

Udforskning i psykologien – De kvantitative metoder

Benny Karpatschof

Udforskning i psykologien

De kvantitative metoder

A

AKADEMISK FORLAG

Udforskning i psykologien – De kvantitative metoder

© 2006 Benny Karpatschof og Akademisk Forlag, København
– et selskab i Bonnier Forlagene A/S

Forlagsredaktion: Vibeke Nørgaard og Andreas Bonnevie

Fagkonsulent: Jan Ivanouw

Omslag: Ida Balslev-Olesen

Typografisk tilrettelægning: Jens Lund Kirkegaard

Sat med:

Trykt hos

Printed in Denmark 2006

ISBN-10: 87-500-3882-6

ISBN-13: 978-87-500-3882-5

Alle rettigheder forbeholdes. Mekanisk, fotografisk eller anden gengivelse af eller kopiering fra denne bog eller dele deraf er *kun* tilladt i overensstemmelse med overenskomst mellem Undervisningsministeriet og Copy-Dan. Enhver anden udnyttelse er uden forlagets skriftlige samtykke forbudt ifølge gældende dansk lov om ophavsret. Undtaget herfra er korte uddrag til brug i anmeldelser.

Akademisk Forlag • Pilestræde 52 • 1112 København K
www.akademisk.dk

Indhold

Kapitel 1

Grundbegreber i psykologisk metodologi 7

Videnskabens udgangspunkt: genstand, praksis og teori 8

Psykologiens placering blandt videnskaberne 12

Den empiriske udforsknings to akser, erkendelsessøgning og erkendelseskritik 15

Metodologi som en før-empirisk eller som en empirisk disciplin 18

Hvilke krav kan stilles til beskrivelse og analyse af metoden i en publikation? 21

Undersøgelsesmetodens bestanddele og processer 22

Genstandsfelterne for psykologi og dens nabovidenskaber 33

Hvor kommer psykologiens metoder fra? 35

Kvalitativ og kvantitativ metode 38

Gyldighedens banemænd – fejlkilderne 60

Intern og ekstern validitet 78

Litteratur 80

Noter 83

Kapitel 2

De eksperimentelle metoder 87

Det eksperimentelle spektrum 87

Demonstrationseksperimentet 89

Det individbaserede (klassiske) eksperiment 98

Det randomiserede gruppeeksperiment 107

Kvasieksperimentet (gruppeeksperiment med naturlige grupper) 127

Den eksperimentelle metodes problemer anskuet symptomalt 133

Litteratur 134

Noter 135

Kapitel 3

Surveys med kvantitative data 139

En karakteristik af surveyundersøgelsen som genre 139

Et eksempel på en surveyundersøgelse 141
Surveydesign 152
Udformning af spørgsmål og svarkategorier 155
Registrering og indkodning af oplysninger og opstilling af database 161
Statistiske sammenhænge 162
Binære fejlslutninger – tilsløring (confounding) 180
Afsluttende bemærkninger om statistiske sammenhænge 192
Survey-metodens indbyggede begrænsninger 192
Litteratur 196

Kapitel 4

Psykometriske metoder 197

Psykometri som genre 197
Psykometriens begyndelse og tidlige udvikling fra Galton til Binet 200
Faktoranalyse 211
Itemanalyse og itemresponsmodeller 223
Validitet og validering 247
Itemresponsteori 253
Den konfirmatoriske faktoranalyse 254

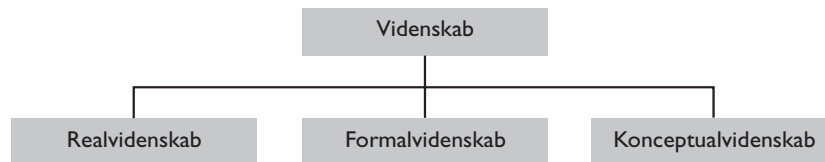
KAPITEL I

Grundbegreber i psykologisk metodologi

Hvis der absolut skal være principper for videnskabelig metode, er det første, der gør krav på vores opmærksomhed, at man bør beskrive fænomener nøjagtigt og lade dem være vejledende for valget af problemstillinger og fremgangsmåder (Asch, 1952, s. ix)

Enhver videnskab må have en metodologi.¹ Den er for forskeren, hvad navigationskunsten er for kaptajnen. Metodologien er den del af videnskaben, der beskæftiger sig med dens undersøgelsesmetoder – med hvad der er betingelserne for at bruge en bestemt, og hvornår den fører en i afgrunden.

Videnskabsteori og metodologi er beslægtede, men hvor videnskabsteori er en særlig videnskabelig disciplin, der studerer de empiriske videnskaber fra et *eksternt* perspektiv, er metodologi en *hjælpedisciplin* for den videnskab, den er en del af. Metodologien har som praktisk opgave at udvikle og kritisk afprøve de empiriske metoder i den pågældende videnskab. Afprøvningen vedrører især en metodes *gyldighed* (validitet), dvs. spørgsmålet om, hvorvidt metoden er egnet til at afdække det stykke af virkelighed, som gøres til genstand for udforskning.



Figur 1.1. Tre typer af videnskab.

Videnskabens udgangspunkt: genstand, praksis og teori

Før vi kommer nærmere ind på grundelementerne i den psykologiske metodologi, kan det være nyttigt at placere selve den videnskab, vi her arbejder inden for, altså psykologien.

Alt det, vi kalder videnskab, kan opdeles i tre kategorier:

1. *Realvidenskab* kalder vi den gruppe af videnskaber, der studerer håndgribelige *genstande*.
2. *Formalvidenskab* er den videnskabsgruppe, hvor man beskæftiger sig med formale størrelser, som fx tal, geometriske former og logiske slutningsregler, denne gruppe omfatter: *matematik* og *logik*.
3. *Konceptualvidenskab* er den videnskab, hvor man funderer over begrebsmæssige problemer; denne videnskab kaldes *filosofi*.

Empiriske undersøgelser forekommer *kun* i den første gruppe, og vi skal i denne bog kun beskæftige os perifert med de to andre.

Inden vi går over til at fokusere på empiriske undersøgelsesmetoder, bør realvidenskaberne placeres i deres fulde kontekst, som udgøres af tre forskellige felter, *genstandsfelt*, *praksisfelt* og *teorifelt*:

Videnskaben og dens kontekst

- *Genstandsfeltet* for en videnskab er de reale størrelser, der kan siges at eksisterende uafhængigt af videnskaben selv, disse størrelser kan opdeles i genstande, fænomener og væsenstræk.
- *Praksisfeltet* knyttet til en videnskab er den del af den almindelige menneskelige virksomhed, som er rettet mod videnskabens genstandsfelt, nemlig den virksomhed, der foreligger i det væsentlige uafhængigt af videnskaben selv.

- *Teorifeltet* er den særlige del af den menneskelige virksomhed, som er erkendelsessøgende.

Genstandsfeltet

Den følgende fremstilling tager udgangspunkt i genstandsfeltet, som er den del af virkeligheden, der udforskes af en bestemt videnskab, men som basalt eksisterer forud for og uafhængigt af denne. Denne bog bygger på den erkendelsesopfattelse, at det ikke er den videnskabelige virksomhed, der skaber genstandsfeltet, men derimod genstandsfeltet som er udgangspunktet for den videnskabelige virksomhed.

Genstand

Ved *genstande* vil jeg forstå forholdsvis sammenhængende, stabile og velafgrænsede entiteter, hvis eksistens har en objektiv status. Det er altså størrelser med tingskarakter. Ordet genstand er på dansk overtaget fra tysk *Gegenstand*, som er noget, man står over for som et erkendende eller handlende subjekt. På engelsk og fransk taler man om genstand som et objekt der foreligger for et subjekt. Herved har ordet fået to betydninger.

De to betydninger af ordet *genstand*:

1. noget tingsagtigt foreliggende uden for og uafhængigt af subjektet,
2. det man står over for i sin virksomhed, det som virksomheden er rettet mod, således kan et landskabsmaleri være genstanden for en maler, og løsningen af et matematisk problem genstanden for en matematiker.

For at skelne mellem disse forskellige betydninger vil jeg i denne bog kalde en genstand i den første betydning for en *real* genstand, og i den anden betydning for en *intentional* genstand. Når det drejer sig om en videnskabs genstand eller et undersøgelsesprojekts genstand, vil jeg bruge betegnelsen: den *epistemiske* genstand. Mens jeg om konkrete undersøgelsesgenstande, fx forsøgspersoner, vil tale om *empiriske* genstande.

Hvis vi ser på et projekt, hvor man følger en række unge mennesker, der har diagnosen Aspergers syndrom, for at undersøge virkningen af et psykoedukativt program, der går ud på at udvikle deres socialitet, så er den epistemiske genstand altså effekten af social færdighedstræning over for unge Aspergere, mens de empiriske genstande er de personer, der gennemgår denne træning.

Fænomen

I modsætning til genstande er *fænomener* størrelser, der har en ustabil, kortvarig og diffus karakter, eller som har den subjektive karakter, at de kun foreligger som noget, der fremtræder for et menneskeligt subjekt. Ordet fænomen bruges både om noget konkret, enkeltstående (singulært) foreliggende og om begrebet for alle singulære fænomener af samme slags. Således kan man tale om en bestemt orkan som et enkeltstående fænomen, der ligefrem kan få tilknyttet et egennavn. Ved siden af denne singulære anvendelse af ordet fænomen, har vi også den generelle (universelle). Således er *begrebet* orkan den universelle klasse af fænomener, der omfatter alle de enkeltstående orkaner. Lige som en genstand kan et *singulært* fænomen kaldes en *entitet*, dvs. noget der objektivt foreligger. Imidlertid er det *begrebslige* fænomen ikke nogen entitet, i hvert fald ikke hvis man ser bort fra dets status i kognitiv eller sproglig henseende.

En epistemisk genstand vil ofte være et begrebsligt fænomen, som vi netop ser det i eksemplet med psykoedukation i forbindelse med Aspergers syndrom.

Væsenstræk

Væsenstræk er de forhold ved enkelte genstande og fænomener, som har en almen karakter. Væsenstræk er således en forudsætning for en dybere forståelse af genstandenes eller fænomenernes natur. Hvor genstande og fænomener er det empiriske udgangspunkt for en videnskab, nemlig det som umiddelbart kan iagttages, er væsenstrækkene målet for videnskabens stræben efter at udvikle teorier, som sigter mod at trænge ind bag genstandenes og fænomenernes umiddelbare, ofte uklare og indbyrdes modstridende fremtrædelsesformer.

Væsen er en oversættelse af det skolastiske ord *essens*, af latin *essentia*, som kommer af *esse* = være. Essens har således samme oprindelse som entitet. Essens er imidlertid netop *ikke* en entitet, men lige som det begrebslige (universelle) fænomen noget, der først fremtræder i vores forståelse efter en indordning eller en analyse af det realt foreliggende. Herved ligner essens det begrebslige fænomen og er egentlig et begrebsligt fænomen, som når vi taler om en persons personlighed, om et samfunds kultur eller om den teoretiske forståelse af kvantefænomenerne som formuleret i kvanteteorien. Men selv om essens altid er begrebsligt, så er ikke alle begreber udtryk

for essens. Det er de først, når de er knyttet til en videnskabelig teoridannelse.

Som allerede nævnt hører psykologien til den videnskabskategori, som omfatter realvidenskaberne. Vi skal nu se på dens placering inden for denne kategori.

Realvidenskaberne kan opdeles i tre hovedfelter:

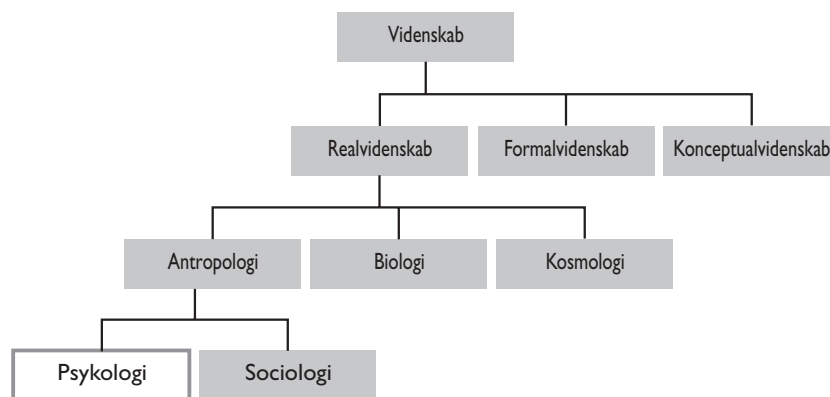
- Den fundamentale naturvidenskab som studerer kosmos, og som består af enkeltvidenskaberne astronomi, fysik og kemi. Da den beskæftiger sig med bogstaveligt talt universelle, altså kosmologiske fænomener, vil jeg under ét benævne denne gruppe af videnskaber kosmologi, en term der i snæver forstand omhandler teorier om universets oprindelse og udvikling, men jeg bruger den i bred forstand som videnskaben om kosmos.
- Livsvidenskab som studerer alle livsformer og livsprocesser og som består af enkeltvidenskaber som zoologi, botanik, biokemi, genetik, palæontologi, evolutionsteori. Under ét skal jeg benævne denne kategori af videnskaber som biologi.
- Videnskaben om mennesket, dvs. human og samfundsvidenskab, som jeg under ét vil kalde for antropologi. Antropologi står traditionelt for en filosofisk og en biologisk disciplin samt for studiet af bestemte folkeslag, men bruges i denne bog som et overbegreb for de enkeltdiscipliner, der beskæftiger sig med forhold specifikke for menneskearten. Antropologi kan opdeles i psykologi, læren om mennesket som person, og sociologi, læren om menneskets frembringelser. Sociologi bruges altså i denne bog som samlebegreb for enkeltvidenskaber som samfundsteori, historie, økonomi, sprogteori og en lang række kulturstudier som kunsthistorie, litteratur, musikhistorie.

Psykologiens placering blandt videnskaberne

Psykologien indplacering fremgår af figur 1.2. ■

Genstand, teori og praksis

Efter denne placering af psykologi inden for videnskabernes familie, vender vi nu tilbage til en videnskabs placering i forhold til dens eksterne kontekst, dvs. til teorifeltets forhold til det genstands- og praksisfelt, den er tilknyttet.

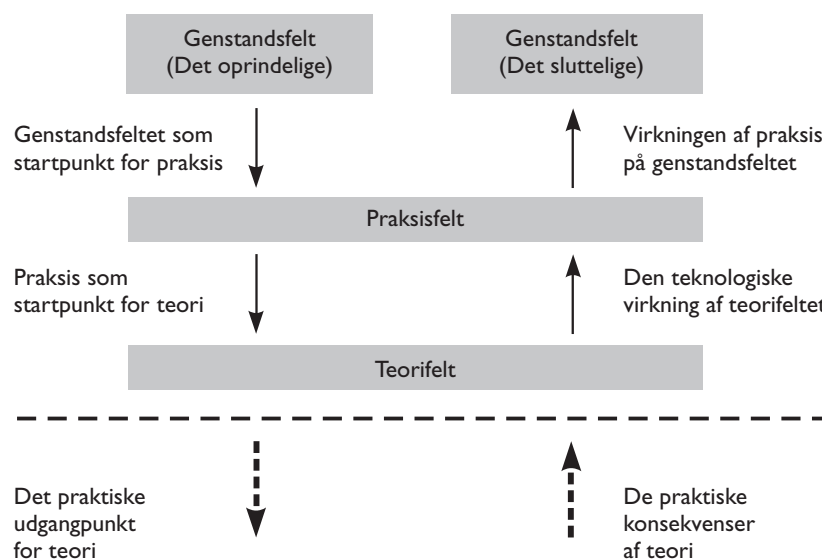


Figur 1.2. Psykologiens placering.

Det kan give anledning til forvirring, som det fx kommer til udtryk i radikale tolkninger af socialkonstruktivismen,² at ganske vist foreligger genstandsfeltet forud for den videnskab, som studerer det, men *opdelingen og afgrænsningen* af genstandsfeltet bliver først afklaret gennem arbejdet i den pågældende videnskab, altså gennem dens udvikling af teorifeltet.

Det kan derfor se ud som om, at det er teorifeltet, der *konstruerer* sit genstandsfelt, og det er i virkeligheden socialkonstruktivismens hovedargument. Imidlertid er det her min påstand, at vi allerede, før der udvikles den målrettede erkendelsesvirksomhed, som vi kalder videnskab, har opbygget en fond af *førvidenskabelig* erfaring gennem den del af vores praksis, der er rettet mod genstandsfeltet. Vi møder altså genstandsfeltet i praksisfeltet, før vi møder det i teorifeltet. Teorifeltet er en *udspaltning* af

praksisfeltet, en fortsættelse af den førvidenskabelige erfaring med andre midler (se figur 1.3).■)



Figur 1.3. Den eksterne relation mellem teori- og genstandsfelt.

Diagrammet illustrerer, hvordan genstandsfeltet er udgangspunktet for praksis, som igen er udgangspunktet for teorifeltet. Dette er, hvad man med en fysiologisk metafor kunne kalde den afferente vej, iagttagelsens, erkendelsens vej. Nu kommer vejen tilbage, den efferente, den virkende, handlende vej. Den efferente vej går fra teorifeltet over til de praktiske anvendelser eller udmøntninger af den videnskabelige erkendelse. Dette er i diagrammet kaldt for *den teknologiske virkning af teorifeltet*. Altså teori-feltets indvirkning på praksisfeltet. Ændringerne i praksisfeltet vil da give sig udslag i et genstandsfelt, men ikke nødvendigvis det samme som vi begyndte i.

Her skal vi imidlertid koncentrere os om videnskaben psykologi.³ Hvad kan man forstå ved en psykologisk genstand og det psykologiske genstands-felt?

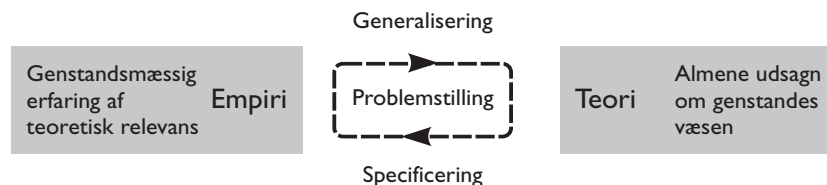
I denne bog defineres en *psykologisk genstand* som et *menneskeligt individ*, vel at mærke i den dobbelte egenskab af at være *menneskeligt* og *individ*. Dermed har vi allerede afgrænset psykologi i forhold til dens to beslægtede

videnskaber. På den ene side er der *biologien*, der nok beskæftiger sig med individer, men ikke specielt med de menneskelige individer. Også de træk ved mennesker, som er så generelle, at vi har dem til fælles med andre dyrearter, falder efter denne definition ind under biologien. Det gælder således hele den *komparative psykologi*. På den anden side har vi *samfundsvidenskaberne*, her under ét kaldt *sociologien*. Disse videnskaber beskæftiger sig ikke med det menneskelige individ som sådant, men med resultatet af den menneskelige virksomhed.

Fælles for de nævnte videnskaber er, at de alle er real-videnskaber, fordi de beskæftiger sig med tinglige størrelser, mens filosofi alene beskæftiger sig med abstrakte begreber og idéer, og formalvidenskaberne alene arbejder med formale abstraktioner af matematisk og logisk art.

Genstand, teori og empiri

Efter denne videnskabsteoretiske indledning skal vi beskæftige os med selve metodologien, som primært vedrører den interne relation mellem teori- og genstandsfelt i en videnskab. Mens vi i diagrammet figur 1.3 så på denne relation ud fra et eksternt perspektiv nemlig praksis, så skal vi nu betragte den internt, inde fra videnskaben selv. Da realvidenskaberne tager udgangspunkt i konkrete tinglige størrelser og foreteelser, må de studere disse gennem iagttagelse, dvs. empirisk (se figur 1.4■).



Figur 1.4. Den interne relation mellem teori- og genstandsfelt.

De centrale begreber i metodologien er altså *teori* og *empiri*. Læg mærke til at ordet *teori* anvendes i to forskellige betydninger. Den mere omfattende betydning af teori svarer til, hvad jeg kalder *teorifelt*, hvor ordet står over for *praksis*. I den mere snævre betydning af ordet teori står det over for *empiri*. I begge tilfælde er der tale om et sæt af tvillingebegreber og ikke om egentlige modbegreber. For lige som teorifelt ikke kan eksistere uden praksisfelt, kan teori ikke eksistere uden empiri (i hvert fald ikke i en

realvidenskab som psykologi, men nok i begrebs- eller konceptualvidenskab som filosofi).

Empiri og teori er altså grundkonstituenterne i teorifeltet, og de skal forstås på følgende måde:

- *Empiri* er genstandsmæssig erfaring af teoretisk relevans.
- *Teori* er almene udsagn om genstandes væsen.

Relationen mellem empiri og teori udspændes af den eller de problemstillinger, der er udgangspunktet for empiriske undersøgelser. Problemstillingen kan være en egentlig hypotese som i Poppers videnskabsteori (1968), der kanoniserer den hypotetisk-deduktive metode. Den kan også være mindre færdig og firkantet, fx nogle undersøgelsesmaal i en eksplorativ undersøgelse.

Den empiriske udforsknings to akser, erkendelsessøgning og erkendelseskritik

Vi har nu indført de tre centrale begreber, teori, empiri og problemstilling. Imidlertid kan vi ikke klare os med disse. For i empiribegrebet ligger der allerede, at vi må have nogle *undersøgelsesgenstande*, dvs. de genstande, vi observerer, således at det, vi observerer vedrørende disse genstande, bliver vores empiri. Men vi kan ikke blot observere nogle tilfældige genstande på en tilfældig måde. I en videnskabelig undersøgelse må man ikke blot kunne redegøre for sine *undersøgelsesgenstande*, men også for sin *undersøgelsesmetode*. Ganske vist er hovedmålet i en realvidenskab at opnå erkendelse om dens genstandsfelt. Denne bestræbelse udgør realvidenskabens første, *erkendelsessøgende akse*. Men for at være en erkendelse, vi kan bygge videre på, må der ud over en løbende indsamling af empiri og udvikling af teori, også være en kritisk kvalitetsvurdering af empirien og dermed af vores grundlag for at danne teorier. Derfor er realvidenskabens anden dimension, den *erkendelseskritiske akse*. Dvs. den aktivitet, hvor vi kritisk reflekterer over den måde, vi opnår vores erkendelse på. Det er den virksomhed, som er metodologi.

Enhver videnskabelig undersøgelse vil således have to komponenter. Dels undersøgelsens resultatorienterede, *erkendelsessøgende* komponent. Dels undersøgelsens reflektive, *erkendelseskritiske* komponent. I den før-

ste, erkendelsessøgende proces fremlægger man sine resultater og giver en konklusion, primært i forhold til den basale problemstilling. I den anden, erkendelseskritiske proces fremlægger man sin metode og argumenterer på den ene side for sine valg, men peger på den anden side også på evt. svagheder ved metoden eller overlader til kolleger på grundlag af metodebeskrivelsen at fremkomme med en ekstern kritik.

Man kan sige at enhver videnskabelig undersøgelse har to genstande og to mål:

1. Erkendelsessøgningens genstand, *undersøgelsesgenstanden* er en del af genstandsfeltet, for psykologiens vedkommende typisk en person eller en gruppe, og erkendelsessøgningens mål er at opnå viden om denne genstand.
2. Den erkendelseskritiske genstand er derimod selve *undersøgelsesmetoden*, og målet er at opnå viden om denne metode.

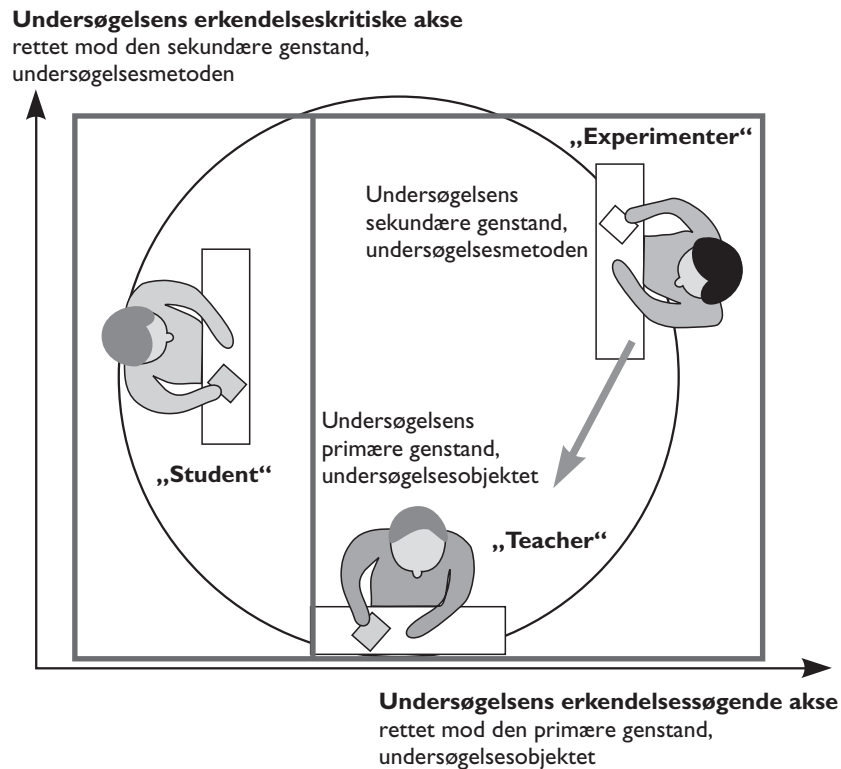
Når man forsker, udforsker man således to forskellige ting: *undersøgelsesgenstanden* og *undersøgelsesmetoden*.

Figur 1.5 ■ illustrerer denne dobbelthed. Det er en fremstilling af Stanley Milgrams (1974) berømte eksperiment om lydighed. Lige som i et andet af psykologiens klassiske eksperimenter, Asch's undersøgelse af gruppepres, var der her tale om en dobbelt manipulation, dels var formålet med undersøgelsen skjult bag den falske angivelse, at der var tale om et indlæringsforsøg, dels var opsætningen manipuleret, idet den person, der var angivet som genstanden for straf, i virkeligheden en medvirkende assistent, der blot illuderede smerte ved afstraffelse.

På tegningen indgår der således tre personer:

1. den falske forsøgsperson angivet som „elev“, som altså faktisk er assistent,
2. den sande forsøgsperson angivet som „lærer“ og
3. forsøgslederen angivet som „forsøgsleder“.

Den manipulativt udpegede „lærer“ fik at vide, at den falske forsøgsperson, „eleven“ var i gang med nogle indlæringsforsøg, der gik ud på at undersøge betydningen af afstraffelse gennem elektriske stød. Den sande forsøgsperson sad foran et voltmeter og skulle eksekvere afstraffelsen i form af gradvist stærkere strømspænding for hver fejl begået af „eleven“, der var



Figur 1.5. Et eksempel – søgning og erkendelseskritik

instrueret i gennem smerteudbrud og vridninger af kroppen at fremvise et passende udtryk for lidelse. På voltmeteret var spændingen angivet på en skala fra 0 til 450 Volt, og virkningen heraf blev beskrevet som gående fra først *let*, så *stærkere smerte*, over en *ekstrem intens smerte* til *livsfare*. „Forsøgslederen“, måtte lægge et stigende pres på „læreren“ for at få denne til at optrappe afstraffelsen i form af voksende spænding. Og eksperimentet havde netop til formål at undersøge, hvor langt dette pres fra en autoritetsfigur kunne drive ganske almindelige mennesker i retning af moralsk uacceptable handlinger.

Her er altså den primære *undersøgelsesgenstand* „læreren“ med påtvungen lydighed som *problemstilling*. Den sekundære genstand er imidlertid hele *undersøgelsesmetoden*, dvs. selve opstillingen med de tre medvirkende. Hvad var det for interaktioner, der forelå, var der tale om observationer, der var *gyldige* (valide) inden for laboratoriet, sådan at de var reproducerbare, og var de også gyldige uden for, på den måde at de modsvarede reelle

lydighedsfænomener, fx i tilfælde af krig? Med hensyn til det første, spørgsmålet om den interne gyldighed, blev der ved forsøgenes offentliggørelse i 1963 rejst stor tvivl om, hvorvidt situationen havde realitetspræg for forsøgspersonen, om de faktisk selv troede, at de var ved at slå „eleven“ ihjel. Tvivlen skyldtes ikke mindst, at over halvdelen indvilgede i at gøre det. Med hensyn til den eksterne gyldighed var spørgsmålet, om den lydighed, der blev udvist, modsvarede fx soldaters adfærd i en krigssituation.

Metodologi som en før-empirisk eller som en empirisk disciplin

Undersøgelsesmetodens „korrekthed“ bliver ofte anskuet som en forudsætning for gyldigheden af undersøgelsesresultaterne. Hvis der er fejl i metoden, vil mange også opfatte resultaterne som fejlbehæftede. Dette kan man kalde *kongruensopfattelsen* i metodologi. Altså at der er fuld overensstemmelse mellem udgangspunktet og slutpunktet for undersøgelsen. En sådan opfattelse fører ofte til en dogmatisk metodologi, der opfatter metodekravene som en absolut lovbog, hvor afvigelser fører til fortabelse, mens det at følge den er en garanti for en pletfri undersøgelse med gyldige resultater.

Kongruensopfattelsen er imidlertid ikke holdbar. En øjensynlig fejlbehæftet undersøgelsesmetode kan sagtens føre til resultater, som er holdbare, og en tilsyneladende fejlfri undersøgelse kan alligevel vise sig at give resultater, der ikke holder vand.

Hvordan kan vi da anskue sammenhængen mellem metode og resultat? Der er ikke tale om en *kausal* relation, som når vi siger uden antændelse ingen brand, men derimod om en *logisk* relation, nemlig den såkaldte gyldighedsrelation.

Lad os som eksempel se på, hvordan Holberg lader Erasmus Montanus foretage den berømte opvisning i logisk skoleridt:

Præmis nr. 1: En sten kan ikke flyve

Præmis nr. 2: Morlille kan ikke flyve

Konklusion: Morlille er en sten.

Vi bliver her præsenteret for et realt og et logisk forhold:

1. Det *reale* forhold er at konklusionen heldigvis er *usand*, selvom morlille af den lærde søn forledes til at tro, den er sand. Hvilket vidner om en høj grad af hensynsløshed over for moderen.
2. Det *logiske* forhold er at selve slutningsformen er *ugyldig*. Hvilket vidner om en lige så stor hensynsløshed over for logikken.

Her er altså præmisserne sande, slutningsformen ugyldig og konklusionen falsk.

Vi kan imidlertid ikke gå ud fra, at når slutningsformen er ugyldig, så bliver konklusionen falsk. Der er nemlig ikke i almindelighed noget simpelt forhold mellem slutningsformens gyldighed eller ugyldighed på den ene side, og konklusionen sandhed eller falskhed på den anden.

Det eneste tilfælde, hvor der er et simpelt forhold mellem præmisser og konklusion, er den konstellation, hvor begge præmisser er sande, og hvor slutningsformen er gyldig; der må konklusionen være sand.

I alle andre konstellationer af præmisser og slutningsform er det slet ikke givet, hvad udfaldet bliver med hensyn til konklusionen.

Det fremgår at følgende eksempel:

Præmis nr. 1: Et menneske kan ikke flyve

Præmis nr. 2: Morlille kan ikke flyve

Konklusion: Morlille er et menneske.

Slutningsformen er ugyldig, hvilket kun indebærer, at vi ikke kan være sikre på, at de to sande præmisser fører til en sand konklusion. Vi kan imidlertid heller ikke være sikre på, at de ikke gør det. Vi kan kort sagt ikke være sikre på noget som helst, når der foreligger en ugyldig slutningsform.

Det samme er tilfældet, når vi har en undersøgelse, som er gennemført med en metode, der ikke er gyldig. Så kan vi heller ikke være sikre på noget som helst. Metoden kan meget vel føre til forkerte resultater, men vi kan på den anden side heller ikke, ud fra metoden alene, på forhånd forkaste undersøgelsesresultatet.

Kongruensopfattelsen er ofte knyttet til en fundamentalistisk metodologiopfattelse, hvor metodologien betragtes som en før-empirisk disciplin, på linje med netop logikken, der i hvert fald indtil opdagelsen af de kvantemekaniske fænomener⁴ blev opfattet som over- og før-empirisk. Denne opfattelse er vel umiddelbart forståelig nok, for hvis ikke metoden

er givet før undersøgelsen, vil der være alvorlig fare for, at en metodologisk argumentation for en undersøgelses gyldighed løber i ring. Hvis argumenterne for metodens gyldighed er bestemt af undersøgelsesresultatet og ikke omvendt, falder øjensynligt enhver metoderedegørelse til jorden.

Imidlertid rejser også opfattelsen af metodologiens før-empiriske karakter store problemer. Hvis vores viden om undersøgelsesmetoder er absolut hævet over det empiriske, hvor kommer denne viden da fra, og kan den så slet ikke ændre sig i tidens løb? Og hvis metodologien er givet uafhængigt af det empiriske, som dettes forudsætning, hvor har vi så argumenterne for, at en given metode er gyldig?

Den opfattelse, jeg vil fremføre over for tesen om metodologiens før-empiriske status, er mindre beroligende for forskeren. Den giver nemlig ingen *sikre* garantier. Til gengæld giver den heller ingen *falske* garantier. Denne metodologiopfattelse vil jeg kalde den *forskningsintegrerede*. I forlængelse af modellen med de to forskningsakser med hver sin genstand og målsætning går denne opfattelse ud på, at den side af forskningen, som er erkendelsessøgende, rettet mod en undersøgelsesgenstand, og den anden side, som er erkendelseskritisk, rettet mod undersøgelsesmetoden, er sammenknyttede aspekter af den empiriske proces. Gennem en empirisk undersøgelse får vi viden om både den primære og sekundære genstand, om såvel undersøgelsesgenstanden som *-metoden*. Erkendelsen af de to slags genstande i forskning sker i princippet parallelt.

Men fører en sådan opfattelse ikke til noget værre rod? Jo, i den udstrækning forskning ofte må være en temmelig rodet affære. Imidlertid er der også grænser for roderiet. Der er nemlig nogle partielle måder, hvorpå vi kan have forholdsvis klare erkendelsesveje fra genstand til metode og omvendt.

1. For hvis vi med en given allerede udviklet metode kommer frem til øjensynligt forkerte eller bogstavelig talt misvisende resultater, er vi tvunget over i en metodologisk analyse for at se, om det kan være *metoden*, der fører os på gale veje. Når vi er blevet overbevist om, at det er tilfældet, så må vi prøve at udvikle en ny og bedre metode. (Spørgsmålet om metodens gyldighed i forhold til genstand og problemstilling vil blive behandlet i kapitlets sidste afsnit).
2. En anden mulighed er imidlertid, at der ikke er noget i vejen med metoden og for så vidt heller ikke med resultaterne, men derimod med vores *teoretiske* forhåndsopfattelse af genstandsfeltet. Når resultaterne

er i modstrid med det, vi forventer, er det ikke nødvendigvis en modstrid i forhold til virkeligheden, men i forhold til vores fejlagtige *opfattelse* af virkeligheden. Og det er netop et eksempel på en måske radikal nyerkendelse inden for genstandsfeltet.

