, or the curriculum may change. Finally changes could simply occur because of changes in the population characteristics.

The possibility of item parameter drift makes it very important to continuously check the item parameters of the items in a test, which can be done by recalculating item parameters at two different points and comparing the values. A shift in the item parameters may require a revision of the test or the elimination of certain items.

The CAT delivery system

Traditionally CATs have been administered at testing locations with the test items stored on specific computers under the watchful eye of test administrators. However this has changed as the widespread use of CAT has been extended to areas outside of large scale test settings. Today many CATs are unsupervised tests administered at examinees' homes via the internet. This method of administering tests decreases the costs associated with staffing and maintaining testing locations. With CATs delivered over the Web, an item is answered and the response is sent to the server for scoring, ability estimation, and item selection, then the selected item is transmitted back to the examinee through the Web. There are some concerns regarding the administration of tests on the Web. According to Wise (2007) the variety of browsers and browser settings can potentially destroy test standardization, and differences in web response time are of concern because some web connections are fast and some may take several seconds to upload. Although this has been a realistic concern in the past, the amount of data that is sent is so small that it takes a fraction of a second to go through this process and is usually not noticeable to the test taker in most operational CATs. There could be some exceptions however, which makes it important to have procedures in place for test takers that experience problems with web based test administration.

Another concern regarding unproctored web test administration is the possibility that cheating may influence the test results in terms of score inflation. Research findings are mixed in terms of the equivalence of proctored and unproctored online tests for occupational selection (e.g. Nye et. al., 2008; Potosky & Bobko, 2004). Therefore additional security measures should be considered for high stakes unproctored tests. Confirming the results of an unproctored test can be accomplished within CAT easier than with a conventional test. The examinees' test score for the unproctored test can be used as a starting value for a short follow-up test that can be administered under controlled conditions. Person fit statistics (Meijer & Sijtsma, 2001) can then be used to assess the probability that the follow-up test was completed by an examinee with the same ability level as the original test.

Increasing the feasibility of CAT development

There is a lot of research designed to increasing the availability of CATs to contexts other than large scale testing settings. Much of this research has focused on decreasing the cost and improving the feasibility of developing and administering CATs. Two promising research areas are related to item cloning and item calibration. Item cloning can be used to increase the number of items available for adaptive testing by reducing the cost of item writing. Item cloning techniques involve the application of an algorithm to generate new test items based on the characteristics of existing test items (Glas, 2005). The term parent is sometimes used to describe the original items, and the term family is sometimes used to describe the group of items that are cloned from the same parent item. A study by Glas & Van der Linden (in press) describes a multilevel IRT model that makes it possible to simultaneously use the unique parameters for the items as well as the parameters from the item families in calculating ability, and calibrating items.

Research related to item calibration has also focused on decreasing the cost and increasing the feasibility of developing a CAT. Research related to automatic online calibration designs (e.g. Makransky & Glas, 2008) provides methods to increase the efficiency of calibrating items when there is a limited number of available examinees in the development phases of a test. Additionally research related to motivation filtering (Wise & DeMars, 2006) improves the calibration of items in settings where it may be impossible to use a sample from the actual examinee population.

Examinee considerations

One of the largest considerations in terms of switching from a traditional paper-and-pencil test to a CAT has been the equivalence of the two versions. A study by Bergstrom (1992) synthesized 20 studies that compared the equivalence of paper-and-pencil and CAT using live examinees. He found equal ability estimates, equal variances, and high correlations between the two administration methods.

While poor and minority children have less access to computers, Wise (1997) noted that the research on equity and CAT is mixed, therefore the issue should be investigated with regard to each test being developed (Rudner, 1998b).

Advantages and Disadvantages of CAT for Occupational Settings

Some of the advantages and disadvantages of using a CAT instead of a conventional test have been covered briefly in this paper. Here the specific

advantages and disadvantages of using CAT in occupational testing will be discussed in more detail. The advantages can be classified in terms of three main groupings: increased precision/a shorter test, additional security, and more flexibility.

Increased precision/A shorter test

CATs can reduce testing time by more than 50% while maintaining the same level of reliability (Rudner,1998b). In terms of occupational testing many companies are interested in using CATs because of the possibility to have a test with increased precision. Since many large business and industry organizations have recently moved toward automation and cost reduction in the hiring process, the amount of time needed for a test is also an important consideration in deciding on an instrument to use in their selection process.

An additional benefit to shorter tests is that lower testing times reduce fatigue, a factor that can significantly affect an examinee's test result. This is specifically important in employee selection procedures where applicants are frequently administered a battery of tests. A related issue is the opportunity to maintain examinees motivation throughout the assessment. With CAT examinees are administered items that match their ability level. This means that they do not have to spend a lot of time and effort on items that are too easy, and may bore them, or items that are too difficult, and may frustrate them. Therefore examinees' focus can be maintained throughout the test which can result in a more accurate outcome.

Test security

Test security may be increased by using CAT because hard copy test booklets are not available. In addition examinees are administered different items, making it more difficult to cheat because particular examinees will not be sure of the items that they will be administered in the test.

Flexibility

The greatest advantage of using CAT in occupational testing may be the flexibility that is possible in terms of: test maintenance, alternatives for the customer, reporting options, and new item formats.

In conventional tests, deleting old items and adding new items is difficult, time consuming, and expensive because it require resources related to changing printed material, recalculating norms, and informing stake holders. These challenges usually mean that test development companies are reserved about revising tests. Conversely CATs are easier to maintain and improve. It is fairly straightforward to include new items in a CAT

and to calibrate them while the operational test is in use. This is called item seeding and it is accomplished by including a few new items in each test that do not count toward the examinees' score. New items can then be scored and integrated into the test, after they have been administered to a number of test takers and data analysis is conducted to confirm that they fit the IRT model. Moreover bad items can easily be removed from an item bank without affecting the scores of subsequent examinees, and without requiring new norms.

An added advantage to using CATs for occupational testing is that they can provide accurate scores over a wide range of abilities, while conventional tests are usually most accurate for average examinees. This means that one test can be used to assess employees from a large variety of jobs within a company. A CAT can also include a flexible test management system that allows customers to prioritize things such as the test content, starting rules, and termination criteria. Therefore instead of having a test with fixed characteristics customers could decide on the characteristics of the test based on the needs for each job within their company.

In addition to a simple ability score CATs can also provide a large variety of reporting options. These can include: problem solving speed, answering pattern, an indication of abnormal responding, or a system that tracks information such as the amount of time spent on reading a passage, or the number of times an examinee changes their answer. All of this can provide additional information about the way in which an examinee responds to the test, which could give an indication of the way that they would react in a working situation.

Finally CAT offers the potential to accommodate a wider range of item types. Because a CAT adapts according to the responses on the test, it is imaginable that future tests could use more dynamic interactive testing procedures. This could improve the content validity of tests by including testing scenarios that have a closer relationship to job tasks.

There are also limitations to using CATs. The greatest challenge is that CATs are more expensive to develop than conventional tests. This is mainly because they require a lot of items, as well as the development of computer algorithms. CATs also require computers, which introduces the possibility of differences regarding examinees' computer familiarity. Hardware limitations may also restrict the types of items that can be administered by computer. Items involving detailed art work and graphs or extensive reading passages may be difficult to present. Additionally the test administration procedures are different, which may cause problems for certain examinees that are used to taking conventional tests. There could also be perceived inequities, because each examinee receives a different set of items.

The final two limitations can be restricted by properly educating and informing stakeholders about CAT, which is a topic that is discussed in

the next section.

Informing and Educating Stakeholders about CAT

Test developers have to be aware of informing and educating stakeholders about the CAT system and how it works, as well as the benefits and limitations to CAT when launching a new test. This is mainly because the test is different from what stakeholders are used to. This paper has discussed several advantages and disadvantages to using CAT. Many of these can be avoided by properly educating test takers about CAT.

A further purpose for educating stakeholders about a new CAT is that its implementation will usually mean the development of a new high quality test system, or the improvement of an existing test system. If stakeholders are not aware of the advantages to using CAT then the extra expenses invested in the system will not pay rewards.

A user friendly website with information about the particular CAT is a useful method to educate stakeholders. Some of the information included on the website should be: the purpose and background of the particular test, an introduction to CAT and the system, the differences examinees may experience between CAT and conventional tests, the benefits and limitations to using CAT, and research articles related to the test and CAT. Some test programs provide a preview or tour of how the test works. It may also be beneficial to have practice items or practice tests available for examinees that want to practice before taking the test. Finally contact information that stakeholders can use if they need more information should be available. In addition to a website, letters and training material or other forms of information to stakeholders should be available.

In conclusion this paper has covered the most important issues related to CAT from a practitioner's perspective. CAT provides test developers with a valid method to increase the precision and flexibility of their tests. Computers have changed the way most people live their lives through creative solutions that have increased the quality of many products that are consumed on a daily basis. Although CAT algorithms and computers also provide the opportunity to increase the possibilities within testing, traditional paper and pencil items continue to be used. This is a consequence of the great amount of research that is available that documents the validity of traditional items and testing methods. Future research should focus on investigating new test formatting possibilities that may provide a richer testing experience, and better testing results.

References

- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.
- Baker, F. B., & Kim, S. H. (2004). *Item Response Theory: Parameter Estimation Techniques*, 741 Marcel Dekker Inc.
- Bergstrom, B. (1992) Ability measure equivalents of computer adaptive and pencil and paper tests: A research synthesis. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Brown, J. M., & Weiss, D. J. (1977). *An adaptive testing strategy for achievement test batteries* (Research Rep. No. 77-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). Reading, MA: Addison-Wesley.
- Daville, C. (1993) Flow as a testing ideal. *Rasch Measurement Transactions* 7:3.
- Gialluca, K. A., & Weiss, D. J. (1979). *Efficiency of an adaptive inter-subtest branching strategy in the measurement of classroom achievement* (Research Report 79-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.
- Glas, C. A. W., Suarez Falcon J. C. (2003) A Comparison of Item-Fit Statistics for the Three-Parameter Logistic Model. *Applied Psychological Measurement*, Vol. 27, No. 2, 87-106
- Glas, C. A. W., (2005). The Impact of Item Parameter Estimation on Computerized Adaptive Testing with Item Cloning. Newtown, PA: *Law School Adminssion Council*.
- Glas, C. A. W., & van der Linden, W. J. (in press). *Modeling adaptive testing using item cloning* (Computerized Testing Report 01-03). New-

town, PA: Law School Admisssion Council.

Hambleton R. K., Swaminathan, H., & Rogers H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park CA: Sage.

Irvine S. H., & Kyllonen P. C. (2002). *Item Generation for Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kingsbury, G., Zara, A. (1991). A Comparison of Procedures for Content-Sensitive Item Selection in Computerized Adaptive Tests. *Applied Measurement in Education*, 4 (3) 241-61.

Linacre, J. M. (2000) Computer-adaptive testing: A methodology whose time has come. *MESA Memorandum*. No. 69.

Makransky, G., & Glas. C. A. W. (2008). *Bootstrapping an Item Bank: An Automatic Online Calibration Design in Adaptive Testing*. Manuscript submitted for publication (copy on file with author).

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149-174.

Maurelli, V. A., & Weiss, D. J. (1981). *Factors influencing the psychometric characteristics of an adaptive testing strategy for test batteries* (Research Rep. No. 81-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Computerized Adaptive Testing Laboratory.

Meijer, R. R., & Sijtsma. K. (2001). Methodology Review: Evaluating Person Fit. *Applied Psychological Measurement*. Vol. 25, # 2, 107-135.

Nye C. D., Do, B. R., Drasgow, F., & Fine, S. (2008) Two-Step Testing in Employee Selection: Is Score Inflation a Problem? *International Journal of Selection and Assessment*. 16, 112-120.

Potosky, D. Bobko, P. (2004). Selection Testing via the Internet: Practical Considerations and Exploratory Empirical Findings. *Personnel Psychology*. 57, 1003 – 1034.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Denmark's Pædagogiske Institut.

Rudner, L. M. (1998a.) *An applied study on computerized adaptive testing*. Rockland, MA: Swets & Zeitlinger.

Rudner, L. M. (1998b.) *An On-line, Interactive, Computer Adaptive Testing Tutorial*, retrieved August 2008 from <http://edres.org/scripts/cat/catdemo.htm>.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*. Monograph Supplement, No. 17.

Sympson, J.B. & Hetter, R.D. (1985) Controlling item exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Military Testing Association*. San Diego, CA: Navy Personnel Research and Development Center.

Thissen D. & Mislevy R. J. (2000). Testing Algorithms. In. H. Wainer. *Computerized Adaptive Testing. A primer*. Second edition. (pp. 101-133). Hillsdale, NJ

Van der Linden, W. J. & Glas, C. A. W. (Eds.) (2000a.) *Computerized Adaptive Testing. Theory and Practice*, Dordrecht: Kluwer Academic Publishers.

Van der Linden, W. J. & Glas, C. A. W. (Eds.) (2000b.) Capitalization on Item Calibration Error in Adaptive Testing. *Applied Measurement in Education*, 13(1), 35-53.

Way W. D. (2005) Practical questions in introducing computerized adaptive testing for K-12 assessments. Research Report 05-03. *Pearson Educational Measurement*.

Wainer, H. (2000) *Computerized Adaptive Testing. A primer*. Second edition. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wise, S. (1997). Examinee issues in CAT. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago.

Weiss, S. (2007). Basics of Computerized Adaptive Testing (CAT). Paper presented at the 2007 Computer Adaptive Testing Conference in, Minneapolis MN.

Weiss, S. (in press). CAT central web site: <http://www.psych.umn.edu/psylabs/catcentral/irtconcepts.htm>.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450.

Wise S. L., DeMars C. E. (2006) Low examinee effort in low-stakes as-

essment: Problems and potential solutions. *Educational Assessment*, 10, 1-17.

Wise S. L., Kong, X. (2005). Response Time Effort: A New Measure of Examinee Motivation in Computer-Based Tests. *Applied Measurement in Education*, 18, 163 – 183.

Kontekst-begrebet i kvantitativ forskning, generaliserbarhed og ekstern *invaliditet*

Af Mette Høyer

Indledning.

Mooks artikel, *In defense of external invalidity* (Mook, 2003), om en ukritisk anvendelse af begreberne *økologisk validitet* og *generaliserbarhed* bærer præg af en skarp argumentation og kritik af det, han finder er en principielt tåbelig anskuelse af disse begreber. Den fik mig til at tænke over, hvad det er Benny Karpatschof især har inspireret til igennem mit uddannelsesforløb. Det drejer sig om tilegnelsen af en særlig logisk tænkning, en stadigt væren på vagt over for mangelfuldt definerede betegnelser for universalialia, en tænkning som konsekvent stiller spørgsmålstegn ved *absolutter*, som f.eks den principielle kontekst eller konteksten som princip om man vil, eller antagelsen om, at der findes en simpel beskrivelse af det metodisk principielt korrekte. For en garvet akademiker er dette muligvis trivielt, men Karpatschof har altid kombineret denne tænkning med en særlig og ind imellem også lidt bidsk humor, måske netop pga. dette trivielle aspekt, hvilket undertiden har været ganske underholdende.

Empirisk forskning er en proces, som indebærer nogle kvalificerede valg i afgrænsningen af en empirisk tilgang til den valgte problemstilling, ihukommende, at der givetvis findes videnskabelige discipliner, som problematiserer de forhold, som netop disse afgrænsninger beror på. Karpatschof har gjort en dyd ud af opmærksomheden på sådanne andre videnskabelige perspektiver, dvs. på de afgrænsende valg, som kommer til at udgøre konteksten, forstået som de valgte empiriske forudsætninger og begrænsninger for en undersøgelse og for den metodiske tilgang til denne. Dette har givet en del digressioner i vores samtaler, nogle frugtbare, nogle mere frustrerende for en yngre kollega.

En refleksion over de metodiske betydninger af begreberne ekstern *invaliditet*, økologisk validitet og empirisk generaliserbarhed er derfor relevant for sådanne overvejelser, både i forbindelse med at designe et empirisk studie, men også i forbindelse med en systematisk vurdering af andres empiriske resultater. Her vil en sådan overvejelse tage udgangspunkt i nogle eksempler fra high-risk forskningen, som bliver præsenteret nærmere neden for. Begreberne belyses og diskuteres i relation til disse eksempler og afslutningsvist relateret til Karpatschofs model af det dialektiske forhold mellem teori og empiri.

Den familiære dynamik og tidlige forældre-barn relation i high-risk studierne

Den egentlige grund til, at jeg overhovedet faldt over denne problemstilling, udsprang af et problem, som jeg stødte på i mit ph.d.-projekt. ph.d.-projektets empiriske del var en undersøgelse af karakteristika ved forældre-barn samspillet hos familier, hvor en eller begge forældre havde en alvorlig mental lidelse inden for hhv. det skizofrene og det affektive spektrum.

I high-risk studierne¹ inden for skizofreniforskningen har fokus helt overordnet været at finde tidlige prædiktorer for en udvikling af lidelser inden for det skizofrene spektrum. Disse studier har været baseret på indsamling af data på børn og deres forældre med sådanne lidelser og børn i en kontrolgruppe. I nogle få af disse studier har estimater af forskellige aspekter af den familiære dynamik været nogle af de mange faktorer, som man har inddraget. Specielt i perioden fra midt 1970'erne og frem til slutningen af 1980'erne blev det aktuelt også at inddrage den familiære dynamik i disse studier på grund af den udvikling, der var sket specielt inden for spædbarnsforskningen i undersøgelser af betydningen af denne familiære dynamik for barnets tidlige udvikling (Ainsworth et al., 1978; Bowlby, 1997; Bråten, 1998; Stern, 2000).

Niemi og kolleger sammenfatter i et review de aktuelle empiriske fund om specifikke og nonspecifikke prædiktorer for udviklingen af skizofrene lidelser. Ud af de i alt 16 kohorter, som de inddrager i deres review, finder de 3, hvor forskerne specifikt har inddraget observation af forældre-barn samspillet. Et af disse, et studie fra Emory University (Goodman & Brumley, 1990), rapporterer fund baseret på en 3 års periode, hos 127 børn af mødre med hhv. en skizofren lidelse, en alvorlig depressiv lidelse eller en kontrolgruppe. Børnenes alder varierede mellem et par måned-

¹ High-risk studierne er pr. definition prospektive og longitudinale studier af specielt udvalgte kohorter af probander, forskningsdeltagere som genetisk er i øget risiko for udvikling af lidelser inden for det skizofrene spektrum. Det bør selvfølgelig nævnes, at denne forskningsstrategi anvendes inden for andre felter også, ligesom den også anvendes i undersøgelser af andre alvorlige sygdommes livsforløb.

er og 7 år ved første dataindsamling. Der er endnu ikke rapporteret om senere follow-up undersøgelser på denne kohorte, hvorfor den ikke skal omtales yderligere her. Dog skal det nævnes, at man i de 3 studier ikke har fundet entydige resultater af observationerne af mor-barn samspillet og ingen af disse studier har kunnet påvise, at aspekter af forældre-barn samspillet, dvs. af disse observationer var *specifikke* prædiktorer for udviklingen af skizofrene lidelser. Hvorvidt de kandiderer til at være non-specifikke prædiktorer er stadig uafklaret, især fordi opfølgning i voksenalderen endnu ikke er gennemført på disse studier (Niemi et al., 2003).

I de to andre studier, et amerikansk, *the Rochester Longitudinal Study* (Sameroff et al., 1982) med opstart i 1970 og et sydsvensk studie, *Offspring of Women with Nonorganic Psychosis* (McNeil et al., 1983) med opstart i 1973 har man gennemført follow-up dataindsamlinger over længere tidsspænd, hvorfor disse vil blive mere grundigt omtalt her. I begge disse studier har man inddraget data om forældre-barn samspillet og forældre-barn relationen i barnets tidlige leveår for at undersøge, hvorvidt den familiære dynamik i barnets tidlige alder kunne prædikere barnets generelle udvikling og tilpasningsevne i den senere barndom og ungdom. Man gennemførte undersøgelser af barnets tilknytningsmønster i relation til mødrene og observationer af mor-barn samspillet². I begge studier fandt man også en markant påvirkning på mor-barn samspillet og -relationen hos de familier, som var belastet af alvorlige mentale lidelser. Men når det er sagt, så var der også markante forskelle på de to studiers resultater og det lykkedes ikke at påvise en entydig sammenhæng mellem det belastede mor-barn samspil eller kvaliteten af barnets tilknytningsmønster og barnets senere udvikling frem til den tidlige voksenalder, hverken i sig selv eller i interaktion med andre faktorer. Forskernes tolkning af deres resultaterne var derfor også delvist modstridende. Dette betyder, at vi faktisk stadig ikke ved nok om betydningen af den familiære dynamik hos familier med alvorlige belastninger på grund af lidelser inden for det skizofrene spektrum, hvorfor en forskningsindsats på dette område stadig er relevant.

Resultaterne i det sydsvenske high-risk studie

I det sydsvenske studie blev i alt 192 gravide kvinder inkluderet, hvoraf 88 havde haft en non-organisk psykose (varierende over mindst 5 forskellige diagnoser). Resultaterne er baseret på hyppige observationer af mor-barn samspillet, på 3. og 6. dag efter fødslen, 3 og 6 uger efter fødslen og 3 og 6 og 12 mdr. efter fødslen (McNeil et al., 1985; Naslund et al., 1985; Persson-Blennow et al., 1984). Alle observationer varede en

² Disse studier kan, i lighed med megen af den tidlige forskning inden for tilknytningsteorien, kritiseres for at være modercentrisk. Imidlertid fandt jeg, i det materiale, som min ph.d. afhandling er baseret på, en markant overvægt af ukendte fædre.

halv time og indrog forskellige situationer, spisning, skiften ble og fri leg. Frafaldet på flere af disse tidspunkter var imidlertid relativt højt og varierende, specielt i relation til 6 og 12 mdr.s observationerne. Hos gruppen af mødre med en lidelse inden for det skizofrene spektrum, var frafaldet på næsten 50% ved 6 mdr.s undersøgelsen (McNeil et al., 1985).

Sammenfattende fandt forskerne, at samspillet især var påvirket i relation til en skizofren lidelse eller til en alvorlig depression, samspillet var mere negativt præget, dette var specielt udtalt i perioden frem til barnets 6 mdr.s alder. Ved 1 års alderen sås en øget frekvens af utryk tilknytning samt en øget frekvens af *manglende* fremmedangst, især hos børn af mødre med en skizofren lidelse (Naslund et al., 1984b; Naslund et al., 1984a; Persson-Blennow et al., 1986). Derimod var mor-barn samspillet hos gruppen af mødre med en alvorlig depressiv lidelse ikke anderledes end hos gruppen af mødre uden kendte psykiatriske lidelser (kontrolgruppen) igennem hele det første leveår. Først da barnets var godt 1 år, sås begyndende problemer hos gruppen af mødre med en alvorlig depressiv lidelse også.

I follow-up studierne på den sydsvenske kohorte har man, mig bekendt, ikke senere inddraget data fra disse tidligere observationer i analyser af graviditets og fødselskomplikationers betydning som potentielle prædikatorer for de skizofrene lidelser. Derimod har man i de senere undersøgelser inddraget mødrenes diagnostiske status og hhv. graviditets- og fødselskomplikationer (McNeil & Blennow, 1988; McNeil et al., 1988b; McNeil & Cantor-Graae, 1999), neurologiske (McNeil et al., 1988b; McNeil et al., 1992; McNeil & Cantor-Graae, 1999; Schubert & McNeil, 2007) og sociodemografiske data (McNeil, 1988; McNeil et al., 1988a) i en forsæt søgen efter signifikante markører for udviklingen af lidelser inden for det skizofrene spektrum. Variationen i mødrenes diagnostiske status i denne kohorte har sandsynligvis bidraget til, at man valgte at lave mange forskellige statistiske analyser på materialet. Fremfor at udføre en kontrol for effekten af de sociodemografiske data på alle disse analysers resultater, har man gennemført denne form for kontrol på nogle udvalgte fund.