3. Der er endvidere den mulighed, at vores erkendelsesudvidelse omfatter undersøgelsesgenstand og metode under ét. Det gælder således ved forskning inden for et hidtil uopdyrket og uudforsket område, hvor vi må forsøge at udvikle undersøgelsesmetoden, samtidig med at vi påbegynder udforskningen, og dermed sker opdagelsen af nye fænomener parallelt med opdagelsen af nye metoder.

Hvilke krav kan stilles til beskrivelse og analyse af metoden i en publikation?

I forrige afsnit har vi taget et skridt tilbage fra et absolutistisk sandhedsbegreb og fra et doktrinært metodebegreb. Men hvilke krav kan man da med rimelighed stille til metodebeskrivelse og analyse i en publikation? Vi kan her betragte en sekvens af stigende krav til metodologisk sofistikation:

1. Eksplicitering
2. Argumentation
3. Dokumentation

I første fase, *eksplicitering*, beskrives den metode, der har været anvendt. Hvis denne eksplicitering er tilstrækkelig præcis, har andre forskere mulighed for dels at replikere undersøgelsen, dels at vurdere dens troværdighed og evt. kritisere metoden, hvilket naturligvis kan være relevant for såvel undersøgelsens forfatter, som for de læsere, der interesserer sig for artiklen.

I den næste fase går vi et skridt videre. Her diskuterer undersøgeren sin metode, sammenligner den evt. med alternative muligheder og *argumenterer* for sit valg. Dermed bliver undersøgerens intention og bevæggrunde udfoldet, og det forøger mulighederne for et frugtbar debat i den videnskabelige offentlighed om såvel undersøgelsesresultater som metode.

I tredje fase udvides argumentationen for den valgte metode med et forsøg på at *dokumentere*, at den har været adækvat i forhold til undersøgelsesgenstand og problemstilling. At den med andre ord opfylder validitetskravet. Selv om at andre forskere ikke nødvendigvis bliver overbeviste,

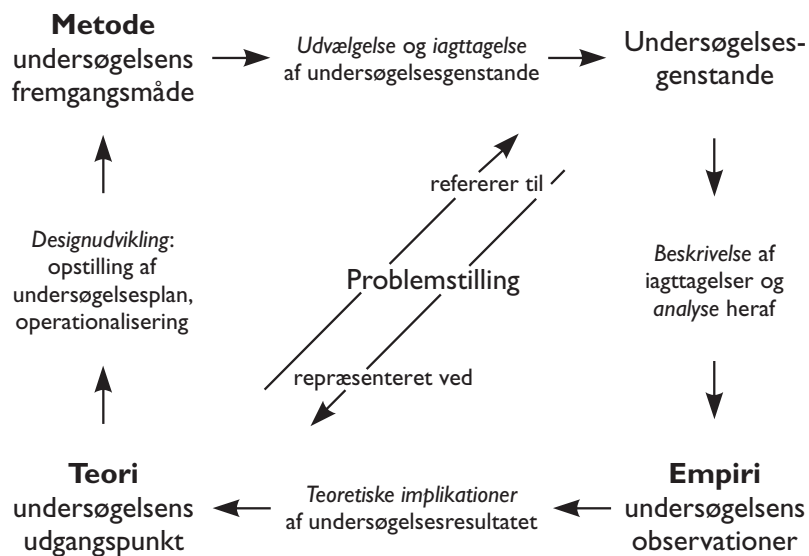
ja, netop når de ikke gør det, er der her de bedste muligheder for at bidrage til en metodologisk afklaring.

Man kan sige at fase 1 betegner den videnskabelige tærskel: uden eksplicitering er der faktisk slet ikke tale om en videnskabelig undersøgelse. Fase 3 indebærer imidlertid den fulde opfyldelse af videnskabelighedskravet.

Efter disse overordnede betragtninger vil vi nu koncentrere os om, hvordan en mere detaljeret metodologisk analyse kan foregå. I den forbindelse kan det være nyttigt at opstille en samlet model for de konstituent og processer, der indgår i undersøgelsesmetoden.

Undersøgelsesmetodens bestanddele og processer

Figur 1.6, som er en udbygning af figur 1.4, viser de bestanddele og processer, der indgår i en empirisk undersøgelse.



Figur 1.6. Model over den empiriske undersøgelse.

I dette diagram er alle metodologiens bestanddele og processer inddraget. Omdrejningspunktet er problemstillingen, som forbinder undersøgelsens teoretiske og empiriske side. Gennem problemstillingen kommer teorien til at vise hen (referere) til undersøgelsesgenstanden, som gennem problemstillingen bliver repræsenteret ved teorien.

Diagrammets forskellige processer er forenet i en cirkel- eller rettere i en spiralbevægelse, sådan som vi skal se det om lidt.

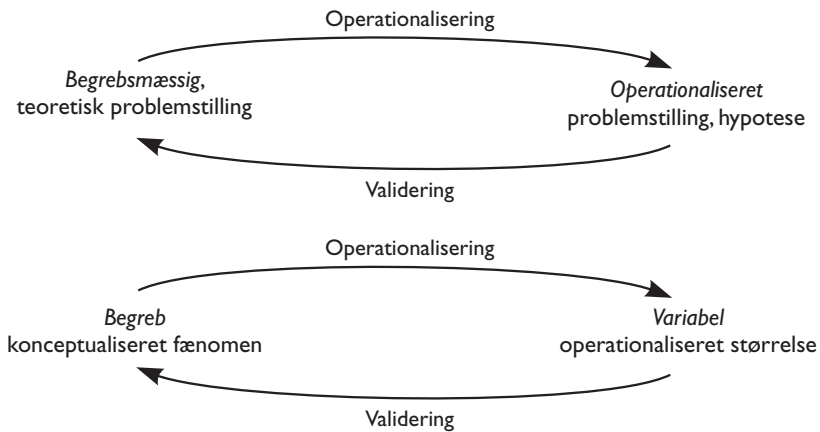
Vi kan fx begynde i *teorifeltet*, altså med en teoretisk baseret problemstilling, som vi gerne vil undersøge empirisk. Først må vi udvikle eller skitsere et *forsøgsdesign*, altså en plan for undersøgelsen, herunder vores måde at udvælge undersøgelsesgenstande og observere dem. Dette indebærer i særdeleshed en konkretisering af vores teoretiske begreber, så vi ved, hvordan vi skal få fat en empiri, der modsvarer disse. Dette kaldes (især i kvantitative undersøgelser) for *operationalisering*.

Lad os nu sige, at vi vil undersøge, om et integrationsprogram for indvandrerbørn i en skole har virket. Så skal vi have et mål for integration eller det modsatte, som vi med et amerikansk begreb kalder for segregering. Her kan vi måske brug et begreb som *social distance*⁵ til at operationalisere vore teoretiske forestillinger om nærhed eller afstand mellem gruppen af såkaldt danske og såkaldt indvandrerbørn. En klassisk metode er sociometrien,⁶ hvormed man kan lave sociogrammer, der er en slags kort over de sociale relationer.

Begrebet *operationalisering* er især knyttet til de kvantitativt orienterede metoder og til mere naturvidenskabelige områder og teoridannelser i psykologien. Den mest konsekvente opfattelse af operationalisering finder man i den videnskabsteoretiske retning, der hedder operationalisme. Ud fra denne opfattelse er videnskabens begreber alene givet gennem de operationer, forskeren udfører.

Da psykologens begreber kan have forskellige funktioner og optræde i flere forskellige teorier, er det imidlertid nødvendigt at have et *begreb* om det, man gør, når man udfører et forsøg, hvor man operationaliserer den problemstilling, som er undersøgelsens udgangspunkt. Der er således en indre forbindelse mellem en problemstillings begrebslige eller teoretiske form og dens operationelle form. Ligesom der er det mellem en begrebslig og en operationaliseret størrelse. Dette illustreres i figur 1.7. ■

Ved operationaliseringen giver vi vores forestilling en håndgribelig, og dermed også en observerbar form. Ved valideringen søger vi at vise, at det vi har operationaliseret faktisk svarer til vores forestilling. Hvis vi fortsætter



Figur 1.7. Forholdet mellem begrebslig og operationaliseret udgave af problemstilling og fænomen.



Figur 1.8. En skala til operationalisering af integration.

vores eksempel med etnisk integration i skolen, kan vi med forsøge at operationalisere dette begreb (og dets modbegreb adskilthed) ved hjælp af den dimension, der fremgår af figur 1.8.

Vi kan nu bruge de førnævnte socialpsykologiske målemetoder, dvs. sociometri og den sociale afstandsskala til at operationalisere integration. Ved hjælp af sociometri kan man spørge børnene i en gruppe om deres positive valg, fx hvem de vil foretrække at sidde ved siden eller arbejde sammen med, eller deres negative valg, hvem de ikke vil have med at gøre. Når man bruger den sociale afstandsskala, spørger man, hvor intime relationer man kan acceptere til en person med en etnisk (eller anden social) oprindelse, der er forskellig fra ens egen. Hvor sociometrien vedrører konkrete sociale relationer, handler den sociale afstandsskala altså om mere abstrakte kategorier.

Når vi så har operationaliseret integration, kan vi forsøge at *validere* den eller de variable, vi er kommet frem til. Vi kan fx prøve at finde en faktisk foreliggende størrelse i vores genstandsfelt, der kan fungere som kriterium

for integrationsvariablen. Lad os nu sige, at vi har mulighed for at følge, hvem der i en undersøgt klasse inviterer hvem til fødselsdag, og der viser sig at være en vis overensstemmelse mellem den operationaliserede variabel og kriteriet, så har vi foretaget en validering af vores operationalisering.

Vi skal i efterfølgende kapitler vende tilbage til spørgsmålet om, hvordan man hhv. operationaliser og validerer. I dette kapitel holder vi os til de mere overordnede emner, og her fortsætter vi med beskrivelsen af undersøgelsesplanen i sin helhed. Vi går nu ud fra, at vi har en operationaliseret undersøgelsesplan, der som angivet i modellen indeholder dels en *udvælgelses-* dels en *iagttagelsesplan* for vores undersøgelsesgenstand. I psykologien vil der hyppigst være tale om *personer* i en bestemt kontekst eller situation eller om primærgrupper, fx en skoleklasse eller en familie, og i organisationspsykologien også om større sociale entiteter, fx en hel erhvervsvirksomhed. Den empiri, der kommer ud af vores iagttagelser, vil, efter at den er blevet beskrevet og analyseret, være vores *undersøgelsesresultat*, der så forhåbentligt har nogle *teoretiske implikationer*, som bringer os tilbage til den teori, vi begyndte med. Denne rundtur var altså den teoretisk initierede version.

Der kan også være en udgave, som ikke tager udgangspunkt i en allerede til formålet udarbejdet teori, således at der alene er tale om teoretiske *forestillinger*, der i hvert fald implicit må være til stede. Der kan være fremkommet en problemstilling, som er udgået af selve praksisfeltet som i eksemplet med integration. Her har man nok et undersøgelsesmål og en problemstilling, som alene er baseret på praktiske erfaringer i en del af genstandsfeltet. Men også her må man gøre sig metodiske overvejelser, og undersøgelsesresultatet må pr. definition være relevant for psykologisk teoridannelse, hvis der er tale om en videnskabelig psykologisk undersøgelse.

I det foregående har vi set på undersøgelser med hhv. et teoretisk og et førteoretisk udgangspunkt. Som regel er der ikke tale om et enkelt gennemløb, men om flere, der blot ikke er rene gentagelser, men som gerne skulle indeholde en udvikling i retning af voksende erkendelse. Dette forhold udtrykkes ofte på den måde, at der ikke foreligger en cirkel-, men en spiralbevægelse. Metaforikken i spiralfiguren er den, at mens en cirkelbevægelse er udtryk for, hvad Hegel kaldte den slette uendelighed,⁷ hvor man i realiteten ikke bevæger sig af stedet, så er spiralen sindbilledet på en udvikling, som gennemløber de samme stadier, men i en bevægelse med en stadig større udstrækning eller på et stadig højere niveau.

Forskningens faser og cyklusser (forskningsspiralen)

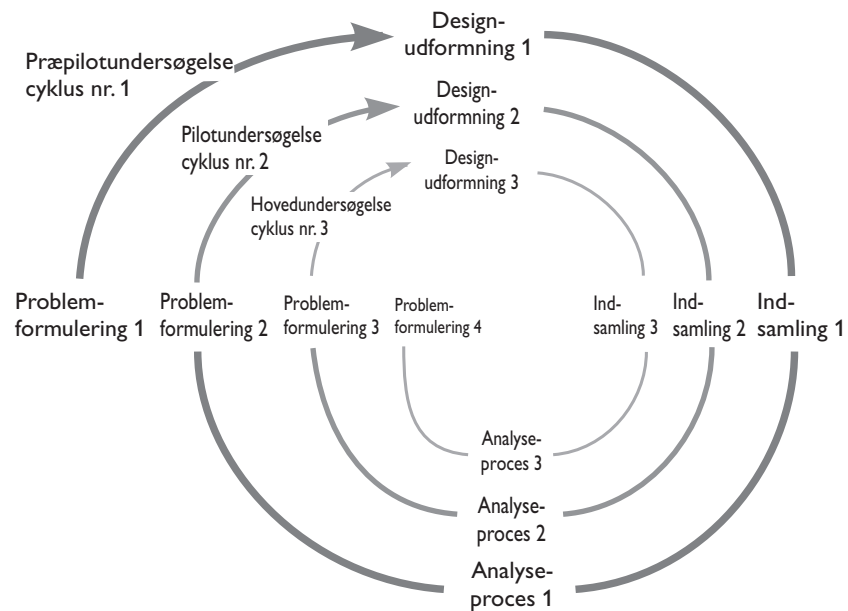
Det fremgår af figur 1.9, at vi i forskningsprocessen må skelne mellem på den ene side *faserne*, dvs. de processer, der blev introduceret i diagrammet over undersøgelsens bestanddele og processer, og på den anden side *cyklusserne*, dvs. de konsekutive gennemløb, som fremgår af figur 1.10.

Hvis vi kun havde faserne, ville der være tale om Hegels slette uendelighed, men hvis vi helt manglede faserne, ville vi bevæge os uden at anvende de tidligere erfaringer.

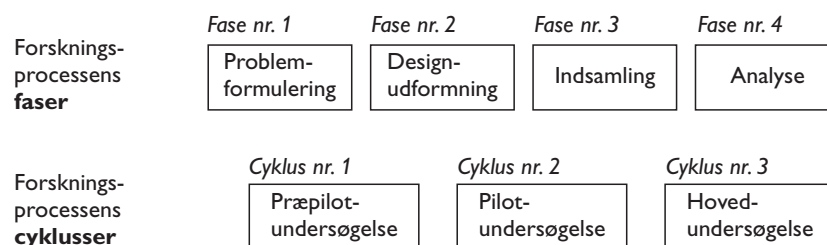
I figur 1.9's forskningsspiral er der angivet tre cyklusser, præpilot-, pilot- og hovedundersøgelse. Her er præpilotcyklussen den meget anbefalelsesværdige allerførste lille undersøgelse, som foretages, når man skal til at arbejde med et helt nyt genstandsfelt. Ved *præpilotundersøgelsen* har man måske slet ikke nogen af faserne afklaret, selv problemstillingen kan være temmelig uklar. En præpilotundersøgelse kan altså være en endnu ikke på forhånd fastlagt feltundersøgelse, hvor man får sit første møde med genstandsfeltet. Resultatet af en vellykket præpilotundersøgelse er da, at man nu har en problemstilling. Med denne kan man gå videre til *pilotfasen*, hvor man operationaliserer problemstillingen til et bestemt undersøgelsesdesign, der afprøves i mindre skala end den påtænkte *hovedundersøgelse*. Hvis man planlægger en interviewundersøgelse med 100 personer, bruger man i pilotundersøgelsen måske kun fem. Endelig kan man så efter fastlæggelse og afprøvning i pilotundersøgelsen gå over til en mere omfattende og forpligtende hovedundersøgelse.

Fordelen ved denne tredobbelte gentagelse af cyklusbevægelsen er:

1. at man får præciseret målsætning og metode inden den egentlige undersøgelse
2. at man får rettet op på fejl og mangler, som kunne invalidere hovedundersøgelsen
3. at man får udviklet den rutine som forsøgsleder, der forudsættes i videnskabelige undersøgelser.



Figur 1.9. Forskningsspiralen.



Figur 1.10. Forskningsprocessers faser og cyklusser.

Undersøgelse af gyldighed og relevans som metodologiens arbejdsopgave

Ud fra denne opdeling i faser og cyklusser kan vi opstille en samlet oversigt over metodologiens opgaver:

Undersøgelsens afgørende valg	Undersøgelses-mål	Undersøgelsesmetode		
Undersøgelsens fokuspunkter	Undersøgelsens relevans	Undersøgelsens gyldighed		
Undersøgelses-fase	Problem-formulering	Design-udformning	Indsamling	Analyse
Undersøgelses-cyklus	Prepilot	Pilot	Hoved-undersøgelse	Hoved-undersøgelse

Figur 1.11. Oversigt over metodologiens arbejdsopgaver.

Når vi ser på den første undersøgelsesfase, problemformuleringen, i præpilotfasen, giver det endnu ikke mening at tale om gyldighed. Hvis man vælger at undersøge etnisk integration hos skolebørn, kommer spørgsmålet om gyldighed først ind under designudformning, hvor vi kan overveje, om det er bedst at bruge feltobservation, gruppeinterview eller enkeltinterview. Derimod er der til problemstillingen knyttet et mindst lige så afgørende spørgsmål, nemlig spørgsmålet om *relevans*.

Her kan vi bruge et umiddelbart, men også tvivlsomt relevanskriterium: nemlig om der er tilstrækkelig mange fagfæller, der synes at problemstillingen er relevant. Et mere relevant relevanskriterium præciserer imidlertid spørgsmålet ved at definere det som en relation mellem problemstilling og implikations- eller relevans-område, det vil sige et genstandsområde, som det har implikationer for. Vi kan sige, at en problemstilling er relevant i forhold til et bestemt genstandsområde, det være sig et snævert eller bredt, prosaisk eller sublimt. Hvis vores genstandsfelt er mænds tanker under morgenbarberingen, er problemstillingen både snæver og prosaisk.

Hvis vi på den anden side ser på Milgrams førømtalte forsøg, er relevansområdet her meget omfattende og sublimt i den forstand, at det dels vedrører et alment menneskeligt forhold, dels et moralsk afgørende spørgsmål. Det er da også karakteristisk, at det i den udbredte og ofte hidsige debat, som rejste sig efter forsøgene, var gyldigheden og ikke relevansen, der blev anfægtet. Netop lidenskaberne i denne debat demonstrerer jo undersøgelsens relevans.

Gyldighedsbegrebet

Vi skal nu forsøge at definere, hvad gyldighed egentlig er. Gyldighed er et begreb, der dækker over tre beslægtede, men dog forskellige betydninger. For det første har det en betydning i logikken, hvor det angives, at en bestemt logisk slutningsform er (logisk) holdbar.

Gyldighed 1 (i logik)

Gyldighed af slutningsform

Slutningsformen er logisk holdbar

For det andet har det en betydning inden for videnskabsteorien, hvor det er den påstand om en teori, at den er sand inden for sit gyldighedsområde, det vil her sige inden for et nærmere specificeret genstandsområde. Vi kan beskrive en teori som en række påstande om en klasse af entiteter, som udgør dens *gyldighedsområde*. Gyldighed er altså den metapåstand, at teoriens påstande er gyldige udsagn om dens gyldighedsområde.

Gyldighed 2 (i metateori)

Gyldighed af teori

Teoriens påstande er gyldige udsagn om dens gyldighedsområde

Den tredje anvendelse af gyldighedsbegrebet ligger inden for metodologien, hvor gyldighed betyder, at en given undersøgelsesmetode er sådan indrettet, at den (korrekt udført) giver et retvisende billede af undersøgelsesgenstanden.

Gyldighed 3 (i metodologi)

Gyldighed af undersøgelsesmetode (i metodologi)

Undersøgelsesmetoden er så godt tilpasset sin undersøgelsesgenstand, at den giver et retvisende billede af denne.

Relevans og gyldighed

I det følgende skal vi hovedsagelig beskæftige os med metateoretisk og metodologisk gyldighed.

Gyldighed i metateorien kan altså karakteriseres som et metaudsagn om relationen mellem et udsagn og den entitet, udsagnet omhandler. For eksempel:

Gyldighed (småbarn NN, utryk tilknytning) = sand

– betyder, at småbarn NN faktisk er karakteriseret ved utryk tilknytning

Gyldighed i metodologisk sammenhæng angår derimod forholdet mellem en undersøgelsesmetode og dens undersøgelsesgenstand. For eksempel:

Gyldighed (Fremmedsituationen eller Strange Situation-proceduren), småbørns tilknytningsmønster) = sand

– betyder, at småbørns tilknytningsmønstre kan afdækkes på korrekt måde ved hjælp af Mary Ainsworths procedure Fremmedsituationen.

Begrebet relevans angår derimod relationen mellem to udsagn i den forstand, at udsagn 1 er af betydning for afgørelsen af, om udsagn 2 er sandt.

Relevans-relationen

Relevans (udsagn 1, udsagn 2) betyder altså:

Hvis udsagn 1 er sandt, har det betydning for vores afgørelse af, om udsagn 2 også er det.

Når vi for eksempel ser på spørgsmålet om genfundne erindringer, så handler det metodologiske *gyldighedsproblem* om, hvorvidt spørgsmålet om sådanne fænomener kan dokumenteres eller ej.

Lad os nu sige, at vi har ti tilfælde af genfundne erindringer om seksuelle overgreb i barndommen – erindringer, der først er dukket op i forbindelse med en psykoterapeutisk behandling. Det metodologiske gyldigheds-spørgsmål er da: Kan der etableres en metode til afgørelse af, om de genfundne erindringer er baseret på faktiske begivenheder af denne art?

Lad os forestille os, at det har været muligt at afgøre i halvdelen af disse cases. Denne empiri vil da ikke i sig selv etablere noget gyldighedsbevis, men den vil være relevant. Det vil sige, at de to udsagn:

- 1: Der foreligger gynækologiske indicier på seksuelt overgreb
og
- 2: Der foreligger en sand genfundne erindring om seksuelt overgreb i barndommen

er forbundet på den måde, at udsagn 1 vil indgå i proceduren for afgørelse

af gyldigheden af udsagn 2. Og gyldigheden af udsagn 2 vil være fastlåst, hvis udsagn 1 kan understøttes med dokumentation af to yderligere sagforhold:

3: Overgrebet har fundet sted i barndommen

og

4: Erindringen om overgrebet har været fortrængt

Vi kan nu se mere overordnet på gyldigheds- og relevansspørgsmålet ved empiriske undersøgelser.

Gyldigheden af en empirisk undersøgelse vedrører spørgsmålet, om hvorvidt undersøgelsens metode kan etablere empiri, der kan bidrage til at besvare problemstillingen. Hvis denne er en hypotese, vil gyldigheden altså betyde, at *undersøgelsesmetoden* sikrer resultaterne udsagnskraft til at fortælle, om hypotesen er sand eller falsk (eller i hvert fald til understøtte et udsagn herom).

Relevansen af undersøgelsen handler om *undersøgelsesresultaterne*, enten som faktisk foreliggende eller forventede. At undersøgelsesresultaterne er relevante vil sige, at de *har* eller *kan få betydning* for afgørelsen om undersøgelsens problemstilling.

Vi kan udvide relevansbegrebet ved at tale om, at et udsagn er relevant for et andet, når det snarere vedrører forståelsen af et fænomen end afgørelsen af teoretiske udsagns gyldighed.

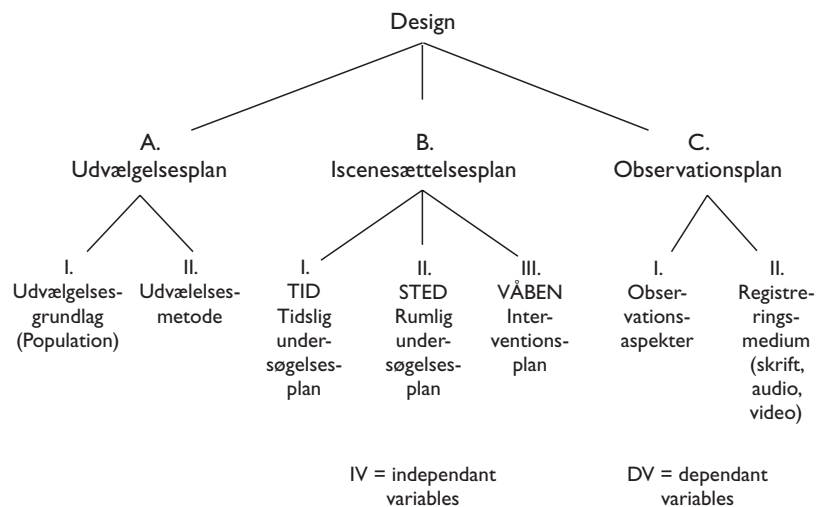
Når vi siger, at tilknytningsteorien er *gyldig* inden for udviklingspsykologien, betyder det, at dens påstande er sande udsagn om børns udvikling. Siger vi derimod, at tilknytningsteorien er *relevant* for udviklingspsykologien, giver vi den en mindre vidtgående karakteristik, som indebærer at teorien bidrager til forståelse af børns udvikling.

Med udsagnet, at Ainthworths Fremmedsituation-procedure er en *gyldig* metode til vurdering af udviklingstruede børn, hævder vi, at denne procedure giver et sandt billede af disse børn. Når vi derimod siger, at udfaldet af Fremmedsituation-proceduren er *relevant* for vurderingen af udviklingstruede børn, mener vi, at den kan bidrage til vores forståelse af sådanne børn.

Empirisk design

En undersøgelsesmetode fastlægges som nævnt gennem sit design, sådan som det er illustreret i figur 1.12.

Designet er en betegnelse for den samlede undersøgelsesplan. Det indeholder tre forskellige underpunkter: udvælgelse, iscenesættelse og observation.



Figur 1.12. Empirisk design.

- *Udvælgelsesplanen* består i en specifikation af *udvælgelsens grundlag* (Hvorfra skal vi vælge vores genstande?) og *udvælgelsens metode* (Hvordan skal vi vælge nogle bestemte genstande blandt udvælgelsesgrundlaget?).
- *Iscenesættelsesplanen* handler om alt det, vi udsætter vores personer for med henblik på at gennemføre undersøgelsen, og den er især betydningsfuld ved eksperimentelle undersøgelser. Iscenesættelsesplanen består med et udtryk fra aftalegrundlaget for dueller i valget af *tid*, *sted* og *våben*. Her defineres vores *uafhængige variable* (eng. *independant variables*, IV).
- *Observationsplanen* handler om, at det, vi agter at observere, fx vores forsøgspersoners reaktioner, adfærd, handlemåde. Her bestemmes vores *afhængige variable* (eng. *dependant variables*, DV).

Hvis vi ser på Milgrams forsøg, bestod *udvælgelsen* i at avertere efter forsøgspersoner i en avisannonce med tilbud om en betaling på 4.50\$ for en times arbejde, hvilket må have været et pænt beløb i begyndelsen af 1960'erne. Udvælgelsesgrundlaget var således læsere af den pågældende avis, og metoden var den noget problematiske, der slet ikke er styret af en forsøgsleder, men bestemt af tilfældigheder.

Men hensyn til *iscenesættelsen* var den *tidslige struktur* et enkeltforsøg af ca. en times varighed. *Stedet* var Milgrams laboratorium udstyret, som det fremgår af figur 1.5. *Våben* er det, forsøgspersonen blev udsat for – helt konkret Milgrams letvægtsudgave af den elektriske stol, mere generelt hele den situation med falsk instruktion, falske responser fra den falske „elev“ og den eskalerende pression fra „forsøgsleder“ over for „lærer“.

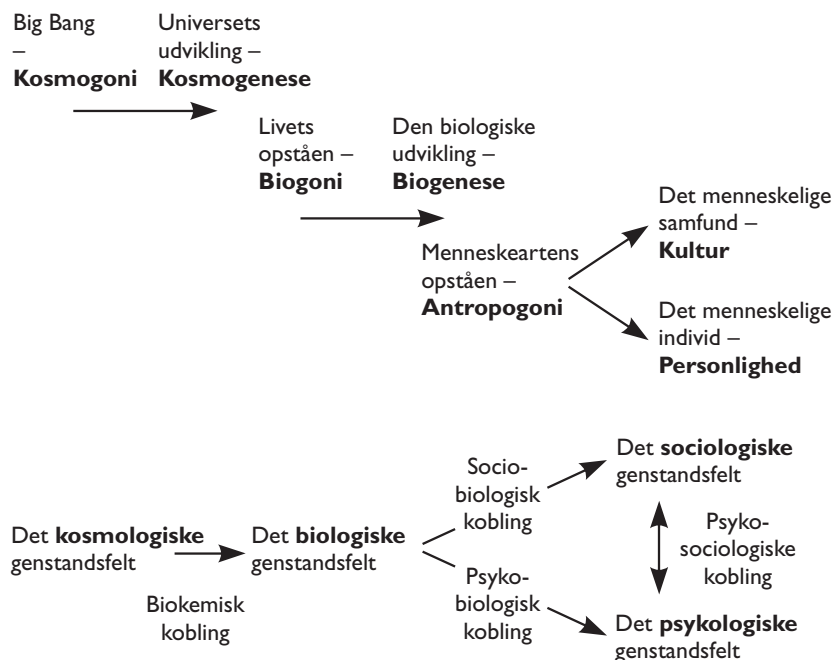
Observationsplanen var en registrering af den egentlige forsøgspersons („lærerens“) reaktioner på de forskellige skalatrin af tilsyneladende spænding, herunder om vedkommende stoppede og i givet fald hvornår, samt et efterfølgende interview, der bl.a. skulle vise, om forsøgspersonen troede på forsøgets virkelighedskarakter.

Jeg vil i det følgende forsøge at placere en række metoder, som anvendes i psykologisk forskning i relation til de teorifelter, hvor de er opstået. Og først vil jeg udbygge opstillingen af hovedgenstandsfelter med en beskrivelse af deres opståen og udvikling, og med en forklaring af de relationer, der er mellem nabofelterne.

Genstandsfelterne for psykologi og dens nabovidenskaber

I forlængelse af afsnittet om psykologiens placering i videnskabernes familie, skal vi se nærmere på denne kontekst med henblik på at lokalisere oprindelsesstederne for psykologiens undersøgelsesmetoder. I figur 1.13 er denne indplacering fremstillet i diagramform.

Øverst til venstre har vi placeret det kosmologiske genstandsfelt, dvs. de fænomener og genstande, som udgør alt det i universet, der ligger forud og til grund for dannelsen af levende organismer. *Kosmogonien* er selve universet opståen, det såkaldte *big bang*, og det efterfølgende forløb gennem en periode, som astronomien sætter til ca. 10 milliarder år, udgør



Figur 1.13. De empiriske genstandsfelters opståen, udvikling og relationer.

kosmogogenesen. Studiet af hele det genstandsfelt kaldes her kosmologien, og det omfatter altså astronomi, fysik og kemi.

På jorden er der opstået levende organismer ved den initiale proces, som kaldes *biogoni*. (En alternativ teori går på, at de er transporteret hertil via meteoriter). Den efterfølgende evolutionære proces kaldes *biogenesen*. Den videnskab, som studerer disse processer og deres genstande, kaldes *biologi*, hvis område altså er det biologiske genstandsfelt.

Homo sapiens sapiens anslås at være udviklet for ca. 100.000 år siden. Denne opståen af vores art kaldes *antropogonien*, og hele vores arts udvikling efter denne oprindelse kaldes *antropogenesen*.

Med antropogenesen opstår der ikke blot en, men to slags nye genstande. For det første nogle særlige individer udstyret ikke blot med bevidsthed, men selvbevidsthed, dvs. personer. Og for det andet nogle sociale dannelser, der ikke blot er gruppedannelser som hos dyr, men organisationsstrukturer, som ud over deres individuelle medlemmer og disses indbyrdes relationer også består af en række frembringelser af såvel materiel som kognitiv art. Indbegrebet af disse frembringelser er samfundets *kultur*.

Den disciplin som studerer det første af disse to genstandsområder, altså de menneskelige individer, personerne, er *psykologi*. Det andet genstandsområde, de menneskelige samfund, tilhører samfundsvidenskaberne, som under et i denne bog kaldes *sociologi*.

Nederst i diagrammet har vi så en oversigt over de nævnte udviklinger, altså kosmogenese, biogenese og antropogenese samt de opståede genstandsfelter med deres indbyrdes relationer.

Hvor kommer psykologiens metoder fra?

I det næste diagram (figur 1.14) er alle videnskabsområderne anført sammen med de af de undersøgelsesmetoder, som er gået over i psykologien (og en enkelt som psykologi har udviklet selv).

1. Det komplette eksperiment	2. Det randomiserede eksperiment	3. Surveymetode	
	7. Casemetode	6. Interviewmetode	
		5. Feltundersøgelse (antropologi)	4. Psykometri
Det kosmologiske teorifelt	Det biologiske teorifelt	Det sociologiske teorifelt	Det psykologiske teorifelt

Figur 1.14. Oprindelsen af psykologiske undersøgelsesmetoder.

Det fremgår, at psykologi i metodologisk henseende nærmest kan karakteriseres som videnskabernes svar på lommetyvenes konge. Af syv forskellige typer af undersøgelsesmetoder har psykologi hugget de seks fra andre videnskaber; kun en enkelt, psykometrien, er hjemmelavet.

Type 1: Det klassiske (komplette) eksperiment

Fra det kosmologiske teoriområde har psykologi taget og videreudviklet det naturvidenskabelige eksperiment, der her er betegnet som det *klassiske*

eksperiment. Vi vil i kapitel 2 behandle denne hovedtype af undersøgelsesmetode mere indgående og skal her blot give en overordnet karakteristik.

I det klassiske eksperiment tilsigtes en fuldstændig kausal reduktion af et bestemt fænomen, sådan at det, som ellers synes specifikt psykologisk – den undersøgte persons individuelle karakteristika og subjektivitet – almindeligvis søges elimineret.

Type 2: Det randomiserede eksperiment

Det randomiserede eksperiment, som ligeledes gennemgår mere detaljeret i kapitel 2, er overført til psykologi fra det, der i lægevidenskaben hedder det *kontrollerede kliniske forsøg*. Her undersøges virkningen af en bestemt behandling gennem en procedure, som i hovedtræk er defineret på følgende måde:

A. *Inklusionskriterier* (definitionen af en bestemt undersøgelsespopulation gennem inklusionskriterier)

B. *Udvælgelse* (tilfældig udvælgelse af et antal forsøgspersoner/patienter)

C. *Randomisering* (opdeling i en forsøgs- og en kontrolgruppe ved lodtrækning)

D. *Selve eksperimentet*

- a. forsøgsgruppen udsættes for den behandlingsmetode, der skal afprøves
- b. kontrolgruppen unddrages denne behandlingsmetode

E. *Blinding* (begge eksperimentets grupper, og så vidt muligt også behandlerne, er blindede, dvs. uvidende om, hvilke forsøgspersoner der modtager hvilken behandling – og hvilken gruppe de tilhører)

Både det klassiske og det randomiserede eksperiment har til formål at gennemføre en kausal reduktion, dvs. gennem manipulation af en eller flere kausale variable og observation af en eller flere afhængige variable at påvise en kausalsammenhæng mellem førstnævnte og sidstnævnte.

Det skal her tilføjes, at en særlig form for eksperiment, hvad enten det er i den klassiske eller den randomiserede form, er *demonstrationseksperi-*

mentet, der i modsætning til de egentlige, eller kausale, eksperimenter, har den funktion at påpege eksistensen og karakteren af et bestemt psykologisk fænomen, uden at sigte på en kausal reduktion.

Type 3: Survey (Spørgeskemaundersøgelsen)

En survey er en masseundersøgelse, der består i udsendelse eller udlevering af et spørgeskema til en række personer og en efterfølgende statistisk bearbejdning af de indsamlede data. Surveyen er således basalt en samfundsvidenskabelig metode, idet sigtet typisk er at få oplysninger om en bestemt befolkningsgruppes (evt. flere befolkningsgruppers) levevilkår, adfærd og synspunkter. Dette sociologiske præg er fastholdt i de fleste psykologiske survey-undersøgelser, der dog i højere grad inddrager psykologiske fænomener og teorier i problemstilling og analyseform.

Survey-undersøgelsen som forskningsmetode gennemgås i kapitel 3.

Type 4: Psykometri

Som nævnt er den psykometriske undersøgelsesmetode den eneste, der ikke har rødder i andre videnskaber, men er udviklet inden for psykologien selv. Umiddelbart ligner den psykometriske metode survey-undersøgelsen på den måde, at der anvendes et eller flere skemaer til selvudfyldning. Det psykometriske skema har imidlertid et andet sigte end spørgeskemaet, idet undersøgelsesmålet ikke er statistisk at abstrahere fra det enkelte individ, men tværtimod at karakterisere nogle psykologiske forhold ved den enkelte person. Survey-metoden skal abstrahere den psykologisk genstand, altså personen, væk for at finde nogle overordnede træk ved den undersøgte gruppe, mens omvendt psykometrien abstraherer de sociologiske rammebetingelser væk for at finde frem til den enkelte persons karakteristika.

Den psykometriske metode gennemgås i kapitel 4.

Type 5: Feltundersøgelsen

Feltundersøgelsen har sin oprindelse i zoologien og antropologien, hvor man under ekspeditioner foretog undersøgelsen i felten af henholdsvis dyrearter og fremmede folkeslag. Feltundersøgelsen udgør dermed et modstykke til eksperimentet. Eksperimentet forsøger at aftvinge naturen dens hemmeligheder ved at lægge den på pinebænken, så den underkastes

eksperimentets tvang, mens feltundersøgeren ydmygt underordner sig de iagttagelsesbetingelser, som foreligger umiddelbart hos den dyreart eller det folkeslag, der skal studeres. Udgangspunktet for den psykologiske feltundersøgelse er således henholdsvis den etologiske og den antropologiske feltundersøgelse.

Type 6: Interviewundersøgelsen

Interviewet udvikles som endnu en uformel observationsmetode inden for samfundsvidenskab og humaniora, bl.a. af historikere og antropologer. En mere standardiseret form antager interviewmetoden, da den bliver industrialiseret af marketing- og opinionsundersøgelsesfirmaer. Når interviewmetoden gennem de sidste årtier for alvor har udviklet sig som en egentlig psykologisk forskningsmetode, hænger det sammen med den massive overvægt, som de kvantitative metoder længe har haft i samfundsvidenskaberne og psykologien.

Type 7: Caseundersøgelsen

Forlægget for den psykologiske undersøgelsesmetode, der hedder casemetoden, er lægevidenskabens sygehistorie, hvor en lægelig behandler fremlægger sine erfaringer fra et forløb med en eller flere patienter, der har været i behandling for en bestemt sygdom. Den psykologiske caseundersøgelse har dermed også sine rødder i den anvendte psykologi, snarere end i grundforsknings-siden af faget psykologi. Som forskningsmetode ligger den i udvekslingen mellem professionens praktiske erfaringsdannelse og den videnskabeliggørelse heraf, som finder sted i grundforskningen.

De tre sidstnævnte undersøgelsesmetoder gennemgås ikke nærmere i denne bog, men vil være hovedtemaerne i en planlagt opfølgende bog om psykologiens kvalitative metoder.

Kvalitativ og kvantitativ metode

I figur 1.14 er der fire metoder, som de fleste vil karakterisere som helt eller overvejende kvantitative:

1. Det komplette eksperiment
2. Det randomiserede eksperiment
3. Surveymetoden (spørgeskema og statistisk opgørelse)
4. Psykometriske (testpsykologiske) metode

Og tre der kan karakteriseres som helt eller overvejende kvalitative:

5. Feltundersøgelse
6. Interviewmetode
7. Casemetode

Hvad er nu baggrunden for en sådan dikotom klassifikation, og hvordan skal den egentlig forstås?

På sin vis er psykologien født som en metodologisk bastard, der fra sin begyndelse som empirisk videnskab i midten af 1800-tallet, har kombineret en naturvidenskabelig stræben efter eksakt måling af objektive størrelser med en kvalitativ undersøgelse af oplevelsesmæssige fænomener. I psykofysikken er den objektive størrelse en fysisk *stimulus*, mens det oplevelsesmæssige fænomen observeres gennem såkaldt *introspektion*, hvorved det angives, hvilken slags kvalitet (af latin *quale*) der fremtræder for forsøgspersonen. Denne angivelse er da *responsen*.

Med denne kombinationsstrategi foretager psykologien en særegen, næsten promiskuøs sammenblanding af de to slags størrelser, der siden naturvidenskabens begyndelse med Galilei i 1600-tallet har været betragtet som uforenelige, nemlig de primære og sekundære kvaliteter. Galilei havde den overbevisning, at videnskab pr. definition må afgrænses til måling af de for videnskabens opfattelse.

Galileis grænsedragning blev på en mærkelig måde respekteret af psykologien i den videre fremfærd efter dens psykofysiske start. Såvel den kognitive psykologi som den psykometriske testpsykologi forsøger enten helt at undgå eller at omgå introspektion, og i det hele taget bliver det kvalitative aspekt helt underordnet en stræben efter at kvantificere psykologien.

Denne kvantitative dominans er ikke uden opposition. Husserl (Zahavi, 2001) er meget skarp i sin kritik af den *naturalistiske psykologi*, og Dilthey (1977) foreslår, at man skal opdele psykologien i en *forklarende* og en *forstående* disciplin, men hovedtendensen i psykologisk forskningsmetodologi er op til omkring 1970 massivt kvantitativt orienteret.

Fra slutningen af 1960'erne begyndte en offensiv mod den såkaldte positivistiske tendens i såvel psykologien som samfundsvidenskaberne. Med Adorno-*eleven* Habermas som frontfigur indledtes i Tyskland fra midten af 1960'erne den såkaldte positivisme-debat (Habermas, 1965), hvor der blev sat spørgsmålstegn ved relevansen og gyldigheden af de kvantitative metoder, der nærmest blev kasseret som ubrugelige til kritisk samfundsforskning og til dybtgående psykologisk analyse. I bogen *Erkenntnis und Interesse* (1968) definerede han ligefrem den tankeform, som den positivistiske videnskabsteori stod for, som en særlig ideologi, der var indsnævrende og undertrykkende. Hermed var der åbnet op for en metodologisk religionskrig mellem strammere (tilhængere af de kvantitative metoder) og slappere: Først udgjorde marxister stødtropper i forlængelse af Habermas' ideologikritik, og senere angreb feminister den fremherskende metode som et udtryk for patriarkalsk ideologi (Keohane et al., 1981).

Denne kvalitative-kvantitative debat var i begyndelsen præget af en vis forvirring om, hvad der var uenighedens kernepunkt. Var det målsætningen, der skilte, sådan at den kvantitative metode var generaliserende, nomotetisk, mens den kvalitative var individrettet, idiografisk? Eller var det fremgangsmåden, der udgjorde forskellen, sådan at den kvantitative metode efterstræbte netop talmæssigt definerede data, mens den kvalitative fokuserede på fænomener af kvalitativ art?

Ud over denne usikkerhed om, hvad man overhovedet skulle forstå ved en kvalitativ metode, var tilhængerne af mindre kvantificerende metodologi hæmmet af, at de kvantitative metoder over en længere periode var blevet præciseret og perfektioneret, hvorimod den nærmere angivelse af fremgangsmåden ved kvalitative metoder var overordentlig uklar. Fra midten af 1980'erne bliver der udgivet mindre polemisk og mere ren metodologisk litteratur om kvalitative metoder.⁸

I dag er vi på mange måder hinsides den skyttegravskrig, der raserede fra 1960'erne til 1980'erne. Selv traditionalisterne i den kvantitative lejr har fået øjnene op for, at der kan laves fremragende kvalitative undersøgelser, og at man i samme undersøgelse kan kombinere kvantitative og kvalitative metoder. Samtidig er moderne udøvere af kvalitative metoder også blevet mere pluralistiske. I Center for Kvalitativ Metodeudvikling, der blev oprettet for en halv snes år siden af Steinar Kvale ved Institut for Psykologi i Århus, har man således argumenteret for den såkaldte triangulering, hvor man ved at kombinere kvantitative og kvalitative metoder dels får en bedre belysning af det undersøgte fænomen, dels får de to slags

metoder til at kaste lys over hinanden (Kvale, 1997, s. 238). Samtidig har man her fremhævet, at de videnskabelighedskriterier, der er udviklet i forbindelse med de kvantitative metoder, fx pålidelighed (reliabilitet) og gyldighed (validitet) bør overføres til de kvalitative metoder (Kvale, 1997, s. 225 ff).

Serialitet og kontekstualitet

Jeg skal nu forsøge at opstille en overordnet model for de grundlæggende forskelle mellem kvantitativ og kvalitativ metode. Til det formål har jeg udviklet begrebet *serialitet*, som sættes over for *kontekstualitet*. Inspirationen stammer fra begrebet *serie* i Sartres *Kritik af den dialektiske fornuft* (Sartre, 1972), hvor begrebet *serie* står i modsætning til *gruppe*. En gruppe er et kollektiv, der handler på en samordnet måde, mens en serie er et antal individer, som handler parallelt, men uafhængigt af hinanden og ganske ensartet, afindividualiseret, fordi de alle er bestemt af nogle snævre institutionelle handlerammer. Som eksempel peger Sartre på de mennesker, der står i en kø og venter på en bus.

Det som karakteriserer personerne i en serie (se figur 1.15), er altså tre forhold:



Figur 1.15. En serie.

1. Personlig kontekstløshed (i serien er personen *løsrevet fra sin naturlige kontekst*, fx gruppetilhørsforhold)
2. Ensartethed (i serien er *personens individualitet ophævet*, personen er blot en position i en kø)
3. Atomiserethed (i serien er individet reduceret til en simpel egenskab, her det nummer personen har i køen).

I modsætning hertil gælder der for personerne i det, Sartre kalder en gruppe, tre kontrasterende forhold:

1. Kontekstualitet (personen er indlejret i en naturlig kontekst, fx gruppetilhørsforhold)
2. Unikhed (personens har en særegen individualitet)
3. Udelelighed (individualitet er en helhed, der ikke kan defineres alene ved enkeltegenskaber)



Figur 1.16. En gruppe.

I figur 1.16 ses netop en gruppe (her en arbejdsgruppe) med fire forskellige personer, som vi delvis kan beskrive ud fra enkeltegenskaber, men kun ved at forråde det helhedspræg, der kendetegner individualiteten.

Denne forskel mellem en person i en serie og i en gruppe er således defineret *ontologisk*, men den kan overføres til en forskel i *undersøgelsesmetode*, hvor serialitet defineres som et perspektiv, som anlægges på en person, der opfattes som en del af en serie, mens kontekstualitet er det perspektiv, der anvendes, når en person betragtes som en del af en gruppe. Som vi skal se, har denne dobbelthed i begreberne den konsekvens, at gyldigheden af en seriel undersøgelsesmetode kan kritiseres, når det er gruppeindlejrede fænomener, der studeres, ligesom en kontekstualiseret undersøgelsesmetodes pålidelighed (reliabilitet) kan anfægtes, hvis det er serielle fænomener, som er den epistemiske genstand.

Før en nærmere gennemgang af henholdsvis den serielle (kvantitative) og den kontekstuelle (kvalitative) metode, skal vi foretage en bestemmelse af disse begreber. Jeg har allerede fremhævet tre aspekter, der adskiller metoderne, i hvert fald i deres psykologiske brug:

1. Kontekstualitet versus dekontekstualitet
2. Specificitet versus generalitet
3. Totalitet versus atomiserethed

1. Aspekt: Kontekstualitet versus dekontekstualitet

De kvalitative metoder er generelt orienteret mod, at den eller de personer, der indgår som undersøgelsesgenstande, bliver beskrevet og begrebet i deres kontekstualitet. Det betyder, at de ikke undersøges som isolerede individer, men betragtes i deres livskontekst, såvel socialt som livshistorisk. Dette giver mulighed for at konkretisere en række forhold ved de livsomstændigheder, som knytter sig til den eller de undersøgte personer. Til gengæld kan det være vanskeligt at abstrahere fra disse særlige livsforhold.

De kvantitative metoder, fx eksperiment eller survey, er i højere grad orienteret mod at isolere forsøgspersonerne eller informanterne fra den kontekst, de indgår i, bortset fra den kategoriale information, der er knyttet til gruppeudvælgelseskriterier eller oplyste baggrundsoplysninger. Resultatet er, at der sker en serialisering af individerne, så de *ikke* bliver anskuet som særegne personer i hver deres specifikke kontekst, men som medlemmer af en serie, hvad enten det nu er en forsøgsserie (eksperimentelt) eller en sociologisk serie (survey). Denne dekontekstuering er dog ikke nødvendigvis en metodologisk svaghed. Ved et velovervejet kvantitativt design, er der derimod tale om et afgørende træk ved metodens *modus operandi* (funktionsmåde).

2. Aspekt: Specificitet versus generalitet

De kvalitative metoder har deres styrke i den enkelte persons specificitet, altså i det idiografiske. En del af denne specificitet kan for så vidt ses som et udtryk for den specifikke individuelle kontekst, men faktisk viser en persons særtræk sig ofte at følge personen, mere eller mindre uændret, når konteksten ændres. Det vil sige, at vi kan opfatte denne del af specificiteten som knyttet til individets særlige personlighed. Denne orientering mod det specifikke ved den enkelte person, gør det til gengæld vanskeligt at nå frem til generelle, individafhængige konklusioner på baggrund af en kvalitativ undersøgelse.

De kvantitative metoder er derimod ofte stærke i deres orientering mod generaliteten, altså det nomotetiske, eller i hvert fald det, som er alment for en bestemt kategori af mennesker. Til gengæld er disse metoder svage mht. specificiteten ved de enkelte personer, der indgår i en undersøgelse.

Dette er jo netop karakteristisk for en serialisering, hvor de undersøgte personer anskues som principielt ensartede (i hvert fald i deres særlige kategori), dvs. som elementer i en abstrakt serie.

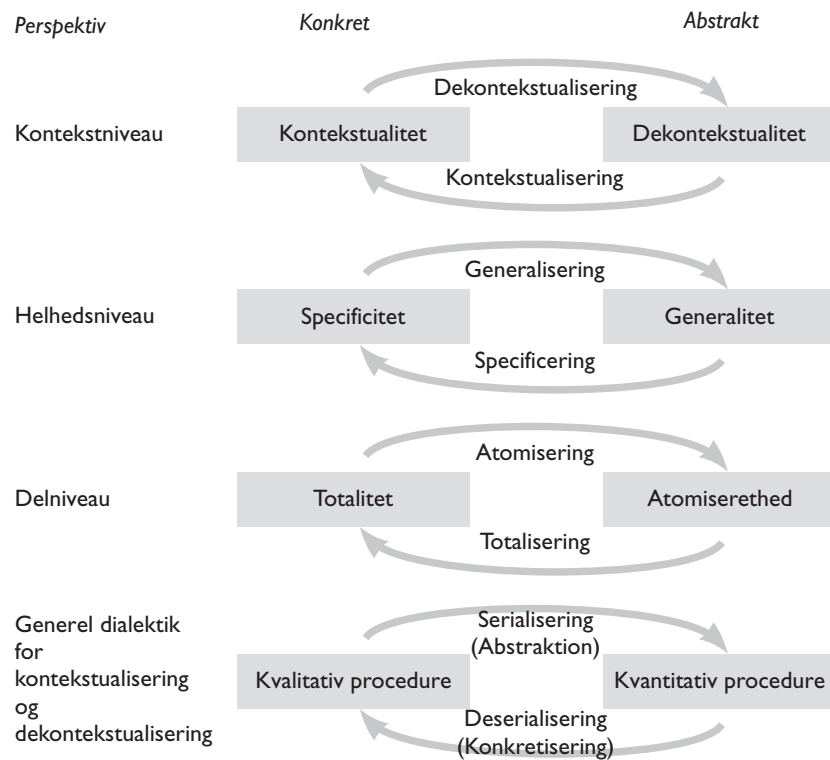
3. Aspekt: Totalitet versus atomiserethed

Ved de kvalitative metoder kan enkeltpersonen betragtes som en helhed, en enhed der er udelelig (på latin *individuus*). Og personen er ikke blot udelelig i ordets oprindelige biologiske forstand, men også som en psykologisk helhed. Dette kommer dels til udtryk i det socialpsykologiske forhold, at andre mennesker betragter en bestemt person som en helhed, ligesom vedkommende også selv gør det. Det sidste forhold er, hvad man i personligheds- og socialpsykologien ofte benævner personens Selv.

Til gengæld vil de kvalitative metoder ofte være svage til at præcisere de enkelte egenskaber ved personen. De kvantitative metoder tenderer omvendt mod at beskrive en person ud fra en liste af adskilte oplysninger vedrørende denne person. Dette er tilfældet ved surveymetoden, såvel som ved den psykometriske metode. Dette gør dem stærke i detaljen, men ofte svage i helhedsbilledet.

De ontologiske gyldighedskriterier for den kvantitative og den kvalitative metode

Som antydnet i indledningen af dette afsnit er det urimeligt at betragte spørgsmålet om valg mellem kvantitativ og kvalitativ metode som et spørgsmål om absolut antropologisk eller videnskabsteoretisk gyldighed. At det med andre ord skulle ligge i enten menneskets eller videnskabens natur, at den ene eller anden metode er den rette. Ideen med de her indførte begreber om serialitet og kontekstualitet er derimod at afvise en sådan ideologiserende metodologi, ud fra den opfattelse, at psykologiske og i øvrigt også andre humanvidenskabelige undersøgelser som epistemisk genstand kan have fænomener af såvel seriel som individuel karakter. Det kan også forekomme, at det ligger inden for en given undersøgelses formål at inddrage såvel fænomener af den ene og den anden type. Det sidste er utvivlsomt det mest almindelige, og det indebærer, at valget af enten en kvantitativ eller kvalitativ metode giver en forskel i fokusering på de undersøgte fænomener. Det gør til en vis grad valget af undersøgelsestype til et spørgsmål om, hvad man ønsker at lægge mest vægt på, den serielle og generelle eller den gruppemæssige og individuelle karakter af de



Figur 1.17. En model for dialektikken mellem abstraktion (serialisering) og konkretisering (deserialisering).

undersøgte fænomener, og det er da igen et spørgsmål om, hvad der mest relevant i forhold til undersøgelsens problemstilling.

For at forstå fremkomsten af først og fremmest den statistiske metode, der dominerer de kvantitative undersøgelsesretninger i psykologi, må man se på de historiske og samfundsmæssige betingelser for statistikens opståen, fremvækst og udbredelse. *Statistik* betyder oprindeligt statskunst og siges at være grundlagt af Machiavelli i renæssancen. I 1600- og 1700-tallet udvikles der inden for denne statskunst en disciplin med navnet *politisk aritmetik*.⁹ Denne diskursive behandling af regeringsanliggender kendetegner samfundsudviklingen i Europa, hvis oprindelige feudale struktur, baseret på organiske (slægtskabs- og traditionsbaserede gruppestrukturer og personlig loyalitet i interpersonelle relationer, bliver afløst af et markedsøkonomisk og statsligt samfundssystem, hvor organisationsstrukturer

og interpersonelle relationer tenderer mod at være upersonlige, bestemt af enten en økonomisk eller en bureaukratisk logik.

I denne udvikling er det karakteristisk, at overgangen fra den organiske, gruppestrukturerede samfundsopbygning til den *serielle* struktur i det moderne samfund, skaber et behov for statistiske oplysninger. Det gælder både den markedsøkonomiske (private), og den statslige (offentlige) sektor. Således udvikles der på den ene side inden for forsikringsbranchen, som var en særlig fremstormende gren af markedsøkonomien, en *aktuarisk* disciplin til risikoberegning, og på den anden side hos statsmagten *statistiske* kontorer, der skal indsamle oplysninger om befolkningsforhold og efterhånden også nationaløkonomiske forhold. I den sidstnævnte sammenhæng har ordet statistik skiftet karakter fra at betegne forhold vedrørende staten i det hele taget til at vedrøre oplysninger af betydning for at kende disse befolkningsmæssige forhold. De tre trin i denne udvikling er altså følgende:

1. Samfundet gennemgår efter middelalderen en udvikling i retning af stigende serialisering.
2. Dette nødvendiggør udviklingen af en markedsøkonomi og en statslig struktur, hvor der for begges vedkommende er behov for at indsamle oplysninger om hhv. de markedsøkonomiske enkeltindivider og de statslige undersåtter, fordi hverken det økonomiske liv eller regeringsmagten kan forlade sig på de tidligere selv bærende organiske organisationsformer, som i stadig stigende grad forvandles til serielle strukturer.
3. Indsamlingen af disse oplysninger af serialiseret karakter kræver en udvikling af metode hertil, hvilket især vil sige spørgeskemaer og statistisk efterbehandling.

Udviklingen forklarer, at sociologien, der er den del af samfundsvidenskaben, som beskæftiger sig med Vestens eget samfundssystem, har satset massivt på den kvantitative metode, mens etnologien eller kulturanthropologien i langt højere grad har foretrukket en kvalitativ metode.

I psykologien findes den kvantitative metode to helt forskellige steder. Dels inden for den kognitivt orienterede eksperimentalsykologi, dels inden for de mere socialpsykologiske områder af psykologien. Dette skyldes, at vi i begge tilfælde har god grund til at abstrahere for den kontekstualiserede og konkret gruppemæssigt forankrede person. I det første tilfælde ser vi på naturbetingede forhold ved individet, som slet ikke,

eller i meget begrænset udstrækning, er bestemt af særlige individuelle eller gruppeforankrede forhold. I det andet tilfælde studerer vi psykologiske fænomener, der netop er samfundsmæssigt forankrede – på niveauer, der ligger over konkrete gruppestrukturer – således at vi kan nøjes med personoplysninger af mere overordnet karakter, knyttet til abstrakte organisationsstrukturer som geografisk region, uddannelsesbaggrund og socioøkonomisk placering. Tilbage bliver da den kvalitative metode, som vi benytter, når vi ønsker at studere en persons særlige individualitet og gruppemæssige kontekstuering.

Det ene af de områder, hvor den kvantitative (serielle) metode synes adækvat, altså den kognitivt orienterede eksperimentalsykologi svarer til det som Bruner og Goodman (1962) kalder for de *autoktone perceptuelle determinanter*. Hertil henregner de egenskaber ved det neurale system, der kommer til udtryk i de basale perceptuelle fænomener. Disse fænomener kan efter forfatterens mening godt udforskes isolerede (abstrakt) i et laboratoriets „mørkerum“.

Det andet område, hvor den kvantitative (serielle) metode har sin gyldighed, er de *serialiserede samfundsmæssigt* bestemte fænomener. Over for disse forskellige serielle fænomener har vi individualitetens og konkretthedsens område, som kræver en kvalitativ metode. Denne dikotomi svarer til neokantianeren Windelbands (1894) opdeling i en nomotetisk (generaliserende) og idiografisk (specificerende) metode.¹⁰

Den kvantitative og den kvalitative metode som modpartens skyggeside

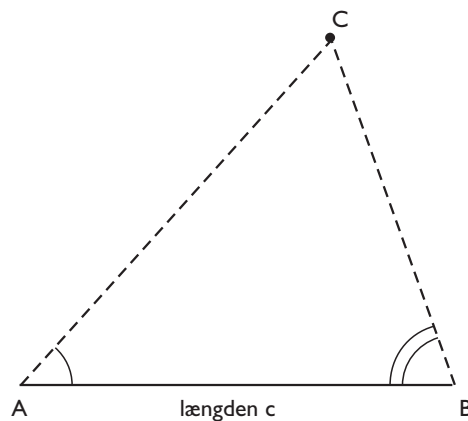
I den debat, som rasede gennem 70'erne og 80'erne om betimeligheden af at anvende henholdsvis kvantitativ og kvalitativ metode overså de fleste fortalere for de to metodeformer ikke blot de ontologiske gyldighedsbetingelser og dermed begrænsningerne ved deres metoder. De to grupper overså også de rent metodologiske vanskeligheder, der er indbygget, og som dermed må inddrages i afgørelsen af, hvilken metode der skal vælges i de tilfælde, hvor den ikke er ontologisk givet i og med undersøgelsens epistemiske genstand.

Dette skyldes, at de to slags metoder har hver sin metodologiske flaskehals. Den kvantitative metode har generaliteten indbygget, og for så vidt udvælgelsesmetoden sikrer repræsentativiteten af den population, som metoden skal belyse, vil en generalisering fra stikprøve til population

alene være begrænset af den statistiske pålidelighed, som er knyttet til stikprøvestørrelse og til usikkerheden af observationsmetoderne. Så det er ikke her flaskehalsen opstår. Den opstår derimod, når man har udledt et generelt undersøgelsesresultat og skal anvende eller tolke dette i en mere specifik og konkret sammenhæng. Lad os sige, at vi i en survey-undersøgelse om trivsel i forskellige beboelsesformer finder ud af, at der er større trivsel i tæt-lav end i åben-lav beboelse. At folk altså trives bedre i rækkehuse end i villaer.¹¹ Så ved vi det, og vi kan endog forudsætte, at vi har taget alle statistisk metodiske former for bias i ed, at der således ikke er tale om falske eller skjulte sammenhænge (se kapitel 3). Men vi ved jo bare ikke rigtig, hvad vores fund egentlig betyder. I så fald kunne man forestille sig, at vi gik i gang med en eksplorativ eller interviewbaseret undersøgelse, hvor vi prøvede at finde ud af, hvordan man lever i henholdsvis et rækkehus og villakvarter. Hvordan det sociale netværk og den sociale aktivitet fx adskiller sig.

Kombination af kvantitative og kvalitative metoder (triangulering)

Da de kvantitative og de kvalitative metoder har dyder og svagheder, som overlapper i meget ringe grad, kan det synes en oplagt idé at kombinere dem. Som nævnt er dette ikke altid ontologisk velbegrundet, men ofte kan der være fornuft i at betragte et fænomen eller et genstandsområde ud fra såvel et serialiseret som et konkret perspektiv, og det vil da være ønskvær-



Figur 1.18. Geometrisk triangulering.

diget at tilstræbe et supplementsforhold gennem en kombination af de to metoder.

En sådan kombination kaldes ofte for triangulering. Udtrykket er hentet fra astronomi og landmåling og betyder egentlig opdeling af et bestemt område i trekanter. Udgangspunktet for den metaforiske transport fra geometriens område til humanvidenskabernes metodologi er den klassiske form for bestemmelse af et område ud fra en allerede kendt base, dvs. et linjestykke, hvis placering og udstrækning er kendt. I nedenstående diagram er således linjestykket AB den base, som er udgangspunktet (se figur 1.18).

Når man ud fra linjestykket AB måler vinklerne i sigtelinjerne fra de to endepunkter A og B, dvs. vinklen A og vinklen B, kan man jo enten ved en geometrisk konstruktion eller ved en særlig trigonometrisk beregning udregne fx længderne af de to sigtelinjer AC og AB, eller C's placering i et koordinatsystem med fx udgangspunkt i A.

Analogien til metodekombination i humanvidenskaber er nu, at man også her har en genstand at bestemme (den epistemiske genstand), og hvad der svarer til de to sigtelinjer i form af to forskellige metoder, typisk en kvantitativ og en kvalitativ.

Argumenterne for triangulering er ofte, at styrken ved de respektive metodetyper kan kombineres, samtidig med at man undgår svaghederne. Dette er imidlertid lettere sagt end gjort, for det er ikke altid, at en sådan kombination kan opnås. Snarere bør man gennemtænke, hvilke undersøgelsesmetoder man har brug for i en bestemt undersøgelse, dvs. i et bestemt forskningsprojekt og på et bestemt tidspunkt. Valget vil afhænge af såvel undersøgelsens problemstilling som af den udviklingsfase, projektet på et givet tidspunkt er inde i. Ud over den simultane kombination af kvalitative og kvantitative metoder er der således følgende sekventielle kombinationsmuligheder:

1. Den kvalitative eksplorative forundersøgelse (præpilotundersøgelse), efterfulgt af den kvantitative hovedundersøgelse (fx eksperiment eller survey eller psykometrisk metode).

Her forsøger man først med en forberedende præpilotundersøgelse at bestemme, hvad det er, der skal observeres og hvordan. Den kvalitative forundersøgelse er da ikke et mål i sig selv, men alene et middel

til at udvikle hovedundersøgelsens design, som da afprøves gennem en pilotfase.

2. Den kvantitative forundersøgelse, typisk en survey-undersøgelse, efterfulgt af en mere dybdegående og kontekstualiserende interviewundersøgelse.

Dette forløb har man fx i tilfælde, hvor man, måske lidt overilet, har foretaget en kvantitativ survey, som giver et større datamateriale, hvor resultaterne enten er noget uklare eller peger på nye problemstillinger ud over dem, der er defineret i undersøgelsens design. Her kan man forsøge at afklare disse forhold gennem en efterfølgende kvalitativ undersøgelse, der så igen kan funderes gennem en ny survey-undersøgelse.

3. Den integrerede undersøgelsesform, der samtidig anvender kvalitativ tekstanalyse og kvantitative (statistiske) analysemetoder, fx gennem kodning af bestemte oplysninger pga. af semistrukturerede interview.

I den integrerede undersøgelsesform har man forberedt en kombineret kvalitativ og kvantitativ del. I to evalueringsundersøgelser¹² har jeg selv anvendt følgende form: Efter en eksplorativ pilotfase designede jeg en semistruktureret interviewundersøgelse, samtidig med at jeg udviklede et kodningssystem, hvor intervieweren foretog en kvantificering af en række statusvariable ud fra interviewudskriften. Ud over at fungere som basis for denne kvantificering blev interviewudskrifterne samtidig udsat for en kvalitativ analyse pga. det kvalitative analysesystem NVIVO.

Vi har i det forrige diskuteret forholdet mellem kvalitativ og kvantitativ undersøgelsesmetode, men hvilke dataformer er karakteristiske for de to slags metoder?

Ud fra den bestemmelse af kvalitativ og kvantitativ fremgangsmåde vil jeg operere med en kvalitativ ikke-operationaliseret, og en kvantitativ operationaliseret dataform, den sidste bestående af tre skalatyper (hvoraf dog den ene underopdeles).

Den kvalitative dataform

Den kvalitative dataform er tekster, der angiver handlinger, adfærd og udtalelser i en form, der evt. kan viderebearbejdes, fx med tekstanalyseprogrammer. At den kvalitative dataform har en almindelig sproglig

udformning betyder dog ikke, at der stilles særlige metodiske krav, som går ud over en dagligdags sproglig fremstilling. Således må mundtlige udtalelser, fx i interview, *transkriberes*, hvorved der i nogen grad sker en tydeliggørelse og eksplicitering af den information, som fandtes i udtalelsernes oprindelige sociale og kommunikative kontekst.

Den kvantitative dataform

Den kvantitative dataform stiller krav om en operationaliseret observationsmåde, hvor registreringen af en observation følger en bestemt på forhånd defineret standard. En sådan standard for dataformen kaldes en skala. Der er her tre forskellige hovedtyper af skalaer, som kan opstilles efter den grad af information, der er indbygget i skalaen.

Kategoriskala (nominalskala)

Ordet skala er det italienske ord for *trappe*, hvilket peger på en metaforisk brug af trappens niveauforskydning. Sådan er skalabegrebet netop anvendt i toneskalaen, hvor de enkelte toner er anbragt efter tonehøjde. Grundbetydningen af skala er altså en størrelsesmæssig dimension.

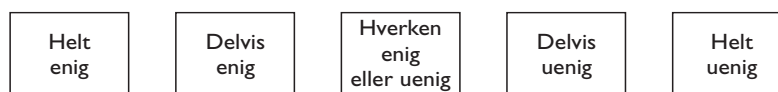
Da psykofysikeren S.S. Stevens i 1940'erne udviklede en teori om psykologiske data, gik han imidlertid ud over denne grundbetydning af ordet skala, idet han som den svageste skalatype definerede, hvad han kaldte nominalskalaen, som i virkeligheden blot er et kategorisystem, hvor de enkelte kategorier er forsynet med et navn. Det, som gør kategorisystemet til en operationaliseret dataform, er de logiske krav, der stilles til forholdet mellem en klasse af observationer og kategorisystemet:

1. entydighed (det må være klart, i hvilken kategori en given observation skal indplaceres)
2. gensidig udelukkelse (en observation må kun placeres i én bestemt kategori) og
3. fuldstændighed (enhver observation må kunne placeres i en af kategorierne).

Det er denne veloperationaliserede kvalitet, som var grunden til, at Stevens indførte begrebet nominalskala, som altså faktisk slet ikke er en skala. Det arketypiske eksempel på en kategoriskala er køn som et kategorisystem med to kategorier.¹³

Ordinalskala

Ordinalskalaen findes i to udgaver, den kategoriske (diskrete) og den kontinuerte:



Figur 1.19. Ordinalskala (Likert-skala).

Den *kategoriske* ordinalskala er egentlig blot en kategoriskala, hvor der er tilføjet en ordningsrelation, sådan at der bliver tale om en „rigtig“ skala. Standardeksemplet er her social klassifikation, som i angelsaksisk terminologi kaldes *socio-economic status*.

En udgave her i landet er Socialforskningsinstituttets sociale klassifikation med fem kategorier fra I til V, hvor I er højeste og V er laveste statuslag, og hvor der i den klassiske udgave var eksplicitte kriterier, baseret på uddannelse og ansættelsesforhold.

Et andet eksempel er den meget anvendte Likert-skala. Ved holdningsundersøgelser bruges ofte en sådan skala med tre eller fem ordnede kategorier.

Den *kontinuerte* ordinalskala adskiller sig fra den kategoriske ved at være en skala uden spring (som fx de intervalskalaer, der omtales i næste afsnit). Den kontinuerte ordinalskala bruges eksempelvis inden for geologien, hvor man har en hårdhedsskala, der ofte anføres som eksempel på denne skalatype. Her er ordningsrelationen for to mineraler bestemt således, at mineral 1 er blødere end mineral 2, hvis mineral 2 kan ridse i mineral 1, men ikke omvendt.