Resultaterne i high-risk studiet fra Rochester

I det amerikanske high-risk studie fra Rochester inkluderedes 337 gravide, heraf havde 177 en mental lidelse, som ligeledes varierede over flere diagnosegrupper. Antallet af de gennemførte observationer af mor-barn samspillet var lavere i dette studie. Disse observationer fandt sted da barnet var henholdsvis 4. og 12. gammelt (Sameroff et al., 1982).

Da barnet var blevet 30 mdr. gennemførtes systematiske interviews af mødrene, for at afklare børnenes udviklingsstatus og mødrenes oplevelse af samspillet med barnet på dette tidspunkt.

Analyserne i dette studie påviste, at mor-barn samspillet var påvirket i relation til mødrenes diagnostiske status, med nedsat spontanitet og mindre nærhed i forhold til kontrol-mødre og deres børn. Dette gjaldt især for gruppen af mødre med en alvorlig depressiv lidelse og gruppen af mødre med en lidelse inden for det skizofrene spektrum, dog var påvirkningen af mor-barn samspillet mindre udtalt for denne sidste gruppe. Dette resultat afkræftede en af forskernes indledende arbejds-hypotese. De fandt dette påfaldende og lagde vægt på, at mor-barn samspillet var dårligt hos mødre med en lidelse inden for det skizofrene spektrum, men ikke dårligt nok til statistisk set at afvige signifikant i forhold til kontrolgruppens mor-barn samspil.

Imidlertid fremgik det også, at det ikke så meget var den kvalitative forskel på mødrenes psykopatologi, angivet ved deres diagnostiske status, som statistisk set kunne forklare variationerne i samspilsobservationerne. Det afgørende var derimod *intensiteten* af deres psykopatologi, *varigheden* af den og i hvor høj grad den kunne betegnes som *kronisk*. Forskerne konkluderede, at det ikke var mødrenes diagnostiske status per se, som prædikerer barnets udvikling, men snarere intensiteten og kroniciteten af mødrenes lidelse, som, sammen med flere andre risikofaktorer, kunne forklare en hæmmet udvikling hos barnet. Af alle de faktorer, som de indekserede for at gennemføre denne sidste analyse, var de mest markante mødrenes sociale status og deres etniske tilhørsforhold. I en logistisk regressionsanalyse var det disse to faktorer samt interaktionen mellem dem, som bedst kunne forklare variationerne i børnenes udvikling frem til deres 30 levemåned – statistisk set.

Forskerne tog i deres follow-up studier konsekvensen af dette fund, fremfor at fokusere på et statistisk link mellem mødrene diagnoser og alvoren af deres lidelser og børnenes udvikling, valgte de både ved barnets 4. og 13. leveår at inddrage deres indeksering af den samlede forekomst af potentielle risikofaktorer (Sameroff et al., 1987a; Sameroff et al., 1987b; Sameroff et al., 1993; Seifer et al., 1992; Seifer et al., 1996).

Variationerne i resultaterne og nogle mulige forklaringer på disse

Begge forskergrupper angiver klart, at de deskriptivt finder at mor-barn samspillet er påvirket/hæmmet hos gruppen af mødre med alvorlige depressive lidelser og lidelser inden for det skizofrene spektrum – påvirket i en grad, så det opleves bekymrende. Men de statistiske analyser kan ikke entydigt bekræfte, at dette forhold skyldes mødrenes diagnostiske status. Resultaterne fra de to kohorter angiver begge, at det ikke er mor-barn samspillet og relationen, som har en prædikativ betydning, hverken i forhold til barnets senere kognitive og socioemotionelle udvikling eller til den øgede forekomst af udviklingsmæssige vanskeligheder og psykopa-

tologi hos børnene.

Derudover er det påfaldende, at deres resultater er forskellige: observationerne af mor-barn samspillet hos kohorten fra Rochester er statistisk set mere påvirkede hos gruppen af mødre med en alvorlig depressiv lidelse end hos gruppen af mødre med en lidelse inden for det skizofrene spektrum, hvorimod i den sydsvenske kohorte er mor-barn samspillet hos mødre med en depressiv lidelse upåfaldende gennem hele barnets første leveår og afviger ikke statistisk fra kontrolgruppen. Disse variationer kan måske tilskrives sociodemografiske forskelle på datamaterialerne, hvilket ville være i tråd med amerikanernes vurdering; det ikke er mødrenes diagnostiske status per se, men derimod alvoren af denne lidelse i samspil med en række andre risikofaktorer, som har betydning for kvaliteten af det tidlige mor-barn samspil og for barnets tidlige kognitive og socioemotionelle udvikling.

Begge kohorter er rimeligt store og begge er som udgangspunkt systematisk udvalgt ud fra nogle velovervejede kriterier. Det er derfor påfaldende, at de begge kommer frem til, at det *ikke* er data fra de observerede samspil, som kan give et statistisk signifikant grundlag for barnets senere udvikling. Det er også slående, at effekten af mødres alvorlige depressive lidelse er forskellig i de to studier.

I den svenske forskergruppe valgte forskerne ved 6 års follow-up undersøgelsen at fokusere på aktuelle neuropsykologiske og motoriske data i relation til data fra neonatalperioden, i forsøget på at detektere meningsfulde prædiktorer for en senere udvikling af skizofrene lidelser. På Rochester kohorten har forskerne siden gennemført follow-up undersøgelser, da børnene var blevet hhv. 4 og 13 år gamle og i begge disse valgte forskerne konsekvent at arbejde videre ud fra et kompositindeks, hvor de adderede antallet af risikofaktorer, men ikke forholdt sig specielt til arten af disse, idet dette komposit-indeks i alle deres undersøgelser kunne forklare den største proportion af variationerne i børnenes udvikling. De involverede risikofaktorer var både mødrenes diagnostiske status, kronicitet og intensitet eller belastningsgrad af lidelsen og social status, uddannelsesniveau, bolig forhold, indkomst, etnisk tilhørsforhold m.m., mens data vedrørende den tidlige mor-barn relation, dvs. barnets tilknytningsmønster og observationer fra de tidlige mor-barn samspil, ikke indgik i de senere risikoindekseringer.

Man kan tolke disse forskelle ud fra flere forskellige vinkler. Umiddelbart er det oplagt at fremhæve, at forskerne bag det sydsvenske studie fastholdt et relativt snævert fokus i forsøget på at detektere distinkte prædiktorer for udviklingen af skizofrene lidelser, altså en fastholdelse af den genstand, som de fra starten af satte ud for at finde – mens forskerne bag Rochester studiet valgte en anden strategi, nemlig at fokusere på de faktorer, som statistisk set bedst kunne forklare variationerne i børnenes udvikling og på denne måde viste sig at være betydningsfulde for deres udviklingsmæssige problemer. De fastholdt altså også genstanden, prædik-

torer for udviklingsmæssige problemer, men tog konsekvensen af deres primære resultater og frigjorde sig fra en søgen efter tidlige markører for udviklingen af lidelser inden for det skizofrene spektrum. Deres fokus blev således først og fremmest at undersøge hvilke faktorer, der specielt havde en hhv. positiv og negativ effekt på børnenes udvikling og generelle intellektuelle formåen. Forskellen har den konsekvens, at de i deres follow-up anvender forskellige strategier: Svenskerne fastholder en søgen efter *specifikke* prædikatorer for de skizofrene lidelser og finder nogle mere non-specifikke udviklingsstier, som fører til en række psykopatologiske lidelser hos børnene i den tidlige voksenalder. Amerikanerne er mere loyale over for det, man måske kan betegne som den partikulære eksterne validitet³ og deres fokus bliver herefter bredt at vurdere risikofaktorerens betydning for udviklingshæmning, herunder også udviklingen af psykopatologi.

Svenskerne fastholdt således en antagelse om, at disse alvorlige lidelser er forbundet med tidligere individuelle udviklingsafvigelser, som, hvis de kunne bestemmes med tilstrækkelig sikkerhed, kunne inddrages i nogle frugtbare interventionsstrategier. Amerikanerne tog konsekvensen af deres empiriske resultater, i den forstand, at de anerkendte nogle kontekstbestemte faktorer, som værende de mest betydningsfulde for børnenes udvikling og udviklingsmæssige vanskeligheder⁴, repræsenteret i deres sociodemografiske data.

Zeanah, Boris & Larrieu (Zeanah et al., 1997) angiver, at de i deres forskning i tidlig udvikling og fejludvikling kun finder meget få faktorer, som har specifik betydning for barnets udvikling ud over det 3. leveår. De konkluderer, at barnets tidlige udvikling må ses *i kontekst af den tidlige omsorgsrelation og det familiære miljø*. Dette specificerer den partikulære eksterne validitet, fordi man ved at se barnets tidlige udvikling i kontekst af den *tidlige omsorgsrelation og det familiære miljø* netop får en mulighed for at nedbryde de kontekstekstuelle forhold, operationalisere dem, således at enkelte antageligt betydningsfulde aspekter, som f.eks. tilknytningen eller dialogiske mønstre eller familiens socioøkonomiske rammer kan inddrages konkret i de empiriske analyser af, hvad der befordrer hhv. hæmmer barnets tidlige og senere udvikling. Og det bør bemærkes her, at denne forståelse af familien som ramme for omsorgsrelationen og familiens miljø, set som udviklingsmæssig kontekst, er blevet adresseret forskelligt i de to ovenfor omtalte high-risk studier. Så lad os se lidt mere på dette kontekst-begreb, som blandt andre Kruglanski (Kruglanski, 1975) relaterer til begrebet ekstern *invaliditet* og til generaliserbarheden af empiriske undersøgelsesresultater.

³ Begrebet, *den partikulære eksterne validitet*, defineres og diskuteres alment i næste afsnit.

⁴ I forlængelse af dette kunne man sige, at man i det svenske studie holder sig inden for en snæver medicinsk forskningsmetode, mens man i den amerikanske i højere grad inddrager både psykologiske og især sociologiske aspekter, påtvunget af de indledende empiriske resultater.

Kontekst-forståelse, økologisk validitet, generaliserbarhed og ekstern invaliditet

Diskussionerne af meningen med begrebet ekstern invaliditet tager udgangspunkt i Campell & Stanley's definition af ekstern validitet fra 1966 (Kruglanski, 1975; Mook, 2003). Campbell og Stanley diskuterer både interne og eksterne problemer, som kan påvirke undersøgelsesresultatet relateret til en række forskellige undersøgelsesdesign. Kruglanski definerer begrebet ekstern invaliditet, som de ikke-planlagte interaktioner mellem et ubekendt forhold og den faktor (variabel), som udgør genstanden for undersøgelsen. Hermed søger han, at bestemme to forskellige betydninger af begrebet: en generel teoretisk og en partikulær eller kontekstuel. Vi gennemfører derfor et eksperiment, hvor vi ønsker at undersøge B for indflydelse af A ($A \rightarrow B$). Vi har en kendt mistanke om, at relationen $A \rightarrow B$ er under indflydelse af C. Altså må der være tale om en statistisk medierende eller modererende interaktion, ($C | A \rightarrow B$). Imidlertid finder vi postanalytisk ud af, at vores resultater har været påvirket af en uerkendt designfaktor, Z. Indflydelsen af C er en baggrundsvariabel, hvis effekt har betydning for B's interne validitet. Z har vi en mistanke til, men vi har ikke nødvendigvis kvantitative estimater af udover det generelle error-estimat, vi kan derfor heller ikke påvise en interaktion mellem $Z | A \rightarrow B$ eller for den sags skyld mellem $Z | C \rightarrow A \rightarrow B$ eller mellem $Z | A \rightarrow C \rightarrow B$. Kruglanski påpeger en fundamental forskel mellem forhold vedrørende målingens interne validitet (måler vi det, vi ønsker at måle?) og målingens generaliserbarhed (under hvilke forudsætninger er dette mål altid et validt estimat af det samme fænomen?). Han påpeger at ud fra en teoretisk betydning, angivet ved den interne validitet, er empiriske interaktioner problematiske i udforskningen af det teoretiske koncept⁵: *Its relentless quest is for maximally general laws, so an interaction (i.e., a departure from generality) exists only as a temporary (even if sometimes protracted) challenge to be overcome in the future by an insightful redefinition of the critical variables.* (Kruglanski, 1975, s. 655). Den partikulære betydning af både interaktionen med C og mistanken om en tilsvarende med Z angiver en begrænset generaliserbarhed givet C og potentielt også Z, hvilket kan reformuleres som et spørgsmål om, under hvilke forudsætninger B altid vil være et mål for det samme koncept. Hermed demonstrerer Kruglanski samtidig, at intern og ekstern validitet *ikke* begrebsmæssigt er dikotome modsætninger, men anddrager nogle kvalitativt forskellige aspekter af den kvantitative undersøgelsesmetodik.

Anderson & Bushman (1997) anvender og definerer begrebet ekstern validitet således: *External validity typically refers to the generalizability*

⁵Kruglanskis krav til B's generelle uafhængighed af empiriske interaktioner er tæt knyttet til det klassiske eksperiment, som afprøvning af en almen og idealistisk teoretisk konstruktion.

of the results of a study to other (usually real world) settings or populations. All else being equal, external validity is assumed to be a function of the similarity of the sample, setting, and empirical realizations of the variables in the target study to the population, setting, and empirical realizations of the variables in the target setting or population to which one wishes to generalize. The level of external validity also is usually assumed to be higher in field studies (correlational or experimental) than in laboratory experiments.(ibid., s. 21). Hermed begår de netop den logiske fejlslutning, som Kruglanski påpeger, idet den eksterne validitet bruges som betegnelse for alt det, som omfattes af begrebet generaliserbarhed. Anderson og Bushmans definition svarer til en hyppigt fremført kritik af specielt de eksperimentelle studiers kvalitet, relevans og videnskabelighed. Dette vender jeg tilbage til i forbindelse med Mooks kritik af den måde begrebet validitet typisk anvendes på.

Som Kruglanski påpeger, er intern validitet et udtryk for et estimats operationelle betydning (som helst skal være så entydigt og objektivt eller unbiased som muligt) og ekstern validitet er en betegnelse, som angiver generaliserbarhed, et problem, som ifølge Kruglanski intet har at gøre med *validitet!* Validitet hører til det indledende teoretiske tolknings- og afgrænsningsarbejde i forbindelse med et eksperiment, generaliserbarheden som problem hører til udkommet af det samme eksperiment, nemlig en refleksion over de konkrete forhold, hvorunder eksperimentets empiriske resultater er gyldige. Hertil kan tilføjes, at begrebet *intern* validitet af natur kun giver mening i relation til eksperimentelle studier, hvor der er én variabel faktor, da validitet altid er et spørgsmål om gyldigheden af den anvendte operationalisering, mens den eksterne validitet eller generaliserbarheden er relevant i forhold til mange forskellige undersøgelsesdesigns.

Mook (2003) angiver ligefrem, at begrebet ekstern invaliditet simpelt hen er noget forvrøvlet sludder. Han påpeger det fejlagtige i antagelsen om, at metodiske regler og ræsonnementer er ideale, at de per se rummer en gyldighed udover den konkrete problemstilling, som de anvendes på. Konkret fremfører han, at ekstern validitet logisk må være modsætningen til en ekstern invaliditet, men mens den eksterne validitet kan betegne en økologisk eller kontekstuel validitet, så kan en ekstern invaliditet ikke tilsvarende dokumenteres. Den økologiske eller kontekstuelle validitet er i og for sig også noget sludder, fordi som Kruglanski påpeger, har dette mest at gøre med generaliserbarhed og eksterne begrænsninger, og den eksterne invaliditet forekommer at være en logisk umulighed. Det er det samme som at sige, at en måling skal være universelt valid – men udsagt ikke med udgangspunkt i det teoretiske koncept eller dets operationalisering – men på baggrund af alt det, som ikke er blevet målt eller på anden måde inddraget i selve undersøgelsesdesignet. Mooks gennemgang af dette begreb problematiserer dels en ukritisk dikotomisering af validitetsbegrebet og mere væsentligt, antagelsen om, at resultaterne

af empiriske studier *altid* vinder ved en højere generaliserbarhed.

Hans væsentligste pointe er at advare, imod en ukritisk generalisering af specificerede metodiske kriterier og en ukritisk antagelse om, at man altid arbejder hypotetisk deduktivt hen imod at teste en antagelse om virkelige empiriske forhold fremfor at teste en teoretisk funderet hypotese i den kvantitative forskning. Han tager udgangspunkt i nogle klassiske og for udviklingspsykologien meget betydningsfulde studier for at eksemplificere sin kritik. Et af disse er Harlow's demonstration af, at abeungers nærhedssøgende adfærd og tilknytning har mere at gøre med tryghed end med mad. Den teoretiske implikation af dette fund var et opgør med teorierne om sekundær drift, at barnets udvikling af emotionelle bånd til moderen beroede på, at hun var kilden til næring, Harlow's fund udgjorde både en kritik af en datidig udbredt antagelse, både inden for den psykoanalytiske tradition og inden for social learning teorierne, men blev også en væsentlig inspirationskilde for Bowlbys udvikling af tilknytningsteorien.

Denne gennemgang er et rigtigt godt eksempel på Kruglanskis teoretiske betydning af begrebet intern validitet, eller mere præcist udtrykt, konstruktions- og begrebsvaliditet. Mooks pointe er imidlertid mere radikal, idet sådanne studier slet ikke er – eller behøver at være generaliserbare per se, de svarer til de eksperimenterelle komparative studier, som i allerhøjeste grad oftest tager udgangspunkt i et meget teoristyret design og et specifikt teoretisk problem. Og resultaterne af dem har ikke umiddelbart en praktisk eller generaliserbar betydning, hvorfor generaliserbarheden ikke i første omgang er et problem, relevansen af sådanne studier er jo netop den klassiske eksperimentelle teori-testning.

Imidlertid kan man af denne diskussion videre udlede, at validitetsbegrebet ikke kun har relevans i forhold til eksperimentel og kvasi-eksperimentel undersøgelsesmetode. Men at tale om intern validitet giver faktisk kun entydigt mening i relation til det klassiske eksperiment – fordi vi i dette taler om det valide fund af eksperimentet, som gyldigheden af den variation, som forekommer på en faktor, nemlig den, som entydigt påvirkes i eksperimentet – og tilsvarende i forhold til de enkelte målinger, som foretaget, f.eks. inden for rammerne af de longitudinale designs. I relation til high-risk undersøgelserne giver begrebet intern validitet først og fremmest mening i relation til de enkelte målingers resultater. En vurdering af resultaterne af disse studier er betydeligt bedre hjulpet ved inddragelse af validitetsbegrebets metodiske kompleksitet: målemetodernes operationalisering vurderet i forhold til begrebs-, indholds- og konstruktions-validitet, og undersøgelsesresultaternes konvergente, divergente og diskriminative validitet (Haynes, 2001; Karpatschof, 2006). Dette er imidlertid områder, som traditionelt undersøges inden for den psykometriske tradition (se f.eks. Karpatschof, 2006). I longitudinale designs, som eksempelvis anvendes inden for high-risk forskningen, er det primære problem ikke at undersøge de enkelte målingers validitet, men derimod at anvende disse

målinger, ud fra antagelsen om, at de kan vise noget substantielt om f.eks. specifikke og nonspecifikke risikomarkører for skizofreni. Og for at kunne dette anvender man målemetoder, og refererer til at andre har demonstreret og dokumenteret god validitet og reliabilitet. Dette forhold udgør en tredje mulig forklaring på resultaterne af de her gennemgåede high-risk studiers fund vedrørende mor-barn samspillet og barnets tilknytningsmønster. I begge studier anvendte man en modificeret version af *the Strange Situation Procedure* til bestemmelse af barnets tilknytningsmønster, godt nok under supervision af de førende forskere på dette område. Og man brugte ligeledes nogle modificerede observationsmetoder – og disse metoder er virkelig omkostningsfulde, hvilket forklarer, at de metoder, som blev anvendt, ikke på samme måde var psykometrisk vel-dokumenterede.

Er et sådan metodisk perspektiv relevant for udviklingsforskningen?

Det klassiske eksperiment har generelt været underkastet en kritik, som netop pointerede en *for tæt relation* mellem (ideal-)teori og eksperiment, hvilket gjorde resultaterne økologisk eller kontekstuel mindre relevante (en pointe, som Mook blandt andet fremhæver). Et af de fremtrædende karakteristika ved *Developmental Psychopathology* (Rutter & Sroufe, 2000; Rutter, 2006; Sroufe & Rutter, 1984) disciplinen er netop inddragelsen af reale faktorer i subjektets livsverden; hvis subjektet er et barn er en sådan inddragelse et uomgængeligt *must* (Zeanah et al., 1997). Menneskelig udvikling foregår ikke i et vakuum, men i et reelt levet liv i relation til familie, andre sociale relationer, de biologiske og fysiologiske individuelle og fælles begrænsninger, og de samfundsmæssige betingelser – det som i diskussioner af generaliserbarhed betegnes med samlebegrebet kontekst, udgør hele grundlaget for at kunne tale om forskellige udviklingsstier og specifikke og nonspecifikke prædiktorer for forskellige udviklingsmæssige udkomme.

I både high-risk forskningen og i alle andre prospektive undersøgelser af mennesker og deres udvikling og livsbetingelser inddrages konteksten typisk, repræsenteret ved nogle hyppigt anvendte og nogle gange også nogle partikulære baggrundsvARIABLE. Men konteksten i sådanne studier kan kun repræsenteres ved et vist antal sådanne kontrolfaktorer samt i den målefejl (måleusikkerhed), som empirisk altid viser sig.

Man kan altså imødegå kontekst-relaterede problemer ved at inddrage data vedrørende dette direkte (f.eks. ved indsamling af sammenlignelige sociodemografiske data); indirekte (f.eks. ved stratificeret sample) eller ved at sammenholde forskningsresultaterne med andre resultater fra kohorte-studier og fra epidemiologiske studier, hvor man i stedet for at inddrage kontekstuelle faktorer empirisk, implicit antager, at disse

er repræsenteret ved selve samplers sammensætning. Man er i netop sådanne studier meget opmærksom både på konteksten og på den genstand, man vil følge og dokumentere udviklingen af over en årrække/livsbane. Dette er selvsagt omkostningsfuldt og dette har selvfølgelig konsekvenser for selve undersøgelsesdesign.

Forskellen på det sydsvenske studie og studiet fra Rochester kan anvendes til at illustrere det meningsløse i distinktionen mellem intern og ekstern validitet og den heraf afledte eksterne invaliditet. Hvad er det i disse to studier, som den interne validitet skal omfatte? Det er ikke bare de diagnostiske kriterier, det er også symptombeskrivelserne, diverse baggrundsdata, samt data fra interviews, observationer og laboratorieprøver af meget forskellig art. Alene at redegøre for den interne validitet af disse studier er en kunst og det er da heller ikke det man gør i sådanne studier, derimod redegør man for de hypoteser, man satte ud for at undersøge og diskuterer de empiriske resultaters validitet på et trinøjere niveau: overensstemmelse med forventningen til resultaterne (arbejdshypoteserne), med resultater fra andre studier, samt hvis der forekommer afvigelser fra disse, hvorvidt man så kan redegøre systematisk for potentielle kilder til disse osv. Grundlaget for dette beror imidlertid på et velgennemtænkt design og valget af metoder, som man enten afprøver psykometrisk integreret i dette design eller som er veldokumenterede i andre tidligere studier.

Afrunding

Mit oprindelige fokus med inddragelsen af specielt disse to studier i mit arbejde, var netop at der var tale om nogle relativt store datamaterialer, som inddrog direkte observationer af det tidlige mor-barn samspil. Imidlertid har de anvendte observationsmetoder varieret, ligesom de tidspunkter, hvorpå man gennemførte observationerne var forskellige. I begge studier har observationsmetoderne været tilpasninger af nogle psykometrisk afprøvede metoder, hvilket betyder, at resultaterne af disse observationer bør betragtes som eksplorativt hypoteseafprøvende, da målingernes, dvs. observationsmetodernes validitet ikke var tilstrækkeligt afklaret forud for disse studier.