I psykologi bruges undertiden en afart af den kontinuerte ordinalskala, idet man i spørgeskemaer kan støde på spørgsmål, hvor respondenter skal rangordne forskellige emner efter præference. Fx:

Hvad foretrækker jeg at se i tv? (Giv en rangordning af nedenstående programmer):

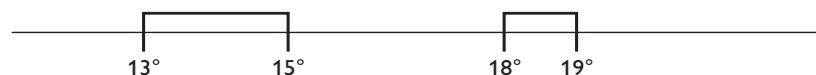
- a. Robinson/ Big Brother
- b. Sport

- c. Krimiserier
- d. Nyheder
- e. Bolig og have

En sådan type spørgsmål forudsætter en underliggende kontinuert præferenceskala: men spørgsmål af denne type må ud fra en metodisk vurdering frarådes, da der ikke er nogen tilgængelige statistiske metoder, der kan anvendes til opgørelse af resultaterne.

Intervalskala

Intervalskalaen definerer Stevens som en ordningsskala med et intervalmål, således at man kan sammenligne ikke bare enkelte værdier, men to intervaller mellem forskellige værdier. Man kan endog sammenligne intervaller med hinanden. Hvis vi således tager Stevens eget eksempel temperatur og holder os til vores hjemlige Celcius-skala, kan vi fx have fire værdier, som vist i figur 1.20.

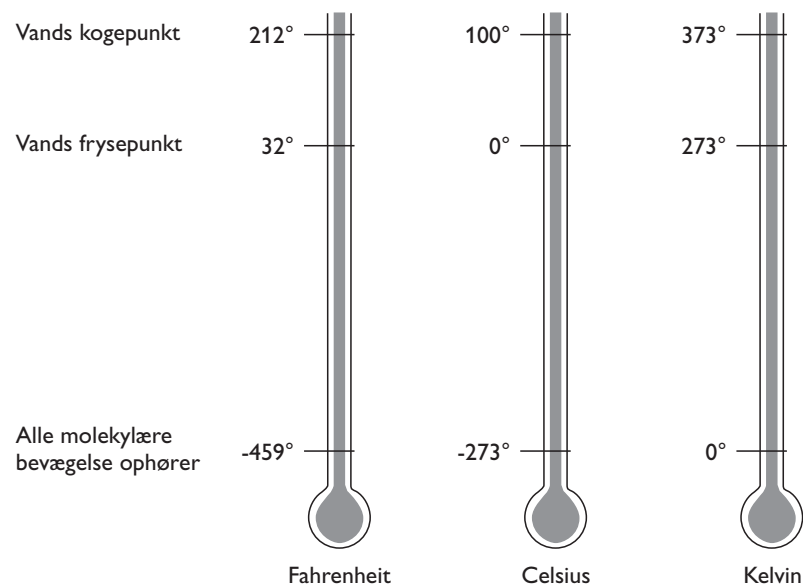


Figur 1.20. Intervalskala.

Her kan vi sige, at det første temperaturinterval er dobbelt så stort som det andet. Da det kun handler om at måle intervaller, er nulpunktet til gengæld en *arbitrært*, dvs. vilkårligt valgt ved intervalskalaen, lige som måleenheden er vilkårlig. Begge dele er jo netop tilfældet ved de tre temperaturskalaer, der benyttes.

Problemet med intervalskalaer i Stevens' definition er, at de har en vis hybridkarakter, deres definition er tilsyneladende videnskabelig og præcis, men faktisk repræsenterer de en videnskabshistorisk forældet position (figur 1.21). De tre praktisk anvendte temperaturskalaer blev udviklet af henholdsvis tyskeren Daniel Fahrenheit i 1715, franskmanden René Antoine Ferchault de Reaumur i 1730 og svenskeren Anders Celsius i 1741. Det var, inden termodynamikken blev udviklet, og alle tre skalaer er derfor begrænset til praktisk anvendelse, idet de alle har et arbitrært nulpunkt.

Fahrenheit satte oprindeligt sit nulpunkt efter frysepunktet for en is-salt-blanding, og reviderede senere sin skala, så vands frysepunkt blev



Figur 1.21. Fahrenheit-, Celsius- og Kelvinskala.

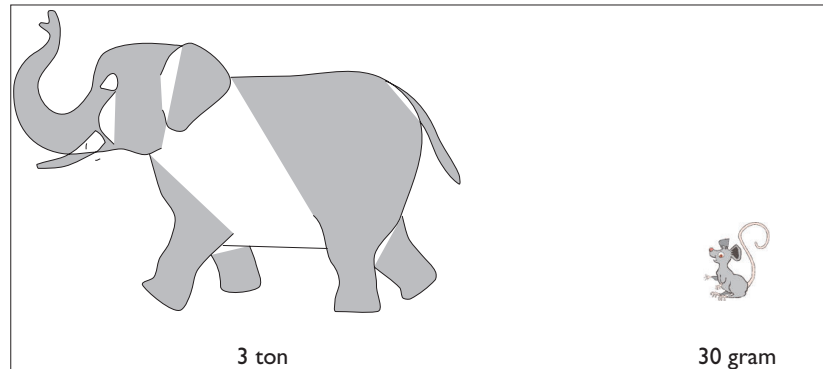
sat til 32 og kogepunktet til 212 grader. Det er denne skala, som stadig er i brug i de angelsaksiske lande.

Reamur og Celsius valgte at sætte nulpunktet efter vands frysepunkt, og placerede så kogepunktet som hhv. 80 og 100 grader. Det synes vi er mere fornuftigt, især i den Celsius-udgave vi selv bruger.

Det er imidlertid helt unødvendigt med et vilkårligt valgt nulpunkt, eftersom der findes et *absolut*. Det er defineret efter englænderen Kelvin og bruger samme måleenhed som Celsius-skalaen. En grad Kelvin er altså det samme som en grad Celsius. Men hvor det absolutte nulpunkt i på Celsius-skalaen er -273 grader, så er vands frysepunkt ifølge Kelvin-skalaen, som jo slet ikke har negative temperaturer, $+273$ grader. Med Kelvin er vi omsider fremme ved en fornuftig udgave af intervalskalaen, nemlig ratioskalaen.

Ratioskala

Ratioskalaen er ifølge Stevens' definition en finere udgave af intervalskalaen, nemlig en skala, hvor nulpunktet er fastsat vilkårligt, men ud fra et videnskabeligt tvingende sagforhold. Ud fra almindelige videnskabelige begreber er det, Stevens kalder en ratioskala, en almindelig *måleskala*. Den



Figur 1.22. To størrelser der sammenlignes på en ratioskala. Elefant og mus.

er karakteriseret ved, at vi har at gøre med en målbar kvantitet, dvs. at vi kan sammenligne to kvantiteter ved at måle, hvor mange gange den ene kvantitet går op i den anden (se figur 1.22).

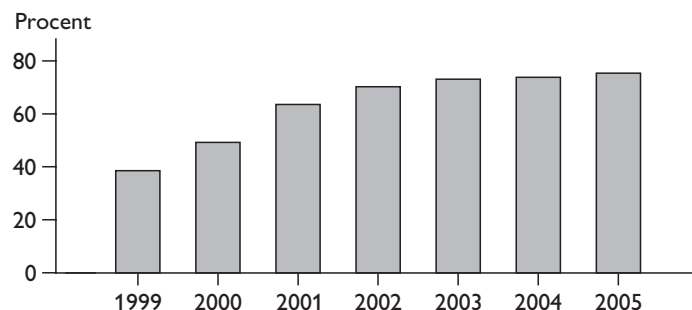
Elefanten og musen udgør netop yderpunkter mht. til størrelse, herunder vægt. Den ene vejer typisk tre ton, den anden 30 gram. Det betyder, at der vægtmæssigt går ca. 100.000 mus på en elefant. Vi kan altså sammenligne kvantiteter, sådan som Stevens siger, men selve måleenheden, hvad enten den nu tilfældigvis er gram, ounce eller ton, gør det muligt at angive, *hvor mange gange den ene er større end den anden*. Der er altså i virkeligheden ikke noget arbitrært nulpunkt. Der er et absolut nulpunkt, nemlig den størrelse, der ikke har nogen vægt, og dette fænomen, vægtløshed er i øvrigt en fysisk realitet.

Den absolutte skala

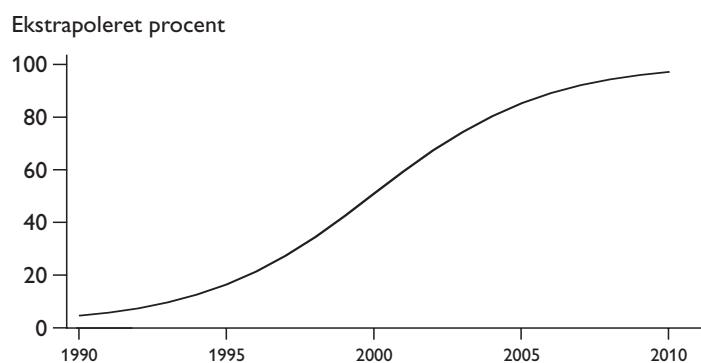
Ifølge Stevens' definition har intervalskalaen to vilkårligt definerede specifikationer, nemlig et arbitrært nulpunkt og en arbitrær enhed. Ratioskalaen har kun den ene af disse, nemlig en arbitrær enhed, mens den absolutte skala slet ikke har nogen vilkårlige specifikationer. Det er derfor den kaldes absolut. Det vigtigste eksempel på en absolut skala, er sandsynligheds- eller hyppighedsskalaen, der går fra 0 til 1 eller fra 0 til 100 procent. Her er der altså dels et absolut nulpunkt, dels en absolut enhed.

I figur 1.23 ser vi udbredelsen af mobiltelefoner i Storbritannien. Her er den lodrette akse en hyppigheds- eller procentskala.

Selv om hyppigheds- og sandsynlighedsskalaerne er intervalskalaer, skal



Figur 1.23. Et eksempel på anvendelse af en hyppighedsskala – udbredelsen af mobiltelefoner i Storbritannien i perioden 1999-2005.



Figur 1.24. Et eksempel på anvendelse af en hyppighedsskala – udbredelsen af mobiltelefoner i Storbritannien i perioden 1990-2010.

man normalt afholde sig fra at sammenligne intervaller. Det skyldes den absolutte skalas begrænsethed.

Figur 1.24 viser en ekstrapoleret kurve, der på basis af talangivelserne i figur 1.23 giver et billede af mobiltelefoners udbredelse i Storbritannien i kurveform.

Det er karakteristisk for den absolutte skala, at vi har såkaldte bund- og loftseffekter. Det vil sige, at den absolutte bund- og topgrænse på henholdsvis 0 og 100 procent gør intervalsammenligninger meningsløse, især i bunden og toppen.

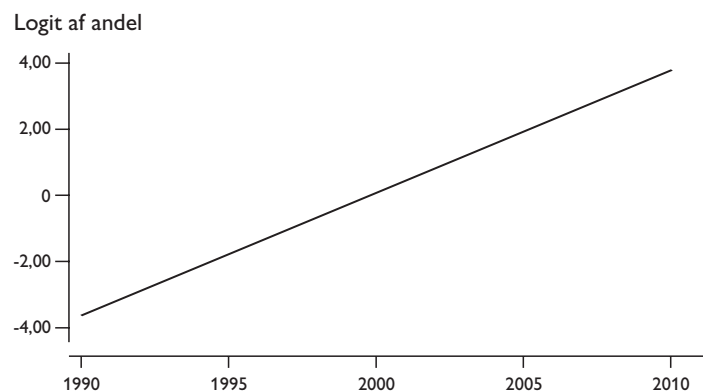
Vi ser, at stigningen i begyndelsen af perioden er meget lavere end i midten af perioden, og så falder den igen i slutningen af perioden. Dette er ikke blot et empirisk forhold, for mens stigningen fra 1998 til 1999 er

Tabel 1.1. Den ekstrapolerede udbredelse af mobiltelefoner i Storbritannien i perioden 1990-2010.

År	Ekstrapoleret procent	År	Ekstrapoleret procent
1990	3	2001	59
1991	4	2002	67
1992	6	2003	74
1993	8	2004	80
1994	11	2005	85
1995	15	2006	89
1996	20	2007	92
1997	26	2008	94
1998	33	2009	96
1999	41	2010	97
2000	50		

8 procent, så må den nødvendigvis være lavere i 2009, hvis procenten her er 96. Når man udvikler modeller for sådanne data, må man derfor foretage en alternativ intervalangivelse, hvor vi ikke udregner differensen, men bruger den såkaldt logistiske omregning, dvs. beregner: $\log(\text{hyppighed}) - \log(1 - \text{hyppighed})$, hvor log er logaritme-funktionen (nærmere betegnet den naturlige logaritme). Se figur 1.25.

Efter omregning er vi sluppet fra bund- og loftseffekter for nu går den lodrette akse fra minus uendelig til plus uendelig.

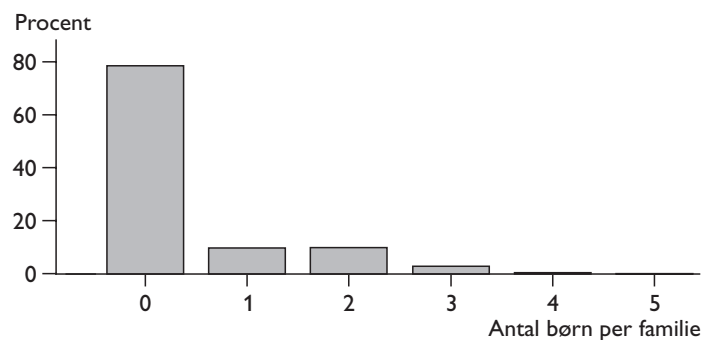


Figur 1.25. Den ekstrapolerede udbredelse af mobiltelefoner i Storbritannien i perioden 1990-2010 – med andel omregnet til logit.

Tælleskala

Lidt på sidelinjen blandt intervalskalaerne står tælleskalaen. Den ligner nemlig den absolutte skala ved på sin vis også at være absolut, dvs. uden vilkårlige specifikation, uden arbitrært nulpunkt eller arbitrær måleenhed. Også tælleskalaen har altså et absolut nulpunkt og en absolut enhed. Den afviger imidlertid fra den absolutte skala og fra de andre intervalskalaer ved at være diskret og ikke kontinuert. Den naturlige enhed er nemlig ikke en måleenhed, men en tælleenhed.

Lad os fx se på fordelingen af danske familier i 2004 efter antal hjemmeboende børn (figur 1.26 og tabel 1.2):



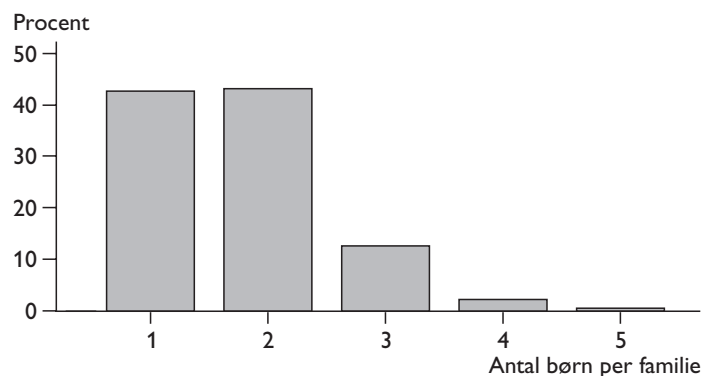
Figur 1.26. Fordelingen af danske familier i 2004 efter antal hjemmeboende børn.

Tabel 1.2. Fordelingen af danske familier i 2004 efter antal hjemmeboende børn.

Antal børn	Procent
0	77
1	10
2	10
3	3
4	1
5 eller mere	0

Vi ser her, at langt de fleste familier slet ikke har nogen hjemmeboende børn, og af dem, som har, er der for hovedpartens vedkommende tale om et eller to børn.

Det, som gør antal børn til en tælleskala, er det forhold, at skalaen er



Figur 1.27. Fordelingen af danske børnefamilier i 2004 efter antal børn.

Tabel 1.3. Fordelingen af danske børnefamilier i 2004 efter antal børn.

Børn	Antal	Procent
1	280.503	42
2	283.467	42
3	83.652	13
4	15.322	2
5 eller flere	4.117	1

diskret, dvs. at vi kun tæller hele værdier. Figur 1.27 viser fordelingen blandt børnefamilier alene, og den samme fordeling fremgår i tabelform af tabel 1.3.

Ved hjælp af en særlig udregningsmetode kan vi ud fra tabellen beregne *det gennemsnitlige antal børn pr. børnefamilie til 1,77*.

Som det gælder for ethvert andet gennemsnit, er der også her tale om en abstraktion. Af alle de 667.061 børnefamilier i Danmark er der ikke en eneste, som har en trekvart barn. Og heldigvis for det. Børn kan kun tælles i heltal, og når vi udregner gennemsnittet af antal børn pr. børnefamilie, går vi således ud over vores umiddelbare begreber. Det har vi imidlertid gjort i så mange år, at vi tillægger det en mening at opgive gennemsnittet for antal børn i danske børnefamilie til 1,77.

Lad os holde os til det, tallet informerer om. At der ikke er særlig mange børn pr. familie. At et stort antal af børnefamilier holder sig til tre eller færre. Lidt mere præcist har de fleste et eller to børn, og der næsten ikke er nogen med fire eller flere. Her holder vi os til fores umiddelbare

begreber. For det er det, tælleskalaen meningsfuldt kan bruges til. At operere med hele tal. Det med brøker er faktisk noget tøjeri, når vi har med tælleskalaen at gøre.

Gyldighedens banemænd – fejlkilderne

Metodologisk set er gyldighed defineret ved sit modbegreb ugyldighed. At en metode er ugyldig vil i realiteten sige, at den er behæftet med fejlkilder.

Men hvornår er en metode gyldig, og hvornår er den det ikke?

Det første spørgsmål kan formentlig slet ikke besvares fyldestgørende, for efter flere tusind års dyrt indhøstede erfaringer er det en udbredt opfattelse, at man slet ikke kan vide, om en bestemt metode er gyldig i det konkrete tilfælde. Imidlertid kan vi nok besvare spørgsmålet mere principielt, definatorisk.

- *Definition af undersøgelsesgyldighed*

At en empirisk metode er *gyldig* vil sige, at den er sådan tilpasset til sin empiriske genstand og til sin videnskabelige problemstilling, at resultatet af den anvendte metode *korresponderer* med det, vi ønsker at undersøge, dvs. at det *billede*, den giver af sin genstand (eller sit genstandsfelt), er, hvad der i bank- og revisionsverdenen hedder *retvisende*, og at det *svar*, der fremkommer mht. problemstillingen, er *korrekt*.

Når en metode er gyldig, har vi altså grund til at tro, at den giver et korrekt resultat, men kun, om jeg så må sige, på det foreliggende grundlag. Vi kan måske bruge retsplejelovens princip om, at en sigtet person anses for at være uskyldig, indtil det modsatte er bevist. På samme måde må vi bruge de bedste metoder, vi har, indtil vi finder ud af, at de har nogle hidtil usete defekter og derfor må revideres. Mht. til det andet spørgsmål om *ugyldighed*, er vi på noget sikrere grund. Hvis vi vel at mærke ikke tror, at manglende gyldighed er det modsatte af gyldighed. Vi kan nemlig *ikke* uden videre slutte fra det forhold, at en metode er ugyldig, og til, at resultatet altid er falsk. Derimod kan vi nok gå ud fra, at når vi anvender en ugyldig metode, er vi på gyngende grund.

- *Definition af undersøgelsesugyldighed*

At en empirisk metode er *ugyldig* vil sige, at den er så *dårligt tilpasset* til

sin empiriske genstand og til sin videnskabelige problemstilling, at resultatet af den anvendte metode *kan være uden* korrespondance eller kun have en *delvis eller fejlbehæftet* korrespondance med det, vi ønsker at undersøge, dvs. at det *billede*, den giver af sin genstand (eller sit genstandsfelt), *kan være misvisende*, og at det *svær*, der fremkommer mht. problemstillingen, kan være *falsk*.

En metode kan være ugyldig på to måder:

1. Den kan i kraft af sin egen opbygning producere *artefakter*, dvs. resultater, der afspejler metoden selv og ikke dens genstand.
2. Den kan give *misvisende* resultater, der nok afspejler virkeligheden, men ikke det i virkeligheden, vi agter at undersøge.

1. *Definition af ugyldighed som undersøgelsesartefakt*

Resultatet er helt eller delvis *et produkt af metoden og ikke genstanden*. Det er fx tilfældet, hvis undersøgelsen er belastet af fejlkilder, som fremkalder artefakter, dvs. resultater, som ikke er bestemt af empirien, men af en fejlkonstruktion ved metoden, fx i form af ledende spørgsmål i et interview.

2. *Definition af ugyldighed som misvisende resultat*

Resultatet giver et i og for sig *informativt billede af virkeligheden, dog ikke af det udsnit, vi tror*, men derimod af andre forhold i virkeligheden, som vi ikke mener at have spurgt om, og som vi måske slet ikke kender. Det kan fx være tilfældet med en intelligensundersøgelse, som tilsyneladende påviser en aldersbestemt reduktion, mens der i realiteten er tale om en kohorteeffekt knyttet til uddannelsesmæssige forandringer.

Vi skal nu i det følgende gennemgå de vigtigste former for fejlkilder og de primære validitetsbegreber.

- *Definition af fejlkilde*

En *fejlkilde* er et forhold ved en empirisk undersøgelse, der influerer på undersøgelsesresultaterne på en måde, der ikke afspejler det fænomen, vi ønsker at studere, men derimod forsøgsomstændigheder, som er irrelevante for undersøgelsens formål.

- *Definition af bias*

Ved bias skal mere specifikt forstås en fejlkilde, som resulterer i en systematisk misvisning, en skævhed, hvis retning er bestemt af forhold ved fejlkilden.

Vi skal i det følgende gennemgå seks forskellige typer af fejlkilder, knyttet til forskellige af undersøgelsesmetodens bestanddele eller processer. De seks vigtigste fejlkilder ved undersøgelsesmetoder:

1. selektionsbias er knyttet til udvælgelsen,
2. instrumentbias er knyttet til forsøgets redskabsinventar,
3. forsøgseffekten er knyttet til, hvordan forsøgspersonerne forholder sig til selve det at indgå i et forsøg,
4. fordringseffekten er knyttet til forsøgspersonernes forestilling om, hvad forsøget går ud på,
5. selvpræsentationseffekten er knyttet til forsøgspersonens ønske om at præsentere sig selv på en bestemt måde,
6. forsøgsledereffekten er knyttet til den måde, hvorpå forsøgslederens viden om forsøgets formål kan styre hans adfærd, så den påvirker udfaldet.

Af disse seks fejlkilder er:

- de to første (1 og 2) knyttet til forsøgets design, det er *designeffekter*,
- de tre næste (3, 4 og 5) udspringer af forsøgspersonen som aktiv medvirkende og undertiden uerkendt modvirkende aktør; det er *forsøgspersoneffekter*,
- den sidste (6) hænger sammen med forsøgslederens egen person som en ukontrolleret (og ukontrollerbar) faktor; det er en *forsøgsledereffekt*.

Selektionsbias (selektivt bortfald)

Der kan være tale om selektionsbias i to tilfælde. For det første i repræsentative undersøgelser, hvor der tages en stikprøve af en bestemt population. For det andet i de gruppebaserede eksperimenter med forsøgs- og kontrolgrupper, der udvælges fra samme population. Her er *selektionsbias* den fejlkilde, som består i, at undersøgelsen kommer til at lide af et *bortfald*, der ikke er tilfældigt i forhold til populationens sammensætning.

Det er ikke i sig selv en katastrofe, at der er et bortfald på fx 30 %, hvis

bare bortfaldets sammensætning ikke afviger fra sammensætningen af de personer, der gennemfører undersøgelsen. Hvis sammensætningen ikke afviger, sker der blot det, at *usikkerheden* ved angivelser af de observerede størrelser bliver *forøget*, altså en rent kvantitativ forringelse svarende til, at man var begyndt med 30 % færre personer.

Helt anderledes stiller det sig, hvis der foreligger et *selektivt bortfald*, dvs. når det er personer med særlige karakteristika, som holder sig væk eller dropper ud undervejs. Hvis det er tilfældet, bliver stikprøven *skæv* på den måde, at den *systematisk afviger fra udgangspopulationen*. Endnu dårligere står det til, hvis der er forskelligt bortfald i forsøgs- og kontrolgruppen i et eksperiment, hvilket let kan forekomme, hvis fx forsøgsgruppen har en højere motivation for at gennemføre.

Selektionsbias forekommer ikke blot som en bortfaldsproblemetik. Det er et lige så alvorligt problem ved selve den indledende selektion i de tilfælde, hvor udvælgelsen er baseret på *frivillig* deltagelse. Dette problem går under betegnelsen *frivillighedsproblemet*.

I det følgende skal vi se på den ret omfattende forskning vedrørende dette problem. Denne forskning kaster samtidig lys over de farer, som er knyttet til det selektive bortfald, da de personligheds- og motivationsfaktorer, som styrer den frivillige tilmelding til en undersøgelse i vid udstrækning spiller den modsatte rolle, når folk dropper ud af undersøgelsen.

Jeg skal her ikke komme ind på de særlige forsøgsdesign, der er udviklet inden for dette forskningsområde,¹⁴ men henvise til monografien *The Volunteer Subject* (Rosenthal & Rosnow 1975). Udgangspunktet for undersøgelserne vedrørende frivillige forsøgspersoner har været en voksende uro inden for psykologien over det faktum, at den overvældende del af alle psykologiske forskningsprojekter har været baseret på frivillige forsøgspersoner rekrutteret blandt psykologistuderende på studiets første del(sophomores).

En ting er, at de altså repræsenterer en relativt snæver befolkningsgruppe, hvis forhold til genstandsfeltet er ret specielt med de særlige kognitive og holdningsmæssige konsekvenser, dette kan indebære. En anden ting er, at det inden for denne snævre kategori af forsøgspersoner, kun er de frivillige, der bliver undersøgt, mens de modvillige ikke gør det.

Er der da forskel på disse to grupper? Ja, det er der massiv evidens for. Jeg skal her blot oprids de mest sikre hovedresultater:

De frivillige er:

1. bedre uddannede,
2. har højere social status,
3. højere intelligens (målt med standard intelligensprøver),
4. har større behov for anerkendelse (fra fx lærernes side),
5. mere sociale (fx målt gennem foreningsaktivitet) (ibid., s. 86).

Andre interessante, men mere specielle fund, er:

1. frivillige til angstprovokerende forsøg med stress, sensorisk deprivation og hypnose, synes at være mere oplevelses- eller arousal-søgende,
2. frivillige til undersøgelser af seksuel adfærd tenderer mod at være mere ukonventionelle (ibid., s. 118).

Mens de første fem fund viser en *generel* bias i retning af, at frivillige er de mere ressourcestærke, peger de to sidste resultater på, at såvel *frivillig selektion* som *selektivt bortfald* kan være knyttet til selve undersøgelsens indhold på en metodologisk særdeles ubehagelig måde, idet denne form for bias bevirker, at resultaterne bliver *systematisk misvisende*. Således vil undersøgelser af seksuel adfærd tendere mod at vise et billede af en mere ukonventionel måde at forholde sig på, end hvad der svarer til hele populationen.

Hvis der er tale om et interventionsforsøg, kan det selektive bortfald især ramme de konventionelle forsøgspersoner, således at den formodede interventionseffekt bliver forstærket på en fejlkildeafhængig måde. I tråd med disse resultater har man fundet en tendens til, at frivillig deltagelse forøges med forsøgspersonens:

1. interesse i forsøgets indhold
2. formodning om en positiv bedømmelse i undersøgelsen
3. vurdering af undersøgelsens betydning (for videnskaben eller offentligheden).

Sådanne systematiske selektionseffekter kan resultere i direkte forkerte empiriske resultater med hensyn de kausalsammenhænge, der er fundet i eksperimentelle undersøgelser. Den forskellige fordeling af personlighedstræk hos frivillige og ikke-frivillige, kan fx i holdningspåvirkningsforsøg indebære, at de ikke-frivillige fremviser den modsatte kausalsammenhæng

af den, man finder hos de frivillige. Således har man fundet (Horowitz, 1969), at angstpåvirkning (fx skrækkfilm om lungekræft) har den ønskede virkning på frivillige (dvs. at en del holder op med at ryge), mens den virker mod sin hensigt på ufrivillige (dvs. at nogle ryger mere end før).

Instrumentbias

Instrumentbias vil sige, at der i selve den redskabsmæssige opbygning af designet er fejlkilder, som kan forårsage en bestemt misvisning. Denne fejlkilde omfatter specifikt det eksperimentelle udstyr til påvirkning og registrering i eksperimenter, spørgsmålenes udformning og placering i survey-undersøgelser og scoringsmetoderne i psykometriske test. Vi skal behandle bias ved spørgsmåludformning i kapitel 3 og ved scoringsmetoder i kapitel 4, således at vi i denne omgang nøjes med selve definitionen af instrumentbias.

Forsøgseffekt (Hawthorne-effekt)

I de to foregående afsnit om designeffekterne, altså selektions- og instrumentbias, har vi set typer af fejlkilder, der trods alt i nogen grad er under forsøgslederens kontrol, for så vidt som de forårsages af svagheder i forsøgslederens egen udformning af undersøgelsen. Den fejlkilde, vi skal omtale nu, er mere vidtgående og foruroligende. Den er nemlig knyttet til selve det forhold, at forsøgspersonerne bliver inddraget i en undersøgelse, og hermed er vi inde på de former for bias, som er bestemt af det særegne for psykologiske undersøgelser, at de empiriske genstande er *personer*, og at de fænomener, som faget studerer, enten direkte eller indirekte består i *interaktioner mellem forsøgspersonerne og forsøgslederen*.

Forsøgseffekten er altså den mest omfattende inden for kategorien forsøgspersoneffekter, og dermed er den også den mest uspecifikke, hvilket igen gør den til den dårligst definerede og dårligst operationaliserede af de fejlkilder, som er relateret til forsøgspersoner.

Når forsøgseffekten har fået øgenavnet Hawthorne-effekten, er det for en gang skyld ikke en navngivning efter opdageren, men efter det sted, hvor opdagelsen skete, eller hvor den i hvert fald hævdes først at være blevet observeret. Selve dokumentationsgrundlaget for opdagelsen er nemlig lige så tåget og problematisk som denne fejlildes gyldighed eller rettere gyldighedsområde.

Som Gillespie (1991) har vist med sin bog *Manufacturing Knowledge*, er historien om, hvordan begrebet *Hawthorne-effekt* opstod, snarere en receptionshistorisk end en metodologisk succeshistorie. Historien har næsten fået lige så sakrosant placering i den elementære psykologiundervisning, som det ny testamente har på fundamentalistiske præsteseminarier. Jeg vil derfor uden tøven tilslutte mig de øvrige narrative metodologer og videreformidle denne grundfortælling, som er blevet en uudryddelig del af psykologiens trosgrundlag – det, som Bourdieu kalder *doxa*.

Midt i 1920'erne emigrerede Elton Mayo, en ukendt og indtil da heller ikke synderlig succesrig australsk psykolog til USA, hvor han blev professor ved Harvard University School of Business. Her blev han en af grundlæggerne af den nye Human Relations-retning inden for arbejdspsykologien. Hans ideer om en slags progressiv og reformistisk kapitalisme fandt støtte hos magnaten John D. Rockefeller, som var en af universitetets sponsorer, og som oprettede det enorme National Research Council (NRC), hvorfra Mayo fik bevillinger til den række af arbejdspsykologiske studier, der fra midten af 20'erne til begyndelsen af 30'erne foregik på Hawthorne-fabrikken, en Western Electric-ejet fabrik for elektriske artikler.

Mayo var optaget af tanken/ideen om en kapitalisme med et menneskeligt ansigt, hvor arbejdsforholdene blev så gode, at arbejderne helt opgav de socialistiske (eller i det mindste fagforeningsvenlige) tilbøjeligheder, der var udbredte i mellemkrigstidens USA. Det var denne idé Mayo ville afprøve med de arbejdspsykologiske felteksperimenter, han satte gang i på Hawthorne-fabrikken.

De første år var udgangspunktet stadig den fysiologisk prægede form for „scientific management“, der var forløberen for Human Relations-retningen og som senere viste et mere professionelt og hårdhudet ansigt med tidsstudierne i 1950'erne og 60'erne. Det indebar, at de første felteksperimenter bestod i at eksperimentere med fordelingen af arbejde og pauser og med at variere fx belysningen. Den specifikke genstand for eksperimenterne var et særligt prøverum på fabrikken, hvor seks kvindelige arbejdere udførte montagearbejde.

Da der var en generel produktivitetsforbedring i prøverummet, som ikke hang sammen med de eksperimentelle variable, siger myten nu, at det arbejdspsykologiske hold, som Mayo ikke var medlem af, men inspirator for, fik den idé at kontrollere forsøgene ved at begynde forfra med en hel serie. Og man skulle da have gjort den epokegørende opdagelse, at produktiviteten fortsatte med at stige, når man foretog øjensynlige forrin-

gelser i arbejdsbetingelser og -miljø. Og voila! Hawthorne-effekten var ikke alene opdaget, men også dokumenteret. Det skulle altså være selve virkningen af at få tildelt en særlig betydning, som motiverede de i øvrigt stærkt underprivilegerede kvinder i prøverummet – hovedsagelig immigranter, især fra Polen.

Umiddelbart kunne denne besynderlige effekt virke som en undergravning af Human Relations-ideologien, der jo gik ud på, at man skulle forbedre arbejdsbetingelser til såvel arbejdsgivernes som arbejdstagernes fordel, men faktisk kan fænomenet jo opfattes som et supplerende element i denne ideologi, nemlig som udtryk for de velsignelsesrige følger af at behandle arbejdere som personer, der er betydningsfulde og interessevækkende.

Nu blev det som nævnt holdet bag Hawthorne-undersøgelserne, der ikke bare opdagede dette generelt psykologiske, og altså også metodologiske fænomen, men som også synes at kunne dokumentere det med deres eksperimenter. Som Gillespie har afdækket, er selve den metodologiske kvalitet af forsøgene, sådan som de fremstår ikke blot i publicerede undersøgelser, men også i materialet fra Hawthorne-fabrikkernes arkiv, særdeles ringe. Dette er i nogen grad en afspejling af arbejdspsykologiens endnu ret umodne karakter på det tidspunkt, forsøgene blev udført.

Således blev der heller ikke taget højde for betydning af aflønningssystemer, dvs. fastløn eller akkord. Desuden skete der undervejs en udskiftning af to arbejdere, som blev fyret på grund af deres formodede socialistiske sympatier, men motivet var dog snarere en måske ikke helt grundløs angst for, at disse arbejdere ville modarbejde den tilsigtede produktivetsforøgelse, der set ud fra arbejderens synsvinkel kunne opfattes som et forsøg på at forcere arbejdstempoet også i den øvrige del af fabrikken.

Som det sidste eksempel antyder, var Human Relations-tankegangen ikke bare fremmed, men direkte fjendtligt indstillet over for en mere arbejds sociologisk tankegang, hvor man ser på de interesse mæssige, for ikke at sige klasse mæssige, modsætninger mellem arbejdere og arbejdsgiver.

Det interessante er nu, at på trods af denne svage dokumentation er Hawthorne-effekten blevet en fetich for såvel arbejdspsykologien som for den psykologiske metodologi. Det skyldes utvivlsomt, at forsøgene og deres begejstrede modtagelse i eftertiden har været båret af en moderat progressiv reformisme, der blev markedsført som beviset på, at hvad der var godt for arbejderne også var godt kapitalisten.

Ser man på den status, denne forsøgseffekt har i nutidens metodologi,

er der ingen tvivl om, at fænomenet har så stor betydning, at man må tage højde for det ved enhver undersøgelse. På den anden side er fænomenets kausale determinanter så uklare, at det kan være svært at gøre det i praksis. Sagen er nemlig, at der ikke altid er nogen nævneværdig forsøgseffekt, og det er ydermere usikkert, i hvilken retning den vil gå. Eksemplet med de fyrede arbejdere fra prøverummet viser fx, at holdningsmæssige og motivationelle forhold kan være afgørende for, hvordan man som forsøgsperson forholder sig til at deltage i et forsøg.

Rent designmæssig kan man måske forsøge at eliminere en eventuel forsøgseffekt ved at have en kontrolgruppe, som ikke udsættes for den specifikke intervention, der er målet for undersøgelsen, men som i øvrigt indgår under forsøgsbetingelser, der skulle give samme forsøgseffekt som i forsøgsgruppen. Altså en metode svarende til brugen af placebo. Det er imidlertid indlysende, at det kan være ikke bare svært, men helt umuligt at skabe lige så stor interesse for et sådan quasi-forsøg, som for den rene vare.

Et af de steder, hvor man meget ofte møder forsøgseffekten, er ved pædagogiske forsøg, der afprøver nye undervisningsformer. Her siger man ofte, at pædagogiske forsøg altid lykkes, *første gang*. Denne særlige udgave af Hawthorne-effekten må siges at være så udbredt, at den i nogen grad har undergravet tilliden til pædagogiske forsøg overhovedet. En metodologisk modgift er da simpelthen at gentage forsøget, så det efterhånden mister nyhedsværdien.

Fordringseffekten (demand characteristics)

En af de mest gennemgribende former for undersøgelsesbias blev opdaget af Orne (1951) i forbindelse med eksperimentelle undersøgelser af hypnose. Bragt på sporet af selve suggestibiliteten ved dette fænomen fik Orne en mistanke om, at visse eksperimentelle fund ved hypnose ikke var forårsaget af den eksperimentelle variabel, men af uerkendte forventningsfaktorer hos såvel forsøgslederen som forsøgspersonerne.

Han foretog en række omhyggeligt designede eksperimenter, hvor han udsatte psykologistuderende for hypnose, efter at de havde overværet et foredrag om hypnose, hvor forsøgsgruppens personer havde fået en falsk information om, at hypnose fremkaldte katalepsi i den dominerende arm, mens kontrolpersoner ikke havde fået denne information. På trods af at såvel hypnotisøren som observatøren ikke vidste, hvem der tilhørte hvil-

ken gruppe, var den falske information i stand til at fremkalde den virkning hos forsøgsgruppen, som var blevet meddelt.

Derefter foretog Orne (1959, 1962) nogle endnu mere raffinerede forsøg, hvor han opererede med dels ægte, dels falsk hypnotiserede. De sidste var instrueret om at forstille sig, dvs. at de skulle opføre sig som om de var blevet hypnotiseret. Med disse to slags forsøgspersoner gentog han så et eksperiment baseret på Bruners berømte mønteksperiment, der oprindeligt var blevet gennemført med henholdsvis rige og fattige børn. Nu blev eksperimentet gennemført med personer, der enten var hypnotiseret til tro at de var henholdsvis rige og fattige børn eller opførte sig, som om de var hypnotiseret til at tro det. Ved den hypnotiske udgave af mønteksperimentet havde man fundet samme resultat, som i Bruners oprindelige, altså at fattige børn indstillede på større værdier på et apparat til angivelse af møntstørrelse end rige børn gjorde. Orne kunne nu vise, at der var samme effekt ikke bare hos reelt hypnotiserede, men også hos de falsk hypnotiserede. De sidstnævnte var instrueret til at opføre sig, som om de var hypnotiseret til først at være fattige, så almindelige og til sidst rige børn. Selvom de ikke havde fået nogen forhåndsviden om den eksperimentelle hypotese, viste en forsigtig efterfølgende udspørgen, foretaget af trænede klinikere, at de havde haft en fornemmelse af forskellen mellem adfærden hos børn af de forskellige kategorier.

Orne opstillede nu den hypotese, at der var tale om et generelt fænomen, nemlig effekten af hvad han kaldte "demand characteristics", hvorved han forstår de træk ved eksperimentet, som forsøgspersonerne tror karakteriserer det, herunder selve den eksperimentelle hypotese, der ikke må være kendt, hvilket forsøgspersoner selvfølgelig også godt ved, men som de alligevel, uden at gøre sig det klart, dels ofte får en fornemmelse af, dels altså uerkendt påvirkes i retning af.

Med hensyn til hvordan disse "demand characteristics" formidles i forbindelse med et eksperiment eller for den sags skyld enhver type undersøgelse, er der efter Ornes opfattelse dels tale om forventninger, som forsøgspersonerne danner sig ud fra selve den synlige indretning af forsøgsrummet eller undersøgelsens lokalitet, dels om nonverbal (og altså særdeles utilsigtet) kommunikation mellem forsøgsleder og forsøgsperson. Det sidste skal vi se nærmere på i afsnittet om forsøgsledereffekten (s. xxx).

Fordringseffekten viser forsøgspersonen som en aktiv medspiller ved en undersøgelse. En anden form for artefakt knyttet til det forhold, at

forsøgspersonen er et menneskeligt subjekt og ikke en empirisk genstand, finder vi i det følgende afsnit om selvpræsentationseffekten.

Selvpræsentationseffekten (evaluation apprehension)

„Evaluation apprehension“ betyder noget i retning af ængstelse for eller optagethed af den måde, ens adfærd bliver vurderet på af forsøgslederen. Rosenberg definerer denne bias på følgende måde:

... selvpræsentationseffekten (*evaluation apprehension*) ... er en aktiv, ængstelsespræget optagethed af, om han (forsøgspersonen) nu også vil få en positiv bedømmelse af forsøgslederen eller i det mindste ikke vil give anledning til en negativ bedømmelse (Rosenberg, 1965).

Rosenberg bemærker, at der typisk er tale om en forsøgsleder, der er psykolog, og at forsøgspersonerne forventer, at deres handle- og reaktionsmåde bliver vurderet, måske på en måde, som er skjult for dem. De fleste mennesker ønsker at gøre et godt indtryk, et fænomen som i testpsykologi kaldes ”social desirability”, og denne optagethed af at fremtræde som intelligente, ærlige og sympatiske mennesker kan indebære en systematisk fejlkilde. Forsøgspersonernes handlinger og formuleringer vil ofte bevidst eller ubevidst være påvirket af denne optagethed af, hvilket indtryk de gør.

I et meget kompliceret og veludført eksperiment, der med rette fremstår som en klassiker i den psykologiske metodelitteratur, har Rosenberg efterprøvet et tidligere eksperiment vedrørende kognitiv dissonans. Rosenberg (1985) gentager her det kunststykke, som vi netop har set Orne udføre, nemlig at efterprøve et klassisk eksperiment på en sådan måde, at det tidligere eksperiment bliver perspektiveret og på afgørende måde modificeret. Og lige som i Ornes tilfælde med vurdering af møntstørrelse slår Rosenberg to fluer med et smæk. På den erkendelsesrettede akse vises det, at det oprindelige forsøg har afdækket sin genstand på en korrekt måde, og at det formentlig er kommet frem til en forkert konklusion om selve forsøgets sagforhold. På den erkendelseskritiske akse afdækkes et helt nyt metodologisk forhold, der så i næste omgang bliver et selvstændigt bidrag til almenpsykologiens erkendelsesmæssige akse. Denne vekselvirkning mellem det genstandsmæssige, erkendelsesrettede på den ene side og det metodologiske, erkendelseskritiske på den anden demonstrerer i øvrigt, at selve denne skelnen er en abstraktion. I den konkrete undersøgelse hænger

de to sider sammen. Men inden vi ser nærmere på denne dialektik mellem den erkendelsesrettede og den erkendelseskritiske forskning, skal jeg forsøge at præsentere Rosenbergs sofistikerede problemstilling og design.

For at afprøve sin teori om kognitiv dissonans og dissonansreduktion foretog Festinger og Carlsmith (1959) et eksperiment med påtvungen accept (forced compliance), hvor forsøgspersonerne blev presset til at gøre sig til talsmænd eller advokere for synspunkter, som de var modstandere af. Brehm og Cohen (1962) gentog eksperimentet med bachelorstuderende ved Yale University.

Udgangspunktet var nogle netop opståede uroligheder i universitetsområdet, hvor det lokale politi havde grebet ind på en åbenbart noget håndfast måde, hvilket den langt overvejende del af de studerende tog kraftigt afstand fra. Forsøgslederen dukkede nu op på tilfældigt udvalgte kollegier på campus og præsenterede sig som medlem af et institut for Human Relations Research, der nu udspurgte en given studerende, der fik at vide, at han eller hun var udvalgt som modstander af politiets aktion. Udspørgeren fortalte, at en god måde at forstå argumenter for og imod et bestemt spørgsmål på var at formulere et essay, der kun fremførte den ene sides synspunkt, og den studerende blev derfor bedt om at skrive et sådant essay, der forsvarede politiets aktion. Der blev nu givet forskellige belønninger for dette arbejde, nemlig beløb på henholdsvis en halv, en, fem og ti dollars. Da det foregik for over 40 år siden, har fem og ti dollars været forholdsvis store beløb for studerende. Efter den eksperimentelle intervention i form af tvungen advokatur og belønning herfor blev de studerendes holdning til politiaktionen målt på en graderet holdningsskala.

Festingers forudsigelse var, at studerende med den lave belønning (en halv og en dollar) ville blive en mental kamplads for modstridende synspunkter, nemlig deres oprindelige politifjendtlige og deres i essayet formulerede politivenlige holdninger. Dette skulle da ud fra Festingers teori føre til en kognitiv proces, som ville tendere mod en ophævelse eller i alt fald en formindskelse af denne dissonans. Derimod ville der ud fra Festingers argumentation ikke foreligge en sådan dissonans hos de studerende, der modtog den store belønning (fem og ti dollars), for her var der jo en rimelig grund til at de (på skrømt) havde argumenteret for noget, de ikke mente. I hvert fald under den forudsætning, at folk generelt finder det rimeligt at føre andre bag lyset, blot de får tilstrækkelig mange penge for det.

Denne forudsigelse blev bekræftet i både Festinger og Carlsmiths og i Brehm og Cohens eksperimenter.

Rosenbergs ræsonnement var nu, at disse empiriske fund kunne være et artefakt, nemlig resultatet af en selvpræsentationseffekt. Han forudsagde (og fik i sit eksperiment bekræftet), at forsøgspersonerne med en høj belønning ville være på vagt over for, at psykologen var ude på at se, om de kunne købes til at skifte standpunkt. For at undgå at fremstå som sådanne moralsk uvederhæftige personer, skulle de da forsøge at fastholde deres standpunkt.

Ud fra denne argumentation skulle studerende, der fik den lave belønning, være fri for mistanke om at være købt, fordi beløbet var så uanseeligt, og de ville da ikke fremvise en sådan præsentationseffekt. Rosenbergs forklaring resulterer altså i samme empiriske resultat, men teoretisk er der tale om modstridende kausalhypoteser. I øvrigt fastholder Rosenberg det oprindelige design, bortset fra at universitetet er Ohio State University, og den aktuelle begivenhed, de studerende skal forholde sig til, i stedet for en politiaktion er universitetsledelsens meget upopulære beslutning om give dette provinsuniversitet et fornemmere image ved at droppe deltagelsen i en traditionel fodboldturnering.

Det såvel teoretisk som metodologisk sofistikerede ved Rosenbergs arbejde er, at han dels opstiller et design til eliminering af selvpræsentationsfaktoren og dermed det empiriske artefakt, dels forudsiger, at det således rensede eksperiment vil fremvise det stik modsatte resultat af de oprindelige. Denne polemiske side af Rosenbergs eksperiment er elegant præsenteret i selve titlen på hans afhandling, der hedder *When Dissonance Fails*, hvilket er en henvisning til en af Festingers mest berømte bøger, nemlig *When Prophecy Fails*, hvor dissonansteorien bliver empirisk afprøvet i et dramatisk feltstudie, nemlig de sidste dage inden jordens forventede undergang ifølge en dommedagskult, som Festinger under dække tilmeldte sig.

Rosenberg søger at minimere selvpræsentationseffekten ved at adskille den første og den anden forsøgsleder. Den person, der spørger forsøgspersonerne om deres holdning, *efter* at de har påtaget sig rollen som talsmænd og modtaget deres belønning, er altså en anden end den, der overtaler dem til at være talsmænd, og som giver dem belønningen. Han kontrollerer også ved en samtale med forsøgspersonerne, efter at hele det egentlige eksperiment er gennemført, at der med den nye design faktisk er tale om, at selvpræsentationseffekten er fjernet.

I det nye rensede eksperiment er den gamle empiriske sammenhæng, der blev så flot forudsagt og forklaret af Festinger ligeledes elimineret. Men der er en faktisk endnu mere opsigtsvækkende virkning af selvpræsentationseffektens fjernelse. Sammenhængen mellem belønningens størrelse og holdningsændring er nemlig nu vendt på hovedet. Rosenberg argumenterer for, at han har fjernet et artefakt, der har haft en sådan styrke, at det ikke bare har maskeret, men direkte spejlvendt den sande sammenhæng.

Rosenberg har da også en helt anden argumentation, der ganske vist stadig er funderet på Festingers dissonansteori, men med en opstilling af præmisser, som er anderledes. Rosenbergs analyse er, at de studerende, der får de høje belønninger, gør sig mere umage end dem, der får de lave belønninger. Denne præmis verificerer han faktisk ved at optælle længden af de forskellige essays og ved lade „blindede“ bedømmere vurdere argumentationens kvalitet. Hans næste præmis er, at når en studerende har udført et ordentligt stykke, ganske vist påtvungen, advokatur, så bliver han eller hun også selv mere overbevist, end når det er sjusket og lidet overbevisende. Dissonansen fremkommer altså mellem den oprindelige holdning og den, man selv har formuleret i essayet, men altså kun ved de gode essays, der har fået den høje belønning.

Hvis vi ser bort fra de øjensynlige vanskeligheder ved at udmønte dissonansteorien i entydige og empirisk gyldige prædiktioner, så er moralen af Rosenbergs eksperiment følgende:

1. at der faktisk eksisterer en selvpræsentationseffekt.
2. at en bias af denne art kan skabe så drastiske former for artefakter, at en bestemt kausalsammenhæng bliver spejlvendt.¹⁵
3. at der er en dialektik mellem den erkendelsesrettede og den erkendelseskritiske akse. Det, som i forhold til den epistemiske genstand, dissonansteorien, er erkendelseskritisk, nemlig selvpræsentationseffekten, bliver i sig selv til en ny epistemisk genstand, i første omgang i den psykologiske metodologi og i næste omgang i selve den almene psykologi, nemlig som et bidrag til en selvpræsentationsteori – et teoriområde, der også omfatter et fænomen som „social desirability“.

Forsøgsledereffekt (Rosenthal-effekten)

Den mest berømte af alle former for undersøgelsesbias er formentlig Rosenthal-effekten. Det skyldes del, at det måske er den, der udgør det

mest ødelæggende angreb på eksperimentalpsykologiens metodologiske uskyldstilstand, dels at den demonstrerer nogle forhold omkring autoritetsudøvelse i fx pædagogiske og arbejdsmæssige sammenhænge, der på sin vis er lige så bekymringsvækkende som Milgrams forsøg.

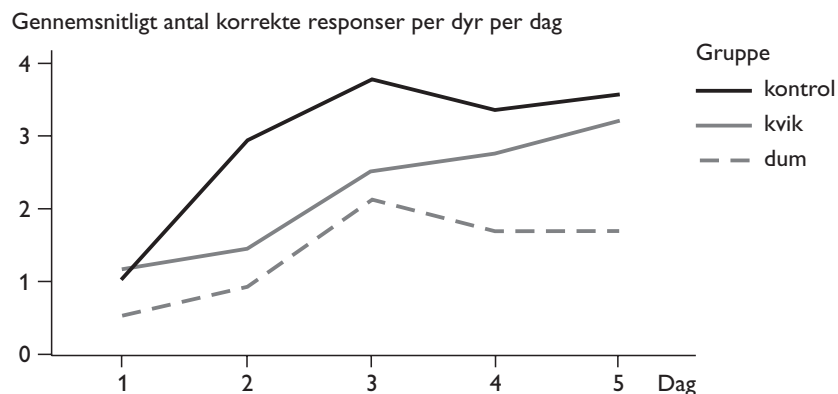
Rosenthal gennemførte gennem 1960'erne nogle eksperimenter med sine medarbejdere, og disse eksperimenter demonstrerede, hvad der i daglig tale hedder Rosenthal-effekten, selv kalder han det beskedent for forsøgslederbias.

Lige som de netop beskrevne former for bias, frivillighedseffekten (ved selektion), Hawthorne-effekten, fordringseffekten og selvpræsentationseffekten udgør Rosenthal-effekten en del af det, som Orne (1962) har kaldt for det psykologiske eksperiments socialpsykologi.

Når Rosenthal-effekten imidlertid i den almindelige bevidsthed er kommet til at overskygge de andre og er blevet en slags overbegreb for alle disse former for bias, så skyldes det nok to ting. For det første, at demonstrationen allerede i første omgang blev udført ved hjælp af dyreforsøg (laboratorierotter) og dermed rammer lige ind i hjertet af datidens fejrede psykologiske reduktionisme, nemlig den skinnerianske læreproces-teori. For det andet, at Rosenthals nyopdagede fænomen i næste omgang, hvor forsøgsindividerne var skoleelever, viste særdeles alvorlige perspektiver for autoritetsbestemte interpersonelle relationer, fx undervisning.

Selve ideen til Rosenthals første eksperiment (Rosenthal & Fode, 1963) angives at være selveste Pavlov, der engang om en assistent bemærkede, at dyreforsøgene snarere var resultatet af assistentens undervisning end af dyrenes indlæring. Lige som i Milgrams, Ornes og Rosenbergs eksperimenter betjener Rosenthal sig af assisterende forsøgsledere, der i dette særdeles raffinerede design faktisk i mindst lige så høj grad som rotterne er de reelle forsøgsobjekter.

Designet har så at sige to niveauer. Nederst er der tale om et banalt læreprocesforsøg, hvor rotter skal reforceres til en såkaldt diskriminationsindlæring, idet de lærer at vælge den mørke gang i en T-labyrint. Faktisk er der tale om en rutinemæssig undervisningsopgave, hvor psykologistuderende bliver trænet i at udføre rotteforsøg med indlæring. Men oven over dette banale niveau bygger Rosenthal så et metaniveau, hvor den eksperimentelle manipulation er rettet mod de psykologistuderende, som udfører rotteeksperimentene. Det er her, eksperimentet skifter fokus fra rotters diskriminationsindlæring til den nyopdagede forsøgsledereffekt, men som i daglig tale altså kaldes for Rosenthal-effekten.



Figur 1.28. Gennemsnitligt antal korrekte responser per dyr per dag i de tre grupper.

På dette metodologiske metaniveau er der tale om et gruppeeksperiment, hvor 12 psykologistuderende blev matchet to og to ud fra deres forventninger til forsøget, således at der blev to forsøgsgrupper af hver seks personer. Derudover var der en 13. studerende, der fungerede som en blanding af en minimal kontrolgruppe og observatør i forhold til de andre.

Alle fik at vide, at de skulle træne nogle rotter i et diskriminations-indlæringsforsøg, ved at de udsultede rotter fik mad, når de valgte den mørke gang, mens de ikke fik noget, når de tog den lyse.

Forskellen mellem grupperne bestod nu i, at „forsøgslederne“ i den ene gruppe fik at vide, at de fem rotter, som de hver især skulle træne, var udvalgt af en så uintelligent stamme, at de ikke skulle forvente nævneværdige fremskridt i løbet af de fem dage, forsøget stod på. „Forsøgslederne“ i den anden gruppe blev derimod instrueret om, at deres rotter var udvalgt på grundlag af deres intelligens, og at man ville kunne forvente fremskridt allerede den første dag, hvorefter præstationerne hurtigt ville forbedres. De studerende blev selv sat til at udfylde en række spørgeskemaer, hvor de vurderede såvel rotternes som deres egne præstationer og derudover også deres tilfredshed og måde at forholde sig på under forsøgene.

De nærmest sensationelle resultater for de tre grupper fremgår af figur 1.28. Kontrol„gruppen“ består af den enlige studerende, som ikke var ført bag lyset, mens den gruppe, som var vildledt til at tro, at deres rotter var intelligente, kaldes kvik-gruppen, og den gruppe, som fik at vide, at rotterne var uintelligente, kaldes dum-gruppen.

Vi ser, at det i de to forsøgsgrupper går ganske som „forsøgslederne“ har

fået at vide, at det ville gå. I kvik-gruppen gør rotterne hurtige fremskridt og ender ca. dobbelt højt som rotterne i dum-gruppen, der kulminerer allerede den tredje dag. Med hensyn til kontrol-gruppen, der selvfølgelig er noget usikker, da den kun udgøres af en enkelt person, ser det ud til at gå endnu bedre end i kvik-gruppen, hvilket i sig selv er en ret interessant tendens. Den tyder nemlig på, at selv om at det er skidt for elever, når en instruktør udsætter dem for negative projektioner, og meget bedre, når der er tale om positive projektioner og muligvis en vis overidentifikation, så er det nu bedst med en saglig instruktør uden projektioner.

Det er altså den berømte Rosenthal-effekt, som vi jo tror, at vi godt kan forstå, når der er tale om et interpersonelt forhold mellem forsøgsleder og forsøgsindivid. Men hvordan mon denne effekt fungerer, når individerne er rotter? Hvad er det for en af forsøgslederen uerkendt overførsel, hvorigennem hans eller hendes forventninger overføres til forsøgsdyret?

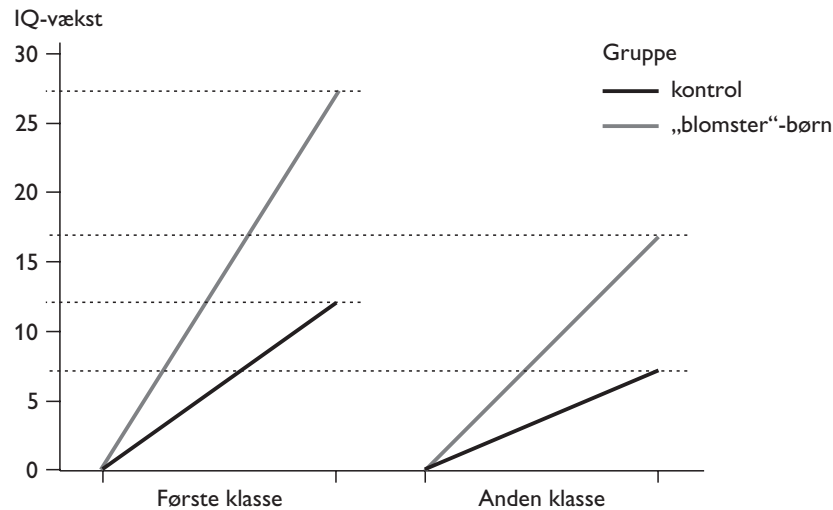
Gennem de spørgeskemaer forsøgslederne udfyldte, får vi løftet en flig af sløret for denne gåde. Forsøgslederne i kvik-gruppen beskrev nemlig deres holdning til deres rotter som meget mere venlig, opmuntrende og interesseret end forsøgslederne i dum-gruppen. Og selv om man anerkender føden som en afgørende forstærker i forsøget, er der altså noget, der tyder på, at forsøgsledernes adfærd er en lige så afgørende motiverende faktor over for en så social dyreart som rotten.

Som nævnt har Rosenthal udført samme type eksperiment med mennesker som empiriske forsøgsobjekter, og her er der oven i købet tale om et felteksperiment, der således ud over en indre validitet pga. det overlegne design, også besidder en ydre (økologisk) validitet (disse begreber bliver nærmere introduceret i næste afsnit, s. xxx).

Som beskrevet af (Rosenthal, 1961, 1968) var der i hans undersøgelser tale om assisterende forsøgsledere i form af skolelærere, der fik falske oplysninger om, at 20 % af en gruppe tilfældigt udvalgte skoleelever var intelligenstestet som usædvanligt begavede, og at de kunne forventes at blomstre intellektuelt i den kommende skoleår. I dette eksperiment med mennesker som basisindivider tog Rosenthal altså fornuftigvis den forskningsetiske beslutning, at der ikke skulle være tale om selvopfyldende profetier, der gjorde nogle elever dummere.

Ved skoleårets afslutning blev alle eleverne igen intelligenstestet. Resultaterne fremgår af figur 1.29.

I første klasse ser man altså hos de ganske tilfældigt udpegede „begavede” børn en mere end dobbelt så stor stigning end hos kontrol-børnene.



Figur 1.29. IQ-fremgang i første og anden klasse.

Også for anden klasses vedkommende er stigningen dobbelt så stor hos de børn, der var blevet udpeget som dem, der ville blomstre i løbet af året.

Mens kontrol-eleverne altså kun har en ubetydelig øget IK, er der tale om en substantiel fremgang for de tilfældigt udvalgte elever, om hvem det var oplyst, at de ville blomstre i løbet af skoleåret. Titlen på 1968-artiklen var *Pygmalion in the Classroom*; den refererer til Shaws skuespil, hvor professor Higgins gennemfører et eksperiment med en simpel blomsterhandlerske, der skal uddannes på højeste niveau. Ud fra denne parallel har Rosenthal også kaldt forsøgsledereffekten for Pygmalion-effekten.

En interessant faghistorisk vignet skal knyttes til denne metodologiske klassiker. Der var som nævnt tale om et målsøgende metodekritisk missil rettet mod behaviorismens hjerte, og efterhånden som forsøgene blev replikeret og valideret skulle man tro, at behavioristerne ville reagere med en blanding af panik og raseri over dette angreb på deres positivistiske videnskabsteori.

Imidlertid viser Suls & Rosnow (1988, s. 178), at dette slet ikke var tilfældet. Den mest almindelige reaktion var slet ingen reaktion overhovedet. En psykodynamisk orienteret person ville måske tillade sig den frækhed at fortolke denne mangel på reaktion som benægtelse eller ligefrem som et perceptuelt forsvar.

En enkelt, til dels sofistikeret behavioristisk opponert (ibid., s. 178)

foreslog, at Rosenthals eksperimenter i sig selv kunne betragtes som artefakter, der var resultatet af Rosenthal-effekten.

Dette ret desperate angreb, hvor fænomenets eksistens er en logisk forudsætning for benægtelsen af denne eksistens, blev i øvrigt imødegået med overlegen ro af Rosenthal, der svarede, at det nok var tænkeligt, at manden havde ret, men at dette kunne undersøges empirisk gennem en designmæssig modifikation, hvor også opdelingen af „forsøgslederne“ var blindet over for den virkelige forsøgsleder.

Måske kan man sige, at selve behaviorismens manglende respons var et symptom på, at denne retning var på vej til at blive et selvkørende paradigme, der ikke viste større interesse for verden uden for Skinner-boksene. Til gengæld står Rosenthals opdagelse af forsøgsleder- eller Pygmalion-effekten som en af de helt store bedrifter i psykologiens historie.

Intern og ekstern validitet

I det foregående afsnit om undersøgelsesbias har vi set eksperimentalspsykologiens kamp for at opdage, afdække og om muligt eliminere den ene form for artefakt efter den anden, indtil en fejlfri metodologi ville fremstå. Det er en kamp, der forekommer lige så heroisk og lige så udsigtsløs som Danaidernes. Her forekommer overgangen fra Rosenthals rotteforsøg til hans skoleforsøg at være tankevækkende. I det første følger Rosenthal en meget eksperimentel fremgangsmåde. I det andet går han over til en mindre ren, men til gengæld også mere virkelighedsnær og relevant form for felteksperiment. Forskellen mellem disse to forsøg kan belyses ud fra begreberne intern og ekstern validitet.

- *Intern validitet* er gyldighed inden for det givne design: den eksperimentelle ramme.
- *Ekstern validitet* er gyldighed af undersøgelsesresultatet uden for det givne design.

Der vil ofte være et modsætningsforhold mellem de eksperimentelt, laboratorieprægede design og de mere virkelighedsnære, felt-prægede. I de første med det strenge, naturvidenskabeligt anlagte design er det meningsfuldt at tale om og tilstræbe intern validitet. Til gengæld vil en afprøvning af resultaternes eksterne gyldighed kræve, at man forlader det strenge

design og går over til en mere virkelighedsnær og dermed også mindre kontrollerbar undersøgelsesmetode.

Imidlertid viser vores gennemgang af de mange former for undersøgelsesbias, at selve kampen for at opnå intern validitet i nogen grad forekommer illusorisk. Efter min opfattelse viser disse artefakter, at selve det mål at udvikle en psykologisk metodologi med samme grad af pålidelighed og gyldighed som i naturvidenskaben kun er muligt inden for den del af psykologiens genstandsområde, som Bruner og Goodman har kaldt det autoktone, mens det for den øvrige og langt større dels vedkommende er dømt til fiasko. Dette undersøgelsesmål og den hertil knyttede eksperimentelle metode indebærer, at den empiriske genstand skal objektiveres lige som en genstand i fysikken, og at forholdet mellem undersøgelsesleder og forsøgsperson skal reduceres til en af designet determineret og af undersøgelseslederen kontrolleret mekanisk virkende envejspåvirkning.

Men som vi har set, unddrager menneskelige forsøgspersoner sig på adskillige måder denne objektivering, ikke fordi de er særlig oprørske, men tværtimod fordi de ofte så loyalt søger at hjælpe undersøgelseslederen til, hvad de tror er det ønskede resultat. Og tilsvarende er det lige så svært at reducere den interpersonelle relation mellem forsøgsleder og forsøgsperson til den tilsigtede mekaniske envejspåvirkning. Sådan er menneskers socialitet simpelthen ikke udformet, og som det fremgår af Rosenthals rotteforsøg er dette åbenbart heller ikke tilfældet for andre pattedyrs vedkommende.

Ud fra denne odysse gennem eksperimentalpsykologiens mange fælder kan vi derfor ikke blot uddrage en lære om psykologisk metodologi, men også om selve naturen af de psykologiske fænomener. Nemlig det centrale faktum, at det menneskelige individ er et subjekt, der aktivt forholder sig til verden og de aktiviteter, det bliver inddraget i.

En anden lige så central kendsgerning er, at forholdet mellem menneskelige individer, og et langt stykke ad vejen også individer blandt de højerestående hvirveldyr, er et forhold mellem to subjekter og ikke et forhold mellem en manipulator og et manipuleret objekt.

Rosnow har sammenlignet de metodologiske problemer i eksperimentalpsykologien med de problemer i kvantemekanikken, der førte til opdagelsen af Heisenbergs usikkerhedsrelation. Sammenligningen er skæv på mange, ja, de fleste punkter, men der er dog også nogle væsentlige paralleller. For det første, at man måtte opgive den præcision, man opererede med i den klassiske fysik, for til gengæld at kunne definere

angivelse af observationsusikkerhed præcist. For det andet, at dette ikke blot var en metodologisk opdagelse, men i høj grad et afgørende bidrag til en helt ny teoretisk forståelse af de udforskede fænomener.

Men hvad betyder det, at pionererne inden for den eksperimentalspsykologiske metodologi i stedet for at finde den fejlfri undersøgelsesmetode fandt nogle advarselsskilte, der fortalte, at noget sådant ikke er muligt inden for størstedelen af det felt, psykologien udforsker?

Den første lære er, at metodologi ikke kan garantere fejlfri resultater, men tværtimod er en måde, hvorpå man gennem opdagelse af metodefejlene kan forbedre undersøgelsesmetoderne. Måske kan man også uddrage den lære, at de tilsyneladende forkromede kvantitative undersøgelsesmetoder er lige så problematiske som de umiddelbart mere løsagtige kvalitative metoder. At der med denne bogs terminologi dybest set er de samme grundproblemer knyttet til de serialiserede metoder som til de kontekstualiserede.

Den anden lære er, at opdagelsen af undersøgelsesfejlene også er en måde, hvorpå man opdager nye træk ved det, vi udforsker. I virkeligheden har nemlig såvel Hawthorne-eksperimenterne som Ornes eksperimenter med fordringseffekten, og Rosenbergs med selvpræsentationseffekten, for slet ikke at nævne Rosenthal med forsøgsledereffekten, alle haft karakter af demonstrationseksperimenter. De er ikke alene metaeksperimenter, der afslører fundamentale forhold ved naturen af det psykologiske eksperiment, men de er samtidig demonstrationer af grundlæggende psykologiske forhold.

Litteratur

- Aerts, D., D'Hondt, E. & Gabora, L. (2000). Why the disjunction in quantum logic is not classical. *Foundations of Physics*, 30 (10): 1-7.
- Asch, S.E. (1952). *Social Psychology*. New York: Prentice Hall.
- Baltes, P.B. (1968). Longitudinal and crosssectional sequences in the study of age and generation effects. *Human Development*, 11: 145-171.
- Bogardus, E.S. (1925). Measuring social distances. *Journal of Applied Sociology* (1925): 299-308.
- Brehm, J.W & Cohen, A.R. (1962). *Exploration in Cognitive Dissonance*. New York: Wiley.
- Bryman, A. (1984). The debate about qualitative and quantitative research: A question of method or epistemology? *The British Journal of Sociology*, 35: 75-92.
- Bruner, J.S. & Goodman, C.C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, 42: 33-44.