Men på det tidspunkt, hvor disse studier blev gennemført, var det relevant at observere samspil og tilknytning hos disse børn, fordi deres livsbetingelser var som de var. Formodningen var, at børnenes udvikling ikke blot var genetisk betinget, men i høj grad afhang af den tidlige omsorgsrelation. Og dette er formodentlig også rigtigt, man kan ikke alene med disse resultater tillade sig at forkaste denne grundhypotese – men spørgsmålet om, hvordan denne relations betydning bedst kan undersøges empirisk, det står stadig åbent. Og det er her Karpatschofs (2006) model af dialektikken mellem teori og empiri med metoden er

relevant. Fordi det er min opfattelse, at et af de væsentligste budskaber denne model rummer, er at kun gennem en stadig specificering af såvel teori og empiri, og en stadig vurdering af forskningsmetode og -design, bliver vi videnskabeligt klogere. Karpatschof begrundet endvidere nødvendigheden af denne proces med flg. ræsonnement: Da mennesket både som forskningsobjekt og som forsker er altfor socialt i sit virke, kan relationen mellem forsker og deltagere per definition ikke være objektiv, unbiased eller fuldstændig uden variationer. Variationer i målingerne er udtryk for målefejl, ja, men de skyldes genstandens særlige karakter, den menneskelige aktivitet. Derfor vil psykologisk kvantitativ forskning altid være belastet af reliabilitetsproblemer. Fremfor at problematisere den kvantitative metodes anvendelighed til studiet af mennesker og deres livsverden, fremføres i stedet behovet for en dialektisk videreudvikling og stadig specificering af både teorier og metoder i lyset af de resultater, som allerede er opnået. Derfor bør det overvejes, om psykologisk forskning mere aktivt skal inddrage spørgsmål om reliabilitet og måleusikkerhed i deres analyser (ibid. s. 14ff. & s. 82). Som Karpatschof fremhæver, så kan netop målefejl og uventede analyseresultater udgøre kilder til helt nye indsigter, som f.eks. Hawthorne effekten eller andre kilder til systematiske bias.

Den her fremsatte korte gennemgang af de to high-risk studier viser netop, at dialekten mellem teori og empiri og valget af metode hverken er entydig eller enkel. Således findes der heller ikke en enkelt *sand* metode til at kontrollere for metodefejl – eller systematiske effekter af den valgte metode, men dette bør ikke afholde os fra at prøve at gennemføre sådan en kontrol, således at vi når til et relativt sikkert resultat. Metodefejl kan vi forholde os til inden for rammerne af et konkret design, det er langt vanskeligere at forholde sig konkret til den eksterne validitet, det kræver både en kontekstuel reformulering af potentielle fejlkilder og en operationalisering af disse.

Anvendt litteratur

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of Attachment. A Psychological Study of the Strange Situation*. Lawrence Earlbaum Ass. Publishers.

Anderson, C. A. & Bushman, B. J. (1997). External validity of "trivial" experiments: The case of laboratory aggression. *Review of General Psychology, 1*, 19-41.

Bowlby, J. (1997). *Attachment and Loss Volume 1 Attachment*. (vols. 1) London: Pimlico.

- Bråten, S. (1998). *Intersubjective Communication and Emotion in Early Ontogeny*. Cambridge University Press.
- Goodman, S. H. & Brumley, H. E. (1990). Schizophrenic and Depressed Mothers: Relational Deficits in Parenting. *Developmental Psychology*, 26, 31-39.
- Haynes, S. N. (2001). Clinical applications of analogue behavioral observation: dimensions of psychometric evaluation. *Psychol.Assess.*, 13, 73-85.
- Karpatschof, B. (2006). *Udforskning i psykologien. De kvantitative metoder*. København: Akademisk Forlag.
- Kruglanski, A. W. (1975). The Two Meanings of External Invalidity. *Human Relations*, 28, 653-659.
- McNeil, T. F. (1988). Women with nonorganic psychosis: psychiatric and demographic characteristics of cases with versus without postpartum psychotic episodes. *Acta Psychiatr.Scand.*, 78, 603-609.
- McNeil, T. F. & Blennow, G. (1988). A prospective study of postpartum psychoses in a high-risk group. 6. Relationship to birth complications and neonatal abnormality. *Acta Psychiatr.Scand.*, 78, 478-484.
- McNeil, T. F., Blennow, G., & Lundberg, L. (1988b). A prospective study of psychosocial background factors associated with congenital malformations. *Acta Psychiatr.Scand.*, 78, 643-651.
- McNeil, T. F., Blennow, G., & Lundberg, L. (1988a). A prospective study of psychosocial background factors associated with congenital malformations. *Acta Psychiatr.Scand.*, 78, 643-651.
- McNeil, T. F., Blennow, G., & Lundberg, L. (1992). Congenital malformations and structural developmental anomalies in groups at high risk for psychosis. *American Journal of Psychiatry*, 149, 57-61.
- McNeil, T. F. & Cantor-Graae, E. (1999). Does preexisting abnormality cause labor-delivery complications in fetuses who will develop schizophrenia? *Schizophr.Bull.*, 25, 425-435.
- McNeil, T. F., Kaij, L., Malmquist-Larsson, A., Naslund, B., Persson-Blennow, I., McNeil, N. et al. (1983). Offspring of women with nonorganic psychoses. Development of a longitudinal study of children at high risk. *Acta Psychiatrica Scandinavia*, 68, 234-250.

McNeil, T. F., Naslund, B., Persson-Blennow, I., & Kaij, L. (1985). Offspring of women with nonorganic psychosis: mother-infant interaction at three-and-a-half and six months of age. *Acta Psychiatr.Scand.*, *71*, 551-558.

Mook, D. G. (2003). In defense of external invalidity. In A.E.Kazdin (Ed.), *Methodological Issues and Strategier in Clinical Research* (3 ed., pp. 379-387). American Psychological Association.

Naslund, B., Persson-Blennow, I., McNeil, T., Kaij, L., & Malmquist-Larsson, A. (1984a). Offspring of women with nonorganic psychosis: fear of strangers during the first year of life. *Acta Psychiatr.Scand.*, *69*, 435-444.

Naslund, B., Persson-Blennow, I., McNeil, T., Kaij, L., & Malmquist-Larsson, A. (1984b). Offspring of women with nonorganic psychosis: infant attachment to the mother at one year of age. *Acta Psychiatr.Scand.*, *69*, 231-241.

Naslund, B., Persson-Blennow, I., McNeil, T. F., & Kaij, L. (1985). Offspring of women with nonorganic psychosis: mother-infant interaction at three and six weeks of age. *Acta Psychiatr.Scand.*, *71*, 441-450.

Niemi, L. T., Suvisaari, J. M., Tuulio-Henriksson, A., & Lonnqvist, J. K. (2003). Childhood developmental abnormalities in schizophrenia: evidence from high-risk studies. *Schizophr.Res.*, *60*, 239-258.

Persson-Blennow, I., Naslund, B., McNeil, T. F., & Kaij, L. (1986). Offspring of women with nonorganic psychosis: mother-infant interaction at one year of age. *Acta Psychiatr.Scand.*, *73*, 207-213.

Persson-Blennow, I., Naslund, B., McNeil, T. F., Kaij, L., & Malmquist-Larsson, A. (1984). Offspring of women with nonorganic psychosis: mother-infant interaction at three days of age. *Acta Psychiatr.Scand.*, *70*, 149-159.

Rutter, M. (2006). *Genes and Behavior. Nature-Nurture Interplay Explained*. (1. ed.) Oxford: Blackwell Publishing.

Rutter, M. & Sroufe, L. A. (2000). Developmental psychopathology: concepts and challenges. *Dev.Psychopathol.*, *12*, 265-296.

Sameroff, A., Seifer, R., Zax, M., & Barocas, R. (1987a). Early indicators of developmental risk: Rochester Longitudinal Study. *Schizophr.Bull.*, *13*,

383-394.

Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: the influence of social and family risk factors. *Child Dev.*, *64*, 80-97.

Sameroff, A. J., Seifer, R., Barocas, R., Zax, M., & Greenspan, S. (1987b). Intelligence quotient scores of 4-year-old children: social-environmental risk factors. *Pediatrics.*, *79*, 343-350.

Sameroff, A. J., Seifer, R., & Zax, M. (1982). Early development of children at risk for emotional disorder. *Monographs of the Society for Research in Child Development*, *47*, 1-82.

Schubert, E. W. & McNeil, T. F. (2007). Neurobehavioral deficits in young adult offspring with heightened risk for psychosis who developed schizophrenia-spectrum disorder. *Schizophr.Res.*, *94*, 107-113.

Seifer, R., Sameroff, A. J., Baldwin, C. P., & Baldwin, A. (1992). Child and family factors that ameliorate risk between 4 and 13 years of age. *J.Am.Acad.Child Adolesc.Psychiatry.*, *31*, 893-903.

Seifer, R., Sameroff, A. J., Dickstein, S., Gitner, G., Miller, I., Rasmussen, S. et al. (1996). Parental psychopathology, multiple contextual risks, and one-year outcomes in children. *Journal of Clinical Child & Adolescent Psychology*, *25*, 423-435.

Sroufe, L. A. & Rutter, M. (1984). The Domain of Developmental Psychopathology. *Child Development*, *55*, 17-29.

Stern, D. N. (2000). *Spædbarnets Interpersonelle Verden*. (2 ed.) København: Hans Reitzels Forlag.

Zeanah, C. H., Boris, N. W., & Larrieu, J. A. (1997). Infant development and developmental risk: a review of the past 10 years. *J Am Acad.Child Adolesc.Psychiatry.*, *36*, 165-178.

Udvikling af velfunderede, relevante og valide spørgeskemaer til sundhedsvidenskabelig forskning

Af *John Brodersen*

Fra lægelig paternalisme til patientautonomi

Den bio-psyko-social sygdomsmodel

I slutningen af 1970'erne og begyndelsen af 80'erne foreslog George Engel, at sundhedsprofessionelle skulle udvide deres sygdomsopfattelse fra den daværende monokausale statiske og reduktionistiske biomedicinske sygdomsmodel til en mere dynamisk bio-psyko-social sygdomsmodel (Engel, 1977; Engel, 1980). Forskningsresultater havde før Engel fremvist tiltagende solid evidens for, at personers helbred og sundhedstilstand ikke kun afhang af den biologiske arvmasse, barnet havde fået tildelt ved undfangelsen. Den afhang også af det miljø, man var opfostret i, og det miljø man som individ levede i. Dette gjaldt ikke kun for fosteret i livmoderen men også for barne-, ungdoms- og voksenlivet, hvor andre faktorer end de biologiske kunne forklare sygdom og levetid.

I den biomedicinske sygdomsmodel betragtes individet som passivt, mens man i den bio-psyko-sociale sygdomsmodel opfatter individet som en aktiv aktør, der påvirkes og påvirker sine omgivelser (Iversen, Kristensen, Holstein, & Due, 2002). Dette gør sig dog ikke gældende ved naturkatastrofer, hungersnød, krig og lignende katastrofer. Omvendt er interaktionen mellem person og omgivelser afgørende for helbredet og sundhedstilstanden i lande med "vestlig" livsstil. Her er den biologiske arvmasse, personens psykiske habitus og aktuelle balance, samt de sociale samspil personen indgår i, afgørende for helbredet og sundhedstilstanden. Forskning har de sidste tre årtier bekræftet Engels hypoteser

om, at helbredet ikke kun er biologisk bestemt men er et samspil af mange faktorer. Den bio-psyko-sociale model præger i stigende grad lægevidenskabens sygdomsopfattelse og begreber. Den har resulteret i talrige forskningsprojekter, der forsøger at belyse og forklare årsagssammenhænge mellem forskellige variabler, domæner og begreber, der påvirker personers helbred. Der er publiceret andre teorier om sygdomsmodeller, som kan opfattes som afledt af den bio-psyko-sociale sygdomsmodel. Det gælder f.eks. den situationelle sygdomsopfattelse, psyko-neuro-immunologien, locus of control, sense of coherence, for at nævne nogle.

Den alment medicinske konsultationsproces

Ud over at lægevidenskaben ved hjælp af naturvidenskabelige, samfundsvidenskabelige og humanistiske forskningsmetoder har været i stand til at skabe et evidensbaseret grundlag for Engels bio-psyko-sociale sygdomsmodel, er der sket et paradigmeskift i lægernes opfattelse af patienterne. Sundhedsvæsener i Vesten har de sidste årtier oplevet et skift væk fra den paternalistiske lægerolle over til en mere dialogorienteret tilgang til patienterne, hvor patientens autonomi er i højsædet. Denne tendens er tydeliggjort i den alment medicinske konsultationsproces med McWinneys patientcentreret kommunikationsmodel. Modellen er blevet fordansket af praktiserende læge Ole Risør og Jan-Helge Larsen og beskrevet af Klaus Witt i lærebogen *Almen Medicin* (2007). Konsultationsprocessen er her beskrevet som en kommunikation mellem patient og læge ind delt i tre faser. I den første fase er den praktiserende læges opgave at være aktivt lyttende og stille åbne spørgsmål for at få patienten til at fortælle sin sygehistorie med sine egne ord, så man får afdækket patientens sygdomsopfattelse, frygt og forventninger til konsultationen. Anden fase markeres ved, at lægen resumerer, hvad hun/han har hørt og uddyber sygehistorien gennem afklarende diagnostiske spørgsmål. Dette fører evt. hen til lægens kliniske undersøgelse af patienten. Konsultationens tredje fase begynder med, at lægen giver patienten en forklaring på patientens symptomer ud fra sygehistorien, lægens fund ved den kliniske undersøgelse og lægens vurdering. I vurderingen inddrager lægen konvergenser og divergenser mellem patientens og lægens sygdomsopfattelser. Lægens vurdering inddrager og relaterer patientens frygt og forventninger. Konsultationen afsluttes med et *sikkerhedsnet*, hvor lægen sikrer sig, at patienten har forstået en evt. behandlingsplan ved hvad hun/han skal gøre, hvis det ikke går som forventet. Således indledes konsultationen i tre faser; en patient-, læge- og fælles-del. Denne model sikrer, at patienten hele tiden er i centrum, og at man løbende sikrer sig, at man forstår hinanden og er enige om de beslutninger, der træffes.

Der er således sket et radikalt lægeligfagligt kulturskifte væk fra en lægelig paternalisme hen i mod patientautonomi. Samtidig er der sket et skift væk fra en monokausal, statisk, reduktionistisk biomedicinsk

tankegang om sygdommes opståen, forværring, lindring, behandling, prognose og evt. helbredelse hen i mod en mere dynamisk helhedsorienteret kompleks bio-psyko-social sygdomsmodel. Der har været en tydelig erkendelse af, at den eneste afgørende faktor for en behandlingssucces ikke kun er sygdommens biomedicinske forløb, men også hvem personen er, der er diagnosticeret med en sygdom, hvilken uddannelse og hvilket liv personen har, samt hvilke ressourcer og belastning personen har. En behandling af et sygt menneske kan ikke adskilles fra patientens livssituation.

Generiske og specifikke spørgeskemaer

Dette paradigmeskift har ført til en udbredt anvendelse af spørgeskemaer til afdækning og måling af patienters og befolkningens opfattelser og oplevelser af bio-psyko-sociale aspekter ved kontakt med sundhedsvæsenet, f.eks. forebyggelsesinitiativer, diagnosticering og behandling af sygdomme. Formålet med spørgeskemaundersøgelserne er at få et indblik i individers og/eller patienters helbreds- og sundhedstilstand. Dette har medført publikationer, hvor f.eks. patienters medicinske problemstillinger er afdækket ved hjælp af spørgeskemaundersøgelser, og hvor patienters livskvalitet, helbredsrelaterede livskvalitet, psykiatriske morbiditet, angst, depression eller lignende begreber er forsøgt kvantificeret.

De spørgeskemaer, der anvendes i medicinsk forskning, kan opdeles i to kategorier; generiske og specifikke spørgeskemaer. Generiske spørgeskemaer er måleinstrumenter, der er udviklet til at opfange og kvantificere et eller flere overordnede helbredsrelaterede begreber. De generiske spørgeskemaer er udviklet til at kunne anvendes på tværs af populationer, landegrænser og kulturer. Modsat generiske spørgeskemaer henvender specifikke spørgeskemaer sig til veldefinerede populationer, personer med samme problemer – samme sygdom. Hovedparten af de medicinske spørgeskemaundersøgelser anvender generiske spørgeskemaer.

Generiske spørgeskemaer

Fordelene ved anvendelse af generiske spørgeskemaer er, at de har en anerkendt og veletableret validitet, at de giver information om betydningsfulde helbredsrelaterede aspekter, og at målinger med det samme generiske spørgeskema kan sammenlignes på tværs af forskellige patientgrupper fra forskellige lande eller kulturer. Så enkelt er det desværre ikke. Forudsætningerne for de generiske spørgeskemaer holder ikke stik. Flere undersøgelser viser, at validiteten ikke er konsistent, når man “flytter” det generiske spørgeskema fra en population til en anden (Hobart, Williams, Moran, & Thompson, 2002; Comins, Brodersen, Krogsgaard, & Beyer, 2007; Hagell, Tornqvist, & Hobart, 2008). Det har vist sig, at

generiske spørgeskemaer af patienter opfattes mindre relevante end specifikke spørgeskemaer. Dette bliver især problematisk, når dette gælder for patienter, der er mest syge (Hagell, 2008). Andre undersøgelser har bekræftet, at spørgeskemaer oversat til flere sprog, eller samme engelske version af et spørgeskema anvendt i to forskellige engelsktalende lande, har et eller flere items, der har differentiell itemfunktion (Brodersen et al., 2007).

Differentiell itemfunktion betyder, at et spørgsmål med tilhørende svarkategorier ikke opfattes og dermed ikke fungerer ens i forskellige respondentgrupper. Hvis et eller flere items i en skala har differentiell itemfunktion, kan det være muligt at justere herfor (Brodersen, Thorsen, & Kreiner, 2007). Det kræver, at itemfunktionen er *uniform*, da *non-uniform* differentiell itemfunktion umuliggør sammenligning mellem respondentgrupper (ibid.). Samlet set er der mange faldgrupper ved anvendelse af generiske spørgeskemaer, hvilket har medført en øget anvendelse af de specifikke spørgeskemaer.

Specifikke spørgeskemaer

En af fordelene ved at anvende specifikke spørgeskemaer er, at de har en højere sensitivitet end generiske spørgeskemaer (Wiebe, Guyatt, Weaver, Matijevic, & Sidwell, 2003). Specifikke spørgeskemaer indfanger i højere grad de relevante helbredsrelaterede aspekter, som den særlige gruppe af patienter/personer oplever og erfarer i forbindelse med sygdom eller andre betydningsfulde hændelser. Dette gør det muligt på et mere forfinet niveau at måle, om der f.eks. er forskelle mellem to medicinske behandlinger. En anden fordel er, at de begreber, som specifikke spørgeskemaer afdækker, opfattes mere relevante for patienterne end de begreber, som generiske instrumenter indeholder. En tredje fordel er, at et specifikt spørgeskema kan virke klinisk relevant for sundhedsprofessionelle, da indholdet i et specifikt spørgeskema svarer til patientens hverdag.

Der er således solide holdepunkter for og viden om, at man bør benytte sig af specifikke spørgeskemaer, hvis man ønsker et validt, reliabelt og konsistent mål for f.eks. bio-psyko-sociale aspekter. Dette paradigmeskift fra generiske mål til mere følsomme og specifikke mål har resulteret i, at der inden for de sidste 10-15 år er sket en stigning i antallet af publikationer om udvikling og validering af sygdomsspecifikke eller andre tilstandsspecifikke spørgeskemaer.

Udvikling og validering af specifikke spørgeskemaer

Det er et godt udgangspunkt først at søge i den videnskabelige litteratur efter et velegnet eller beslægtet specifikt spørgeskema (Streiner & Norman, 1995). Inden søgningen må man have sit forskningsspørgsmål helt præcist og klart formuleret. Hvis man ønsker at måle, hvilken livskvalitet

patienter med leukæmi har, mens de får kemoterapi, må man have afklaret, hvad man forstår ved livskvalitet. I litteraturen findes der flere forskellige teorier om, hvad livskvalitet betyder som begreb, og det bør tydeliggøres, hvilken teori man anvender i et spørgeskema til måling af livskvalitet (McKenna & Doward, 2004). Uanset om man vælger at ville afdække et overordnet begreb som livskvalitet, eller man ønsker at måle et mere afgrænset og specifikt begreb som f.eks. smerte, så skal den teoretiske forståelse af begrebet og begrundelse for valg af definition være afklaret og beskrevet (ibid.). Først herefter vil en søgning efter et velegnet spørgeskema kunne begyndes.

Oversættelse og tilpasning af spørgeskema

Ved en søgning i relevante databaser kan man ofte finde et anvendeligt spørgeskema, der muligvis dækker forskningsspørgsmålet eller et spørgeskema, som afdækker et beslægtet område. Mange af de spørgeskemaer, man identificerer i den videnskabelige litteratur er ikke udviklet på dansk – men på engelsk. Hvis man beslutter sig for at bruge et udenlandsk spørgeskema, skal det oversættes til danske forhold. Der findes mange forskellige oversættelsesmetoder af spørgeskemaer (Acquadro, Conway, Hareendran, & Aaronson, 2008). Formålet med en “oversættelse” er ikke at få en bogstavelig og ordret oversættelse af instruktioner, spørgsmål og svarkategorier i et spørgeskema. Formålet er derimod at bevare meningen af indholdet i de originale instruktioner, spørgsmål og svarkategorier i det oversatte spørgeskema (McKenna & Doward, 2005). Derfor tales der også i nogle videnskabelige kredse om både oversættelse og tilpasning af et spørgeskema fra et sprog til et andet – fra en kultur til en anden (*cultural adaptation*). I tilpasningen indgår, at et oversat spørgeskema skal testes for om meningen af indholdet er bevaret, om det oversatte spørgeskema er forståeligt, og om sproget i et spørgeskema er udtrykt i et klart utvetydigt sprog, der forstås af almindelige mennesker – også personer, der har en kort eller næsten ingen boglig uddannelse (ibid.). Desuden bør indholdet i den oversatte version testes for, om den tilsvarende population i det nye land finder indholdet relevant (Swaine-Verdier, Doward, Hagell, Thorsen, & McKenna, 2004). Man kunne for eksempel spørge, om det også er relevant for danske patienter med leukæmi at blive spurgt til deres sexlyst. Måske viser det sig, at danske patienter med leukæmi har andre værdier og/eller anden moral end patienter med leukæmi i spørgeskemaets oprindelsesland. Hvis dette var tilfældet, kunne man risikere, at et item om sexlyst ville virke stødende og/eller irrelevant i Danmark og dermed ødelægge validiteten af dette item. Det kunne også medføre, at validiteten af en skala blev ødelagt, hvis det pågældende item indgik i en dimension, der målte bredere aspekter af seksualitet end blot sexlyst.

Uanset om man tager udgangspunkt i et oversat spørgeskema, eller

man begynder helt fra grunden, gælder det, at spørgeskemaets dækningsgrad skal afklares (Streiner et al., 1995; Brodersen, 2006). Det vil sige om de items, spørgeskemaet indeholder, er dækkende for alle aspekter af forskningsspørgsmålet - får man svar på de spørgsmål, man stiller?

Sprogligt ekspertpanel og lægpanel

Ved en oversættelse og begyndende tilpasning af et spørgeskema fra et sprog til et andet kan det anbefales at benytte to sproglige paneler (Swaine-Verdier et al., 2004). Det første panel er et sprogligt ekspertpanel bestående af mindst tre dobbeltsprogede personer. Det andet panel er et lægpanel med 3-5 personer uden særligt kendskab til det sprog, som tales i det land, hvor det originale spørgeskema er udviklet fra. Deltagerne i de to paneler bør ikke have nogen medicinsk baggrund, og de bør ikke være repræsentanter for målgruppen, som senere skal besvare spørgeskemaet (ibid.).