- Dilthey, W. (1977). *Descriptive Psychology and Historical Understanding*. The Hague: Martinus Nijhoff.
- Festinger, L. (1956). *When Prophecy Fails*. New York: Harper and Row.
- Festinger, L. & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58: 203-210.
- Gillespie, R. (1991). *Manufacturing Knowledge*. Cambridge: Cambridge University Press.
- Habermas, J. (1965). Analytische Wissenschaftstheorie und Dialektik. I: Ernst Topitsch (Hg.): *Logik der Sozialwissenschaften*. Köln/Berlin: Kiepenheuer & Witsch, 291-311.
- Habermas, J. (1968). *Erkenntnis und Interesse*. Frankfurt: Suhrkamp.
- Hegel, G.W. (1969 [1830]). *Enzyklopädie der philosophischen Wissenschaften*. Hamburg: Verlag von Felix Meiner.
- Karpatschof, B. (2001a). *En evaluering af brugernes oplevelse af SKP-ordningen i Københavns kommune*. Familie- og arbejdsmarkedsforvaltningen, Københavns kommune.
- Horowitz, I.A. (1969). Effects of volunteering, fear arousal, and number of communications on attitude change. *Journal of Personality and Social Psychology* 11: 34-37.
- Karpatschof, B. (2001b). At blive behandlet ordentligt – eller bare blive behandlet. Om brugerstyret kontakt i socialpsykologien og om psykopatologisk reduktionisme. *Psyke & Logos* 22 (2): 695-716.
- Karpatschof, B. (2003a). *Evaluering af Vil:kan – Egmont Fondens forsøgs- og udviklingsprojekt for børn der har mistet forældre eller som har forældre med en livstruende sygdom*. København: Egmont Fonden.
- Karpatschof, B. (2003b). Når sorgens fugle flyver bort – Evalueringen af Egmont Fondens forsøgs og udviklingsprojekt med hjælp til børn i sorg. I: *Egmont Fondens Årsskrift 2002*. København: Egmont Fonden.
- Karpatschof, B. (2004). En person at støtte sig til. I: Kjeld Høgsbro (red.), *Socialpsykiatriens kompleksitet*. Frederiksberg: Samfundslitteratur.
- Karpatschof, B. & Helweg, E. (2003). En tilknytning hinsides døden – om det sørgende barns kamp for en fortsat tilknytning til en afdød far eller mor. *Psyke & Logos* 24 (2):588-620.
- Karpatschof, B. & Rue, K. (2003). En støtte der er brug for – Brugernes oplevelse af støtte-kontakt-person-ordningen i Københavns Kommune. I: Steen Bengtsson (red.), *Handicap, kvalitetsudvikling og brugerinddragelse*. København: AKF.
- Keohane, N.O., Rosaldo, M.Z. & Gelpi, B.C. (eds.) (1981). *Feminist Theory: A Critique of Ideology*. Chicago: University of Chicago Press.
- Latour, B. (1987). *Science in Action*. Cambridge: Open University Press.
- Lincoln, Y. & Guba, E. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage Publications.
- Milgram, S. (1974). *Obedience to Authority*. New York: Harper and Row.
- Millingen, J.G. (1841). *The History of Dueling*, 2 vols. London: Richard Bentley.
- Moreno, J.L. (1953). *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*. New York: Beacon House.
- Orne, M.T. (1951). The mechanism of hypnotic age regression: An experimental study. *Journal of Abnormal and Social Psychology* 46: 213-225.
- Orne, M.T. (1959). The nature of hypnosis: Artifact and essence. *Journal of Abnormal and Social Psychology*, 58: 277-298.
- Orne, M.T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17: 776-783.

- Pedersen, D.O. (2003). *Trivsel, kvalitet og sociale vilkår i by- og boligområder. Evaluering af 15 boligområder i Ballerup og Odense kommuner*. København: Statens Byggeforskningsinstitut.
- Popper, K.R. (1968). *Conjectures and Refutations: The Growth of Scientific Knowledge*. New York: Harper & Row.
- Rosenberg, M.J. (1965). When dissonance fails: On eliminating evaluation apprehension from attitude measurement. *Journal of Personality and Social Psychology*, 1: 28-42.
- Rosenthal, R. (1966). *Experimenter Effects in Behavioural Research*. New York: Appleton Century Croft.
- Rosenthal, R. & Fode, K.L. (1963). The effect of experimenter bias on the performance of the albino rat. *Behavioral Science* 8: 183-189.
- Rosenthal, R. & Jacobson, L. (1966). Teachers' expectancies: Determinants of pupils' IQ gains. *Psychological Reports*, 19: 115-118.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the Classroom*. New York: Rinehart and Winston.
- Rosenthal, R. & Rosnow, R.L. (1969). *Artifacts in Behavioural Research*. New York: Academic Press.
- Rosenthal, R. & Rosnow, R.L. (1975). *The Volunteer Subject*. New York: Wiley.
- Sartre, J.P. (1972 [1960]). *Kritik af den dialektiske fornuft*. Udvalg og indledning ved Peter Kemp. København: Vintens Forlag.
- Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, 103: 677-680.
- Stevens, S.S. (1959). *Measurement, psychophysics and utility*. I: C.W. Churchman & P. Ratoosh (eds.), *Measurement: Definitions and Theories*. New York: John Wiley.
- Galileo, G. (1957 [1623]). The Assayer. I: Stillman Drake, *Discoveries and Opinions of Galileo*. New York, NY: Doubleday.
- Suls, J.M. & Rosnow, R.L. (1988). Concerns about artifacts in psychological experiments. I: J. Morawski (ed.), *The Rise of Experimentation in American Psychology* (s. 163-187). New Haven, CT: Yale University Press.
- Windelband, W. (1894). *Geschichte und Naturwissenschaft*. Internet: http://www.fh-augsburg.de/~harsch/germanica/Chronologie/19Jh/Windelband/win_rede.html (Fachhochschule Augsburg).
- Zahavi, D. (2001). *Husserls fenomenologi*. København: Gyldendal.

To close kinship by marriage

To my club as personal chums

To my street as neighbors

To employment in my occupations in my country as citizen Citizenship occupations

As visitors only to my country Citizenship occupations

Would exclude from my country my country Citizenship occupations

Noter

Note 1, s. 7 Metodologi er dannet af de to græske ord *methodos* og *logos*, der betyder henholdsvis metode og tanke. Methodos er atter dannet af *meta* og *hodos*, hvor meta er et forholdsord, der betyder med, mens hodos betyder vej.

Note 2, s. 12 Socialkonstruktivismen er den videnskabsteoretiske doktrin, at videnskabelige størrelser som empiri og teori blot er sociale konstruktioner, lige som andre samfundsmæssige frembringelser. Sandheds- og objektivitetsbegrebet må herudfra opgives (Latour, 1987). – Læseren henvises i øvrigt til en mere samlet videnskabsteoretisk fremstilling, fx Harnow Klausen (2005) eller Christensen (2002).

Note 3, s. 13 Betegnelsen går tilbage til den klassiske græske filosofi, idet ordet er sammensat af *psyche* = sjæl (oprindelig åndedræt) og *logos* = tanke.

Note 4, s. 19 Opdagelsen af den såkaldte superposition i kvantemekanikken, hvor to modstridende tilstande kan foreligge samtidigt, var ikke bare et chok for vores opfattelse af fysiske fænomener, men også mht. logikkens status. Fra at være noget principielt ikke-empirisk og uforanderligt, så man her et eksempel på, at ingen af delene er tilfældet. Se fx (Aerts, D'Hondt & Gabora, 2000).

Note 5, s. 23 Bogardus (1925) foreslog følgende skala til angivelse af den sociale afstand mellem en person fra et givet land og en person fra et opgivet andet land: [figur se manus!! når den skal fremstilles].

Note 6, s. 23 Moreno, der var psykoterapeut og bl.a. opfandt psykodrama, udviklede i 1936 den sociometriske metode, som ved hjælp af simple angivelser af indbyrdes preferencerelationer mellem medlemmer i en gruppe søger at finde (og grafisk illustrere) gruppens sociale struktur.

Note 7, s. 25 Hegel (1969, s. 112) indførte i sin logik en skelnen mellem den slette og den sande uendelighed, hvor den førstnævnte netop står for en cirkelbevægelse, en penduleringen mellem fx to tilstande uden nogen udvikling. Et særlig grelt eksempel vil være det bipolære manio-depressive syndrom, hvor man skifter mellem dyb fortvivelse og eksalteret begejstring. Den sande uendelighed er derimod karakteriseret ved, hvad han kaldte en ophævelse, dvs. at der skabes en ny kvalitet ud over eller oven over de to oprindelige.

Note 8, s. 40 Bryman (1984) peger på feltundersøgelsen som kvalitativ metode, og Lincoln og Guba (1985) understreger som særlige værdier ved kvalitativ forskning, troværdighed, mulighed for overførsel og bekræftelse mellem forskellige studier.

Note 9, s. 45 Således udgives i 1690 lægen og økonomen Sir William Pettys efterladte bog *Political Arithmetick* med den imponerende (og karakteristiske)

undertitel: *or a Discourse Concerning, The Extent and Value of Lands, People, Buildings: Husbandry, Manufacture, Commerce, Fishery, Artizans, Seamen, Soldiers; Publick Revenues, Interest, Taxes, Superlucration, Registries, Banks, Valuation of Men, Increasing of Seamen, of Militia's, Harbours, Situation, Shipping, Power at Sea, &c. As the same relates to every Country in general, but more particularly to the Territories of His Majesty of Great Britain, and his Neighbours of Holland, Zealand, and France.*

Note 10, s. 47 "Erfaringsvidenskaberne søger i erkendelsen af det virkelige enten det almene i form af naturlovene eller det specifikke i dets historisk bestemte form; de betragter på den ene side den form, der altid forbliver den samme, på den anden side det unikke bestemte indhold i den virkelige hændelse. Den ene form for erfaringsvidenskaber er lovvidenskaberne, den anden begivenhedsvidenskaberne; førstnævnte lærer, hvad der altid er, sidstnævnte hvad der var engang." Fra en tale, Windelband holdt om historie og naturvidenskab (Windelband, 1894).

Note 11, s. 48 Pedersen (2003) anfører faktisk det modsatte resultat, at trivsel er større i enfamilies- end i rækkehuse, men denne undersøgelse arbejder med en ret primitiv multivariat statistisk metode, der indebærer, at resultatet kan være misvisende på grund af skjulte faktorer (eng. confounders), fx at enfamiliehuse er ejerboliger, mens en del af rækkehusene er lejerboliger.

Note 12, s. 50 A. Evaluering af Støttekontaktpersonordning (Karpatschof, 2001a; Karpatschof & Rue, 2003; Karpatschof, 2004). B. Evaluering af rådgivning for børn i sorg (Karpatschof 2003b, Karpatschof & Helweg 2003).

Note 13, s. 51 Når køn er så yndet et eksempel, hænger det sammen med, at denne kategorisering ikke har de indbyggede fortolkningsproblemer, som de fleste psykologiske og samfundsvidenskabelige kategoriseringer. Og så er det den simpleste form for kategorisystem, nemlig den dikotome, der kun har to kategorier. Som biologisk bestemt synes denne kategorisering uproblematisk, men faktisk deler den kategoriseringsproblemerne med dens åbenlyst upræcise psykologisk-samfundsvidenskabelige slægtninge. Der er nemlig også her en mellemkategori, de transseksuelle, hvor der ikke blot er konkrete praktiske problemer med afgørelsen af, hvilken kategori der skal anvendes, men hvor der faktisk i visse tilfælde er flere forskellige måder at tilskrive en person et bestemt køn på, og hvor det derfor måske er mest rimeligt at indføre en tredje kategori, nemlig de transseksuelle. Et sådan skridt er imidlertid også problematisk, fordi det tydeligvis er en heterogen rodekasse.

Note 14, s. 63 En af metoderne anvender universitetets lister over alle studerende på et undervisningshold, og når der sker tilmelding til at deltage i et forsøg, sammenlignes de studerende som melder sig med dem som ikke gør det.

Note 15, s. 73 Dette gælder, om end Rosenberg muligvis selv bliver offer for andre artefakter.

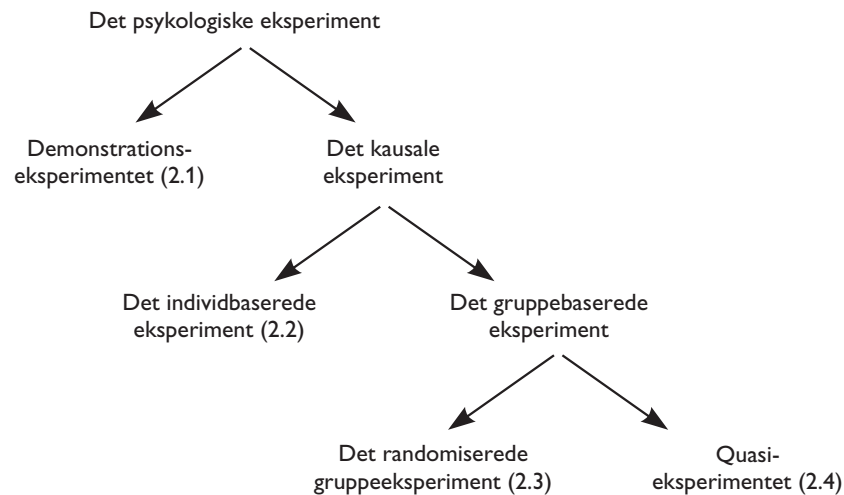
KAPITEL 2

De eksperimentelle metoder

Det eksperimentelle spektrum

Den eksperimentelle metode har en særlig position i psykologien. Det var den metode, der i midten af 1800-tallet definerede faget som en empirisk videnskab, og derfor har eksperimentet også i en lang periode været opfattet som den bærende metode i psykologien. Det er ikke min opfattelse, da jeg ser de forskellige metoder som ligeværdige bidrag inden for en videnskabelig arbejdsdeling, nødvendiggjort af psykologiens meget store bredde og heterogenitet.

På grund af sin særstilling som det videnskabelige adelsmærke, der skabte psykologien, betragtes det psykologiske eksperiment ofte som en



Figur 2.1. Psykologiens spektrum af eksperimentelle metoder.

entydig metodologisk størrelse knyttet til den mest traditionstunge, naturvidenskabelige del af psykologien. Dette er imidlertid en misforståelse. Ordet *eksperiment* bruges om et spektrum af empiriske fremgangsmåder med meget forskelligt sigte og lige så forskelligartede metodologiske problematikker. Jeg vil her følge Baade (1916) med hans opdeling i kategorierne *demonstrationseksperimentet*, som primært søger at påvise og synliggøre et bestemt psykologisk fænomen, og *det kausale eksperiment*, som systematisk søger at afdække de kausale betingelser for fremkomsten af bestemte kvaliteter knyttet til et givet problemkompleks. Ud fra denne hovedopdeling, skal vi derefter gå videre til en underopdeling af den sidstnævnte eksperiment-type, idet der også inden for den kausale eksperimenttype er tale om meget forskellige former (figur 2.1).

Baade var som tysk psykolog fra begyndelsen af 1900-tallet både skolet i den eksperimentelle og den introspektive tradition. Han var ligesom Edgar Rubin elev af G.E. Müller. I sine to metodologiske artikler skelner han mellem *das Darstellungseksperiment* og *das kausale Eksperiment*, idet han gør opmærksom på, at der er tale om en begrebsmæssig, snarere end om en kategorisk opdeling. Torsten Ingemann Nielsen (1978) har oversat de to begreber til henholdsvis *demonstrationseksperiment* og *kausalt eksperiment*. Ved *demonstrationseksperimentet*¹ er hensigten at påpege et psykologisk fænomen for kolleger, studerende eller offentlighed. De fleste af gestaltpsykologernes eksempler på gestaltlovene og Rubins figur-grund-motiver har denne karakter af at fremvise et psykologisk fænomen, som ikke har været begrebsliggjort tidligere, eller hvor nye aspekter ønskes fremdraget.

Heroverfor sætter Baade det kausale eksperiment, der måske nok også undertiden kan fungere som en demonstration af et bestemt psykologisk fænomen, men hvis primære målsætning går ud over dette, idet målet for eksperimentet er en *kausal afdækning*, for ikke at sige en *kausal reduktion* af fænomenet. Her viser det kausale eksperiment klart sin oprindelse i de naturvidenskabelige, sansefysiologiske eksperimenter, hvorigennem psykologi blev grundlagt som en eksperimentel videnskab omkring midten af 1800-tallet.²

Ved en kausal afdækning vil jeg forstå en begrebsmæssig og operationel rekonstruktion af et psykologisk fænomen, således at dette kan fremstilles som virkningen af nogle årsagsfaktorer i en principielt deterministisk virkende kausalmekanisme. For så vidt det lykkes at understøtte en fuldstændig kausal determinisme, kan man tale om, at der ikke blot foreligger en kausal afdækning, men en kausal reduktion. Ved det kausale eksperi-

ment er det altså entydighed og ikke mangetydighed, der er forsknings-idealet. Da dette kapitel fokuserer på det kausale eksperiment og dets undertyper, vil specifikationen *kausal* her blive udeladt, således at det kun er ved *demonstrationseksperimentet* et præciserende præfiks vil blive anvendt.

Jeg skal her opdele det kausale eksperiment ud fra, hvad der er dets empiriske basis. Det klassiske psykologiske eksperiment er *individbaseret*, og i denne form anvendes det stadig i den kognitive psykologi. I sine senere faser, hvor nye områder udviklede sig inden for psykologien, fx socialpsykologi og udviklingspsykologi, har en ny metodologisk tilgang, som er inspireret af lægevidenskabens *gruppebaserede* eksperimenttype vundet frem.

Demonstrationseksperimentet

Hvis man overser Baades distinktion mellem demonstrationseksperimentet og det kausale eksperiment, kommer mange af de mest banebrydende af psykologiens eksperimenter let til at fremstå som metodologisk underlødige, sjusket gennemført og derfor også fejlbehæftede. Det skyldes, at demonstrationseksperimentet har en radikalt anden målsætning, og at der følgelig må stilles andre metodologiske krav til det. Målet er nemlig slet ikke en kausal reduktion, men derimod en demonstration. Ikke en dokumentation af alle de *bestemmende årsagsfaktorer* for et fænomen, men derimod en *anskuelliggørelse af selve fænomenet*. Dette udelukker ikke overgangsformer, der har et dobbelt sigte, således at man på en gang søger at demonstrere et fænomen og at afdække nogle afgørende kausalfaktorer, men i så fald vil der ikke i samme grad som ved det rene kausaleksperiment være tale om en systematisk afdækning.

Milgrams lydighedsforsøg

Hvis vi vender tilbage til Milgrams (1974) eksperiment, der blev omtalt i kapitel 1, er det tydeligt, at han faktisk ikke udvikler en operationalisering af nogle teoretiske begreber og begrebssammenhænge, men snarere foretager en dramaturgisk iscenesættelse af den måde, hvorpå lydighed og specielt lydiggiværksættelse af grusomheder kommer til udtryk i autoritære systemer. I forhold til de egentlig kausale eksperimenter, som skal omtales

nedenfor, er der ved demonstrationseksperimentet ofte tale om komplekse psykologiske fænomener, for hvilke der hverken foreligger en stringent teori, en præcis operationalisering eller et forsøg på en kausal reduktion. Ganske vist foretog Milgram en slags operationalisering af grusomhedens niveau og alvor, nemlig en voltangivelse af strømmens spænding, således at der kunne foretages en slags tærskelbestemmelse af, hvor langt en forsøgsperson ville gå i retning af smertefuld og evt. livsfarlig afstraffelse. Men selv denne kvantificerede variabel har primært en påpegende, dramaturgisk³ funktion. Der er nemlig ikke tale om yderligere operationaliserede kausalfaktorer, som hvis fx også presset fra forsøgslederen var graderet efter styrke. Dette var øjensynlig ikke relevant for Milgram. Hans intention har været at demonstrere, i hvor høj grad iværksættelse af grusomheder er et alment socialt fænomen, og hvor let det er at få psykologisk upåfaldende personer til at udføre overgreb på sagesløse mennesker. En detaljeret afdækning af lydighedens kausale struktur har Milgram med rette opfattet som både uigennemførlig og perspektivløs.

Det er her karakteristisk, at det ved det typiske demonstrationseksperiment er den eksterne, økologiske og ikke den interne validitet, som er afgørende. Da der ikke er tale om et kausalt filigran af årsagsfaktorer, men om et både kompliceret og livsnært fænomen, er eksperimentets hensigt ikke at bruge laboratoriet til at reducere virkeligheden til en kausal mekanisme, men derimod at instrumentere en dramatisk forenklet og fortættet opførelse af et psykologisk skuespil. Et skuespil, der lige som hos de store dramatikere, viser os en side af virkeligheden, der er så påtrængende og indlysende, at den åbenbarer nogle væsensforhold, der ellers drukner i det mangefacetterede og uoverskuelige dagligliv, hvor det demonstrerede fænomen har sit udspring.

Bruners møntforsøg

Bruner udførte fra slutningen af fyrrerne en række forsøg så banebrydende, at de blev kaldt for New Look-psykologien.⁴ Det første i denne række var Bruner og Goodmans forsøg (1947).⁵

Som nævnt i forrige kapitel skelner Bruner mellem de *autoktone* og de *adfærdsmæssige* psykologiske fænomener, og det foreliggende forsøg er et demonstrationseksperiment i den dobbelte forstand, at det både belyser et psykologisk *fænomen* og en psykologisk *undersøgelsesmetode*. Bruners eksperiment demonstrerer for det første et uhyre vigtigt psykologisk

fænomen, nemlig motivationens betydning for vores perception. For det andet bryder det nyt land inden for psykologisk forskning ved lige som Asch, der omtales i det følgende, at opstille et kreativt design, der har intern gyldighed og samtidig økologisk relevans. Begge disse forsøg bryder også mure ned mellem hidtil adskilte områder af psykologien. Hvor Asch kombinerer kognitions- og socialpsykologi, undersøger Bruner forholdet mellem perception og motivation. Ja, faktisk viser han, at selve adskillelsen af de to områder er en abstrakt adskillelse inden for det psykologiske område, der omfatter menneskets opfattelse og bedømmelse af verden.

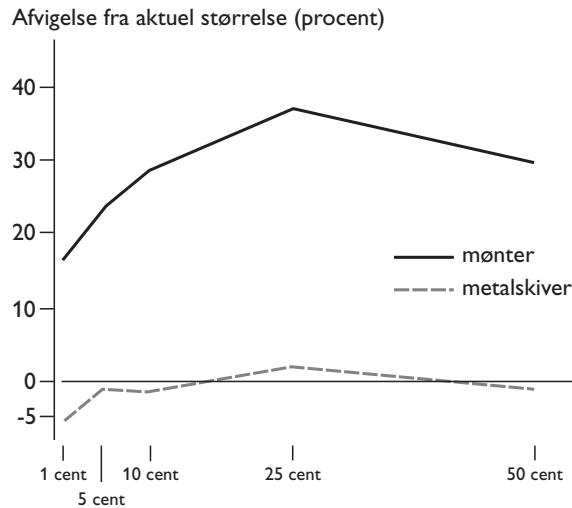
Det forsøg, Bruner og Goodman udfører i 1947, tager på tilsyneladende ortodoks positivistisk vis udgangspunkt i nogle almene hypoteser, der skal afprøves eksperimentelt. Men hvor der ofte inden for eksperimentel psykologi arbejdes med snævert operationaliserede empiriske hypoteser, er der her tale om teoretiske teser af stor relevans og vidtgående rækkevidde. Når vi ser bort fra en tredje hypotese af mere teknisk karakter, kan disse hypoteser formuleres på følgende måde:

1. *Jo store social værdi et objekt har, i desto højere grad vil opfattelsen af det blive styret af adfærdsmæssige determinanter.* Objektet vil blive *udvalgt* perceptuelt blandt andre perceptuelle objekter, vil blive *fikseret* som en perceptuel responstendens og vil blive perceptuelt *accentueret*.
2. *Jo store det individuelle behov for et socialt værdifuld objekt er, desto mere udpræget vil de adfærdsmæssige determinanter gøre sig gældende* (Bruner & Goodman, 1947).

Til begge disse teser udviklede Bruner og Goodman et apparat bestående af en kasse, hvor man på skift så et objekt og en lysende ring, hvis størrelse kunne indstilles ved hjælp en knop, så den svarede til, hvordan forsøgspersonen opfattede objektets størrelse. Forsøgspersonerne var 10 drenge på 10 år.

Hypotese 1 afprøves nu ved først at lade forsøgspersonerne indstille lysringen, så den svarer først til nogle metalskiver og derefter til mønter af forskellig værdi, nemlig 1, 2, 5, 25 og 50 cent (i dagens danske møntfod, ca. 50 øre, 1 kr., 2,50 kr., 12,50 kr. og 25 kr.).

Figur 2.2 viser resultatet af det første delforsøg.



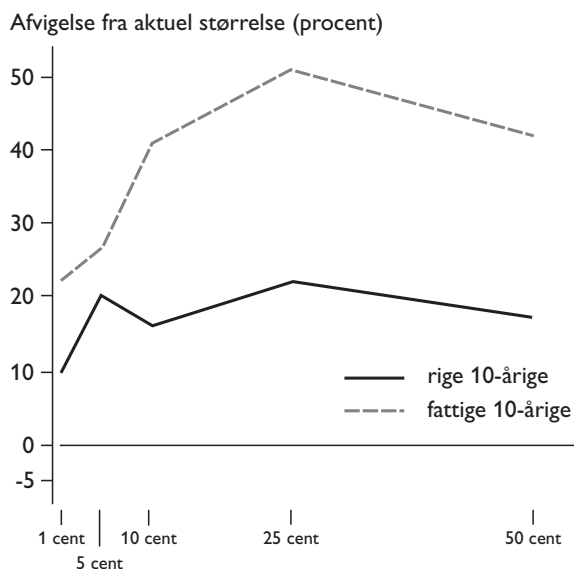
Figur 2.2. Den sociale værdi som organisationsprincip.

Vi ser her, hvordan drengene indstiller lysringen til en størrelse, der med ganske små afvigelser svarer til de værdimæssigt neutrale metalskiver, hvorimod mønternes størrelse bliver væsentlig overvurderet, og denne overvurdering vokser med værdien af mønten.

Der er dog en enkelt undtagelse fra denne tendens, idet der faktisk er et fald fra en kvart til en halv dollar. Hertil har forfatterne en forklaring, der ganske vist *post hoc*, for ikke at sige *ad hoc*,⁶ men som lyder ganske plausibel. I denne sparsommelige efterkrigstid er nemlig en halv dollar, som svarer til 25 kr. i dagens værdi, en ret ukendt og urealistisk størrelse for de fleste 10-årige drenge.

Hermed er den første tese afprøvet, og det med stort held. Bruner og Goodman gik nu over til den anden forsøgsserie. Her var der tale om to forsøgsgrupper udvalgt ud fra en baggrundsvariabel, nemlig drengenes økonomiske baggrund. Gruppe 1 var drenge fra et velhaverkvarter i Boston, hvor de gik på en såkaldt "progressiv skole", hvad der nærmest svarer til en dansk lilleskole. Gruppe 2 var fra et slumområde i samme by.

Den specifikke hypotese udledt af tese 2 var nu, at de fattige drenge, som havde et større behov for penge end de rige, ville opfatte mønterne som større end de rige drenge. Resultatet fremgår af figur 2.3.



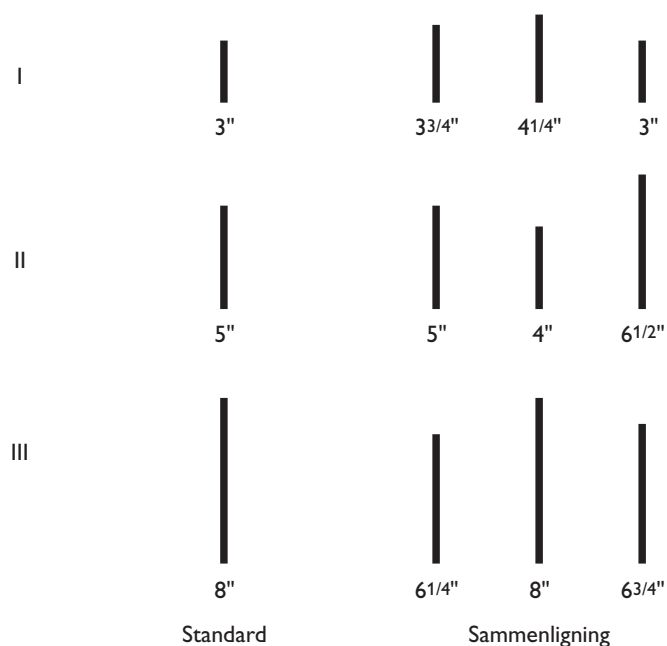
Figur 2.3. Det personlige behov som organisationsprincip.

Vi ser, at der ligesom før er tale om en overvurdering af mønterne i begge grupper. Desuden er der en betydelig højere grad af overvurdering blandt de fattige drenge. I denne gruppe ser vi en gentagelse af den voksende grad af overvurdering gennem rækken fra 1 til 25 cent, hvorefter der er en mindre nedgang til den højeste værdi på en halv dollar, som angiveligt ligger uden for disse drenge ikke blot fysiske, men også forestillingsmæssige rækkevidde. Denne ad hoc-forklaring fra eksperimentators side svækkes dog af det forhold, at den aftagende overvurdering også ses hos de velhavende drenge, for hvem en halvdollarmønt ikke i samme grad er uden for rækkevidde.

Asch's eksperiment med gruppepres

Et banebrydende eksperiment blev i 1951 udført af Asch, som med en yderst manipulativ, men i øvrigt ret simpel opstilling demonstrerede virkningen af gruppepres. Han lod en gruppe studerende foretage sammenligninger af linjestykker (se figur 2.4). Opgaven bestod simpelthen i at angive, hvilket af de tre linjestykker til højre, som matchede standardstykket til venstre. Af de tre sammenligningsstykker til højre var den ene identisk med standardstykket. Af de to øvrige var det ene kun moderat

forskelligt, mens det sidste var „ekstremt“ forskelligt. At der faktisk er tale om så ekstreme forskelle, at intet menneske med normalt syn under almindelige omstændigheder vil være i tvivl, fremgår af, at den maksimale forskel i øverste række af figuren faktisk er på godt 40 %, mens den moderate forskel er på 25 %.



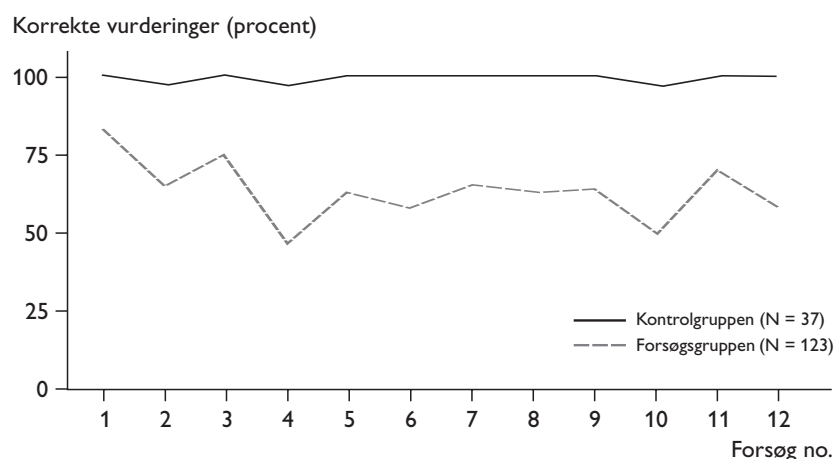
Figur 2.4. Tre eksempler på stimulus i Asch's forsøg.



Figur 2.5. Nærbillede af øverste række i ovenstående figur.

Gruppen blev delt i mindre hold på mellem seks og ni personer, men kun en person i hver gruppe var den egentlige forsøgsperson. De øvrige (kaldet flertallet) var falske forsøgspersoner, der var instrueret i at afgive på forhånd aftalte svar, som blev afgivet *inden* den ægte forsøgsperson udtalte sig. Alle de 123 forsøgspersoner blev (efter to prøveforsøg) præsenteret for i alt 16 forsøg. Heraf var de fire neutrale forsøg, hvor flertallet gav korrekte svar. I de øvrige 12 forsøg gav flertallet henholdsvis moderat falske svar (seks forsøg) og ekstremt falske svar (seks forsøg).

Ved siden af denne forsøgsgruppe var der en kontrolgruppe på 37 personer, der ikke var udsat for dette manipulerende gruppepres. Som det fremgår af figur 2.6, var fejlprocenten her praktisk taget nul (0,7 procent), og hele 95 procent af disse kontrolpersoner begik slet ingen fejl.



Figur 2.6. Forskellen på korrekthedsprocent i kontrol og forsøgsgruppe.

I forsøgsgruppen var den samlede fejlprocent derimod 37 (figuren viser en kurve, der ligger omkring 63-niveauet). Hvad der måske er endnu mere sigende, var der her kun 24 procent af personerne, som ikke begik nogen fejl. Langt de fleste i forsøgsgruppen var således påvirket af gruppepresset i større eller mindre grad.

Det skal her indskydes, at Asch faktisk havde en yderligere eksperimentel variabel, nemlig en selektionsvariabel, idet forsøgspersoner kom fra tre sociale strata, som han antog ville vise uafhængighed i forskellig grad. Den *ex hypothesis* mest uafhængige gruppe var studerende fra et lille udsøgt privat

college og med velstående, højtuddannede familier, der længe havde boet i USA.

Disse studerende var meget intelligente og opdraget til selvstændighed og tro på sig selv. Den anden gruppe bestod af studerende fra et middelklassepræget omegnsscollege og var i vid udstrækning 2. generationsindvandrere. Den sidste gruppe studerende kom fra lavere middelklassehjem, og deres uddannelsesmæssige og intellektuelle udvikling var angivet som dårligere end de to andre grupper. Der er ganske vist også en tendens i retning af en lidt større uafhængighed i den første, privilegerede gruppe, idet fejlprocenten her kun er på 33 procent mod 41 procent for de to øvrige grupper under ét. Men denne forskel er ikke signifikant, og i øvrigt var der heller ikke den forventede tendens til, at den socialt lavest placerede gruppe klarede sig dårligst mht. modstand mod gruppepres.

Dette er de kvantitative resultater, men da Asch var nært knyttet til den gestaltpsykologiske skole, foretog han også en kvalitativ analyse på baggrund af et interview med forsøgspersonerne efter forsøget.

Her fandt han, at forsøgspersonerne havde en række forskellige måder at forholde sig på til det massive gruppepres, de havde været udsat for. Disse måder at forholde sig på har han inddelt i to hovedgrupper med hver sine undergrupper. Hovedgrupperne var *uafhængighed* og *eftergivenhed*. Undergrupperne i disse hovedgrupper er vist nedenfor:

A. Uafhængighed

1. Uafhængighed præget af styrke

Forsøgspersonen er temmelig uanfægtet af flertallet.

2. Uafhængighed uden sikkerhed

Forsøgspersonen fastholder sin vurdering, men er rystet over uenigheden.

B. Eftergivenhed

1. Eftergivenhed på det perceptuelle plan

Forsøgspersonen er så motiveret for enighed, at han/hun ser det samme som de andre.

2. Eftergivenhed på bedømmelsesplanet

Forsøgspersonen er så selvusikker, at han/hun mener, de andres dom er korrekt.

3. *Eftergivenhed på handlingsplanet*

Forsøgspersonen er så bange for eksklusion, at han/hun undgår konflikt.

Også en stor del af de forsøgspersoner, der modstod gruppepresset, var altså meget følelsesmæssigt påvirket. I gruppen af forsøgspersoner, der gav efter, havde nogle særdeles kraftige psykologiske reaktioner, ja, nogle af reaktionerne ville man umiddelbar bedømme som patologiske, hvis man ikke vidste, at der var tale om helt normale reaktioner på gruppepres hos helt normale studerende. Den enkle og hårdtslående konklusion af Asch's forsøg er altså, at de fleste af os, uanset vores sociale baggrund, kan blive ofre for gruppepres i større eller mindre grad.

Asch's demonstration af virkningen af socialt pres, som trods den manipulative form havde et ret uskyldigt indhold, blev, som vi allerede har omtalt, nogle år senere suppleret med Milgrams påvisning af, at presset også virker, når det ikke kun drejer sig om uskyldige linjestykker, men præsenteres som et spørgsmål om liv og død. I øvrigt er Asch's pointe blevet påvist så mange gang i livet uden for laboratoriet, at den må siges at være særdeles godt eksternt valideret.

Asch's forsøg er sammen med Bruners valgt som paradigmatiske eksempler på et demonstrationseksperiment og er med rette blevet en af psykologiens klassikere i sin forening af følgende kvaliteter:

1. et for sin tid overraskende og dristigt, men i øvrigt ret *enkelt design*,
2. en dramatisk og overbevisende iscenesættelse – *dramaturgisk overbevisningskraft*,
3. en psykologisk grænsesprængende forening af et kognitivt og et socialpsykologisk område – *teoretisk betydning*,
4. en *livsnær relevans*, der rækker langt ud over den fagpsykologiske kreds, og med en problemstilling, der indskriver sig i en politisk og moralsk sammenhæng.⁷

Schachter og Singers undersøgelse over emotionernes natur vil blive behandlet i det følgende som et eksempel på et gruppebaseret eksperiment. Det er således primært et egentligt kausaleksperiment, men skal alligevel nævnes her, da det også har nogle af de træk, der karakteriserer et godt demonstrationseksperiment nemlig en klar problemstilling, der kombinerer almenpsykologisk og almenmenneskelig relevans.

Det individbaserede (klassiske) eksperiment

Ved psykologiens etablering som erfaringsvidenskab stod den klassiske form for eksperiment i centrum. Opbygningen er hentet fra naturvidenskaben, først og fremmest fysikken, og det har følgende karakteristika:

1. Som empirisk genstand anvendes en eller flere (oftest meget få) forsøgspersoner. Da denne undersøgelsestype typisk foretages inden for den eksperimentelle del af den almene psykologi, ofres der sjældent mange tanker på udvælgelse. Man skal afdække et alment psykologisk fænomen, der antages at forekomme på principielt samme måde hos alle normale personer.
2. Hver forsøgsperson udsættes for en række enkeltforsøg, som hver især er karakteriseret af en bestemt påvirkning. I sin simple form er der tale om præcis samme forsøgsrække for alle forsøgspersoner.
3. Designet tilsigter dels en fuldstændig kontrol over påvirkningen i det enkelte forsøg, dels en systematisk variation i de eksperimentelle variable, som operationaliserer eksperimentets kausalfaktorer. Denne kontrol af påvirkningssituationen er således bestemt ad to veje:
 - a. den *fikserede ramme*, dvs. de *konstante* aspekter, altså forhold der fastholdes gennem alle delforsøg.
 - b. de *eksperimentelle*, uafhængige, variable, som *manipuleres* af forsøgslederen efter et kombinatorisk mønster fastlagt i eksperimentets design.
4. I hvert enkeltforsøg observeres forsøgspersonens respons i form af en *tilsigtet handling* eller en *utilsigtet adfærd*. Fastholdelsen af responsen sker ved anvendelse af et på forhånd opstillet registreringssystem, fx en maskinel registrering eller et kategorisystem, hvorigennem forsøgspersonen skal angive sin adfærd. Hvis der foretages en simultan registrering af flere aspekter ved responsen, vil der foreligge flere responsdata pr. enkeltforsøg, som dermed udgør eksperimentets afhængige variable.

Den *erkendelseskritiske* (se s. xx) antagelse er nu, at den eller de afhængige (observerede) variable er fuldstændigt bestemt af de eksperimentelle variable. Eksperimentet må således i første omgang foretage en afprøvning af denne antagelse:

- *Når den fikserede ramme fastholdes, vil de eksperimentelle faktorer forårsage en respons, specificeret ved eksperimentets observationsvariabler.*

At den erkendelseskritiske antagelse er korrekt er ensbetydende med, at eksperimentet besidder intern validitet. Omvendt indebærer en mangel på intern validitet, at der foreligger kausalt bestemmende forhold *uden* for eksperimentets kontrol.

Det *erkendelsesrettede* (se s. xx) aspekt af eksperimentet er enten en eksplorativ fastsættelse af en kausal hypotese eller en afprøvning af en allerede opstillet teori om en bestemt kausal mekanisme.

Hvis vi ser på beskrivelsen i figur 1.12 (s. xx), kan det komplette eksperiments karakteristika indordnes på følgende måde:

A. Udvalgelsesplan

- Normalt er kun udvalgelsesgrundlaget beskrevet, og ofte kun gennem eksklusionskriterier, fx kravet om, at forsøgspersonerne skal have normal hørelse og normalt syn.
- Udvalgelsesmetoden betragtes som irrelevant, da individuelle karakteristika antages at være uden betydning.

B. Iscenesættelsesplan

- I. Tid: Typisk er der ingen specifik tidslig struktur, da præsentationen er helt underlagt konstellationen af de eksperimentelle variable.
- II. Sted: Den rumlige iscenesættelsesplan er her defineret ved den fikserede ramme.
- III. Våben: Interventionen er givet ved de eksperimentelle variable.

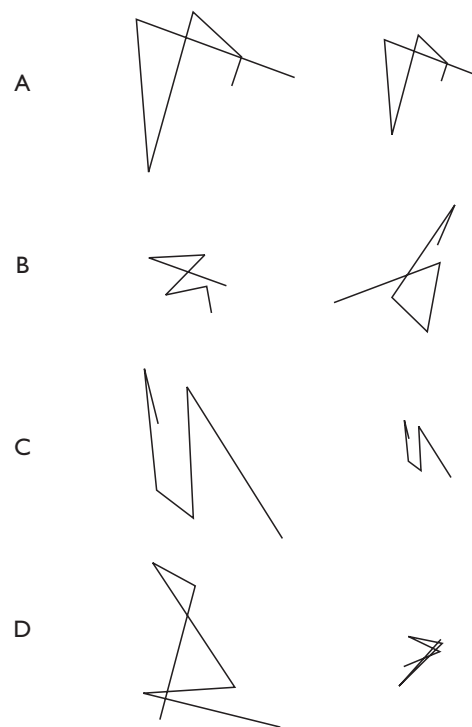
C. Observationsplan

- Observationsplanen er defineret gennem observationsvariablene.

Et eksempel på et individbaseret eksperiment – Mental transformation af visuelle figurer

Som eksempel på et individbaseret (klassisk) eksperiment skal vi her se på et forsøg med mental transformation udført i første halvdel af 1970'erne af Claus Bundesen og Axel Larsen (Bunden, 1975, 1981). Det undersøgte fænomen er mentale transformationer af visuelle figurer. Det teoriområde,

der forskes i, er altså visuel perception med henblik på at finde de basale neuropsykologiske mekanismer, som styrer figurgenkendelse. Forsøget koncentrerer sig om *formgenkendelse*, idet der præsenteres to forskellige figurer efter hinanden, og forsøgspersonens opgave er at afgøre, om de har samme form og samme placering eller forskellig form.



Figur 2.7. Eksempler på figurpar i eksperiment nr. 1 (Bundesen, 1975:215).

I figur 2.7 ses fire eksempler på figurpar i den første forsøgsrække (eksperiment nr. 1):

1. A består af et par med samme form og samme placering (drejningsvinkel), men forskellig størrelse. Her er det rigtige svar altså, at figurerne er ens.
2. B indeholder et par med såvel forskellig form, placering og størrelse. Det rigtige svar er her, at figurerne er forskellige.

3. C har lige som A et par med samme form og samme placering, men forskellig størrelse. Her er det rigtige svar, at figurerne er ens.
4. D har lige som B et par med forskellig form, placering og størrelse. Det rigtige svar er, at figurerne er forskellige.

Den teoretiske problemstilling, forsøgets epistemiske genstand, er som sagt en basal teori for formgenkendelse. Her var der på forsøgets tidspunkt fremsat to meget forskellige teorier, nemlig en *feature*-teori og en *skalering*-teori, i afhandlingen kaldet F- og S-teorien:

F-teorien

Denne teori antager, at det, der betinger opfattelsen af form, er størrelsesuafhængige geometriske egenskaber eller træk (eng. *feature*, deraf F-teori) ved en figur. For eksempel om den er rund eller kantet, og i det sidste tilfælde, hvor mange kanter den har, hvor spidse vinklerne er, og hvor lange linjestykkerne er. For figur A's vedkommende ville man ud fra feature-teorien måske definere formen, som bestemt af følgende træk:

- *en central mod højre let skrånede linje, med en stor, spidsvinklet trekant nedenunder og en lille stumpvinklet trekant ovenover.*

Teorien opererer altså med, hvad vi kan kalde højere kognitive processer, hvor der ikke er tale om autoktone, men om adfærdsmæssige processer i Bruners forstand (se s. xx)

S-teorien (skalering)

Denne teori er i modsætning til den forrige ikke kvalitativt, men kvantitativt formuleret. Man kan også sige, at vi går fra det digitale til det analoge. Ifølge S-teorien finder der nemlig en simpel analog hjerneproces sted, hvorved den ene figur eller begge figurer (efter en evt. drejning, så de vender i samme retning) bliver skaleret sådan, at de får samme størrelse, så de mentalt kan anbringes oven på hinanden, for at se om de dækker hinanden. Hvis de efter skaleringen dækker hinanden, har de samme form. Hvis de derimod efter skaleringen ikke dækker hinanden, har de forskellige form.

Til empirisk afprøvning af disse teorier, udledes der nu hver sin konsekvens af de to konkurrerende teorier:

- Ifølge F-teorien må bearbejdnings- og dermed reaktionstiden være den samme uanset størrelsesforholdet mellem figurer.
- Ifølge S-teorien må bearbejdningstiden derimod være mindst, når figurerne har samme form, og derfor ikke skal skaleres, mens den vokser med størrelsesforskellen er mellem figurerne.

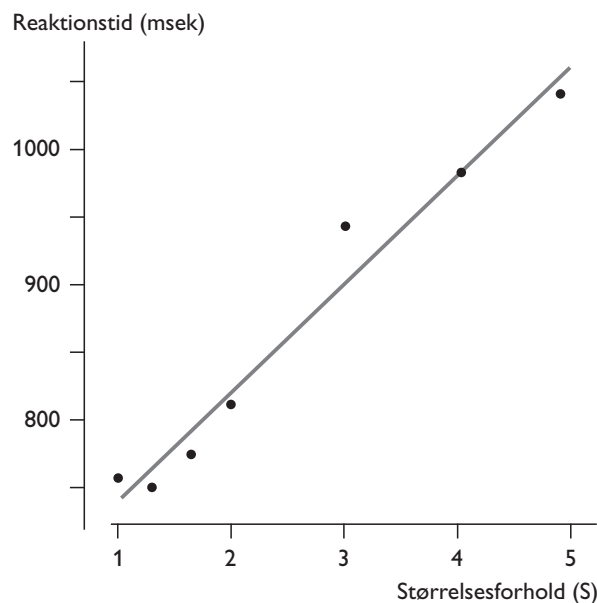
Vi har altså et hypotetisk-deduktivt eksperiment, hvor vi ud fra to konkurrerende teorier kan udlede *modstridende* hypoteser, der er testet gennem et afgørende eksperiment (eng. *crucial experiment*).

Det betyder, at vi som uafhængige variable må have:

1. figurparrets *formrelation* (ens, forskellig),
2. deres *størrelsesforhold*

Som afhængig variabel, observationsvariabel, må vi have *reaktionstiden*.

Det sidste indebærer, at vi skal anvende en såkaldt kronometrisk metode, dvs. at man benytter målinger af reaktionstid som indikator for underliggende kognitive processer. Der er nemlig her tale om et af de sjældne tilfælde i psykologien, hvor vi kan betjene os af et rigtigt intervalmål, der



Figur 2.8. Resultaterne af eksperiment nr. 1.

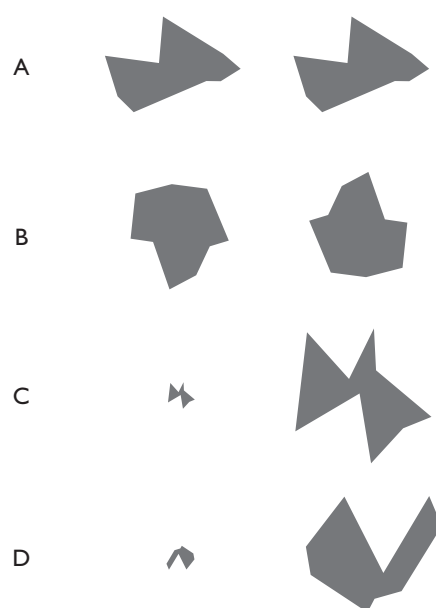
oven i købet er af samme hard-science karakter som i naturvidenskaben, nemlig tidsmåling.⁸

Med hensyn til forsøgspersoner spiller udvælgelsen som nævnt i karakteristikken af det klassiske, individbaserede forsøg en underordnet rolle. Faktisk medvirkede der i denne og i de næste forsøgsrækker kun syv personer, herunder forfatterne.

Resultatet af dette hypotese- og dermed teoriafprøvende eksperiment fremgår af figur 2.8. X-koodinaten angiver størrelsesforholdet, og Y-koodinaten medianværdien for reaktionstiden på samme figurpar. Vi ser, at punkterne ganske tæt følger en ret linje, hvilket understøtter S-teorien, mens det er helt i modstrid med F-teorien, ud fra hvilken størrelsesforholdet ville være aldeles irrelevant.

Efter dette heldigt afprøvede eksperiment står S-teorien meget stærkt. Ikke alene vokser reaktionstiden med størrelsesforskellen, men den gør det altså også lineært, hvad der peger på en simpel, mekanisk virkende hjerneproces for skalering af figurer.

Imidlertid giver de meget specielle og ikke særlig dagliglivsnære stregtegninger en ringe økologisk validitet, så i næste forsøgsrække, eksperiment

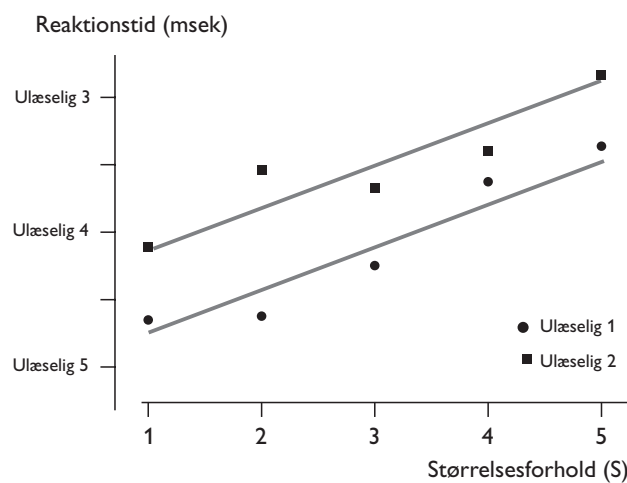


Figur 2.9. Eksempler på figurpar i eksperiment nr. 2.

2, arbejdes der med – stadig meningsløse, men nu mere solide – todimensionale figurer (se figur 2.9).

I hver præsentation vises der to figurer, der kan have forskellig størrelse, men samme form, og som enten vender samme vej eller er drejet 180 grader i forhold til hinanden

Med samme instruktion som i forrige eksperiment fremvises der to figurer ad gangen. Som det fremgår, er de to figurer i præsentationerne A og C ens (testpar), mens de er forskellige i B og D (kontrolpar).



Figur 2.10. Resultaterne af eksperiment nr. 2.

Figur 2.10 viser lige som figur 2.8, at reaktionstiden stiger lineært med størrelsesforholdet, og at kontrolparrene kræver ekstra tid i forhold til testparrene. Også her får vi altså understøttet S-teorien og afvist F-teorien.

Imidlertid stopper forfatterne ikke ved konstateringen af, at S-teorien har fundet støtte. S-teorien er nemlig så bredt formuleret, at den ikke fortæller, hvilken *mekanisme* der er på spil, når sammenligningen af de to figurer finder sted. For at indkredse denne mekanisme opstiller forfatterne to forskellige underudgaver af S-teorien:

1. N-teorien (normaliseringsteori)

Ifølge denne teori skal der være en bestemt hypotetisk (men som vi skal se ikke reelt eksisterende) normalstørrelse for alle figurer. Teorien antager at alle figurer, vi præsenteres for, bliver skaleret, så de *efter*

normaliseringen har den samme størrelse, og derefter kontrolleres det, om de kan dække hinanden.

2. *T-teorien (transformationsteori)*

Ifølge denne teori transformeres den første figur (erindringsbilledet) i størrelse til den anden figur (den aktuelt perciperede), og herefter foretages sammenligningen som ifølge N-teorien.

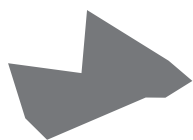
Der kan ud fra denne hypoteseopstilling foretages et nyt afgørende eksperiment ved at opdele stimulusmaterialet i to serier: en serie A, hvor den anden figur er en *formindsket* udgave af den første, og en serie B, hvor den anden figur er en *førstørrelse* af den første (figur 2.11).



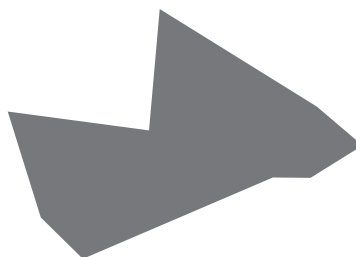
1. billede i A-serie



2. billede i A-serie



1. billede i B-serie



2. billede i B-serie

Figur 2.11. Forholdet til normalstørrelse i serie A og B.

Da vi ud fra empirien i de to første eksperimenter må antage, at bearbejdningstiden for en given proces afhænger af størrelsesforholdet mellem udgangsfiguren og den normaliserede eller transformerede figur, kan vi forudsige følgende:

- Ud fra N-teorien må der med hensyn til reaktionstid almindeligvis være en forskel mellem serie A og B. Lad os nemlig antage, at der eksisterer

en normalstørrelse, sådan som denne teori påstår. Hvis fx denne normalstørrelse ligger tæt op ad den mindste figur, dvs. 2. billede i øverste serie (A-serien), så vil den samlede bearbejdningstid for dette billede være minimal, mens den for det første billede i A-serien vil være summen af denne minimumstid og bearbejdningstiden for at skalere dette billede, der er dobbelt så stor som det andet, ned til halv størrelse.

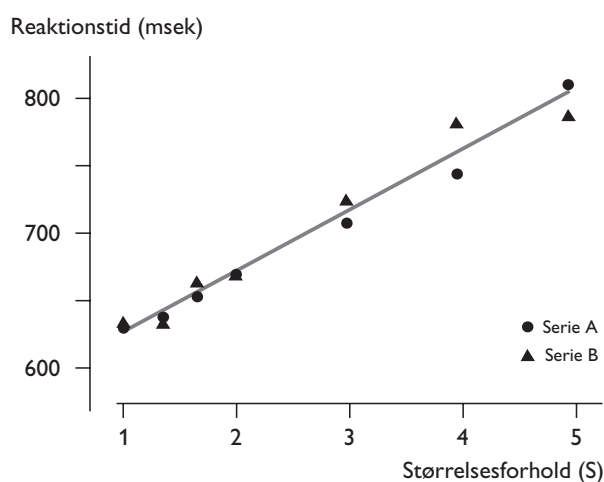
Når vi ser på B-serien, er det første billede ligeledes dobbelt så stort som den hypotetiske normalstørrelse, mens andet billede er fire gange så stort.

Resultatet af vores tankeeksperiment bliver da, at B-serien må tage længere tid, fordi vi her ikke bare har brug for en nedskalering til det halve som i A-serien, men skal bruge både en nedskalering til det halve og til en fjerdedel.

Man kan nu matematisk vise, at uanset hvor stor eller lille det skulle behage vores hypotetiske normalstørrelse at være, så vil reaktionstiden almindeligvis være en længere i B-serien end i A-serien.

Problemet med denne N-teori er på forhånd af samme art som problemet med Nessie, den berømte søslange i Loch Ness-søen. For det første ved vi ikke, om den eksisterer, og for det andet ved vi ikke, hvilken størrelse den i givet fald har.

Ud fra T-teorier ser det meget enklere ud. For det første skal vi ikke antage, at der eksisterer nogen normalstørrelse, hvis specifikke størrelse vi



Figur 2.12. Resultaterne af eksperiment nr. 3.

på forhånd ikke kender. For det andet er T-teoriens forudsigelse af forholdet med A-serien og B-serien meget enkel: Der skal nemlig slet ikke være nogen forskel mellem dem, når blot forholdet mellem den første og den anden figur er givet. Bearbejdnings- og dermed reaktionstiden afhænger simpelthen af forholdet mellem de to figurer og ikke af, om den store eller den lille kommer først.

Hvad empirien siger, fremgår af figur 2.12.

Her er ingen tvivl. Der er ikke, sådan som N-teorien ville forudsige, forskelle mellem i hvert fald en del af reaktionstiderne i A-serien og B-serien. Reaktionstiderne følger alene forholdet mellem den største og den mindste figur – som forudsagt af T-teorien.

Vi har her nærmest et mønstereksempel på et klassisk individbaseret eksperiment (eller rettere en serie af eksperimenter). Designet svarer nøje til definitionen – med et mindre antal ikke specielt udvalgte forsøgspersoner, med nogle få eksperimentelle variable og nogle få observationsvariable. Formålet er en kausal reduktion ud fra disse variable. I det foreliggende tilfælde er der ydermere tale om en hypotetisk-deduktiv metode med en nøje sammenhæng mellem teoriudvikling og teoriafprøvning.

Det her beskrevne eksperiment er meget enkelt udformet, hvilket er karakteristisk for vellykkede kausale forsøg. For at være overbevisende skal det være gennemskueligt og kun forholde sig til det, som er den eksperimentelle hypotese.

Det randomiserede gruppeeksperiment

Til trods for sin logiske stringens og sin psykologisk adækvate definition af, hvad der er den empiriske genstand, har det individbaserede eksperiment dog en afgørende begrænsning. Det forudsætter nemlig, at man kan udsætte samme forsøgsperson for hele den række af enkeltforsøg, som udgør eksperimentets kombinatoriske design. Dette begrænser denne undersøgelsesform til påvirkninger, der ikke væsentlig påvirker forsøgspersonens indstilling eller reaktionsmønster. Forudsætningen om en sådan principiel upåvirkethed af de forudgående påvirkninger vil ofte være opfyldt i kognitive eksperimenter, men den vil slet ikke holde i undersøgelser med mere indgribende påvirkninger, sådan som det typisk er tilfældet inden for personligheds- og socialpsykologi samt i de anvendte discipliner. Når man skal undersøge effekten af en behandling eller et

pædagogisk program, kan man derfor ikke sammenholde en sådan intervention med ikke-intervention for de samme forsøgspersoner. For uanset, om den givne intervention har den ønskede virkning eller ej, vil den i reglen have en sådan indflydelse, at en forsøgsperson ikke vil forholde sig på samme måde i anden omgang over for ikke-intervention som de upåvirkede forsøgspersoner. Dette gælder også, selv om man begynder med "ikke-intervention", fordi et kontrolprogram, udformet så det er generelt sammenligneligt med interventionsprogrammet, må formodes ligeledes at have en indvirkning på forsøgspersonerne. Af denne grund har lægevidenskaben udviklet det, der kaldes det klinisk kontrollerede forsøg, og det er denne forsøgsform, der er overtaget af psykologien som det randomiserede gruppeeksperiment.

En afgørende forskel er, at det randomiserede gruppeeksperiment erstatter forsøgspersonen med den randomiserede gruppe som sin empiriske genstand. Dermed udskifter man ét problem, nemlig forsøgspersonens påvirkning fra de tidligere enkeltforsøg, med et andet problem, som er forskellen mellem de enkeltpersoner, der indgår i forsøget. Problemet ved at tage højde for eventuelle individfaktorer blev i det individbaserede eksperiment løst ved at bruge den enkelte forsøgsperson som sin egen kontrol, idet designets forskellige påvirkninger kunne sammenlignes inden for en forsøgsrække med *samme* forsøgsperson. I det gruppebaserede eksperiment erstattes denne mulighed for at sammenligne enkeltforsøg udført over for den samme person med sammenligningen af forskellige delforsøg udført med *forskellige*, men principielt ækvivalente grupper. Ækvivalensen bliver sikret ved lodtrækning, således at et givet antal af forsøgspersoner, der er udvalgt til forsøget inden for et givet udvælgelsesgrundlag, opdeles i to eller flere forsøgsgrupper gennem lodtrækning, også kaldet randomisering.⁹ Allerbedst er det, hvis der kan foretages to lodstrækninger efter hinanden, således at man ud fra den udpegede population først udtager dem, der skal indgå i forsøget, og dernæst gennemføre en anden lodtrækning, hvor opdelingen i de forskellige grupper finder sted.

Skiftet fra individ- til gruppebaseret har afgørende metodologisk betydning. Ved dette skift opgiver man nemlig strengt taget muligheden for at inddrage selve individualiteten, hvad der for metodologiske purister kan betragtes som et skridt ud af psykologien, som en overgang til medicinsk eller samfundsvidenskabelig forskning. Hovedvægten bliver nu lagt på interventionens betydning for, hvad man ville kunne kalde det kategoriale individ, dvs. det typiske individ inden for en given kategori. Det behøver

ikke at være et alment gennemsnitsindivid; det kan også være en række forskellige typiske individer, som hver repræsenterer en bestemt kategori. Dette er ikke nødvendigvis noget tab, hvis det netop er bestemmelsen af en given interventions generelle indvirkning på en bestemt slags gennemsnitsindivid, der er forsøgets formål.

Det individbaserede eksperiment har den særlige stringens, at man principielt kan inddrage egenskaber, som er helt specifikke for det enkelte individ – de såkaldt individualspecifikke parametre. En sådan parameter kan fx være en persons reaktionshastighed. På denne måde kan man ikke alene bestemme betydningen af de eksperimentelle variable, men også betydningen af de individualspecifikke forhold. Noget sådant vil ganske sjældent være relevant og endnu sjældnere muligt uden for et kognitivt grundforskningsområde. Til bredere brug inden for personligheds- og socialpsykologi og inden for den anvendte psykologis områder er det gruppebaserede eksperiment således den eneste fremkommelige vej, når det gælder eksperimenter.

Vi fastslog før, at det gruppebaserede eksperiment som epistemisk genstand har virkningen af de eksperimentelle variable på det eller de kategoriale individer. Dette var jo også tilfældet i det individbaserede eksperiment. Der er imidlertid en forskel: I det individbaserede eksperiment bliver hver enkelt forsøgsperson (den empiriske genstand) ud fra sine *reaktions- og forholdemåder over for de eksperimentelle variable* betragtet som bærer af *det almene*, mens det i gruppeeksperimentet er hele gruppen. I det individbaserede eksperiment holdes forsøgspersonen konstant gennem serien af forskellige påvirkninger, således at forskelle i respons alene kan tilskrives forskellene i påvirkningerne. I det gruppebaserede eksperiment holdes *grupperne* konstante, således at forskelle mellem grupperne alene kan tilskrives forskellene i påvirkningerne over for de forskellige grupper. (Dette forudsætter naturligvis, at forskellene mellem grupperne er signifikante).

Hele denne sofistikerede metodologi, som først er udviklet inden for lægevidenskaben med henblik på kliniske eksperimenter (se Grossman, 2005) og derefter videreudviklet til brug i psykologien af Campbell og Stanley i 1960'erne (Campbell & Stanley, 1966), har som sigte at garantere validiteten dels af forsøgsgruppernes ækvivalens, dels af interventionens afgørende betydning over for disse gruppers responser. Inden vi går over til den raffinerede fejlkilde- og kontrolteori, som er udviklet her, skal vi

imidlertid først resumere hovedtrækkene i det randomiserede gruppeeksperiment.

Designet for det randomiserede gruppeeksperiment

A. Udvælgelsesplan

Ofte sker udvælgelsen i to tempi:

1. *Primærudvælgelse* Blandt samtlige forsøgspersoner fra udvælgelsesgrundlaget, som kan være en bestemt population, der defineres ud fra specifikke in- og eksklusionskriterier, udvælges, om muligt gennem lodtrækning, en eller flere stikprøver. Der er tale om en enkeltudvælgelse, hvis der kun opereres med én population, men om flere udvælgelser, hvis forsøget skal sammenligne forskellige populationer.
2. *Gruppeopdeling* Gennem lodtrækning foretages en opdeling af forsøgspersonerne i grupper. Der er tale om mindst to grupper, nemlig en forsøgs- og en kontrolgruppe, men der kan også være flere, hvis designet opererer med flere slags interventioner og/eller flere slags kontrol, fx *ingen intervention* og *pseudo-intervention*, dvs. *placebo*.

B. Iscenesættelsesplan

I. Tid

Den tidslige struktur har generelt følgende struktur for interventionsgrupperne:

X-gruppe: O1 X O2

Mens den i kontrolgruppen er

C-gruppe: O1 O2

O1 er observation før interventionstidspunktet og O2 er observationen efter interventionstidspunktet, mens X er selve intervention. X bruges ofte som betegnelse for eksperimentdelen, mens C tilsvarende står for kontroldelen.

Denne tidslige struktur kan med fordel udvides til fx:

X-gruppe: O1 X O2 O3

C-gruppe: O1 O2 O3

Hvor O3 er et senere observationstidspunkt end O2. Denne udvidede tidsstruktur gør det muligt at skelne mellem korttids- og langtidseffekt i forsøgsgruppen.

II. *Sted*

Den rumlige iscenesættelsesplan i form af en fikseret ramme har kun afgørende betydning i de egentlige grundforskningsorienterede eksperimenter. I de anvendte discipliner er den rumlige iscenesættelse simpelthen knyttet til de institutionelle rammer, hvori interventionen finder sted, fx en uddannelses- eller behandlingsinstitution.

III. *Våben*

Specifikationen af interventionen eller (hvis der er flere forsøgsgrupper) interventionerne. De uafhængige variable omfatter ud over disse interventionsvariable (de eksperimentelle variable) også selektions- og baggrundvariable. Selektionsvariable er de størrelser, der definerer de udvalgte grupper. Baggrundvariable er de egenskaber ved forsøgspersonerne, der indgår i dataanalysen, fordi de kan have en betydning for udfaldet. Typiske baggrundvariable er køn, alder og uddannelsesniveau.

C. Observationsplan

Observationsplanen i det valide gruppeeksperiment er nøje fastlagt gennem de observationsvariable (afhængige variable), der skal vise, om interventionen har den *tilsigtede* virkning over for forsøgspersonerne, hvad angår de aspekter, der ønskes intervenseret over for, og om nødvendigt påvise, at den *ikke har nogen utilsigtet* indvirkning på andre aspekter. Således vil observationsvariable i et pædagogisk forsøg kunne være elevernes status på de områder, som er genstand for undervisningen. I et forsøg med psykologisk behandling kan det fx være de psykopatologiske symptomer, som søges afhjulpet.

Et eksempel på det randomiserede gruppeeksperiment

Schachter og Singer om kognition og emotion

Lige som Bruners eksperiment (se s. xx) var overskridende ved at vise den nøje sammenhæng mellem motivation og perception, banede det eksperiment af Schachter og Singer (1962), vi nu skal omtale, nye veje inden for *kognition* og *emotion* ved at relatere emotion ikke blot til psykofysiologi, hvor der var en velkendt sammenhæng, men også til kognition – et område, man på dette tidspunkt betragtede som ikke bare erkendelsesmæssigt, men nærmest ontologisk uden forbindelse med emotion.

Den teoretiske baggrund for forsøget er den modifikation af James-Lange-teorien, som opdagelsen af adrenalin-hormonet bragte med sig. Den opfattelse, der oprindeligt var bred enighed om, gik ud på, at det var den emotionelle tilstand, der fremkaldte de kropslige ændringer. Således at man fx græd, *fordi* man var ked af det. Lange og James (1922) reviderede denne opfattelse, idet de antog, at emotionelle reaktioner var fremkaldt af en fysiologisk tilstand, således at den emotionelle reaktion kunne betragtes som et epifænomen. Herved forstås et fænomen, som ikke har en selvstændig status, men som blot forekommer i forbindelse med et andet, primært fænomen, der er anledning til epifænomenet. Således er skyggen af en genstand et epifænomen, der alene forekommer i kraft af den genstand, der kaster skygge. Walter Cannon (1929) afdækkede som den første, hvordan hormonerne adrenalin og noradrenalin fungerede. De var knyttet til det, han kaldte "kæmp/flugt-syndromet" ("fight-or-flight-syndrome"). Cannon reviderede James-Langes teori, som reducerede den emotionelle oplevelse til et epifænomen, idet han antog, at en given situation, der blev perciperet og dermed klassificeret, efterfulgtes af nogle processer i thalamus, hvortil affekterne er knyttet, og disse processer førte så dels til nogle kropslige ændringer af fx muskulær og hormonal karakter, dels til en emotionel oplevelse, fx frygt eller vrede.

Som udgangspunkt for deres eksperiment foreslog Schachter og Singer nu en tofaktor-teori for emotionerne, som gik ud på, at en bestemt følelse ud over at blive sat i gang af en fysiologisk proces også kræver en kognitiv komponent, idet der til en veldefineret følelse også er knyttet et betydningsindhold.

De forskellige emotionsteorier er fremstillet skematisk nedenfor:

James' udlægning af common sense:

perception > emotionel oplevelse > kropslige reaktioner

Common sense-opfattelsen går altså ud på, at det, man oplever, udløser en følelse, som fremkalder kropslige reaktioner.

James' og Langes teori

perception > motorisk reaktion > fysiologisk aktivering (arousal) > emotionel aktivering

James' og Langes teori indebærer altså, at den emotionelle tilstand var sekundær, et epifenomen, fremkaldt af den kropslige reaktion på det, som blev perciperet.

Cannons synspunkt

> emotionel oplevelse
 perception > processer i thalamus
 > kropslige ændringer

Cannon modificerede James-Lange-teorien, så den emotionelle oplevelse og de kropslige ændringer var sideordnede og forekom samtidig.

Schachter og Singers teori

hormonal påvirkning
 > emotionel oplevelse
 kognitiv information

Schachter og Singer præciserede Cannons teori ved at operationalisere perception i form af *kognitiv information* og ved at gøre denne til en del af den samlede stimulus, sammen med en evt. hormonal påvirkning.

Ud fra denne tofaktorteori afprøvede Schachter og Singer nu den hypotese, at adrenalin *ikke i sig selv* producerer en bestemt emotionel

tilstand, men at subjektets opfattelse af den foreliggende situation er afgørende for, hvilken sindstilstand han eller hun oplever at befinde sig i. Deres eksperiment operationaliserede denne tofaktorteori ved at gøre såvel hormonale som kognitive påvirkninger til interventionsvariable: Den hormonale påvirkning gennem en injektion af enten adrenalin eller en fysiologisk virkningsløs saltopløsning; den kognitive påvirkning gennem dels en information eller manglende information om injektionens virkning, dels en falsk forsøgsperson, der tilsyneladende indgik i forsøget på lige fod med den ægte forsøgsperson, men som i virkeligheden var instrueret til at illudere en bestemt emotionel tilstand: enten eufori eller vrede.