Det er klogt at adskille oversættelsesprocessen og den begyndende tilpasningsproces fra processen med test af relevans og dækningsgrad. Patienter bør ikke bruges i denne fase, da de har svært ved at forholde sig til ordlyden, forståeligheden og meningen med et item i et spørgeskema. De vil være interesseret i, om indholdet af det pågældende item har relevans for dem og deres sygdom. Når man derfor skal teste, om en oversættelse af et spørgeskema er forståelig og tydelig, og om sproget er almindelig dansk, er et lægpanel at foretrække frem for et panel af patienter (ibid.). Relevans og dækningsgrad kan derimod kun testes i målgruppen for spørgeskemaet - i dette tilfælde patienterne (Streiner et al., 1995; Brodersen, 2006).

Enkelt- og fokusgruppeinterviews

Finder man ikke et egnet spørgeskema, må man begynde fra grunden. Uanset om man begynder fra grunden, eller man begynder ud fra et oversat spørgeskema, skal der udvikles en interviewguide ud fra den relevante videnskabelige litteratur. Især kvalitative undersøgelser af samme eller lignende problemstillinger kan rumme værdifulde oplysninger til udformning af en interviewguide. Når den ligger klar, kan man vælge at udføre enkelt- eller fokusgruppeinterviews - eller begge dele - til test af relevans og dækningsgrad.

Fordele og ulemper ved enkelt- og fokusgruppeinterviews

Der er fordele og ulemper ved både enkelt- og fokusgruppeinterviews. Fordelene ved enkeltinterviews er, at det er nemmere at få oplysninger om intime og tabuiserede emner. Ulemperne er dog, at intime oplysninger sjældent kan anvendes til at udforme items til et spørgeskema, idet så-

danne items kan virke stødende på nogle personer. Tabuiserede og intime emner i et spørgeskema kan komme for tæt på respondenter, og nogle af dem vil føle sig stødt eller vrede (Brodersen, 2006). Hvis respondenter bliver vrede eller føler deres grænser gået for nært af et eller flere items i et spørgeskema, vil det kunne få respondenter til ikke at besvare det pågældende item. Måske vil respondenterne endda undlade at besvare resten af spørgeskemaet. I værste fald kan dette resultere i, at respondenter ikke returnerer spørgeskemaet. Som udgangspunkt bør spørgeskemaer derfor ikke anvendes til måling af tabuiserede og intime aspekter. Hvis forskningsspørgsmålet handler om tabuiserede og intime begreber, bør det overvejes at anvende kvalitative metoder. En anden ulempe ved enkeltinterviews er, at de er tidskrævende sammenlignet med gruppeinterviews. Omvendt er en af fordelene ved fokusgruppeinterviews, at der på kort tid indsamles megen information. En anden fordel ved fokusgruppeinterviews er, at deltagerne "smitter" hinanden. Når gruppedeltagerne lytter til de andre gruppedeltageres historie om deres oplevelser og erfaringer, ligner denne historie måske på visse områder de lyttendes egne oplevelser og erfaringer. Dette kan udløse mange frugtbare associationer, som netop bringes frem i samspillet mellem deltagerne. De emner og områder, som uden problemer kan diskuteres i en gruppe af mennesker, som ikke på forhånd kender hinanden, er oftest passende i intimitetsniveau for, hvad man kan tillade sig at spørge om i et spørgeskema.

Relevans og dækningsgrad

De første enkeltinterviews, eller det første fokusgruppeinterview, vil bekræfte, om de begreber, man har identificeret i litteraturen, er relevante for den pågældende målgruppe. Desuden vil åbne spørgsmål til gruppedeltageres oplevelser og erfaringer kunne afdække nye ikke tidligere beskrevne begreber relevante for målgruppen. I de efterfølgende interviews skal informanterne herefter diskutere hvilke forskellige aspekter ved de identificerede begreber, der er relevante for dem (Brodersen, 2006). Dette medfører, at der vil fremkomme flere forskellige udtryk og kvaliteter, som alle ligger inden for det samme begreb. For eksempel er de tre udtryk/kvaliteter "ked af det", "trist" og "deprimeret" sandsynligvis alle tre forskellige aspekter af et begreb, man kunne kalde "nedtrykthed". De tre kvaliteter udtrykker forskellige sværhedsgrader af nedtrykthed, og de er nævnt i en rækkefølge med stigende styrke af nedtrykthed. Det næste og måske vigtigste ved udvikling af nye items bliver at få formuleret items, hvori man inddrager de meningsbærende ord fra de forskellige udtryk og kvaliteter, som er fremkommet i interviewene. Det er vigtigt dels at forsøge at undgå formuleringer af items, der umiddelbart virker redundante, mens det samtidig er vigtigt at have items, der både dækker milde, middel og svære grader/styrker af begrebet, som nævnt ved eksemplet om nedtrykthed.

Svarkategorier

Udover at spørgsmålene eller udsagnene i items inden for det samme begreb kan have forskellige sværhedsgrader og styrker, så kan svarkategorier med forskellige sværhedsgrader og styrker anvendes, f.eks. "nej, slet ikke", "ja, lidt", "ja, noget" og "ja, meget". Dette giver i det nævnte eksempel med nedtryktthed mulighed for ni forskellige "ja"-kvaliteter af nedtryktthed. Der er dog ingen garanti for, at f.eks. udtrykkene "noget ked af det" og "lidt trist" har forskellige sværhedsgrad af nedtryktthed. Dette bliver ikke mindre komplekst af, at de tre kvaliteter af nedtryktthed; "ked af det", "trist" og "deprimeret" sandsynligvis befinder sig på en ordinalskala, da der ikke nødvendigvis er samme kvantitative afstand i sværhedsgraden mellem at være "ked af det" og "trist" og mellem at være "trist" og "deprimeret". Det samme gælder afstandene mellem de fire svarkategorier. Dermed bliver formodningen om at have ni forskellige "ja"-sværhedsgrader af nedtryktthed meget vanskelig at overskue. Hvis man omvendt havde formuleret de tre forskellige kvaliteter af nedtryktthed som dikotome items, hvor respondenterne skulle svare "ja" eller "nej", så ville man med stor sandsynlighed have indfanget tre forskellige sværhedsgrader af nedtryktthed. Om disse tre kvaliteter ville dække måling af hele nedtryktthedsbegrebet, ville være vanskeligt at vide alene ud fra indholdet, meningen og formuleringerne af de tre items. Der er selvfølgelig omvendt en øvre grænse for, hvor mange svarkategorier mennesker er i stand til at skelne imellem. At forestille sig, at det f.eks. skulle være muligt at skelne nedtryktthed (eller andre lignende begreber) i 101 forskellige nuancer kan ud fra almindelig sund fornuft virke latterligt og useriøst. Ikke desto mindre er der 101 svarkategorier i de meget anvendte visuelle analogskalaer, der består af en streg på 10 centimeter inddelt i millimeter.

Statistisk skalavalidering

På et tidspunkt vil den information, man indhenter i de efterfølgende interviews, ikke længere være ny information, men indholdsmæssigt og i mønster ligne information fra tidligere interviews. Dermed har man opnået fænomenet (teoretisk) datamætning. Man har fået afdækket alle de relevante begreber, som har betydning for målgruppen. Man har dermed også fået afdækket de begreber, som indeholdes i forskningsspørgsmål. Om de udviklede og formulerede (nye) items rent faktisk måler forskellige aspekter af samme begreb, skal herefter testes statistisk. Ligeledes skal det afklares statistisk, om de items, man har udviklet, har tilstrækkelige forskellige sværhedsgrader til at kunne dække og dermed måle hele det begreb, man ønsker at kvantificere. Det ideelle er at have tilstrækkeligt mange ikke-redundante items, der hver især opfanger forskellige grader af begrebet. Desuden skal de skalaer, man har udviklet, indeholde items i toppen og bunden af skalaen, der er for nemme eller for svære

(eller attraktive/ikke-attraktive) for de fleste af respondenterne. Dermed undgås gulv- og/eller lofteffekt i den udviklede skala. Desuden skal man statistisk afklare, om de svarkategorier, man har udviklet, fungerer efter hensigten. Man kan statistisk teste, om de items, der indgår i spørgeskemaets skalaer, besidder differentiell itemfunktion, og om der er lokal afhængighed mellem items i en skala. Alle disse statistiske test kan udføres ved brug af forskellige psykometriske teorier og modeller. Hvis man ønsker information om de enkelte items sværhedsgrader, placeringer og rækkefølge på skalaen, skal man benytte sig af item respons teori. Blandt de forskellige item respons teori-modeller adskiller Rasch modellerne (Rasch, 1960; Andrich, 1978; Masters, 1982) sig ved at være de eneste, der besidder specifik objektivitet (Rasch, 1967) (invarians) og sufficiens (Andersen, 1977). Med specifik objektivitet forstås, at uanset hvilket sæt af items, man anvender fra en skala, der fitter Raschmodellen, så vil et vilkårligt sæt af items dækkende skalaen måle den samme forskel (eller mangel på samme) mellem to personer. Det betyder, at items, der besidder differentiell itemfunktion, ødelægger skalaens specifikke objektivitet. Ved sufficiens forstås, at den score, en person opnår på en skala, indeholder al nødvendig information. Man opnår ikke yderligere information ved at få at vide, hvordan personen har besvaret de enkelte items i skalaen. Den information rummer personens skalascore allerede.

Referencer

- Almen medicin* (2007). København: Munksgaard Danmark.
- Acquadro, C., Conway, K., Hareendran, A., & Aaronson, N. (2008). Literature Review of Methods to Translate Health-Related Quality of Life Questionnaires for Use in Multinational Clinical Trials. *Value in Health*, *11*, 509-521.
- Andersen, E. B. (1977). Sufficient Statistics and Latent Trait Models. *Psychometrika*, *42*, 69-81.
- Andrich, D. (1978). A Rating Formulation for ordered Response Categories. *Applied Psychological Measurement*, *2*, 581-594.
- Brodersen, J. (2006). Measuring psychosocial consequences of false-positive screening results - breast cancer as an example. *Månedsskrift for Praktisk Lægegerning*, 87-88638-36-7.
- Brodersen, J., Meads, D. M., Kreiner, S., Thorsen, H., Doward, L., & McKenna, S. P. (2007). Methodological Aspects of Differential Item Functioning in the Rasch Model. *Journal of Medical Economics*, *10*, 309-324.

Brodersen, J., Thorsen, H., & Kreiner, S. (2007). Validation of a condition-specific measure for women having an abnormal screening mammography. *Value in Health, 10*, 294-304.

Comins, J., Brodersen, J., Krogsgaard, M., & Beyer, N. (2007). Rasch analysis of the Knee injury and Osteoarthritis Outcome Score (KOOS): a statistical re-evaluation. *Scand. J. Med. Sci. Sports, Online Early Articles*.

Engel, G. L. (1977). The need for a new medical model: a challenge for biomedicine. *Science, 196*, 129-136.

Engel, G. L. (1980). The clinical application of the biopsychosocial model. *Am. J. Psychiatry, 137*, 535-544.

Hagell, P. (2008). Whose Quality of Life? Ethical Implications in Patient-Reported Health Outcome Measurement. *Value in Health, in press*.

Hagell, P., Tornqvist, A. L., & Hobart, J. (2008). Testing the SF-36 in Parkinson's disease. Implications for reporting rating scale data. *J Neurol., 255*, 246-254.

Hobart, J. C., Williams, L. S., Moran, K., & Thompson, A. J. (2002). Quality of life measurement after stroke: uses and abuses of the SF-36. *Stroke., 33*, 1348-1356.

Iversen L, Kristensen T.S., Holstein B, & Due, P. (2002). *Medicinsk sociologi*. (First edition ed.) København: Munksgaard.

Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika, 47*, 149-174.

McKenna, S. P. & Doward, L. C. (2005). The translation and cultural adaptation of patient-reported outcome measures. *Value Health, 8*, 89-91.

McKenna, S. P. & Doward, L. C. (2004). The Needs-Based Approach to Quality of Life Assessment. *Value in Health, 7*, S1-S3.

Rasch, G. (1967). An Informal Report on a Theory of Objectivity in Comparisons. In L.J.Th. Van der Kamp & C. A. J. Vlek (Eds.), *Psychological Measurement Theory* (pp. 1-19). Leyden: University of Leyden.

Rasch, G. (1960). *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielsen & Lydiche.

Streiner, D. L. & Norman, G. R. (1995). *Health measurement scales: A practical guide to their development and use*. (Second ed.) Oxford: Oxford University Press.

Swaine-Verdier, A., Doward, L. C., Hagell, P., Thorsen, H., & McKenna, S. P. (2004). Adapting Quality of Life Instruments. *Value in Health*, 7, S27-S30.

Wiebe, S., Guyatt, G., Weaver, B., Matijevic, S., & Sidwell, C. (2003). Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol.*, 56, 52-60.

Intelligenstests og Flynn-effekten i Danmark

Af *Tom Teasdale*

Intelligenstestning er nu over hundrede år gammel, idet den kan dateres tilbage til Binets første intelligenstest udviklet til undersøgelse af skolebørn i 1905. Binets tests målte en række forskellige sproglige evner, som for eksempel at kunne opremse årets måneder i omvendt rækkefølge eller at nævne ord af modsat betydning. De blev hurtigt oversat til en engelsk version af Terman ved Stanford universitet i Californien og kom til at hedde Stanford-Binet-testen, som blev revideret flere gange og stadig er bredt anvendt i dag. En dansk oversættelse af den første Stanford-Binet blev udgivet i 1940'erne.

Opgørelsen af Binets tests indebar en sammenligning af det enkelte barns præstation på de forskellige tests med gennemsnittet for børn i forskellige aldre. På denne måde kunne et barn have en "mental alder", der var større eller mindre end dets kronologiske alder. Forholdet eller "kvotienten" mellem de to, almindeligvis gange 100, gav en såkaldt "intelligenskvotient" eller IQ. Denne beregning ville klart ikke passe for voksne. Da han udviklede en intelligenstest for voksne, definerede Wechsler i 1930'erne en IQ-score ud fra, hvordan individet lå i forhold til andre voksne individer, og IQ-scoren blev indrettet så den passede til en "normalfordeling". Gennemsnittet var igen 100, og cirka 2/3 af alle individer skulle have en IQ mellem 85 og 115. En anden af Wechslers væsentlige udviklinger var at inkludere tests for ikke-sproglige evner, for eksempel at kunne løse geometriske opgaver. Men lige som Binet mente Wechsler, at det gav mening at kombinere præstationer på tværs af alle prøver for at opnå en global IQ-score (Jensen, 1998). De nyeste versioner af Wechslers tests er de bredest anvendte intelligenstests internationalt, inklusive i Danmark

En anden væsentlig historisk udvikling i intelligenstests kom med den skotske psykolog John Ravens konstruktion af et sæt af matricer, hvor en ukomplet figur skulle gøres komplet ved selektionen af en ud af flere alternativer, såkaldte "multiple-choice"-tests (Raven, 2000). Ravens tests

er også stadig bredt anvendt i Danmark i dag.

Den danske matematiker Georg Rasch argumenterede på et teoretisk niveau i 1940'erne for, at elementerne i en intelligenstest skulle placeres efter stigende sværhedsgrad, fra opgaver som næsten alle kunne løse, til opgaver, som kun de bedst begavede kunne løse. Hans teknik vandt international anerkendelse blandt mange psykometrikere i udviklingen af intelligenstests og er nu kendt som en "Raschskala" (Rasch, 1980).

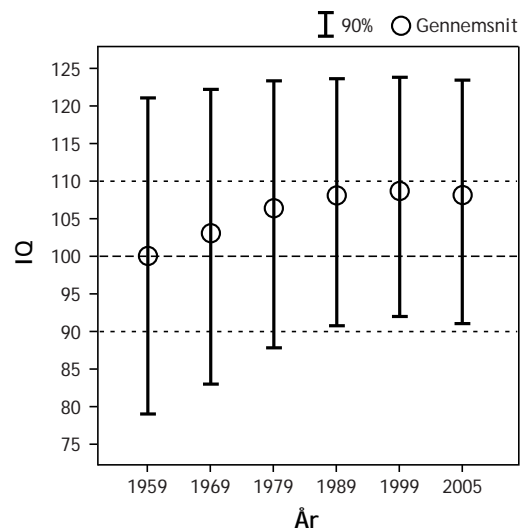
Den praktiske anvendelse af Raschskalaen dukkede først op i Danmark. Efter genindførelsen af værnepligt efter Anden Verdenskrig, mente man, at det var nødvendigt at udvikle tests for at vurdere de mentale evner hos de unge mænd, der blev indkaldt. Psykologen Børge Prien fik opgaven at udvikle testene. Han valgte at anvende fire tests, den ene lignede Ravens matricer, og de tre andre målte sproglige, numeriske og rumlige evner. Han opdagede hurtigt, ligesom Binet og Wechsler, at de forskellige tests korrelerede med hinanden, og at det derfor gav mening at udregne en total score ved at kombinere de fire tests.

Ved at anvende værnepligtige lykkedes det Prien at gennemføre meget omfattende afprøvninger i udviklingen og raffineringen af sine tests, og den endelige version blev officielt taget i brug i de tidlige 1950'ere. I 1957 blev testen inkluderet i sessionen, hvor alle danske mænd, med få undtagelser, skal møde op, når de fylder 18 år. Det er et påfaldende vidnesbyrd om Priens arbejde, at hans test, Børge Priens Prøve (BPP), fortsat bruges uændret ved session den dag i dag, 50 år efter. Det vurderes, at over 1,5 million danske mænd, hvoraf flertallet er mellem 18 og 65 år, har taget hans test. På grund af den stærkt stigende sværhedsgrad af opgaverne i de fire delprøver, har der i alle de år kun været en håndfuld af mænd, der har løst alle prøvens 78 opgaver korrekt. Det har vist sig, at BPP korrelerer højt med Wechslers intelligenstest (Mortensen, Reinisch, & Teasdale, 1989).

Ser vi igen på den internationale udvikling, opsummerede James Flynn i 1980'erne evidens, der tydede på, at præstationer på intelligenstests var blevet forbedret over generationer. Dette viste han først ud fra data publiceret i USA (Flynn, 1984), men dernæst dokumenterede han den samme tendens for data, der stammede fra 14 forskellige lande, dog ikke inklusive Danmark (Flynn, 1987). Hvor store forbedringer der var, varierede fra undersøgelse til undersøgelse, men Flynn konkluderede, at IQ steg med cirka 3-4 point per årti. Flynns stærkeste evidens kom fra de lande, som for eksempel Holland og Norge, hvor der var værnepligt, og som derfor havde testdata fra successive generationer af unge mænd. Flynns arbejde har vakt megen opmærksomhed inden for psykologien, og det fænomen, som han beskrev, er kommet til at hedde *Flynn-effekten*.

En undersøgelse af resultaterne fra den danske BPP har vist stigning i præstationerne fra de sene 1950'er til de sene 1970'er svarende til cirka 6 IQ-points, det vil sige meget i overensstemmelse med, hvad Flynn havde rapporteret for andre lande (Teasdale & Owen, 2000). Dette betyder, at

cirka 70 % af mænd født omkring 1960 præsterede bedre end gennemsnittet for de mænd, der var født omkring 1940. Flynns data var overvejende begrænset til år-for-år gennemsnit, men det har været muligt at undersøge hele spredningen af scores på den danske BPP nærmere. Som det kan ses i figuren, var der en betydelig større forbedring i den lave ende af fordelingen end i den høje ende. Mens proportionen af mænd, der løste 54 eller flere af BPPs i alt 78 opgaver kun steg fra 8-9 % til 11-12 % over de næste 20 år, faldt på samme tid antallet af dem der løste 22 eller færre opgaver fra 15-16 % til 6 %. Forbedringen i præstationerne fortsatte gen-



Figur 1. Resultater fra BPP-testen for danske værnepligtige mænd

nem 1980'erne, men i et reduceret tempo. Stigningen igennem disse ti år svarede til cirka 1,5 IQ-points, og næsten hele denne forbedring fandtes i den laveste halvdel af fordelingen. Over 1990'erne faldt stigningen til mindre end et halvt IQ-point, og fordelingen for 1999 var næsten den samme som for 1989. Siden de sene 1990'ere har der faktisk været et mindre tilbagefald i BPP scores, og i 2006 var gennemsnittet reduceret til niveauet i 1989 (Teasdale & Owen, 2005).

Det er værd at kommentere, at stort set det samme mønster af stigninger, altså hovedsagelig i den lave ende af IQ skalaen, over 1960'erne og 1970'erne, efterfulgt af en stagnation i dette århundrede, også for nyligt blevet rapporteret fra Norge, også med data fra værnepligtige (Sundet, Barlaug, & Torjussen, 2004).

De fire individuelle tests, som tilsammen udgør BPP, bliver ikke registreret separat i standardproceduren ved sessionen, kun totalen af de fire tests. I nogle år har det imidlertid været muligt at få resultaterne for de fire tests hver for sig. En sammenligning af scores for de sene 1980'ere

med de sene 1990'ere afslørede, at den eneste test, hvor der overhovedet var en forbedring, om end lille, var testen for rumlige evner. Gennemsnittet for de andre tre tests var stort set det samme over de ti år (Teasdale et al., 2000).

Hvad var årsagen til stigningen i testscores i den årrække, hvor den stod på? En første mulighed, som skal overvejes, er motivation. Det er velkendt, at motivation for en god præstation ikke nødvendigvis er så høj på session som den ville være under andre omstændigheder, hvor for eksempel en test bruges til at vurdere ansøgere til et job. Man har dog konstateret, at præstationen ved BPP har forbavsende lidt at gøre med personens holdning til at aftjene værnepligt. Ydermere skete stigningerne i BPP i en periode, 1960'erne og de tidlige 1970'ere, hvor Vietnamkrigen var på sit højdepunkt, og der var stor modstand mod militærtjeneste i Danmark, og et rekordstort antal af unge mænd valgte at blive militærnægtere.

En mere generel hypotese, som er blevet foreslået som forklaring på fænomenet, er indflydelse fra biologiske faktorer. Man har observeret, at der i perioden med IQ-stigninger også var væsentlige forbedringer med hensyn til ernæring og folkesundhed, i hvert tilfælde i de økonomisk udviklede lande, hvorfra Flynn hentede sine data (Lynn & Vanhanen, 2002). Der er indirekte dansk evidens for at støtte denne hypotese. Højde er en god indikator for ernæring, og højden hos voksne mænd er steget jævnt gennem de fleste årtier i det forrige århundrede, dog ses kun en meget lille stigning fra de tidlige 1980'ere op til i dag. Imod den hypotese, at Flynn-effekten er forårsaget af stigningen i ernæring og folkesundhed taler den kendsgerning, at der — igen i de udviklede lande — ikke er påvist nogen særlig forbindelse mellem dette og IQ. Korrelationen mellem højde og IQ er specielt lille.

Mange forskere har ment, at sociale og især uddannelsesmæssige ændringer er årsagen til Flynn-effekten. I Danmark behøver man kun at tænke på den meget store ekspansion af uddannelse i bredeste forstand, som har fundet sted siden 1950'erne, fra pædagogisk orienterede børnehaver til hjælpeklasser, specialundervisning og den længere skolegang til den meget store vækst af unge, der går på gymnasier o.l. Kvalitative ændringer i skolefagernes indhold fra udenadslære til problemløsningstænkning kunne også bevirke at opgavetyperne i BPP blev mere velkendte, end det var i de første årtier, hvor prøven blev brugt. Eftersom IQ uvægerligt korrelerer højt med uddannelse, virker det intuitivt antageligt, at noget af denne korrelation skyldes indflydelsen af uddannelse på IQ.

Hypotesen om uddannelsens indflydelse er også kompatibel med de to andre aspekter af Flynn-effekten, som den har manifesteret sig i Danmark. For det første ville mange af de uddannelsesmæssige ændringer her beskrevne have været mere til fordel for de socialt dårligst stillede end for de bedre stillede. Dette ville mindske afstanden mellem ekstremerne, netop hvad man har set i testresultaterne. For det andet fandt de store æn-

dringer og ekspansion af uddannelsessystemet sted i 1960'erne og 1970'erne og er aftaget noget siden. For eksempel har andelen af studerende, der går på gymnasiet o.l., været forholdsvis stabil gennem de sidste 15 år.

Men i sidste ende må forklaringen af Flynn-effekten forblive hypoteser. Hvad der er sikkert, er, at kurven er knækket. Undersøgelser af resultaterne efter år 2000 indikerer ikke, at intelligens, som den er målt af BPP er faldende, men snarere at den er stabiliseret. Det er svært at spå om fremtiden, men det synes usandsynligt, at en væsentlig ændring opad snart skulle komme igen. Det er dog endnu mere usandsynligt, at intelligensen skulle tage et dyk.

Afslutningsvis bør et forbehold nævnes. De her omtalte resultater handler, i sagens natur, udelukkende om mænd, og ikke om kvinder. Men da der er meget lidt evidens for betydningsfulde forskelle i generel intelligens mellem mænd og kvinder (Halpern & LaMay, 2008), er der nok heller ingen grund til at formode, at Flynn-effekten ikke også påvirkede intelligensen hos kvinder i Danmark i samme grad som det gjorde hos mænd.

Referencer

- Flynn, J. R. (1984). The Mean IQ of Americans - Massive Gains 1932 to 1978. *Psychological Bulletin*, *95*, 29-51.
- Flynn, J. R. (1987). Massive IQ Gains in 14 Nations - What IQ Tests Really Measure. *Psychological Bulletin*, *101*, 171-191.
- Halpern, D. F. & LaMay, M. L. (2008). The Smarter Sex: A Critical Review of Sex Differences in Intelligence. *Educational Psychology Review*, *12*, 229-246.
- Jensen, A. R. (1998). *The G Factor: The Science of Mental Ability*. Westport CT: Greenwood Press.
- Lynn, R. & Vanhanen, T. (2002). *IQ and the wealth of nations*. New York: Praeger Publishers.
- Mortensen, E. L., Reinisch, J. M., & Teasdale, T. W. (1989). Intelligence As Measured by the Wais and A Military Draft Board Group Test. *Scandinavian Journal of Psychology*, *30*, 315-318.
- Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. (2nd ed.) Chicago: University of Chicago Press.
- Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, *41*, 1-48.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32*, 349-362.

Teasdale, T. W. & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences, 39*, 837-843.

Teasdale, T. W. & Owen, D. R. (2000). Forty-year secular trends in cognitive abilities. *Intelligence, 28*, 115-120.

Vores indtryks udtryk – et metodesyn på vore “sprog”vaner

Af *Jesper Hermann* ¹

Ledetråd

William James' *Principles of Psychology* er mit udgangspunkt for at se sammenhængende på vore psykologiske funktioner som vores mentale forankring i verden. Hvordan kan vi med støtte hos sprogforskeren Otto Jespersen, dernæst undersøge 'vores indtryks udtryk' (sprogpsykologiens genstand) som fænomener?

I

Ved nu at gå over til at tale om sproglighedens psykologi istf. "sprogets" (som vi hidtil har gjort det, fx ved at bruge betegnelsen "sprog"psykologi), så undgår vi at blive fastholdt i en lang tradition som særliggør det sproglige på bekostning af det almenpsykologiske i vores omgang med verden (jf. Hermann, 1975, 2007b).

Det giver vanskeligheder hvis vi uden at begrunde det (sic), på forhånd skiller 'det sproglige' ud som noget særligt (jf. Naur, 2002, 59f). Derfor ser jeg det sproglige som udtryk for nogle af de mange vaner vi har erhvervet os, vanerne som udfylder størstedelen af vort vågne liv (jf. Jespersen, 1905 og Hermann, 2005). De vaner vi har med det sproglige er ligesom vore andre vaner først og fremmest noget psykologisk, knyttet til den eller de personer som har dem. Det er vanerne vi bruger hver gang vi siger noget, og hver gang vi forstår noget på den baggrund, livet har udstyret os med. Nogle tilsvarende vaner ytrer sig i skriftlige udgaver når vi skriver eller læser noget, det skrevne udformer vi forskelligt alt efter hvilken genre²

¹Let ændret version af foredrag for Department of Psychology, University of Hertfordshire, 2005 May 26.

² Dvs. tilvante sæt fastlagte, ensartede forventninger som er knyttet til bestemte typer udtryk, fx vittigheder, eventyr eller kriminalromaner. Peter Stray Jørgensen fra Akademisk Skrivecenter ved Kbh. Univ. har fortalt om genreforventninger der knytter sig til videnskabelige tekster som denne (Jørgensen, 2004). Frans Gregersen giver et

vi bruger. Det jeg fremhæver som almene træk ved alle mulige sproglige fænomener skal forstås i sammenhæng med resten af vort sindsliv, med den indre helhed vi normalt oplever i sindet. Noget af det vigtigere ved denne oplevede helhed er: *der er aldrig noget sprogligt først*. Før det sproglige kommer frem og blive foldet ud, må der allerede være **en person** (i hvert fald os selv) på banen. Men personer befinder sig altid i **situationer**, som er med til at give betydning til de fornemmelser og det sprog der optræder i personerne, og som de **vil** eller **skal** noget med.

Selv nu mens du læser dette er der hele tiden også andet i gang i dit hoved end de ord jeg skrev og de forestillinger de vækker. For hele tiden har hver af os en **tankestrøm**³ kørende indeni os som vi følger og falder hen i. For det vi alle erfarer hvis vi vender vor opmærksomhed indad i et af vore vågne øjeblikke, det er et mylder af forskellige forestillinger, fornemmelser med følelser og førsproglige tankeobjekter der er i gang i stadig vekslende strømmen. Nogle af dem kommer som sagt lige nu fra det du læser, de fleste af dem kommer fra dit eget sinds hukommelsestråde.

Sådan blev menneskers mentale liv allerede beskrevet for over 100 år siden. Som sådan har det ikke forandret sig, vi går stadig alle rundt med hver sit tankeliv af denne karakter. Men sådan taler vi ikke om det, vi tror vi kun rummer et enkelt tankeobjekt ad gangen. Fordi vi bruger ord for at viderebringe vore fornemmelser til hinanden, ledes vi til at tro at vi har hovedet og resten af kroppen fuld af ord. Det passer bare ikke. Allerede i 1836 skrev Poul Martin Møller i sine strøtanker: "Husk, at vi ej tænker i ord blot."

Vi kan ikke i tankestrømmen⁴ skelne mellem skarpt afgrænsede dele eller elementer af nogen art. Hvert øjeblik er vor tanke opfyldt af noget der er sammensat, *men som vi oplever som en helhed*. Psykologen William James (1842-1910) bruger udtrykket *tankeobjekter* om de helheder. Inden for hver af dem kan vi tale om noget der opleves som mere centralt, nemlig det der i øjeblikket er opmærksomhedens genstand i fokus, og noget der udgør en *bræmme* uden om som vi distraheres mere eller mindre af.

Når vi er vågne og fungerer mentalt foregår der en bestandig 'strømmen' af alle mulige ideer og fornemmelser igennem os. Vi er vant til at det er sådan og tænker ikke over det (jf. Csikszentmihalyi, 2005). For at kunne beskrive det sproglige under denne synsvinkel må vi iagttage det hvor det befinder sig, i vor tankestrøm. Dén bliver udgangspunkt for

godt, alment genreoverblik (Gregersen, 2008).

³ Det hører til en anden sammenhæng at give detaljerede begrundelser for hvorfor der ikke kan være tale om nogen "*bevidsthedsstrøm*"; lad det her række at det udtryk er for tingsliggørende og giver misvisende litterære associationer (jf. slutningen af gennemgangen af selvets fremtrædelsesformer, James, 1890, 401, hvor tænkeren ('subjektet') ikke **er** andet eller mere **end skiftende tanker**.)

⁴ '...forståelsen af den sproglige ytring består af tanker og følelser i en persons tankestrøm. (...) Det er således klart at det at opfatte talte ytringer eller skrevne tekster foregår ved perception og at det der derved fremkommer er tankeobjekter i tankestrømmen.' Naur 2002: 48.

beskrivelsen. I tankestrømmen oplever vi strømmevis af fornemmelser, følelser, tendenser og ord. Af alt dette tanke- og følelsesliv er det kun ordene og deres tonefald vi deler med hinanden. Det er forbløffende så langt vi når med det. Så langt at vi ofte glemmer eller overser den meget rigere mentale virkelighed vi taler ud fra. Dette er under normale omstændigheder ikke noget vi tænker over; det vi siger forholder sig med Peter Naurs metafor således

Metafor 3: *Et menneskes ytringer har samme forhold til menneskets indsigt som sprøjtet over bølgerne har til det gyngende hav under dem.* Denne metafor sigter mod at angive vore verbale ytringers forbigående karakter, dette at de formes, ikke som en kopi af indsigt der allerede har verbal form, men som resultatet af en formuleringsaktivitet der finder sted i det øjeblik ytringen fremkommer." (Naur, 1999, 89).

Med den malende metafor viser Naur hen til vigtige indsigter om vore sprogvaner. En har at gøre med at det er *noget nyt der sker* hver eneste gang vi siger noget⁵, selv om vi altid bruger vante udtryk, ord vi har fra andre⁶. En anden peger på at vores indsigt altid *omfatter mere* end det vi giver udtryk for. Tankestrømme i vore sind bevirker at vi hver især altid og nødvendigvis får noget andet ud af det vi hører eller læser, end dem vi taler med eller som har skrevet det. Vi skaber os hver sin udgave af det der bliver sagt. Selv om vi taler om det samme, er det om hver vores individuelle version af det.

Det vigtigste jeg har at sige er – let paradoksalt – at det sproglige lever på en baggrund af noget andet, **førsprogligt**. Som i sagens natur ikke kan beskrives, men kun udpeges. Uden dette førsproglige ville vi ikke kunne klare os, eller "være os selv". Herved bliver det begribeligt hvorfor drømme og meditation virker som de gør, helsebringende: de giver tankestrømmen mulighed for at 'komme sig' (jf. James, 1906, 57f.).

II

Når nu det sproglige viser som nogens vaner, hvilke følger får dette for metoder vi kan udforske det med?

'*Bevisbyrden*' med hensyn til at udforske vort liv som sprogbrugende væsner - altså de argumenter vi kan kræve af forskerne for valg de har foretaget *før* forskningen kommer i gang, den bevisbyrde ligger både hos sprogforskere og hos psykologer. For begge fags traditioner tager - uden at give nogen som helst argumenter for det - *adskillelsen* af det sproglige fra resten af vor verden (separationen eller udskillelsen ('segregationen' som Roy Harris (1998) kalder det) *for en given sag*. Roy

⁵ Jf Roy Harris om at hver sprogsituation er unik, Roy Harris 1996a.

⁶ og som kan bære ekko fra dem og de sammenhænge de har optrådt i. Se hertil om Bakhtins begreb *det fremmede ord* hos Nina Møller Andersen 2003. Og bemærk hvorfor og hvordan *Politikens* "At tænke sig" er morsom.

Harris (1996b) viser hvordan den samme u-argumenterede, u-bevidste omgang med det sproglige sker ensartet inden for både sprogforskning og filosofi, jf. Peter Naur (1995; 1999).

Denne artikel udfolder et anderledes perspektiv. Det henter vi hos William James' og Peter Naurs (1928-) sammenhængende måde at opfatte forholdet mellem på den ene side det vi er ude for når vi tænker, føler eller gør noget, og så på den anden side de ord vi bruger når vi beskriver det. De ser begge dette som en klar dualisme (James, 1890, VI & 185; Naur, 1999, diverse opslag om "sprog-vildfarelser" og "ord-som-noget-vildfarelsen"). Niels Bohr (1985,147-154) fremhævede også hvordan vi kun har ét problem - at bruge ordene rigtigt (jf. Zinkernagel (1992,15).

III

Men hvordan tager det sig ud når vi så ser sprog og verden under ét, og begge som ligefuldt integrerede ved hjælp af vore handlinger? Vi ser om lidt på to almindelige erfaringer som enten

'A) Jeg har skrevet noget. Hvorfor skal jeg så have en anden til at sige hvad der står?' eller

'B) Sommetider har vi svært ved at udtrykke noget vi godt véd; hvad betyder dette?' (-)(-)

I følge Peter Naur indebærer *videnskabeligt arbejde* at ordene bruges til *sammenhængende beskrivelser* (Naur, 1995, 317ff). Derfor har forskere over for deres forskningsområder den dobbelte opgave både at respektere fænomenerne de undersøger, og vælge de rigtige (jf. Bohr ovenfor) ord at omtale dem med.

Det vil sige udtrykke sig *så tilhørerne kan flytte sig mentalt i den retning*, forskerne mener de skal. Eller som en dansk essayist for 250 år siden udtrykte et stadigt gyldigt princip for kommunikation: "Du begriber ikke tingene ret, før du kan beskrive dem så andre forstår dem." (Sneedorff, 1744). For 100 år siden skrev sprogforskeren Otto Jespersen et flere sider langt opslag om **Sprog** i *Salmonsens Leksikon* hvorfra der citeres:

(...) for rigtig at forstaa de sproglige Teelser må man bestandig spørge hvad der virkelig foregår eller foretages af de talende og lyttende mennesker. (...) (Salmonsens Konversationsleksikon 1905, bd. XVI: 552.)

40 år senere supplerer han det med en teoretisk konstatering som eksplicit går ganske imod den berømte sprogteoretiker F. de Saussures beskrivelser af "sprog" som værende *systemer og deres strukturer*: "Men husk at sprog altid betegner vaner; og hvis sproget er et system, så er det et system af vaner" (Jespersen, 1941, 177).

I forlængelse heraf må vi tænke anderledes på vort liv med det sproglige end vi plejer at gøre, eller snarere ikke gøre, for de fleste af os skænker sjældent det sproglige nogen tanke når vi fx snakker sammen. Kun for sprogforskere og filosoffer eksisterer sproglige udtryk som noget der kan

tages frem og behandles for sig. Som nævnt ovenfor ligger bevisbyrden hos de veletablerede forskningstraditioner som uden argumenter har udskilt og separeret menneskers sproglige liv fra deres mentale liv i øvrigt. Dette vil indebære at menneskers sproglige liv ('the verbal life of human beings') *mister sin videnskabelige særstatus* som det opnåede i forrige århundrede gennem sprogvidenskabens selvstændiggørelse og strukturalismens sejrsgang.

Vort sædvanlige separerende perspektiv som placerer *det sproglige* for sig i forhold til vore andre færdigheder ændres, sådan at vi for fremtiden i stedet beskæftiger os med de sproglige færdigheder som endnu en gruppe blandt de **vaner** som både vort legemlige og vort åndelige liv udgøres af. Hvis vi derfor ser alle færdigheder som brug af indøvede vaner, så ses vor udøvelse af sproglige færdigheder mere sammenhængende og på linje med vore andre færdigheder. Dette fremhæver jeg med henvisning til William James' opfattelse af psykologis formål og af hvorfor vi studerer vore 'tanker', - og hvorledes 'tænkeren' opløses undervejs i beskrivelserne (James, 1890, 224f og 401).

Nedenfor argumenterer jeg yderligere for denne nye og 125 år gamle måde at anskue vor åndelige virksomhed som forankret i 'tankestrøm' og indbefattende 'det sproglige'.

IV

A) Jeg har skrevet noget. Hvorfor skal jeg så have en anden til at sige mig hvad der står?

Hvad sker der når vi læser noget? Med en grov skitse kan det beskrives således: vi danner nogle af vore sædvanlige associationer til ordene vi ser på siden, på denne facon ". . . det dær 'læser noget' i første linje lige før og ovenfor, bringer mig i tanker om 'det og det' (fx i modsætning til 'skriver noget') . . ." Og her må læseren selv fortsætte for det vi kommer til at tænke på må nødvendigvis altid være individuelt og privat jf. (James, 1890, 226f; Hermann, 2008); på dette beskrioveniveau sker der når vi læser i princippet det samme som når vi lytter: der dukker associationer op i forbindelse med vort *tankeobjekt*, d.v.s. i tankestrømmens nuværende fokus.

Men når vi læser en side **vi selv** har skrevet, så vil vi flere gange før den nuværende læsesituation have været igennem de associationer som disse ord vanemæssigt fremkalder hos os. Dette bevirker ofte at vi uvilkårligt kommer i tanker om vore egne hensigter og ideer som de stod for os, da vi skriver. . . skrev den nævnte side (jf. Hermann, 2008). Disse vore egne tidligere associationer stiller sig så at sige *i vejen for* at vi kan danne os den slags ikke-vidende og sædvanlige associationer som den nævnte side vil fremkalde hos **andre** læsere.

De har jo kun adgang til de faktiske ord som står der på siden, ikke

til de mange fascinerende ideer som vi fik mens vi skrev dem, og som førte os videre til at skrive dem. **Derfor vil de være uhildede af vore private associationer fra undervejs i skriveprocessen**, og af den grund kan de fortælle os hvad sidens ord fremkalder associativt hos andre; på samme måde som Nykritikerne (The New Criticism) fra 1950'erne betegnede et digt som 'det objektive korrelat' til de følelser og tanker som kan fremkaldes ved dem.

Når vi derfor vil læse korrektur på en tekst vi selv har skrevet, må vi bære os sælsomt ad, læse den bagfra og nedefra opad for at ødelægge dens sammenhæng for os. Hvis vi læser på den almindelige måde, kommer vi til at overse eventuelle trykfejl.

Dette blev jeg tilfældigvis klar over da jeg læste en tidligere version af denne tekst højt for en kammerat som sad med en kopi og kontrollerede mig. Jeg læste uvilkårligt 'skrev' i stedet for det 'skriver' som står 13 linjer oven for denne. Dette fører til følgende spørgsmål:

'C) I hvilken udstrækning er al læsning projektiv?' (kan besvares sammen med to andre spørgsmål til slut)

B) Hvad kan vi lære af at vi sommetider har svært ved at udtrykke noget vi godt véd for andre?

(Dette har NB intet at gøre med det som i mere end en snes år har været kendt under dæknavnet TAVS VIDEN; for i eksempel B) kæmper du for at udtrykke noget du allerede véd, modsat mens jeg ved min tavse viden intet behov har for at verbalisere min praksis).

Vi har af og til svært ved at udtrykke noget vi meget vel véd, som nævnt fordi "vi ej tænker i ord blot" - for hvis det var tilfældet at vi tænkte med ord, ville de altid komme løbende og frejdigt stille op når vi ville meddele os til andre mennesker. - Vi véd udmærket hvordan en pæresmag føles i munden eller en osteagtig lugt mærkes i næsen (James, 1890, 195), eller hvad vi var ude for i går, men vi kan få svært ved at beskrive det.

Vi kender af vor egen erfaring en situation hvor vi kæmper med at udtrykke denne viden. Men hvorfor havner jeg sommetider i den, hvad kan den fortælle os om det mentale liv og hvordan kommer jeg ud af den? Ved at undersøge den nærmere kan vi finde ud af mere om hvordan vi fungerer mentalt og med ord. Fordi vi er forskellige personer vil vi komme ud for det 'at mangle ord for noget vi godt véd' under forskellige omstændigheder.

Her er det vigtigt at mærke sig at de ord vi kæmper med at udtrykke ikke er forlagt eller glemt. Et indtryk har påtvunget sig dig, eller du har skabt dig en opfattelse af noget - du kender dens retning og dens værdiladning. Noget af det påfaldende ved disse 'ord-løse' tilstande er at de afskærer os fra fejlagtigt at tro at tænkning er noget som sker med ord; for hvis det var tilfældet, så ville vi aldrig komme ud for dét som et mentalt kors for tanken, ikke at have ord for hvad som ligger os på sinde!

Ligesåvel som den menneskelige 'Tankestrøm's eksistens bekræftes

af den anførte, oplevede sprogløshed, således kan sanseindtryk være en indirekte måde at opleve tankestrømmen på, fordi dele af den ved ikke at være ordformet, kan volde os vanskeligheder med at udtrykke vore oplevede sansninger i ord (vore sansede oplevelser som findes for os i tankestrømmen).

Vi har helt sikkert en vag idé om hvad det er det vi prøver at udtrykke med ord. Fordi vi ufortøvet forkaster de første ord eller udtryk som viser sig for os. Vi synes de vildleder. Men selv om vi ikke kan finde de ord som vil udtrykke vort indtryk af fx et landskab eller af en smag, så ændrer dette forhold ikke ved den kendsgerning at **vore indtryk er bestemte og nøjagtige**. Selv om vi udtrykker os med et stort ordforråd, kan det glippe. Ofte vil en person med kun et grundlæggende ordforråd kunne udtrykke sig mere fyndigt, og derved vinde mere genlyd for sit synspunkt.

Det fælles for disse situationer er dette: det som vi tænker og føler og ikke kan sige, det kender vi alligevel så godt at vi tillidsfuldt **forkaster** visse formuleringer som tilbyder sig eller bliver foreslået af andre. For det klare indtryk vi har dannet af erfaringen foreligger i vor **tankestrøm**. Ved at sige det således betjener jeg mig af den måde at beskrive vort mentale liv på som William James foreslog.

V

(**Mellemspil**: Introspektion Genovervejet)

Der er noget mærkeligt ved at bruge introspektion til at skaffe data med. For selv om metoden har været bandlyst fra akademisk psykologi siden forrige århundredes begyndelse, så kan *ingen observationer overhovedet* foretages på noget område uden at bruge introspektion (fx til at blive klar over hvad måleinstrumentet viser når vi ser på det, jf Naur, 1999,40).

Grunden til jeg kommer ind på introspektion som forskningsinstrument ligger i at man ikke kan anbefale William James' resultater uden at forholde sig til dette spørgsmål. Et moderne argument for brug af introspektion finder vi hos Peter Naur.

Introspektion: med introspektion menes at en person retter sin opmærksomhed mod sin *tankestrøm* (se dette), snarere end mod det som tankeobjekterne kender. Introspektion forekommer regelmæssigt hos de fleste mennesker. For eksempel er det kun efter introspektion at en person kan give udtryk for behag eller ubehag ved noget som personen har opfattet gennem sanserne syn, hørelse, smag berøring, lugt. Kun efter introspektion bliver vi i stand til at udbryde, åh hvilken herlig vin! Kun efter introspektion bliver vi bekendt med tankestrømmen og dens altid skiftende indhold (...) (Naur, 1999, 39f.)

På dette sted får den amerikanske lingvist Wallace Chafe noget at skulle have sagt. Han beklager hvordan William James' indsigter om

brug af introspektion gik i glemmebogen og blev forladt allerede få årtier efter William James beskrev dem. Han kalder dette

... en ulykkelig udvikling fordi den skød meget om sindet ud af synsfeltet, så det aldrig blev forstået videnskabeligt. Det tyvende århundrede gav behaviorismen dens chance, og det kom der kun en begrænset forståelse af sindet ud af. En mere afbalanceret tilgang ville anerkende egenobservationer som ikke bare vanskelige, men også som i høj grad gyldige. Derved ville man slå følge med William James' spøgelse når man ledte efter hvordan de kan forbindes med systematiske undersøgelser.

Der ligger en interessant ironi i den kendsgerning at en stor del af moderne lingvistik bygger på introspektive data. . . Uden man er klar over hvad man "tænker på" når man bruger en datid eller en flertalsform, så kunne man næppe drive semantik overhovedet, og uden semantik ville lingvistik bestemt have formindsket interesse og betydning. (Chafe, 2000, 14, oversat af JH).

Her er det på sin plads at høre hvordan William James selv begrundede sin afhængighed af introspektion. Det skete i kapitel VII, "Psykologiens metoder og faldgruber" hvor det første underafsnit "Undersøgelsesmetoder" begynder således

Introspektion er det som vi først og fremmest og til hver en tid må stole på. Ordet introspektion behøver næppe blive defineret - det betyder selvfølgelig at se ind i vore egne sind og fortælle hvad vi opdager der (...). Og eftersom resten af dette bind ikke er meget mere end en samling illustrationer af vanskelighederne ved gennem direkte introspektion at opdage nøjagtig hvordan vore følelser og deres indbyrdes forhold er, så vil vi ikke foregribe de kommende enkeltheder, men blot slå den almene konklusion fast at *introspektion er vanskeligt og fejlagtigt; og den vanskelighed deler den med alle iagttagelser af ligegyldig hvilken slags.* (James, 1890, 185 oversat af JH, originalens kursiv).

Hovedproblemet med introspektion som kilde til viden er ikke noget der er særligt for introspektion, fordi den skulle være ekstra "vanskelig og fejlagtig". *Tværtimod*, som William James nævner det i næste sætning: introspektion befinder sig netop i denne henseende **på linje med enhver anden slags observation!** Selvfølgelig er introspektion besværlig og der er i kraft af selve dens natur alvorlige fejlmuligheder indbygget i den.

Man overser imidlertid to vigtige forhold når man rutinemæssigt fejler den til side: a) at disse vanskeligheder har at gøre med almene menneskelige vilkår (som vi overhovedet er underlagt i denne verden); og ved at forholde os til det introspektive opnår vi også den gevinst at vore opdagelser også bliver relevante for den slags psykologiske begivenheder *som lægmænd oplever i deres hverdag.*

Disse indvendinger forholdt Niels Bohr sig også til, da han var inde på samme tema i sit foredrag ved en "International kongres for antropologi og etnografi" i København august 1938. Han omtalte nogle mulige konsekvenser for psykologi som følge af fysikkens nye resultater. Han fastslog

først hvordan hans arbejde med atomteori også var *en undersøgelse af betingelser for konsekvent brug af vore sproglige udtryksmidler*.

... Jeg er overbevist om at den situation jeg netop har antydnet vil have mindet mange af jer om *velkendte vanskeligheder* ved psykologiske observationer. Det er næppe for meget sagt at en hovedretning inden for moderne psykologi kan karakteriseres som en reaktion mod tidligere tiders forsøg på at analysere psykologiske erfaringer i elementer. Elementerne skulle så kombineres på lignende måde som man gjorde med målinger man foretog inden for klassisk fysik.

Ved introspektion viser det sig umuligt at skelne mellem fænomenerne selv og vor bevidste erfaring af dem. I dagligsproget taler vi ofte om at have sin opmærksomhed rettet mod den ene eller den anden side af en psykisk erfaring. Selv om dette kan være tilfældet vil nærmere undersøgelse vise at vi i sådanne sager har at gøre med situationer som gensidigt udelukker hinanden.

Vi kender alle de gamle visdomsord om at når vi prøver at analysere vore egne følelser, så har vi dem næsten aldrig mere tilbage, og i denne forstand må vi anerkende at der meget vel kan findes et lignende komplementært forhold mellem psykiske erfaringer af den slags vi beskriver med ord som "tanker" og "følelser", som der er mellem bestemte atomare forhold.

Det vil sige forhold som kan gøre sig gældende under forskellige eksperimentelle betingelser og som kan beskrives ved at bruge forskellige analogier fra vor vanemæssige fantasi. Ved denne sammenligning ønsker jeg bestemt ikke at antyde nogen nærmere forbindelse mellem atomvidenskab og psykologi. Men jeg ønsker at understrege *en bestemt lighed* under en rent erkendelsesteoretisk synsvinkel. Herved kan vi blive opmuntret til at undersøge i hvilken udstrækning løsningerne på de forholdsvis enkle fysiske problemer vil kunne hjælpe med at afklare nogle af de mere sammensatte psykologiske problemer vi møder i livet. (Bohr, 1957, 39f, med tilføjede understregninger af JH)

(**Mellemspil** slut)

VI

Sammenhængs betydning for videnskabelighed

Når vi fungerer det bedste vi kan og glemmer os selv undervejs i processen (så vi befinder os i det som førnævnte Mihaly Csikszentmihalyi (2005) kalder et *flow*), betyder det også at vi opfatter alting **sammenhængende**. Det vi tænker og føler, siger eller gør bliver sammenflettet med hinanden. Denne slags indre sammenhæng bør også blive målet for beskrivelser af vore tanker og følelser. Som vi finder det hos William James (1890) i hans klassiske 1364 siders *The Principles of Psychology*. Hos Peter Naur gøres det til kernen i al slags videnskab, at det går ud

på, videnskabeligt at tilvejebringe sammenhæng med andre beskrivelser af fænomener i verden (Naur, 1995, 317ff).

Jeg sigter mod en tilsvarende sammenvævning. Vort liv med det sproglige kan beskrives med William James'sesk sammenhæng i lyset af fire temaer som er indbegrebet af hans psykologi. De blev opdaget og formuleret i 2002 af Peter Naur som i 2005 beskriver dem således

William James' nye beskrivelse af det mentale liv understøttes af fire søjler der tilsammen udgør essensen af hans bidrag. I den yderste sammenfatning (de beskrives udførligere senere i Peter Naur 2005) er de følgende:

Vane: 'levende væsner er . . . bundter af vaner.' (James, 1890, ch. IV)

Tankestrøm: 'den første kendsgerning for os psykologer er at der foregår en eller anden slags tænkning.' Tænkning som vi alle erfarer den frembringer et bestandigt foranderligt kontinuum af tanker og følelser, som kaldes tankestrømmen. (James, 1890, ch. IX og X).

Association: loven om vanemæssig forandring af tankeobjekterne i tankestrømmen. (James, 1890, ch. XIV)

Bekendthed: 'sindet kan altid sigte til, og vide hvornår det tilsigter, at tænke på det samme'. (James, 1890, ch. XII)

De øvrige kapitler i William James' bog beskriver vort mentale liv ved hjælp af disse fire søjler. I disse beskrivelser fejles alle de såkaldte filosofiske spørgsmål i forbindelse med det mentale liv af bordet som irrelevante og værdiløse. William James' *Principles* har kun skønhedspletter få steder, og ved enkelte beskrivende udtryk. Chapter XXIV "Instinct" ville have været bedre placeret ved siden af chapter IV, "Habit". Det som William James betegner 'conception' kaldes bedre 'acquainting' ("bekendthed") som i denne bog. (som citatet er hentet fra, Peter Naur (2005) *An anatomy of human mental life*, 29f, oversat af JH).

Dette citat bringes for at få læseren til at læse mere bogstaveligt, og anvende William James mere end man plejer at gøre. For at opnå sammenhæng i beskrivelsen, må brugen af introspektive data først legitimeres.

I denne forbindelse holder jeg mig til 'association' og 'vane' for de hører begge med til vore dages psykologis *blinde plet*. Prøv bare stikordet "vane" i nyere fremstillinger som fx. Koester og Frandsen (2005). Det fænomen som Peter Naur kalder 'bekendthed' er vigtigt for vor navigation i tankestrømmen. Den mor jeg tænker på i dag er den samme som hende jeg kan huske jeg havde og tænkte på i går –.

VII

'Det sproglige' som vaner er derfor personforankret

Sproglige fænomener skal afmystificeres og personforankres. Kun således kan vor forskning i dem respektere Otto Jespersens ontologiske 1905-

påstand om at sprog kun findes når der *er* nogen, for hvem det er der. Og det skal den fordi jeg ovenstående med støtte i William James' beskrivelse af tankestrøm har givet gode psykologiske grunde for at det sproglige *kun findes* som NOGENS VANER.

Sådan at vi derfor altid må have de personer med i billedet, som giver 'noget sprogligt' dets eksistens, jf. Otto Jespersen citatet oven for i afsnit I; dette sted læser jeg sådan at **personers tilstedeværelse, på et eller andet sted (i en situation), er en ontologisk nødvendighed** for at vi overhovedet kan tale om at der *findes* noget sprogligt, (fx. nogen sproglige tegn) på dette selvsamme sted (jf. Hermann, 2008)!

Hvis vi ser på fænomenet tegn i dette perspektiv - som er radikalt nyt og anderledes end det traditionelle Niels Erik Wille (2007) har som tema - bliver det vigtigt for tegnenes "eksistens" at de netop ikke *hår* nogen egeneksistens, men *for overhovedet at være til* må høre med blandt nogle bestemte menneskers vaner.

(...) et eller andet S er, eller virker som, et tegn for en bestemt person når dette S i personens tankestrøm ved vanemæssig association (perception) fremkalder noget andet, som så siges at betegnes ved S. (...)

Det der gør et bestemt sæt af tegn nyttige under samvær i et bestemt samfund er at en del af den betydning der i samfundsmedlemmernes tankestrømme ved association knyttes til hvert tegn stemmer sammen. Her er det vigtigt at betone 'en del af' i denne forklaring. Da det handler om tankeobjekter i de enkelte personers tankestrøm kan betydningernes samstemmighed kun være delvis. (Naur, 2005b)

Dette med at kun den ovennævnte *en del* af tegnenes betydning er fælles for dem hvis vaner de tilhører, har jeg fået bekræftet gennem uformelle undersøgelser. Hvis man beder forsøgspersoner nævne det første ord de associerer til ved almindelige danske ord (som er blandt de 1100 hyppigste) vil personernes svarord være ens for ca. 40% vedkommende.

Vi kan først — i Sneedorffs forstand — være sikre på vi har en rigtig forståelse af noget, i det øjeblik vi er i stand til at gøre den begribelig for andre personer, sådan at de kan give os den tilbage igen. På denne måde bliver *de andre* vor pragmatiske garanti for at det virkelig er en forståelse af noget som vi har, og ikke en hallucination: *forståelser af noget lader sig sætte i omløb*. De forståelser som Sneedorff (1744) skriver om, kan vi derfor også kalde for "forståelser vi får retur," eller som bliver tilbagetransporterede til os.

Læseren kan nu overveje følgende lumske spørgsmål:

- a) Hvordan er det muligt at begrunde vi skal skelne psykologisk mellem vor perception af noget sprogligt og vor perception af andre ting?
- b) Hvorfor ville det være misvisende at tale om *kodningsvanskeligheder* da jeg i eksempel B) havde svært ved at udtrykke et indtryk?
- c) I hvilken udstrækning er al læsning projektiv?

Referencer

- Andersen, N.M. (2003) Det tavse sprog. In Larsen, G. & Thobo-Carlsen, J. red: *Modernismens betydende former*. København: Akademisk forlag.
- Bohr, N. (1957) *Atomfysik og menneskelig erkendelse*. København: Schultz Forlag.
- Bohr, N. (1985) *Atomer og kerner*. Udvalgte foredrag og artikler. København: Rhodos.
- Chafe, W. (2000) A Linguist's perspective on William James and The Stream of Thought, *Consciousness and Cognition*, 9, 618–628.
- Csikszentmihalyi, M. (2005) *Flow og engagement i hverdagen*. København: Dansk Psykologisk Forlag.
- Gregersen, F. (2008) Dansk som genrer. in Inge Dalsgård et al., *Midt i ræset* red., København: Amanda, Dansk lærerforeningens forlag.
- Harris, R. (1996a) *Signs, Language and Communication* London: Routledge.
- Harris, R. (1996b) *The Language Connection*. Bristol: Thoemmes Press.
- Harris, R. (1998) *Introduction to Integrational Linguistics* London: Pergamon.
- Hermann, J. (1975) "Sprog" som upraktisk bevidsthed i undervisningen in Gregersen & Reisby ed.: *Børn, sprog og undervisning*. København: Gyldendal.
- Hermann, J. (2005) Det sproglige som vaner. in Cramer et al., red.: *Teorien om alt* Århus: Wessel og Huitfeldt.
- Hermann, J. et al., red. (2005: *På sporet af sprogpsykologi* København: Frydenlund. 7-14, 27-38, 157-164.
- Hermann, J. (2007) Visse vrangforestillinger om ordet "sprog" in *MUDS 11*, Århus: Aarhus Universitetsforlag.
- Hermann, J. (2007b) Fordomme vi deler – kan vi gøre op med dem? in *Vandfanget*, 12(1), 27-29.

- Hermann, J. (2008) Uvilkårlige ord eller forestillinger? in *Bag ordene*, Kbh. Dansk Sprognævn Skrifter nr. 41, 56-67.
- James, W. (1950 [1890]) *The Principles of Psychology*, bd. I. Dover: Henry Holt.
- James, W. (1906) *Religiøse Erfaringer*. København: Pios boghandel.
- Jespersen O. (1905) *Sprogundervisning*. København: Det schubtheske Forlag.
- Jespersen O. (1924) *The Philosophy of Grammar*. London: George Allen & Unwin.
- Jespersen O. (1933) *Essentials of English Grammar*. London: George Allen and Unwin.
- Jespersen O. (1941) *Sproget*, København: Gyldendal.
- Koester, T. og Frandsen, K. (red.) (2005) *Introduktion til psykologi*. København: Frydenlund.
- Kjørup, S. (1996) *Menneskevidenskaberne*. Roskilde: Roskilde Universitetsforlag.
- Møller, P.M. (1848-50) *Efterladte skrifter* bd. 2. København: Reitzel.
- Naur, P. (1995) *Knowing and the Mystique of Logic and Rules* Dordrecht: Kluwer Academic Publishers.
- Naur, P. (1999) *Antifilosofisk Leksikon*. København: Eget Forlag. www.naur.com
- Naur, P. (2001) *Antiphilosophical Dictionary*. København: naur.com publishing.
- Naur, P. (2002) *Psykologi i videnskabelig rekonstruktion*. København: Eget forlag. www.naur.com
- Naur, P. (2005) *An anatomy of human mental life - Psychology in unideological reconstruction, incorporating The synapse-state theory of mental life*, naur.com publishing, www.naur.com.
- Naur, P. (2005b) Tegnvaner og sprog. in Hermann, J. et al., red. 2005, 157.
- Rasmussen, S.E. (1979) *Berlingske Aftenavis 21/9*. Citeret efter Fafner

1985.

Sneedorff, J.F. (1744) citeret efter Steen Eiler Rasmussen 1979, jf. Hermann, J., 2005.

Jørgensen, P.S. (2004) Videnskabelige ord - sproglig rådgivning i videnskabelighed. In: Jørgensen og Stray Jørgensen, red.: *På godt dansk*. Århus: Wessel og Huitfeldt.

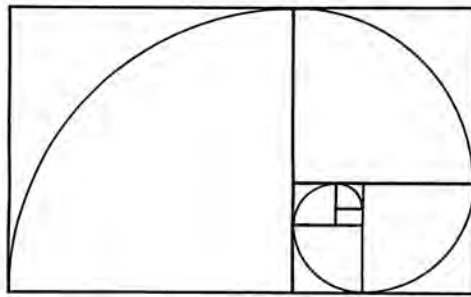
Wille, N.E. (2007) *Fra tegn til tekst*. København: Forlaget Samfundslitteratur.

Zinkernagel, P. (1992) *Virkelighed*. København: Munksgaard.

Phi for livets skønhed

Af *Bjarne Sode Funch*

Skønhed har til alle tider og i alle kulturer haft en plads i menneskets tilværelse. Alligevel er det aldrig lykkedes de videnskaber der beskæftiger sig med skønhed, at afdække de principper der konstituerer det skønne. Et af de mest omdiskuterede principper er *det gyldne snit* (fig. 1). Et



Figur 1 *Det gyldne rektangel* i flere størrelser med grundlinje delt i *det gyldne snit* og med indskreven spiral formet efter *det gyldne snits princip*.

snit der deler en linje i to dele således at forholdet mellem hele linjens længde og det største linjestykke svarer til forholdet mellem det største og mindste linjestykke. Dette lader sig kun gøre hvis hele linjens længde er lig med 1,618... gange det største linjestykke og det største linjestykke er lig med 1,618... gange det mindste linjestykke. Konstanten 1,618... der er et irrationelt tal, betegnes inden for matematikken med det græske bogstav φ (phi). En betegnelse den har fået så sent som i begyndelsen af det 20. århundrede efter det første græske bogstav i navnet på den græske billedhugger Phidas (490-430 fvt.) der er kendt for at have udført frisen i Parthenontemplet i Athen og Zeusstatuen i templet i Olympia. I begge tilfælde skulle han have benyttet det gyldne snit i figurernes proportionering. Det skal endvidere nævnes at Fibonacci-talrækken, 0,

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, ... er nært knyttet til den gyldne proportion ved at ethvert tal i rækken divideret med det nærmest forudgående i stigende grad op gennem rækken er en tilnærmelse til phi.

Phi der også kaldes *det gyldne tal*, anses af mange for at være det smukkeste talforhold i verden da det er fundet mange steder i naturen. Blandt andet i træer og planter hvor bladenes opbygning og placering i mange tilfælde er proportioneret efter det gyldne snits princip. På en rose der er symbolet på harmoni og kærlighed, kan kronbladernes placering i forhold til hinanden beskrives ved en matematisk regel der involverer det gyldne snit (fig. 2). Havsneglens hus som inden for hinduismen opfat-



Figur 2 Rosens kronblade er placeret efter det gyldne snits princip.

tes som symbol på skabelse, følger i sin vækst det gyldne snits princip. Det findes ligeledes i proportioneringen af menneskets og dyrs krop og skelet samt i blodårer og nervebaners forgreninger. Det gyldne snit optræder med overraskende hyppighed i alt lige fra de mindste krystaller til universets spiralgalakser som vores egen Mælkevej.

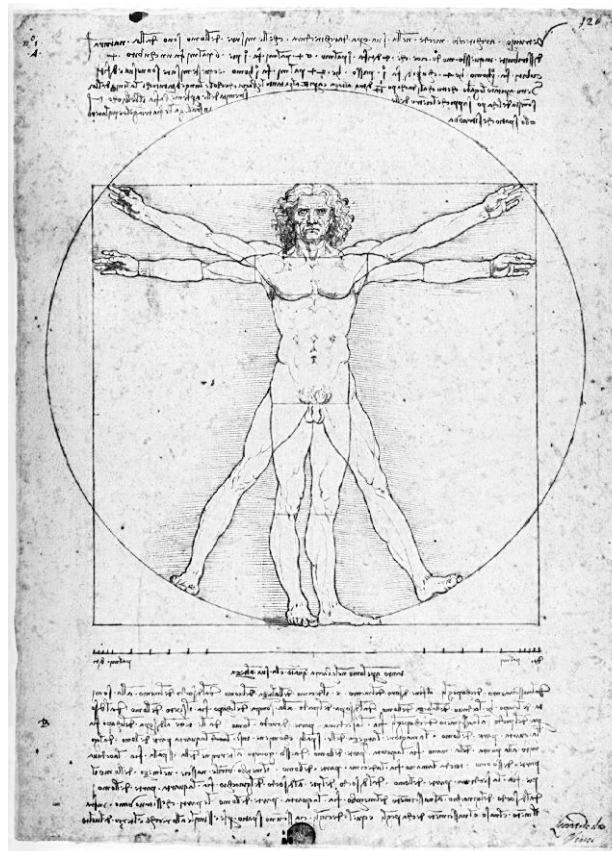
Phi som tallet for det gyldne snits princip, har fundet vej ind i arkitektur, kunst og design. Mennesket har gennem århundreder bevidst eller ubevidst benyttet sig af det gyldne snit i deres skabende arbejde. Renæssancens kunstnere, blandt andre Leonardo da Vinci, anvendte målrettet det gyldne snit som et ideal for at opnå kunstnerisk skønhed. Blandt nyere tids arkitekter og billedkunstnere kan nævnes Frank Lloyd Wright, Le Corbusier, Mario Botta, mange af Kubismens malere, Salvador Dali og Mario Merz, blot for at nævne nogle få blandt de mange der har ladet sig inspirere af det gyldne tal phi.

Det er omdiskuteret hvorvidt ægypterne kendte det gyldne forhold da de konstruerede de store pyramider idet forholdet indgår i for eksempel Keops-pyramidens proportioner med en sådan nøjagtighed at det næppe kan være ren tilfældighed. Det siges ligeledes at pythagoræerne i det 6. århundrede før vor tidsregning var bekendt med den gyldne pro-

portion, men det er først med Euklids matematik ca. 300 år fvt. at der foreligger skriftlig dokumentation for kendskabet til phi. Euklid omtaler et talforhold ud fra en simpel opdeling af en linje i, hvad han kalder *ekstrem* og *middel* forhold. Han skriver: “En ret linje siges at være delt i ekstrem og middel forhold når hele linjen forholder sig til den store del som den store del forholder sig til den mindre” (Livio, 2004, s. 9). Som allerede nævnt har utallige arkitekter, kunstnere og videnskabsfolk siden den tid været optaget af dette princip for skønhed. Et forhold som i franciskanermunken Luca Paciolis værk *De Divina Proportione* fra 1509, får betegnelsen *den guddommelige proportion*. Betegnelsen *det gyldne snit* optræder første gang, ifølge den amerikanske astrofysiker Mario Livio (2004, s. 11) så sent som i 1835 i en matematikbog af en tysk matematiker ved navn Martin Ohm. Det fremgår dog at Ohm kendte betegnelsen andetsteds fra, men hvorfra fortaber sig foreløbigt i det uvisse.

Phi for rektangulær skønhed

Selv om det guddommelige tal phi for skønhed har været kendt gennem århundreder, er det først i andet halvdel af det 19. århundrede at det gyldne snit blev underkastet systematiske, empiriske undersøgelser. Det var den tyske forsker Gustav Theodor Fechner (1876) der foretog de første psykologiske undersøgelser. Han anvendte flere forskellige metoder. I en periode foretog han opmålinger overalt hvor to linjer står i forhold til hinanden. Han målte for eksempel det kristne kors i kirker og på kirkegårde, på smykker og salmebøger. Han målte døre, vinduer, kakler, billedrammer, kuverter og meget andet, og i alle tilfælde foretog han beregninger af de to linjers forhold til hinanden for at se hvor tæt kvotienten var på phi. Resultatet var at aflange rektangulære ting med en bemærkelsesværdig hyppighed er proportioneret efter det gyldne snits princip, mens det tilsvarende ikke gjorde sig gældende med det kristne kors. Disse iagttagelser fik Fechner (1876) til at foranstalte det først beskrevne psykologiske eksperiment inden for det æstetiske område. Hans forsøgsmateriale var ti stykker karton der alle var firkantede og havde samme areal. De adskilte sig fra hinanden i format. Et var kvadratisk og de øvrige ni forskelligt proportionerede i en serie der gradvist gik op til forholdet fem til to. Heriblandt var ét proportioneret efter det gyldne snit. Fechners forsøgspersoner, i alt 347 personer i alderen fra 16 år og opefter, fik forevist disse ti kartonstykker og bedt om at udpege det de fandt størst behag ved. Resultatet var slående. Cirka 35 procent af de adspurgte pegede på det gyldne rektangel og hvis de to rektangler som ligger nærmest det gyldne rektangel medregnes, foretrak syv-otte forsøgspersoner ud af ti det rektangel der var proportioneret efter det gyldne snits princip. Fechner udtaler at rektanget proportioneret efter det gyldne snits princip blev foretrukket af de personer som han selv anså som mennesker med



Figur 3 *Vitruvius' menneske* (1492) af Leonardo da Vinci. Et menneske indskrevet i både et kvadrat og en cirkel fikserer enhver bevægelse udadtil.

god smag. Han antyder at der måske er en korrelation mellem uddannelse og præference for det gyldne rektangel. Blandt otteogtyve håndværkere foretrak syv det gyldne rektangel og fem kvadratet, flere med en bemærkning om at kvadratet var det mest regulære. Blandt kulturelt dannede forsøgspersoner blev kvadratet ofte anset som simpelt, tørt og kedeligt. En veluddannet kvinde omtaler sin oplevelse af kvadratet som "oplevelsesfattig tilfredsstillelse" (bausbackenen Befriedigung).

Lige siden Fechners undersøgelser og indtil vore dage har utallige forskere gentaget hans eksperiment eller udført tilsvarende undersøgelser, og de fleste er kommet frem til et resultat der giver det gyldne snit en prioritet i spørgsmål om skønhed. Resultaterne har på den anden side aldrig være tilstrækkelig overbevisende til at udpege det gyldne snit som et generelt princip for skønhed og derfor er ideen om det gyldne snit gen-

tagne gange og senest i 1997 blevet erklæret for død og begravet af den tyske psykolog Holger Höge. I en undersøgelse hvor Fechners forsøgsprocedure blev repliceret i detaljer, fandt Höge ingen præference for det gyldne rektangel, men derimod en lille præference for kvadratet. Höge rejser spørgsmålet om der skulle være sket en ændring i folks smag i perioden fra 1860'erne til 1990'erne, men erkender at hypotesen er for generel til at teste. Selv om Höge ikke er i tvivl om at hypotesen om det gyldne snit som et princip for skønhed må afvises én gang for alle, hersker der ingen tvivl om at den med tilsvarende sikkerhed vil genopstå. Vi må se i øjnene at psykologiske eksperimenter, uanset resultaterne, aldrig vil kunne mane det gyldne snit i jorden.

De fænomenologiske iagttagelser er det svage punkt i alle disse psykologiske undersøgelser. Almindeligvis anmodes forsøgspersonerne om at udpege den firkant som de foretrækker, og det ligger implicit eller eksplicit i denne anmodning at det drejer sig om at vurdere firkantens skønhed. Men hvad er skønhed og kan en simpel firkant overhovedet være smuk? Svarene er ikke enkle og entydige. En firkant i form af et stykke pap eller som fire linjer tegnet vinkelret på hinanden er et meget enkelt design som mange ikke vil forbinde med skønhed i dagligdags forstand. Alligevel gennemfører alle forsøgspersoner, stort set uden undtagelse og tilsyneladende uden indvendinger, proceduren og udpeger den firkant de finder smukkeste. Men på hvilket grundlag de træffer afgørelsen, forbliver ukendt. Udtalelser om at kvadratet opleves som værende "simpelt, tørt og kedeligt" kan lede til en antagelse om at skønhed i disse eksperimenter drejer sig om noget der er mere komplekst og udfordrende. Det kan antages at en form der virker ekspressiv og kraftfuld også opleves som smuk. Med andre ord, de forsøgspersoner der foretrækker det gyldne rektangel, foretrækker den fordi den virker mest "levende". I modsætning til kvadratet der er centreret og stabilt, måske endog harmonisk, giver det gyldne rektangel indtryk af bevægelse. Det betyder ikke at det gyldne rektangel rent faktisk bevæger sig, men at det har potentiale til at bevæge sig og derfor vanskeligt lader sig fastholde som noget stabilt.

Hvordan kan det så være at der ikke er større konsensus om det gyldne rektangel som det smukkeste? Fechner antyder i sine undersøgelser at uddannelsesniveau kan være en influerende variabel. Men måske er det mere oplagt at tale om frygten for henholdsvis livet og døden som en grundlæggende faktor i præferencen for en simpel firkant, forstået på den måde at den personlighed der på et grundlæggende psykisk niveau er ængstelig og gerne vil fastholde og kontrollere tilværelsens udfordringer foretrækker den centrerede og symmetriske harmoni (fig. 3), hvorimod den personlighed som er åben for nye udfordringer, foretrækker det ekspressive. I dette perspektiv kan Höges resultater fra 1997 med præference for kvadratet virke foruroligende. Det skal her bemærkes at alle firkanter er symmetriske og derfor rummer et moment af stillestående harmoni. Det gyldne rektangel indeholder derfor både ro og uro, mens kvadratet

repræsenterer absolut ro.

Med en eksistentiel faktor som grundlag for vurdering af firkanter skønhed, bliver spredningen af forsøgspersonernes præferencer mere forståelig. For det første, er alle mennesker i højere eller mindre grad præget af både livslyst og dødsangst. De har således både en evne til at hengive sig og en evne til at forholde og distancere sig, og derfor vil deres personlighed være præget af begge sider, blot i varierende grad. Det er samtidig en tilgang til tilværelsen der kan være stærkt påvirkelig af forhold i den aktuelle situation. I dette tilfælde hvorledes den pågældende befinder sig i det eksperimentelle miljø. Man kan endog antage at kulturelle, subkulturelle og historiske omstændigheder generelt kan være afgørende for om livslysten eller dødsangsten bliver den mest aktive faktor i forhold til skønhedsoplevelse. Som antydnet afspejles tidsånden måske på afgørende vis i de psykologiske undersøgelser af det gyldne snit. Helt afgørende er dog det forhold at de visuelle design, ligesom variationen mellem de enkelte figurer der anvendes i de psykologiske eksperimenter, er så minimale at det er en vanskelig opgave for forsøgspersonerne at træffe en absolut afgørelse. Endvidere må man være opmærksom på at en figurs skønhed influeres af den kontekst hvori figuren optræder, samt de associationer betragteren knytter til figuren. Forhold som Fechner allerede i forbindelse med sine første forsøg, pegede på. Men uanset de mange variable og en stor spredning i forsøgsresultaterne er det tankevækkende at se hvordan den gyldne proportion dukker op igen og igen som udtryk for skønhed.

Phi for bevægelse

Omtrent halvtreds år efter Fechners undersøgelser af det gyldne snit gør en anden tysk psykolog, Max Wertheimer, en epokegørende iagttagelse. I en række eksperimenter der udføres i efteråret og vinteren 1910, påviser Wertheimer (1912) at oplevet bevægelse ikke er afhængig af fysisk bevægelse, men derimod under særlige omstændigheder kan frembringes af stationære figurer der ikke bevæger sig i rum og tid. Når for eksempel to ens figurer vises i to forskellige positioner lige efter hinanden, vil forsøgspersonen opleve at det er den samme figur der har flyttet sig. Det bemærkelsesværdige er at forsøgspersonen oplever bevægelsen som om den rent fysisk finder sted. Wertheimer kalder denne bevægelse *phi* og fænomenet kendes i dag som *phi-fænomenet*. Det har desværre ikke været muligt at opklare hvorfor Wertheimer vælger netop betegnelsen *phi* for oplevet bevægelse, men det er påfaldende at en amerikansk matematiker ved navn Mark Barr næsten samtidig vælger betegnelsen phi for det matematiske talforhold der refererer til oplevet skønhed. Det behøver ikke at være andet end resultat af tilfældighedernes spil, men det kan også tages som et udtryk for at der er en lighed mellem oplevet bevægelse og

oplevet skønhed som allerede påpeget i forbindelse med Fechners forsøg med det gyldne rektangel.

Bag Wertheimers lille og enkle eksperiment gemmer sig en grundlæggende erkendelsesteoretisk problemstilling i spørgsmålet om hvorvidt den bevægelse der opleves i forbindelse med phi-fænomenet er forskellig fra den der opleves i forbindelse med fysisk bevægelse. Ved bevægelse i rum og tid kan det konstateres at et objekt har flyttet plads. I Wertheimers forsøg ved vi at de figurer der giver anledning til den oplevede bevægelse ikke har flyttet sig, men er og forbliver stationære. Derfor taler Wertheimer også om en "illusion". Men i begge tilfælde opleves bevægelse, og spørgsmålet er om fysisk bevægelse giver større indsigt i bevægelsens natur end phi-fænomenet gør. Inden for fysikkens domæne er det mulig at forudsige og beregne hvilken retning et bestemt objekt under givne omstændigheder vil bevæge sig i og med hvilken hastighed; men disse beregninger siger intet om bevægelsens natur i den forstand at vi får en dybere indsigt i hvad bevægelse er. I Wertheimers eksperiment opleves bevægelse når to figurer eksponeres efter hinanden under særlige rumlige og tidsmæssige forudsætninger. Men den oplevede bevægelse er den samme som under såkaldt fysisk bevægelse. Den oplevede bevægelse i forbindelse med phi-fænomenet kan endog ifølge Wertheimer, forekomme mere prægnant end ved tilsvarende fysisk bevægelse. I begge tilfælde må det konstateres at der opleves bevægelse og at den ene form for bevægelse ikke giver mere indsigt i bevægelsens natur end den anden. Med Wertheimers iagttagelse bliver det åbenbart at bevægelse er noget der er synligt for vores sanser, men at en yderligere forståelse af hvad der konstituerer oplevet bevægelse, uanset om det er fysisk bevægelse eller phi-fænomen, forbliver ubegribeligt for vores forstand. Vi må erkende at vi dybest set ikke ved hvad bevægelse er.

De to synspunkter der gemmer sig i denne problemstilling står stejlt over for hinanden, og det argument der vil kunne afgøre striden mellem det empiristiske og fænomenologiske synspunkt, findes næppe. Spørgsmålet er om man tror på at verden kan begribes til fulde gennem menneskets intellekt eller ej. Tror man på livet som noget der i sidste instans er ubegribelig, tror man også på den umiddelbare oplevelse som sandhedsvidne.

Phi for liv

Ingen har bedre end de eksistentielle tænkere forstået at bevægelse og liv er to sider af samme sag. Den menneskelige eksistens udmærker sig ifølge Søren Kierkegaard (1846) ved sin tilbliven. Med andre ord, det mest grundlæggende træk ved livet er at det er i evig forandring. Dette gælder ikke blot livet som det udfolder sig her og nu. Biologisk som en krop der ånder og uafbrudt pulserer og ændrer sig. Psykologisk i form

af en spontan og umiddelbar bevidsthed af følelser, iagttagelser, tanker, erindringer og andre bevidsthedsaspekter. En bevidsthed der i sin rene form kaldes ved så forskellige navne som "flow" (Csikszentmihalyi, 1990), "levelse" (Funch, 1974) eller "stream of consciousness" (James, 1892). Men også i det der konstituerer det enkelte menneske som individ. *Hin enkelte* som Kierkegaard ville sige, der biologisk udmærker sig ved en krop der vokser, udvikles, ældes og til slut dør og psykologisk ved en personlighed der gennem erfaringer udvider sine horisonter og mærkes af de udfordringer der gennemleves. Den menneskelige eksistens i sin tilbliven kan ligesom phi-fænomenet ikke beskrives på anden måde end ved de positioner som åbenbarer livets bevægelse. Enten i form af de oplevelser som fastholdes gennem refleksionen eller de stadier på livets vej som kun kan defineres gennem psykologisk analyse. Den bevægelse der kendetegner livet, kan i sig selv ikke begribes. Med Kirkegaards (1846/2002, s. 281) ord: "Existents lader sig ikke tænke uden Bevægelse, og Bevægelse lader sig ikke tænke sub specie æterni"¹ Han fortsætter: "Existents er ligesom Bevægelse en saare vanskelig Sag at omgaaes. Tænker jeg den, saa hæver jeg den, og saa tænker jeg den ikke. Det kunde da synes rigtigt at sige, at der er Noget, som ikke lader sig tænke: det at existere." Men dette er alligevel ikke den fulde sandhed. Ligesom phi-fænomenet frembringes gennem successiv eksponering af stationære figurer, begribes den menneskelige eksistens gennem refleksionen. Det vil sige, gennem refleksionen fastholdes livets bevægelse i oplevelsen og kan dermed gøres til genstand for analyse og fortolkning. Kierkegaard (ibid., s. 300) skriver: "Naar det at existere ikke lader sig tænke, og den Existerende dog er tænkende, hvad vil saa det sige? Det vil sige, han tænker momentviis, han tænker forud og han tænker bag efter. Den absolutte Continueerlighed kan hans Tænken ikke faae."

Den menneskelige eksistens er bevægelse og derfor kun synlig for sanserne. Det er kun eksistensens kontekstuelle vilkår der er begribelig for refleksionen og tænkningen. På samme måde som det spæde udtryk for liv i det gyldne rektangel bliver begribelig i relation til kvadratets ro, og på samme måde som bevægelsen i Wertheimers phi-fænomen bliver begribelig i relation til de stationære figurer, således bliver eksistensen begribelig i forhold til "døden". Refleksionen er en psykisk funktion der fastholder det levende i "dødens greb", forstået på den måde at livet kun forstås i form af det oplevede, mens livet som det liv det er, undslipper og fortsætter ufortrødent.

Eksistensens kontinuerlige bevidsthedsstrøm har ligesom enhver anden bevægelse en retning, men hvilken? Her peger Kierkegaard på lidenskab. Han (1846/2002, s. 283) skriver: "At existere, naar dette ikke skal

¹ Sub specie æterni; egentlig sub specie aeternitatis, lat. "under evighedens synspunkt" eller "ud fra det eviges synspunkt". I Kirkegaards *Afsluttende uvidenskabelig Efterskrift* udtryk for den spekulative idealismes standpunkt.

fortaaes om saadan at existere, lader sig ikke gjøre uden Lidenskabem.” Som det fremgaaer af Kierkegaards formulering er han meget vel klar over at blot ved at anføre lidenskabem som livets drivkraft, bringer han sig i et dilemma idet enhver begrebsliggørelse samtidig er en fastholdelse af livet. Dette bekræftes yderligere naar han anfører Evigheden som eksistensens egentlige lidenskab, mens den jordiske lidenskab forvandler eksistensen til det øjeblikkelige og dermed er med til at forhindre det i at eksistere. Ikke desto mindre må den eksistentielle tænkning leve med dette dilemma i en erkendelse af at livet ikke lader sig begribe, men alligevel leverer alt stof til eftertanke. Derfor må man ogsaa være opmærksom på at naar Kierkegaard taler om lidenskabem som eksistensens drivkraft, kender vi den kun i form af lidenskab på samme måde som vi kun kender livets bevægelse som bevægelse. Enhver ide om en endelig målsætning er resultat af tænkning og står derfor i modsætning til eksistensen. Lidenskabem er derfor tilstedeværende som en integreret del af den menneskelige eksistens, men den kendes kun på sit nærvær og ikke på sin retning og mål.

Der er dog andet end Evigheden der retningsgiver livets bevægelse, nemlig skønheden. Ikke i den forstand at vi efterstræber skønhed. Det gør vi ganske vist ofte, men vi gør det af den simple grund at vi i særlige lykkelige øjeblikke har oplevet skønheden som manifestation på lidenskabens absolutte nærvær. Når vi smykker vores omgivelser er det for at minde os selv om at eksistens er lidenskab. Det er i sig selv ikke et udtryk for lidenskabens nærvær. Lidenskabens absolutte form melder sig kun i de uforudsete øjeblikke hvor den enkeltes eksistens kompletteres. Det vil sige, naar de erfaringer livets vilkår bringer, bliver integreret som afsæt for en ny bevægelse.

Livet *er* og som sådant hverken smukt eller grimt, men det manifesterer sig i skønhed i de sjældne øjeblikke hvor livets bevægelse og lidenskab forenes.

Selv om skønheden i phi er meget spæd, kommer den til syne i de proportioner der kendetegner kimen til liv. Selv om phi-fænomenet er en illusion, åbenbarer den et ubegribeligt træk ved den menneskelige eksistens og det følsomme øje aner skønheden bag bevægelsen. Den absolutte skønhed, livets skønhed, åbenbares derimod kun i lidenskabens nærvær.

Referencer

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper and Row.

Fechner, G. T. (1876). *Vorschule der Aesthetik*. Leipzig: Breitkopf und Hartel.

Funch, B. S. (1974). *Æstetikens psykologi: Dens maya og prolegomena*. København: Psykologisk Laboratorium, Københavns Universitet.

Höge, H. (1997). The golden section hypothesis – its last funeral. *Empirical Studies of the Arts* 15 (2): 233-255.

James, W. (1892). *Psychology*. London: MacMillan, 1904.

Kierkegaard, S. (2002). *Afsluttende videnskabelige Efterskrift*. København: Gads Forlag. Originaludgave 1846.

Livio, M. (2004). *Det gyldne snit*. Oversat af Poul G. Hjort. København: Nyt Teknisk Forlag. Originaludgave på engelsk 2003.

Wertheimer, M. (1912). Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie* 61: 161-265.

Universitetspædagogisk praksis og videnskabelighed

Af *Tine Nielsen*

Mit valg af titel og emne for denne artikel undrer måske nogle af de læsere, der kender Benny og hans arbejdsfelt, og jeg kunne da også have valgt at skrive om mange andre emner. Når jeg alligevel har valgt universitetspædagogisk praksis og videnskabelighed, skyldes det, at Benny var den første, der for alvor var en inspiration for mig på begge områder, og at denne inspiration var så stor, at dette nu er mit arbejdsfelt. Inspirationen startede med mit første møde med Benny, da han på første år af psykologistudiet underviste i faget "Psykologiens historie og videnskabsteori". Her oplevede jeg for første gang undervisning, der både var indholds- og formmæssigt inspirerende, og ikke mindst undervisning, der var udfordrende og særdeles lærerig. Inspirationen og videnstilegnelsen fortsatte gennem en mangeårig ansættelse som hjælpelærer for Benny i faget "Statistik" på psykologistudiet, gennem to ekstraordinære vejledningsforløb i forbindelse med speciale og senere ph.d.-afhandling, og gennem en række fælles kurser og konsulent jobs inden for feltet kvantitativ metode og metodologi. Det er derfor for mig naturligt at vælge at skrive inden for det arbejdsfelt, som Benny både har været medvirkende til, at jeg har valgt, og som Benny har bidraget med så meget viden til, i dette bidrag til en hyldest til Bennys livslange virke som universitetsunderviser, vejleder, forsker, kollega, inspirator ved Institut for Psykologi, Københavns Universitet, og ikke mindst som ven.

Indledning

"§ 2. Universitetet har til opgave at drive forskning og give forskningsbaseret undervisning indtil højeste internationale niveau inden for sine fagområder. Universitetet skal sikre et ligeværdigt samspil mellem forskning og uddannelse, foretage løbende strategisk udvælgelse, prioritering og udvikling af sine forsknings- og uddannelsesmæssige fagområder og udbrede kendskab til videnskabens metoder og resultater" (Ministeriet for

Videnskab, Teknologi og Udvikling, 2007b). Sådan står der i “universitetsloven”, og man kan gøre sig mange overvejelser omkring betydningerne og udmøntningerne af dette i praksis. Nogle af de overvejelser, jeg har gjort mig om ovenstående paragraf i forhold til universitetspædagogisk praksis, er: Hvilke andre styrende vilkår er der for den universitetspædagogiske praksis i Danmark? Hvad er forskningsbaseret undervisning? Hvordan kan forskeres (der underviser) videnskabelige kvalifikationer bidrage til en stadig udvikling af den universitetspædagogiske praksis? Dele af disse overvejelser gengives i denne artikel om videnskabelighed og universitetspædagogisk praksis.

Hensigten med denne artikel er således at give et bud på, hvorledes de allerede eksisterende videnskabelige kvalifikationer, som besiddes af universiteternes videnskabelige personale (læs forskere, der underviser, og vice versa), kan bringes i anvendelse på en ny måde i forhold til en række af de krav, der stilles til universiteternes virke, såsom at yde forskningsbaseret undervisning, at udvikle viden om universitetspædagogik og -didaktik, samt at sikre kvaliteten af undervisning, pædagogik og tilrettelæggelse.

Til start vil jeg kort opridse nogle af de vilkår, der er styrende for den universitetspædagogiske praksisfelt og udviklingen af dette. Dernæst vil jeg se lidt nærmere på, hvad forskningsbaseret undervisning er for en størrelse, og diskutere dette i forhold til videnskabelighed i den universitetspædagogiske praksis. Og sidst vil jeg præsentere en såkaldt *scholarship of teaching and learning*-model, samt hvorledes vi arbejder med denne model i adjunktuddannelsen i universitetspædagogisk kompetence ved Copenhagen Business School (CBS).

Nogle vilkår for det universitetspædagogiske praksisfelt

At de danske universiteter både skal bedrive forskning og forskningsbaseret undervisning er i dag uomtvisteligt slået fast både lovmæssigt og i praksis. Det er således ikke bare “universitetsloven” (jf. ovenfor), som foreskriver dette. Også de bekendtgørelser, der er kommet i forbindelse med oprettelsen af en dansk akkrediteringsinstitution, slår fast, at en del af vurderingsgrundlaget for de danske universitetsuddannelser er, hvad der kaldes “forskningshøjden”, som dækker over, at “*uddannelsen er forskningsbaseret*”, at “*uddannelsen er baseret på et aktivt forskningsmiljø*” samt “*kvaliteten og styrken af det bagvedliggende forskningsmiljø*” (Ministeriet for Videnskab, Teknologi og Udvikling, 2007a, bilag 1, Kriteriesøjle II). Ses der på nogle af de undersøgelser og rapporter, der er lavet gennem de sidste 20 år, er der heller ikke nogen tvivl om, at der blandt universiteternes videnskabelige personale er en udbredt enighed om, at undervisningen skal være forskningsbaseret, og at den i høj grad

er det (Jensen, 1986; Jacobsen, 1990; Fibæk Lauersen, 1994; Ministeriet for Videnskab, Teknologi og Udvikling, 2006).

Ifølge arbejdsgruppen om undervisningskvalitet nedsat af videnskabsministeriet, betyder det, at universiteterne både skal yde og yder forskningsbaseret undervisning, imidlertid *ikke*, at den pædagogiske og didaktiske forskning har "*stor indflydelse på den forskningsbaserede undervisning, der udbydes på universiteterne. I dag er det kun adjunkterne*"¹, der skal deltage i universitetspædagogiske og didaktiske kurser. (Ministeriet for Videnskab, Teknologi og Udvikling, 2006, s. 34).

At bestemte kategorier af videnskabeligt personale ved universiteterne skal deltage i pædagogiske/didaktiske kurser, er således også et af de vilkår, der er styrende for det universitetspædagogiske praksisfelt. Dette ses tydeligt udmøntet i, at alle danske universiteter i dag har et pædagogisk/didaktisk uddannelsesprogram for adjunkter (de såkaldt pædagogiske adjunktuddannelser) med det sigte undervisningsmæssigt at kvalificere adjunkter, forskere og i nogle tilfælde ansatte i post doc.- og lektorstillinger efter et frivillighedsprincip (Kirketerp, 2007).

Der er således lovgivningsmæssigt lagt rammer for universitetspædagogisk og -didaktisk udvikling, og udviklingen er i fuld gang på de enkelte universiteter (jf. Kirketerp, 2007). Men som der også nævnes i rapporten *kvalitet i undervisningen*, har den universitetspædagogiske forskning ikke stor indflydelse på, hvordan den forskningsbaserede undervisning praktiseres eller udvikles. Der er således *ikke* lovgivningsmæssigt stillet krav til videnskabeligheden i det universitetspædagogiske felt.

Forskningsbaseret undervisning = Videnskabelighed i universitetspædagogisk praksis?

Ifølge Ministeriet for Videnskab, Teknologi og Udvikling er uddannelsernes forskningsbaseret, og hvorvidt de er baseret på et aktivt forskningsmiljø, en del af vurderingen af den enkelte uddannelses forskningshøjde i forbindelse med akkreditering. I akkrediteringskriterierne defineres forskningsbaseret undervisning som:

"Kriterium 3: Uddannelsen er forskningsbaseret

- *Uddannelsen skal give de studerende viden, færdigheder og kompetencer baseret på forskning inden for det pågældende fagområde og for de fagområder, hvor det er relevant, baseret på interaktionen mellem forskning og praksis.*
- *Uddannelsen skal give de studerende fagspecifikke og almene akademiske kvalifikationer og kompetencer.*

¹ Dette er i 2007 ændret til at omfatte både adjunkter og forskere (Finansministeriet, 2007, s. 25 hhv. 26) [Forfatterens indskud i citat].

- *Uddannelsen skal give de studerende viden om videnskabelig teori og erfaring med at vurdere og anvende videnskabelige metoder.*

Kriterium 4: Uddannelsen er baseret på et aktivt forskningsmiljø

- *Uddannelse er tilrettelagt af aktive forskere.*
- *De studerende undervises i udstrakt grad af aktive forskere.*" (Ministeriet for Videnskab, Teknologi og Udvikling, 2007a, bilag 1).

Ministeriets definition af "forskningsbaseret" overlapper en hel del med de fem forståelser Fibæk Laursen (1994) finder anvendt i debatten om universiteter og i hverdagen på institutionerne (se nedenfor). Den første pind i ministeriets definition af forskningsbaseret undervisning ovenfor er således en parallel til Fibæk Laursens forståelse af forskningsbaseret undervisning, omend ministeriets forståelse er noget bredere. Den fjerde pind i ministeriets definition svarer til Fibæk Laursens tredje forståelse af forskningsbaseret undervisning, mens ministeriets femte pind svarer til Fibæk Laursens fjerde forståelse af forskningsbaseret undervisning.

- *"Undervisningens indhold består af forskningsresultater. (...)*
- *Undervisningen er tilknyttet et forskningsmiljø, således at der forskes i samme faglige og institutionelle miljø. (...)*
- *Undervisningen ledes, men udføres ikke nødvendigvis, af forskere. (...)*
- *Læreren er aktiv forsker i det fagområde, han underviser i, og sørger for at lade forskningen befrugte sin undervisning. (...)*
- *... at de studerende lærer at arbejde forskningsmæssigt gennem samarbejde med praktiserende forskere."* (Fibæk Laursen, 1998 s. 94-95)

Hvad der ikke indgår i budene på forskningsbaseret undervisning ovenfor, er angivelser af forholdet mellem forskningsbaseret undervisning og almen- og universitetspædagogisk og -didaktisk forskning. Den forskningsbaserede undervisning defineres i stedet gennem angivelser af de studerendes faglige og almenakademiske udbytte, samt hvem der bør lede, tilrettelægge og mestendels gennemføre den forskningsbaserede undervisning. Her kan det undre – det undrer i hvert fald mig – at der i definitionerne på forskningsbaseret undervisning ikke også lægges vægt på, at tilrettelæggelse, undervisningsformer og læringsaktiviteter og processer (læs didaktik og pædagogik) bør være forskningsbaseret eller i det mindste forskningsunderstøttet. Udtalelsen fra arbejdsgruppen om undervisningskvalitet om, at den pædagogiske og didaktiske forskning ikke

har den store indflydelse på den forskningsbaserede undervisning (jf. forrige afsnit), tyder på, at jeg ikke er den eneste, der undrer mig over dette forhold.

Spørgsmålet er så, hvordan denne side af den forskningsbaserede undervisning fås med? I rapporten *kvalitet i undervisningen* (Ministeriet for Videnskab, Teknologi og Udvikling, 2006) fremhæves det, at universiteterne til stadighed har til opgave at udvikle ny viden om universitetspædagogik og -didaktik, og at universiteternes ledelse således må prioritere undervisningen højere. I rapporten anbefales det desuden, at en sådan udvikling styrkes gennem tiltag på forskellige niveauer:

1. Formelle tiltag omkring sammensætning af undervisergruppen i forhold til pædagogiske og didaktiske kompetencer.
2. Tiltag med henblik på at skabe en kultur, der fremmer motivationen for at arbejde professionelt med undervisning.
3. Udnyttelse og videreudvikling af fagdidaktisk forskning på områder hvor denne er stærk, og på områder, hvor denne er svag, anbefales det desuden, at der samarbejdes med eksperter i universitetspædagogik.
4. Gennemførelse af ny forskning, der bl.a. tager højde for fag- og kulturspecifikke forhold.
5. Styrkelse af metodisk kontrollerede undersøgelser af effekten af undervisnings-, eksamens- og evalueringsformer i den danske universitetskontekst.

Gennem en generel styrkelse af den universitetspædagogiske og -didaktiske forskning i Danmark, gennem tiltag som de ovenstående, vil der også ske en påvirkning af den forskningsbaserede undervisning. Særligt tror jeg, at tiltag, der sigter mod pædagogisk/didaktisk kvalificering af undervisere og mod ændringer af undervisningskultur, vil have effekt i den retning, idet det jo netop er den enkelte underviser/grupper af undervisere/institutter og enheder, der skal arbejde med udmøntningen af den pædagogiske/didaktiske forskning i den forskningsbaserede undervisning. Jeg vil derfor også foreslå, at listen af tiltag udvides med:

6. Tiltag, der fremmer gennemførelsen af videnskabeligt funderet praksisforskning i egen og andres undervisning og således også påvirker denne undervisning.

I sådanne tiltag ville udgangspunktet således blive en "venden tilbage" til en anskuelse af forholdet mellem forskning og undervisning som et

reelt dobbelttrettet forhold², hvor det ikke “bare” er undervisningen, der indholds-, metode- og bemandingsmæssigt påvirker undervisningen, som i de tidligere nævnte bud på forskningsbaseret undervisning. I stedet ville forskningen således også kunne bidrage form- og procesmæssigt til undervisningen, og undervisningen ville på forskellig vis kunne bidrage til forskningen. Sådanne tiltag ville desuden betyde, at universitetsunderviseren i sin rolle som praksisforsker (og underviser i en kontinuerlig og effektiv pædagogisk efteruddannelse/professionel udviklingsproces) ville kunne overføre (læs anvende) sin videnskabelige erfaring og praksis samt metodiske og metodologiske viden inden for det fagfelt, han/hun forsker og underviser i rent emnemæssigt, til det sideliggende felt, der udgør deres undervisningspraksis (med alt, hvad dette indeholder af planlægning, tilrettelæggelse, udførelse og evaluering). På denne måde kunne vi komme nærmere en forskningsbaseret undervisning, som også indebærer videnskabelighed i den universitetspædagogiske praksis.

I det følgende præsenteres et enkelt bud på et tiltag i forhold til at fremme videnskabeligt funderet praksisforskning i egen undervisning. Det er et beskedent bud, faktisk en dråbe i havet, og jeg håber således, at flere dråber vil blive beskrevet og praktiseret af andre – der er nok at tage af.

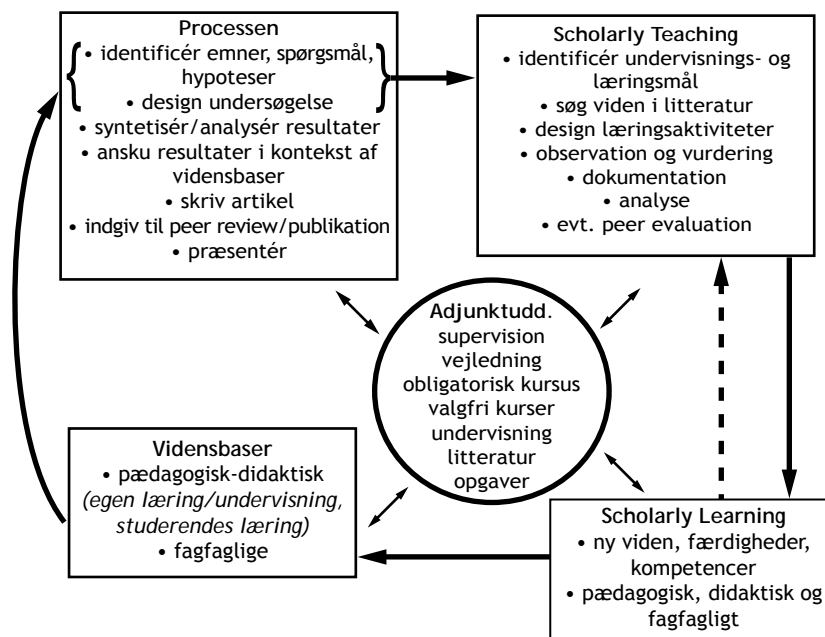
Et bud på udnyttelse af den eksisterende videnskabelighed ind i den universitetspædagogiske praksis

I dette afsnit præsenteres, hvorledes vi³ i adjunktuddannelsen i universitetspædagogisk kompetence (herefter adjunktuddannelsen) ved Copenhagen Business School (CBS), ved hjælp af en såkaldt *scholarship of teaching and learning*-model, har forsøgt at implementere en “udnyttelse” af den eksisterende videnskabelighed i forhold til praksisforskning i egen undervisning. Den *scholarship*-model, vi tager udgangspunkt i (jf. Figur 1⁴), er en bearbejdning af Richlins (2006) SoTL-model. STL-modellen (figur 1) introduceres for adjunkterne på den første dag af det

² At dette kaldes en “venden tilbage” skyldes blot at en sådan dobbelttrettethed i forholdet mellem forskning og undervisning var almindelig i tidligere skrivelser omkring undervisning og forskning på danske universiteter - f.eks. Jensen (1986) og Jacobsen (1990), mens senere tekster synes at have taget en drejning mod “kun” at se på den ene side at dette forhold, nemlig forskningens indflydelse på undervisningen.

³ Det skal bemærkes, at når der flere steder i dette afsnit flere steder skrives “vi”, henviser det til, at adjunktuddannelsen involverer flere undervisere. De meninger, der lægges for dagen, også i dette afsnit, er det dog udelukkende forfatteren, der kan stilles til regnskab for.

⁴ Modellen i figur 1 inkorporerer Richlins (2006) model for *scholarly teaching process* samt modificerede udgaver af Richlins modeller for *scholarly teaching and learning* og *blueprint for the scholarly cycle of course (re)design* og sammensætter disse modeller på en procesmæssig ny måde.



Figur 1. Model af adjunktuddannelsen som scholarship of teaching and learning (bearbejdning af af Richlins (2006) SoTL-model)

fire dage lange obligatoriske adjunktspecifikke kursus ved CBS Learning Lab, således at adjunkterne allerede fra starten af adjunktuddannelsesforløbet kan overveje at tilrettelægge forløbet eller dele af det som et *scholarship of teaching and learning*. Introduktionen af STL-modellen finder sted efter en kort gennemgang af adjunktuddannelsens mange forskellige elementer og krav (jf. CBS Learning Lab, 2007) og består af en trinvis opbygning af STL-modellen, hvor hvert trin gennemgås detaljeret og sættes i forhold til adjunktuddannelsens elementer. Vi starter således med *proces-elementet* og fortsætter med at opbygge modellen i urets retning, endende med at indsætte adjunktuddannelsens elementer i midten af modellen. Vi slutter af med at se på fordelene ved at anskue adjunktuddannelsen som et *scholarship of teaching and learning* i lyset af resultaterne af en større undersøgelse af, hvad der karakteriserer de bedste universitetsundervisere:

- *Proces-elementet* dækker egentlig over hele *scholarship of teaching and learning*-processen og kan simpelt beskrives som en videnskabelig tilgang til undervisning og læring, som i og for sig følger den "samme" opskrift som andet videnskabeligt arbejde. Dermed følger STL-processen den "samme" opskrift som adjunktens fagfaglige videnskabelige arbejde og forskning, og giver dermed mulighed for

anvendelse (læs overførsel) af adjunktens videnskabelige viden og kvalifikationer ind i det universitetspædagogiske og -didaktiske felt.

- *Scholarly teaching*-elementet udgør den del af den samlede STL-proces, hvor undervisningen tilrettelægges, gennemføres, dokumenteres og evalueres i forhold til den undersøgelse, der er designet. Dermed ikke sagt, at dette element nødvendigvis ligger strengt tidsmæssigt efter undersøgelses-designet. I de fleste tilfælde vil der nok snarere være tale om en vekslen mellem de to. Dette element knytter sig på den ene side til de dele af adjunktuddannelsen, som har med pædagogisk og didaktisk teori at gøre – det kunne f.eks. være viden om, hvordan man lærer, undervisningsdifferentiering, målstyring, undervisningsformer, mv. – altså kursus- og litteratur-delene af adjunktuddannelsen. På den anden side knytter *scholarly teaching*-elementet også an til de dele af adjunktuddannelsen, som mere eksplicit har med færdigheder og opnåelse af disse at gøre, dvs. supervisions-, vejlednings- og undervisningsdelene af adjunktuddannelsen.
- *Scholarly learning*-elementet udgør underviserens læringsdel af STL-processen. En del af den læring, der er placeret i dette element udspringer naturligvis direkte af adjunktuddannelsens videns- og færdighedsorienterede elementer (jf. ovenfor). Men selve det at engagere sig i *scholarly teaching* betyder, at adjunkten, ud over den teoretiske viden om læring og undervisning mv., også opnår erfarings- og afprøvningsbaseret viden – det kunne f.eks. være viden om, hvorledes pædagogiske og didaktiske tiltag fungerer i praksis eller viden om studerendes læreprocesser og -udbytte (lærte de det, vi regnede med? – hvorfor/hvorfor ikke?) – og dermed både ny viden, nye færdigheder og nye kompetencer. Denne type læring vil, qua dens erfaringsmæssige element, på den ene side have en direkte tilbagevirkende kraft på adjunktens undervisningspraksis (uanset om denne tilrettelægges som *scholarly teaching* eller ej) og vil på den anden side, qua systematiseringen og evalueringen af disse erfaringer, bidrage med ny viden til adjunktens eksisterende vidensbaser (*vidensbaser*-elementet).

På denne måde når vi i præsentationen af STL-modellen tilbage til udgangspunktet: *proces*-elementet. Denne gang med den pointe, at processen således kan gentages, men nu med udgangspunkt i udvidede vidensbaser og erfaringer. Desuden præsenteres adjunkterne på dette tidspunkt for mulighederne for artikelskrivning som afløsning for den første adjunkt opgave (jf. CBS Learning Lab, 2007) samt eksempler på problemstillinger, som adjunkterne har arbejdet med.

Afslutningsvis præsenteres adjunkterne for, hvad det umiddelbare udbytte for den enkelte adjunkt af at tænke og gennemføre adjunktuddannelsen som et *scholarship of teaching and learning* kan være:

- øget sammenhæng i adjunktuddannelsens elementer (som er mange),
- højnet kvalitet af læreprocesserne (både adjunktens og de studerendes),
- nye muligheder for at arbejde med sin undervisning,
- nye publikations- og konferencemuligheder,
- bane vejen for udvikling mod at blive en *excellent teacher*.

I Bains (2004) kvalitative og mange-facetterede undersøgelse af, hvad der karakteriserer *excellent college teachers* ved 24 amerikanske colleges⁵, blev det (naturligvis) dels fundet, at disse *excellent*⁶ undervisere var forskellige på en del områder, dels at de havde fælles karakteristika på seks områder. Det er disse seks fælles-karakteristika, som drages frem som særligt interessante, idet de afspejler en videnskabelig, professionaliseret tilgang til universitetsundervisning og udviklingen af denne som den, der ligger i *scholarship of teaching and learning*-modellen. De seks fælles-karakteristika for *excellent* undervisere er gengivet i kort form (Bain, 2004):

- *Viden og forståelse*: Underviserne var alle meget velbevandrede i deres fagfelt, og de fulgte fagfeltets udvikling og studerede andres bidrag til fagfeltet. Underviserne bedrev alle forskning og havde originale ideer samt en stor interesse i fagfeltets bredere emner (historier, epistemologiske diskussioner, uenigheder, osv.). Underviserne var kendetegnet ved, at de selv intellektuelt såvel som fysisk og følelsesmæssigt kunne præstere det de forventede af deres studerende. Underviserne havde desuden brugt deres fagfaglige viden til at udvikle teknikker til at forstå og organisere grundliggende principper og begreber, som kunne bruges af studerende til at udvikle deres egen forståelse. Underviserne havde også (mindst) en intuitiv forståelse af menneskelig læring som kun givende mening, hvis den yder indflydelse på, hvordan personen tænker, agerer og føler!

⁵ Disse colleges var forskellige, idet de varierede fra "open admission colleges to highly selective research universities" (Bain, 2004, s. 4). I undersøgelsen indgik nogle og tres undervisere fra mange forskellige fagfelter (sundheds-, samfunds- og naturvidenskab samt humaniora, udøvende kunstarter, management og jura). Af disse undervisere blev 36 studeret indgående, de resterende mindre indgående.

⁶ Det skal bemærkes, at *excellent* i undersøgelsen var defineret som undervisere, "der havde opnået bemærkelsesværdig succes med at hjælpe studerende på måder, der havde en vedvarende, substantiel og positiv indflydelse i forhold til hvordan disse studerende tænker, agerer og føler". (Bain, 2004, s. 5, forfatterens oversættelse).

- *Forberedelse:* Underviserne opfattede og behandlede forelæsninger, diskussioner, problembaserede sessioner og andre undervisningsaktiviteter som seriøse intellektuelle bestræbelser på niveau med deres forskning og fagfaglige bestræbelser. Underviserne tog udgangspunkt i læringsmål for de studerende, frem for hvad underviseren skal gøre i planlægning af undervisning.
- *Forventninger:* Underviserne undgik mål, som var tilfældigt knyttet til et kursus eller fag, og foretrak mål, som relaterede sig til den tænkning og adfærd, som kræves i livet.
- *Undervisning:* De undervisningsmetoder, som de excellente universitetsundervisere benyttede sig af, var ganske varierede. De havde dog det til fælles, at de ofte forsøgte at skabe “et naturligt kritisk læringsmiljø”, hvor der læres ved at konfrontere spændende eller vigtige problemer og virkelige opgaver, som udfordrede til at tumle og arbejde med ideer, gentænke forudantagelser og undersøge opfattelser af virkeligheden.
- *Samspil med studerende:* Underviserne udviste en høj grad af tillid til de studerende – de troede på, at studerende vil og kan lære, indtil andet var bevist. Underviserne udviste en høj grad af åbenhed over for de studerende – også i forhold til deres egen intellektuelle rejse. Desuden fremhæves det, at underviserne alle behandlede studerende anstændigt.
- *Udvikling og evaluering af indsats:* Underviserne havde alle en mere eller mindre sofistikeret metode til at evaluere deres egen indsats og gennemføre passende ændringer i forhold til undervisningen. Underviserne holdt desuden deres egen indsats for øje, når de evaluerede studerende, og undgik således at bedømme studerende efter vilkårlige standarder, fordi de samtidigt evaluerede egen indsats. Bedømmelse af studerende foregik snarere i forhold til de opsatte læringsmål.

Slutteligt gøres det til en pointe i sig selv, at de *excellent* undervisere, der indgik i Bains undersøgelse, ikke altid havde succes med deres undervisning og at de selv var i rollen som *lærende*.

Ud over udbyttet for adjunkterne selv ligger der også en mulighed for institutionelt udbytte ved at indføre muligheden for at gennemføre adjunktuddannelsen som et *scholarship of teaching and learning*, nemlig det, at der gennemføres mere universitetspædagogisk praksisforskning i egen undervisning samt flere, hvad der i seminarieverdenen ville være benævnt, pædagogiske og didaktiske udviklingsarbejder – begge dele lige så nødvendige for den universitetspædagogiske udvikling og kvalitet, som

formelt anlagte større pædagogiske og didaktiske tiltag og forskningsprojekter (jf. det tidligere). Hvorvidt dette institutionelle udbytte vil opstå, er endnu for tidligt at sige noget om, idet STL-modellen første gang blev præsenteret for et hold adjunkter ved CBS i 2007. I skrivende stund har 37 danske og udenlandske adjunkter⁷ således stiftet bekendtskab med STL-modellen og dermed også fået muligheden for at gennemføre deres adjunktuddannelse som et *scholarship of teaching and learning*. Det er således endnu for tidligt at sige noget generelt – for ikke at tale om videnskabeligt baseret – om effekten af dette tiltag, da ingen af disse adjunkter har færdiggjort deres adjunktuddannelse. En hurtig og ganske uvidenskabelig rundspørge blandt de supervisorere, der er tilknyttet adjunktuddannelsen på CBS, afslører, at blandt de adjunkter, som er blevet præsenteret for scholarship-modellen, forholder det sig sådan, at en enkelt adjunkt har indleveret en pædagogisk artikel, syv adjunkter planlægger eller er i gang med at gennemføre eksperiment i undervisningen og af disse adjunkter har fem udtrykt, at de har til hensigt at skrive en artikel derom.

Der er dog al mulig grund til optimisme mht. det institutionelle udbytte af at indføre en *scholarship of teaching and learning*-model på den måde, som vi har valgt at gøre det, frem for med en mere direkte og ovenfra styret form. Det at arbejde med og udvikle sin undervisning på den beskrevne måde, må nødvendigvis være drevet af individuelt oplevet lyst, behov og udbytte hos den, der gør det.

Afsluttende bemærkning – eller tilbage til artiklens årsagsgenstand

Afslutningsvis vil jeg gerne vende tilbage til den person, som er anledningen til, at artiklen blev skrevet. Benny er nemlig den, af alle de undervisere jeg er stødt på, både som studerende, som kollega og som underviser, der bedst lader sig beskrive med fælles-karakteristika for excellente undervisere, som Bain (2004) afdækker i sin undersøgelse. Benny har for mig været en bemærkelsesværdig underviser, der altid var parat til med entusiasme at drage med på læringsrejser.

Referencer

Bain, K. (2004). *What the best college teachers do*. Cambridge, Massachusetts, Harvard University Press.

CBS Learning Lab (2007). *Regelsæt for adjunktuddannelsen på CBS*. http://frontpage.cbs.dk/ll/01_Kurser/2008marts_Regelsaet_adjunktud-

⁷ Størstedelen adjunkter ved CBS, men også enkelte adjunkter fra Københavns Universitet og IT-Universitetet, som deltager i adjunktuddannelsen ved CBS.

dannelsen2007_08_og_bilag.pdf

Fibæk Laursen, P. (1998). Forskningsbaseret undervisning og læring. I Per Fibæk Laursen & Tone Gabrielsen (red.) *At undervise i humaniora*. København, Samfundslitteratur. s. 93 – 109.

Finansministeriet (2007). *Cirkulære om stillingsstruktur for videnskabeligt personale ved universiteter*. København, Personalestyrelsen, Finansministeriet.

Jacobsen, B. (1990). *Universitetsforsker i Danmark*. København, Nyt fra Samfundsvidenskaberne.

Jensen, J-J. (1986). *Forskning og undervisning på universiteterne – et samspil?* Esbjerg, Sydjysk Universitetsforlag.

Kirketerp, S. B. (2007). *Undersøgelse af adjunktuddannelser anno 2007*. http://frontpage.cbs.dk/ll/01_Kurser/Adjunktuddannelser_ind_udland.doc

Ministeriet for Videnskab, Teknologi og Udvikling (2006). *Kvalitet i undervisningen*. København, Ministeriet for Videnskab, Teknologi og Udvikling.

Ministeriet for Videnskab, Teknologi og Udvikling (2007a). *Bekendtgørelse om kriterier for universitetsuddannelsers relevans og kvalitet og om sagsgangen ved godkendelse af universitetsuddannelser – BEK nr 1030 af 22/08/2007*.

Ministeriet for Videnskab, Teknologi og Udvikling (2007b). *Bekendtgørelse af lov om universiteter (universitetsloven) – BEK nr 1368 af 07/12/2007*.

Richlin, L. (2006). *Blueprint for learning. Constructing college courses to facilitate, assess, and document learning*. Virginia, Stylus.

Forfatterliste

John Brodersen

Afdeling for Almen Medicin
Institut for Folkesundhedsvidenskab
Københavns Universitet

Bjarne Sode Funch

Institut for Psykologi og Uddannelsesforskning,
Roskilde Universitet

Jesper Hermann

Institut for Nordiske Studier og Sprogvidenskab,
Københavns Universitet

Tytte Hetmar

Sociologisk Institut,
Københavns Universitet

Mette Høyer

Institut for Psykologi
Københavns Universitet

Jan Ivanouw

Center for Ledelse
Folke Bernadottes Allé 45
2100 Ø

Afdeling for miljø og sundhed
Institut for Folkesundhedsvidenskab,
Københavns Universitet

Boje Katzenelson

Psykologisk Institut,
Aarhus Universitet

Svend Kreiner

Biostatistisk afdeling,
Institut for Folkesundhedvidenskab,
Københavns Universitet

Simo Kørpe

Institut for Psykologi
Københavns Universitet

Guido Makransky

Master Management International
Hillerød

Jens Mammen

Psykologisk Institut,
Aarhus Universitet

Erik Lykke Mortensen

Afdeling for miljø og sundhed
Institut for Folkesundhedvidenskab,
Københavns Universitet

Tine Nielsen

CBS Learning Lab,
Copenhagen Business School

Nini Praetorius

Psykologisk Institut
Københavns Universitet

Jan Rattleff

Knud Rasmussens Vej 17,
2870 Dyssegård

Tom Teasdale

Psykologisk Institut,
Københavns Universitet