Uafhængige variable

A. Fysiologisk

1. Injektion
 - a. adrenalin
 - b. saltopløsning (placebo)

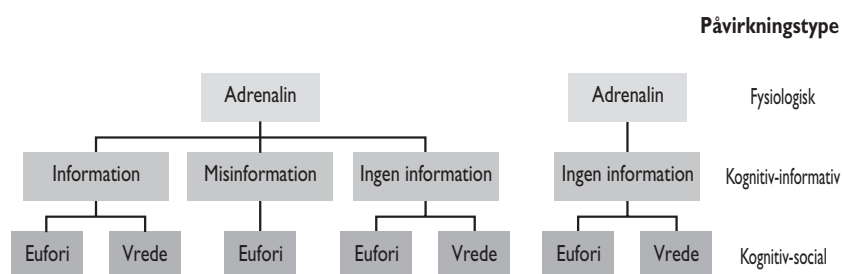
B. Kognitiv

1. Information om injektion
 - a. information (korrekt beskrivelse af injektionens fysiologiske virkning som vrede)
 - b. fejlinformation (falsk beskrivelse af injektionens fysiologiske virkning som glæde)
 - c. ingen information
2. Social påvirkning
 - a. eufori (en falsk forsøgsperson illuderer overgiven glæde)
 - b. vrede (en falsk forsøgsperson illuderer vrede)

Afhængige variable

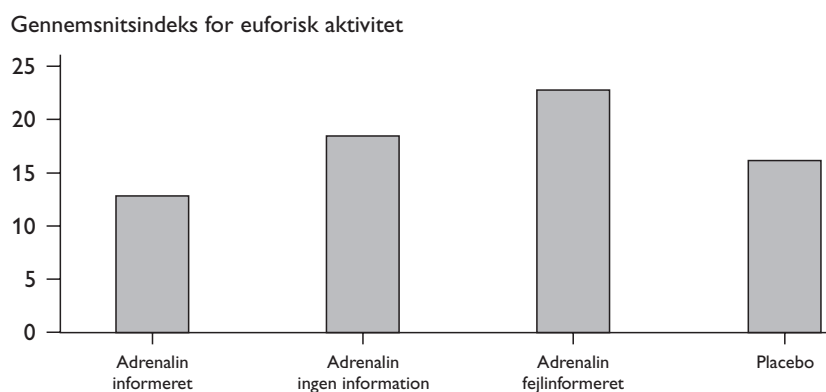
1. Observerede tegn på glæde
2. Observerede tegn på vrede
3. Selvrapporteret glæde eller vrede

Der er tale om et gruppebaseret eksperiment med én forsøgsgruppe pr. forsøgsbetingelse, dvs. for hver konstellation af de uafhængige variable. Disse kombineres ikke frit, men efter det logiske design, der illustreres af figur 2.13.



Figur 2.13. Forsøgsgrupper i Schachter og Singers forsøg. Figuren til venstre viser de grupper, som udsættes for reel fysiologisk påvirkning. Figuren til højre viser de to grupper som ikke fik nogen fysiologisk, men kun en kognitiv (social) påvirkning.

Her er der altså syv forsøgsgrupper, der ved randomisering er gjort sammenlignelige. Vi skal ikke her gennemgå alle resultaterne, men blot vise observationerne for de fire grupper, hvor en falsk forsøgsperson illuderede eufori (se figur 2.14).

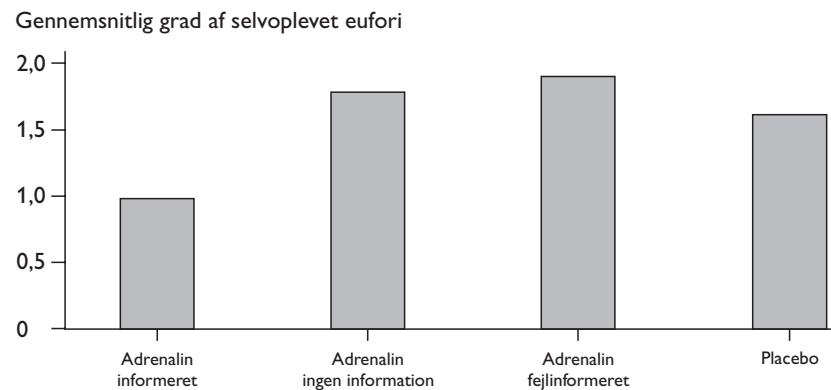


Figur 2.14.

Da der kun var omkring 25 personer i hver gruppe, er ikke alle forskellene signifikante, men der er det signifikante hovedresultat, at den korrekt informerede gruppe (som forventede vrede) kun fremviste et

gennemsnit på det halve omfang af observeret eufori sammenlignet med den fejlinformerede (som forventede glæde).

Med hensyn til selvrapporteret oplevelse fremgår gruppernes reaktion af figur 2.15:



Figur 2.15.

Her var resultaterne endnu mere overbevisende, idet den korrekt informerede gruppe, som forventede at reagere med *vrede*, rapporterer den mindste grad af oplevet glæde, mens både personerne i den ikke-informerede og den fejlinformerede gruppe beskriver deres ændrede tilstand som en stemning af eufori – en oplevelse, der øjensynligt induceres gennem den falske forsøgsperson.

Schachter og Singer får således bekræftet deres hypotese om kognitionens afgørende betydning for emotioner, såvel adfærdsmæssigt som oplevelsesmæssigt.

Metodologisk er der den raffinerede detalje ved designet, at der er indført ikke-informerede og fejlinformerede grupper som en kontrol af fordrings-effekten (se s. xx).

Betingelser for intern validitet

Der er to grundlæggende betingelser for, at kravet om intern gyldighed i det randomiserede eksperiment kan være opfyldt:

1. Forsøgsgrupperne skal ved forsøgets start være ækvivalente.
2. På nær den eksperimentelle variabel, interventionen, skal den fikserede

ramme for påvirkningssituationen være den samme for de forskellige forsøgsgrupper.

Jeg foreslår at kalde disse to betingelser for henholdsvis *udvælgelses-* og *påvirkningsækvivalens*.

Randomiseringen sikrer ganske vist ikke, at de to grupper er helt ens, men den sikrer, at de forskelle, der er mellem grupperne ved forsøgets begyndelse, følger sandsynlighedsteoriens regler for tilfældig variation, og det indebærer, at vi kan foretage statistiske signifikanstest for at sammenligne grupperne, også *efter* interventionen.

Påvirkningsækvivalens indebærer sammen med udvælgelsesækvivalensen, at de forskelle, der evt. opstår mellem grupperne efter interventionen, alene kan tilskrives den eksperimentelle variabel, altså de specifikke træk ved interventionen. Dette er ikke nogen garanti for, at en opstået signifikant forskel bliver fortolket gyldigt, altså i overensstemmelse med virkeligheden, da sandsynlighedsteorien jo ikke beskytter os mod fejl, men blot gør det muligt at forudsige, hvor stor sandsynligheden er for, at vi kommer til at begå fejl. Dette forhold fremgår af det følgende afsnit om forsøgsplanlægning.

Kontrollen med påvirkningsækvivalens er især udviklet til perfektion i forbindelse med lægevidenskabens klinisk kontrollerede forsøg, hvor placebobehandling og blinding er blevet en del af den forskningsmæssigt ortodokse procedure.

Ved placebobehandling er behandlingen af kontrolgruppen ikke blot negativt defineret ved, at personerne i denne gruppe ikke udsættes for den intervention, som foretages over for forsøgsgruppen, men også positivt, dvs. ved at forsøgspersonerne i det hele taget behandles på en måde, der er gjort så lig interventionen, at det faktisk skal være umuligt at opdage forskellen, som omhyggelig holdes skjult over for forsøgspersonerne og om muligt også over for behandlerne.

Kun gennem en hemmelig kode kan det efterfølgende konstateres, hvem der modtog hvilken behandling. Det er denne hemmeligholdelse, der kaldes *blinding*. Hvis det kun er forsøgspersonerne, der holdes i uvidenhed, taler man om enkelt blinding, hvis det også er behandlerne (og evt. observatører), taler man om dobbeltblinding: forsøget er „dobbeltblindt“.

En trussel mod den interne validitet er *selektivt bortfald*. Herved forstås som tidligere nævnt, at de personer, der afbryder forsøget i utide, har bestemte karakteristika, der adskiller dem fra de personer, som gennemfø-

rer. Da det fx kan være motivationelle faktorer, der er afgørende for selektivt bortfald, kan en evt. forsøgseffekt være begrænset til de personer, som gennemfører, således at den omhyggelige udvælgelsesprocedure ikke medfører, at resultatet er gyldig for den undersøgte population.

Forsøgsplanlægning

Gruppeeksperimentet sigter mod at dokumentere, at en bestemt påvirkningseffekt er til stede i forsøgsgruppen – og ikke er det i kontrolgruppen. Det er derfor afgørende, at størrelsen af grupperne er tilstrækkelig til at garantere en signifikant forskel efter interventionstidspunktet. Man må altså dimensionere grupperne med dette for øje.

Denne del af designprocessen kaldes forsøgsplanlægning. Der er her to statistiske begreber i spil: *signifikans* og *styrke* (power). Signifikansen er den statistiske evidens, der er for et empirisk udfald. Signifikans udtrykkes ved sandsynligheden for, at det observerede kan opstå ved et rent tilfælde. Når denne sandsynlighed er lille, er signifikansen stærk. Man plejer at angive denne sandsynlighed for, at der kan tænkes at være tale om et tilfældigt udfald som enten 5 % (almindelig signifikans), 1 % (stærk signifikans) og 0.1 % (1 promille eller meget stærk signifikans). Når man planlægger størrelsen af forsøgsgrupperne, kan man udregne en anden sandsynlighed, nemlig sandsynligheden for, at man med givne gruppestørrelser får en bestemt signifikans. Denne sandsynlighed kaldes for forsøgsdesignets styrke. Her kan man altså fastsætte gruppestørrelserne således, at hvis man antager, at der er en bestemt forsøgseffekt, så vil sandsynligheden være 99 % for at få en signifikans på mindst 5 %. Denne sandsynlighed kaldes for styrken, som altså er et hypotetisk tal, idet man forudsætter en forsøgseffekt, som man jo ikke kender på forhånd, men som man måske nok kan angive en plausibel størrelse for.

Som et eksempel skal vi se på et planlagt forsøg med kognitiv terapi for unge mellem 15 og 18 år, der efter have begået voldskriminalitet og være blevet dømt til ungdomsfængsel har indvilget i at deltage i et randomiseret forsøg til afprøvning af effekten af kognitiv terapi. Begge grupper får et tremåneders kursus i psykoedukation, herudover gives der kognitiv terapi forsøgsgruppen. Efter løsladelse følges begge grupper gennem et år, hvor der indsamles oplysninger om fornyet voldskriminalitet, altså recidiv- eller tilbagefaldshyppighed.

Ud fra tidligere undersøgelser kan vi i dette konstruerede eksempel

antage, at en ubehandlet gruppe vil vise et tilbagefald til voldskriminalitet på 30 %. Hvis vi yderligere antager, at der ikke er nogen effekt i kontrolgruppen, mens effekten i forsøgsgruppen består i et fald fra 30 til 15 %, kan vi beregne det approksimative t-test med dette udfald (tabel 2.1):

Tabel 2.1. Dimensionering af grupper ved afprøvning af behandling for unge kriminelle.

Gruppe	Antal	Tilbagefald til vold (recidiv)
Forsøg	N	15 %
Kontrol	N	30 %
Ialt	2N	22,5 %

Med de antagne værdier kan vi beregne, at hvis vi vil have en sikkerhed på 95 % for at opnå et signifikant resultat (på 5 %-niveau), skal vi have gruppestørrelser på 150 personer i hver gruppe, altså i alt 300 forsøgspersoner.

Der er to årsager til, at man skal bruge så store gruppestørrelser som her:

1. Vi abstraherer fra individuelle forskelle, således at vi alene har gruppedata.
2. Disse gruppedata er hyppigheder, som har en meget lav informationsgrad, idet de er baseret på 0-1-oplysninger, nemlig tilbagefald/ikke-tilbagefald. Der er altså tale om en variabel med et meget lavt informationsindhold.

I det næste afsnit vil det netop beskrevne design blive sammenlignet med et eksperiment, hvor disse to begrænsninger er ophævet, således at der for det første anvendes gentagne målinger, og for det andet bruges mere informationsrige observationsvariable, nemlig intervalvariable. På denne måde kan vi nøjes med langt lavere gruppestørrelser.

Det faktorielle gruppeeksperiment

Som nævnt kan vi i gruppeeksperimentet have mere end to grupper. Man kan således operere med multiple eksperimentelle faktorer, der kombineres lige som ved det individbaserede eksperiment. Hvis vi altså forestiller os,

at vi supplerer designet i forsøget med kognitiv terapi over for unge voldskriminelle med en opdeling af populationen i to kategorier – en med etnisk danske børn og en med børn af indvandrere, og derefter fra begge kategorier udvælger forsøgs- og kontrolgrupper som før, kan resultatet blive som vist i tabel 2.2.

Tabel 2.2. Dimensionering af differentieret afprøvning af behandling for unge kriminelle.

Population	Gruppe	Antal	Tilbagefald til vold (recidiv)
Danske unge	Forsøg	N	15%
	Kontrol	N	30%
Indvandrerunge	Forsøg	N	50%
	Kontrol	N	50%

Her finder vi med tilstrækkelige gruppestørrelser dels, at forsøget kun har en effekt i kategorien med danske unge, dels at niveauet for tilbagefald er højere i gruppen af indvandrerunge. Dette giver to væsentlige informationer om tilbagefald. For det første, at det afprøvede forsøg nok har værdi over for danske unge, men er aldeles virkningsløst over for indvandrerunge, for det andet, at de sidstnævnte repræsenterer et endnu større problem end almindelige unge danske voldskriminelle. Et problem som interventionen øjensynligt er aldeles magtesløs over for. Et sådant differentieret gruppeeksperiment kaster altså lys over de nærmere betingelser for effekt, hvilken kan muliggøre end forbedring af interventionen, så den bedre tilpasset de kategorier, hvor virkningen er svag eller helt mangler. Den type af eksperimentel variabel, som vi her kan indføre, kaldes en *modifikator*, fordi den har en modificerende virkning på interventionseffekten.

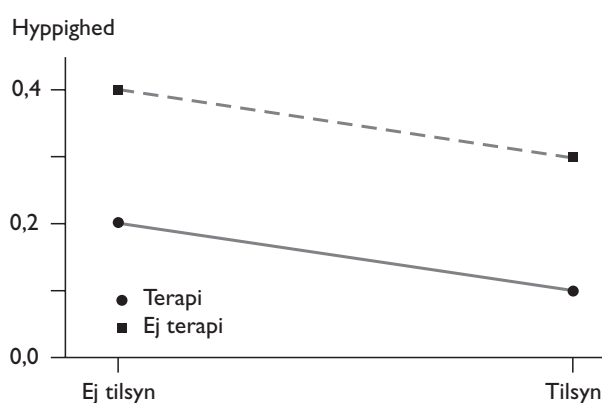
I et sådant faktorielt design er den ene faktor altså en interventionsvariabel, mens den anden er en udvælgelsesvariabel. Vi kan også have faktorielle designs med flere interventionsvariable og/eller udvælgelsesvariable. Eksempelvis kunne vi forestille os, den allerede nævnte interventionsvariabel i form af kognitiv terapi suppleret med en anden bestående af bedre tilsyn i en prøveperiode efter løsladelse som vist i tabel 2.3.

Den tabelmæssige opstilling kan fremstilles grafisk som i figur 2.16.

I det angivne eksempel ser vi en effekt af begge interventionsfaktorer, som her indgår på en måde, der synes at være additiv, således at kognitiv

Tabel 2.3. Faktoriel afprøvning af behandling for unge kriminelle.

Kognitiv terapi	Tilsyn	Antal	Tilbagefald til vold (recidiv)
Ja	Ja	N	10%
Ja	Nej	N	20%
Nej	Ja	N	30%
Nej	Nej	N	40%



Figur 2.16. Tilbagefald til voldskriminalitet.

behandling har den største effekt, nemlig 20 %, mens tilsyn har en noget mindre effekt, nemlig 10 %. Imidlertid fungerer procentskalaen ikke på en sådan måde, at den kan bruges til at addere effekter. Det hænger sammen med dens bund- og lofteffekt, hvilket blev gennemgået i kapitel 1 (se s. xx). Det går imidlertid nogenlunde i dette tilfælde, hvor de to effekter er begrænsede, men hvis de fx er på henholdsvis 50 og 70 %, ryger vi op gennem loftet. Derfor bør man ved statistisk analyse af sådanne sammensatte effekter på en procentskala anvende den såkaldte logistiske regression.

Gruppeeksperiment med gentagen måling

Jeg har tidligere fremhævet, at designet i det simple gruppeeksperiment er svagere end det individbaserede, fordi man i gruppedesignet mangler information om konsistente individuelle forskelle. Argumentet for ikke at bruge et individbaseret design er almindeligvis, at et sådant vil blive ugyldiggjort ved gentagne forsøg, dvs. at gentagen udsættelse for påvirk-

ningerne i et antal enkeltforsøg vil influere på de efterfølgende. Hvis man imidlertid kan regne med, at denne såkaldte retesteffekt er ubetydelig, eller at den designmæssigt kan neutraliseres, vil det være fordelagtigt at anvende et såkaldt *gruppeeksperiment med gentagen måling*.

Denne form for gruppeeksperiment bruges først og fremmest ved effektundersøgelser, hvor vi ønsker at undersøge om en given intervention har en virkning på en bestemt statusvariabel. Statusvariable er typisk mål for egenskaber, der enten er succes- eller fiaskokriterier i den forstand, at de er knyttet til interventionen som indikatorer for en tilsigtet ændring, der altså ikke kan forventes uden denne intervention.

Lad os nu se på et gruppeeksperiment med gentagen måling som alternativ til det netop beskrevne simple gruppeeksperiment. Vi kan her supplere de tidligere tilbagefaldsdata med en test for manglende impuls-kontrol. Denne kunne fx bestå af 20 spørgsmål af typen Ja-Nej vedrørende situationer med krænkelse af forskellig art. Vi kan endvidere forestille os, at scoren, som altså er et mål mellem 0 og 20, har vist sig at være et gyldigt mål for impulsivitet, således at 0 udtrykker fuldstændig kontrol og 20 total mangel på kontrol (tabel 2.4).

Tabel 2.4. Dimensionering af grupper ved afprøvning af behandling for unge kriminelle.

Gruppe	Antal	Ændring (reduktion)			
		Gennemsnitligt impuls-kontrolindeks		Gennemsnitlig forøgelse	Spredning
		før	efter		
Forsøg	N	12,7	10,7	2,0	0,9
Kontrol	N	13,0	13,3	-0,3	1,0
Gruppe-differens	2N			2,3	

Når vi nu anvender et signifikanstest for at kontrollere, om der er en væsentlig større nedgang i forsøgsgruppen end i kontrolgruppen, udnyttet vi den information, vi nu har om individuelle forskelle, idet vi regner på *ændringen* mellem før- og efter-målingen.

Tabel 2.5 viser en forhåndsberegning ud fra plausible gennemsnit og spredningsmål i forsøgs- og kontrolgruppe.

Tabel 2.5. Eksempel 1 – på udfald af eksperiment med gruppestørrelse 8.

Gruppe	Antal	Gennemsnitligt impuls-kontrolindeks		Ændring (reduktion)			
		før	efter	Gennem- snitlig reduktion	Spredning på reduktion	Middelfejl på gennem- snitlig reduktion	t-værdi p-værdi
Forsøg	8	12,7	10,7	2,0	0,9	0,32	
Kontrol	8	13,0	13,3	-0,3	1,0	0,35	
Gruppe- differens				2,3		0,45	5,0 0,0001

I dette eksempel vil signifikansen på forskellen mellem forsøgs- og kontrolgruppe med en gruppestørrelse på blot 8 være overordentlig stærk: $p < 0,0001$. Med denne lave gruppestørrelse vil gennemsnittene imidlertid svinge meget, og vi kan også få et udfald, som vist i tabel 2.6. Her bliver signifikansen altså betydelig svagere.

Tabel 2.6. Eksempel 2 – på udfald af eksperiment med gruppestørrelse 8.

Gruppe	Antal	Gennemsnitligt impuls-kontrol-indeks		Ændring (reduktion)			
		før	efter	Gennem- snitlig reduktion	Spredning på reduktion	Middelfejl på gennem- snitlig reduktion	t-værdi p-værdi
Forsøg	8	12,7	10,7	1,7	1,1	0,40	
Kontrol	8	13,0	13,3	-0,6	0,9	0,32	
Gruppe- differens				1,1			2,2 0,01

Hvis vi vil have en sikkerhed på 95% for at opnå en signifikans på 5%, kan gruppestørrelsen beregnes at være 8. Det nye design reducerer således antallet af forsøgspersoner fra 300 til 16. Denne dramatiske forbedring i forskningsøkonomi er imidlertid ikke alene et resultat af den gentagne

Tabel 2.7. Dimensionering af grupper ved afprøvning af behandling for unge kriminelle med tre slags design.

Forsøgets designtype	Nødvendig gruppestørrelse med 95% sikkerhed for 5% signifikans
Enkeltmåling og hyppighed	N = 150
Enkeltmåling og score	N = 35
Gentagen måling og score	N = 8

måling, men også af overgangen fra procentmål til et intervalmål i form af en score.

Hvis vi alene ser på overgangen til det mere informationsrige intervalmål, men uden gentagen måling, skal vi have en gruppestørrelse på 35 i stedet for 8. Her er antallet af forsøgspersoner reduceret fra 300 til 70.

Mens overgangen fra det informationsfattige 0-1-mål til scoring giver en reduktion til ca. en fjerdedel af stikprøvestørrelsen, bevirker gentagen måling yderligere en reduktion til ca. en fjerdedel. I alt får vi således en reduktion til ca. en sekstendedel.

Ekstern validitet og generalisering

Vi har allerede været inde på, at gruppeeksperimentet afviger fra det individbaserede eksperiment ved at have gruppen i stedet for individet som empirisk genstand. En anden forskel er knyttet til den eller de eksperimentelle variable. Hvor det klassiske individbaserede eksperiment har en påvirkningssituation, der typisk er defineret gennem bestemte fastsatte værdier af fundamentale stimulusvariable, som er entydigt og præcist definerede, er der såvel i de personligheds- og socialpsykologiske som i de anvendte områder af psykologien oftere tale om komplicerede påvirkningstotaliteter.

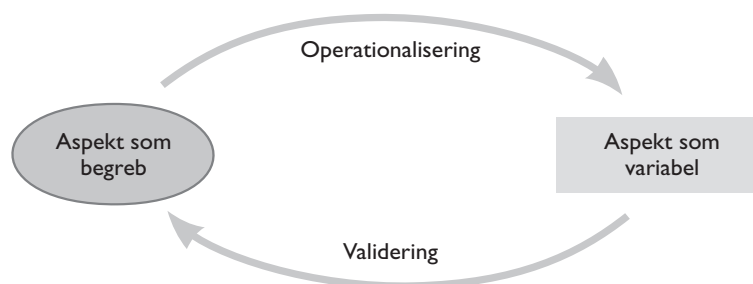
Disse mere komplicerede påvirkninger forsøges afprøvet ved at anvende dem over for forsøgsgruppen, mens de ikke anvendes over for kontrolgruppen. Problemet ved denne metode er for det første, at en præcis definition af den anvendte påvirkningssituation ikke er mulig. For det andet er ækvivalensprincippet heller ikke opfyldt, da det ikke er muligt at isolere en simpel eksperimentel variabel, som alene elimineres i kontrolgruppen.

Denne vanskelighed viser sig i øvrigt allerede i det lægevidenskabelige kliniske forsøg i forbindelse med blindingsbegrebet, som kun fungerer uproblematisk, når man ved interventionen kan isolere en kausal faktor, som det er let at tilføje og at fjerne, og som uden megen besvær kan holdes skjult. Denne simple form for påvirkningsækvivalens indebærer da, at en såkaldt placeboeffekt kan adskilles fra den evt. rene virkning af den isolerede kausale faktor. Det bliver således sikret, at mulige forventningsfaktorer såvel som resultatet af den uspecifikke behandling, der er knyttet til den konstante ramme, vil forekomme i samme udstrækning i begge grupper. Imidlertid kan blindingsiden af påvirkningsækvivalensen faktisk kun opnås, når der er tale om et farmakologisk præparat, som kan tilføres eller ikke tilføres den pille, forsøgspersonen får.

Allerede ved mere komplicerede interventioner, fx kirurgiske, er det dels meget vanskeligt at udføre placebobehandling, dels ofte etisk uantageligt. Ved behandlingsformer som forudsætter en mere interaktiv kontakt af fysisk eller psykisk art er blinding imidlertid udelukket. Dette er således tilfældet i de fleste psykologiske gruppeeksperimenter.

Den anden begrænsning ved den eksperimentelle variabel i gruppeeksperimentet er knyttet til selve dens operationalisering. Da der i reglen enten er tale om en teoretisk uskarpt defineret størrelse eller, som i de anvendte discipliner, om en påvirkningssituation, der er praktisk snarere end teoretisk defineret, er der gennemgående knyttet store problemer både til interventionens operationalisering og validering.

Når det foreliggende begreb for interventionen er ret upræcist, bliver operationaliseringen af interventionsvariablen det også meget let, og omvendt når det skal påvises, at den anvendte interventionsvariabel er et



Figur 2.17. Operationalisering og validering.

gyldigt udtryk for det tilsigtede begreb, opstår der tilsvarende vanskeligheder (se figur 2.17).

Disse operationaliserings- og valideringsproblemer kommer også til udtryk som svagheder ved den eksterne validitet og dermed ved forsøg på at generalisere resultatet af et interventionsforsøg til en bredere fænomenkreds eller et udvidet anvendelsesområde. Rent praktisk er der altså store vanskeligheder forbundet med at overføre erfaringer med en heldigt afprøvet fremgangsmåde fra én konkret eksperimentel eller praktisk kontekst til en anden.

Etiske og praktiske problemer med det randomiserede gruppeeksperiment

Anvendelse af det randomiserede gruppeeksperiment til interventionsafprøvning forudsætter, at behandlere og forsøgspersoner accepterer en opdeling i forsøgs- og kontrolgruppe efter lodtrækning. Dette vil ofte indebære, at man forholder sig på en etisk kritisabel måde over for de personer, der kommer i kontrolgruppen, og som derfor ikke kan opnå den formodede gunstige effekt.

I situationer hvor der er ventetid i forbindelse med en behandling, fx ved psykoterapi, kan man undertiden få etableret en kontrolgruppe på en etisk forsvarlig måde. Hvis udvælgelsen til henholdsvis behandling og venteliste sker pr. lodtrækning, kan man for så vidt tale om en „naturlig“ kontrolgruppe og sammenligne ændringen hos de behandlede personer med dem, der været på venteliste.

Ud over dette forskningsetiske problem kan opdelingen også medføre et rent metodologisk problem, nemlig en indstillingseffekt, hvor forsøgsgruppen har positive forventninger og holdninger til forsøget, mens kontrolgruppen bliver mere negativt indstillet.

Et andet problem ved randomisering er, at denne udvælgelsesprocedure kan være enten praktisk umulig eller i hvert fald meget omkostningskrævende i de tilfælde, hvor de rumlige rammer for forsøget kræver en inddeling af populationen i naturlige grupper. Dette er eksempelvis tilfældet i forbindelse med de fleste pædagogiske forsøg. Når sådanne etiske og praktiske problemer gør randomisering uoverkommelig eller uacceptabel, kan man vælge den næste og sidste form for gruppeeksperiment, nemlig det såkaldte *kvasieksperiment*. Således kaldet af de puristiske foregangs-

mænd Campbell og Stanley (1966), som afviser at indlemme det i de egentlige eksperimenters eksklusive selskab.

Kvasieksperimentet (gruppeeksperiment med naturlige grupper)

Kvasieksperimentet afviger alene fra det randomiserede gruppeeksperiment, ved at randomisering er opgivet, således at der er tale om naturlige (forud for forsøget eksisterende) grupper. Herved opgives også en afgørende gyldighedsbetingelse, nemlig udvælgelsesækvivalens, og det er grunden til at dette design har fået præfikset *kvasi* (af latin *quasi* = som om).

Campbell og Stanley retfærdiggør denne udelukkelse fra klassen af eksperimenter ved at opstille ikke mindre end syv fejlkilder, syv uheldsvangre konsekvenser af den manglende randomisering, som hver for sig kan invalidere et forsøgs gyldighed:

1. Historie
2. Modning
3. Testning
4. Redskabsbrug
5. Regression
6. Udvælgelse
7. Bortfald

Vi skal nu se på disse fejlkilder.

Historie

Hele det metodologiske apparat med kontrolleret udvælgelse og påvirkning er i høj grad rettet mod at eliminere indflydelsen af de historiske begivenheder, der uden eksperimentators kontrol indvirker på forsøgspersonerne efter interventionen. Da interventionen, uanset hvor rimelig og veludført den ellers er, almindeligvis kun udgør en beskedent andel af alt, hvad der influerer på forsøgspersonerne, lader det sig selvfølgelig ikke gøre at gennemføre en egentlig elimination af alle eksterne faktorer. Imidlertid er hele ideen med en kontrolgruppe at sikre en i hvert fald relativ neutralisering af de eksterne forhold, idet disse eksterne begivenheder må

antages at virke på samme måde i forsøgs- og kontrolgruppe. Denne antagelse er imidlertid allerede kompromitteret i det ægte gruppeeksperiment, hvis der er en interaktion mellem intervention og historie. Lad os således forestille os, at der i den periode, hvor en gruppe unge voldskriminelle gennemgår en kognitiv terapi rettet mod forbedret impuls kontrol, er en af personerne i forsøgsgruppen, som foretager en særlig alvorlig voldshandling. Denne kan i så fald tænkes at virke så nedslående på de øvrige deltagere, at behandlingseffekten bliver forringet.

Denne fejlkilde i form af eksterne begivenheder er yderligere forstærket ved kvasieksperimentet, hvor vi jo har naturlige grupper, som har et fælles socialt liv og dermed hver sin særlige gruppehistorie. En forskel mellem grupperne kan således principielt lige så vel hidrøre fra forskelle i gruppehistorie som i intervention.

Modning

Ved modning forstår Campbell og Stanley den naturlige aldersbetingede udvikling, som finder sted over forsøgsperioden, og det er således en faktor, der vil optræde som fejlkilde, hvis dens virkning ikke kan adskilles fra interventionsfaktoren. Det er netop en sådan adskillelse, opdelingen i forsøgs- og kontrolgruppe sigter mod. Hvis modningen i sig selv er upåvirket af interventionen, så er modningsfaktoren den samme i begge grupper, og vi kan derfor isolere interventionsfaktoren. Er der imidlertid tale om en interaktion mellem intervention og modning kan en sådan isolation ikke gennemføres. Det er således tilfældet, hvis interventionen forstærker modningsfaktoren, hvilket eksempelvis vil kunne være tilfældet med en vellykket kognitiv terapi over for ungdomskriminelle.

Testning

Det er karakteristisk for det psykologiske forskningsområde, at det i de fleste tilfælde ikke bare er egentlige påvirkninger, men også selve observationen, der har en indflydelse på forsøgspersonen. Dette gælder, som vi allerede har set, for den form for observation, der består i en testning af egenskaber hos forsøgspersonen. Og specielt har vi ofte det særlige problem, som hedder retesteffekt, hvor resultatet af en efterfølgende testning, uanset om der er sket en intervention i mellemtiden eller ej, bliver påvirket

af, at forsøgspersonen genkender testen, og måske oven i købet husker sine tidligere svar.

Også denne fejlkilde søges modvirket ved hjælp af et forsøgsdesign med kontrolgruppe. Hvis retesteffekten er den samme i begge grupper, hvis de øvrigt er udvalgt ækvivalent, og hvis der ikke forekommer andre forskelle, må en evt. forskel mellem grupperne alene tilskrives interventionen. Når der imidlertid forekommer naturlige grupper, kan vi frygte, at retesteffekten fungerer forskelligt i de forskellige grupper.

Redskabsbrug

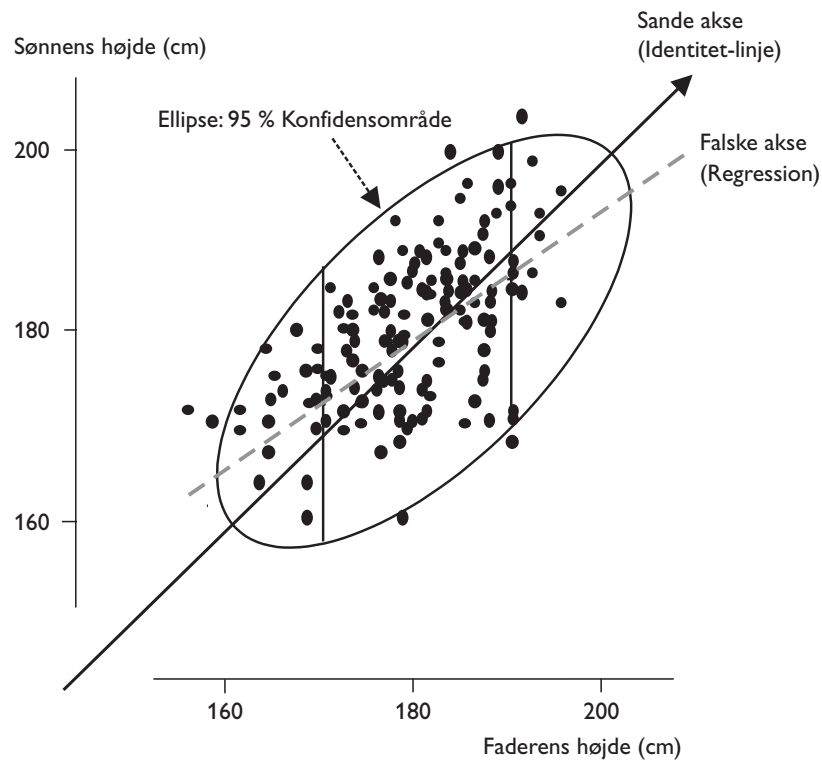
Redskabsbrug er en ret speciel fejlkilde, der består i, at redskabsbrugen varierer over tid eller i forskellige grupper. Man kan således tænke sig, at proceduren ved en testning varierer med de forskellige testere, der indgår i eksperimentet. Hvis der således sker en udskiftning af test i en forsøgsserie, eller hvis der bruges forskellige procedurer i forbindelse med forskellige testere/testpersoner i forskellige grupper, kan denne fejlkilde forekomme.

Regression

Regression er et meget alvorligt metodeproblem, hvis farlighed er knyttet til, at det ikke er umiddelbart psykologisk indlysende, men knyttet til et statistisk artefakt, som netop hedder *regression*. Ordet betyder tilbagegang, og da fænomenet blev opdaget af Francis Galton (1889), troede han først, at der var tale om et rent biologisk fænomen. Han sammenlignede biometriske størrelser af forældre og afkom og opdagede bl.a., at sønner af små fædre blev større end deres respektive fædre, mens omvendt sønner af store fædre blev mindre end disse. Dette kunne jo på et tidspunkt, hvor den genetiske viden var meget ufuldstændig, ses som en naturlig regulering til forhindring af, at træerne vokser op i himlen henholdsvis ned i jorden.

Baggrunden for regressionsfænomenet er, at hvis vi ser på genotypen i det anførte tilfælde, vil der slet ikke være forskel på faderens og sønnens højde, men fænotypen vil være påvirket af tilfældige vækstfaktorer, sådan at hvis vi tager en far med en fænotypisk lav højde, så vil denne sandsynligvis være mindre end det, hans genotype betinger, og da sønnen ikke er udvalgt efter lav højde, men kun bestemt af genotypen samt eventuelle tilfældighedsfaktorer, så vil middelhøjden for en søn være lig med faderens genotype og dermed højere end faderens tilfældigvis relativt lave fænotype.

I figur 2.18 er dette forhold fremstillet i grafisk form.



Figur 2.18. Det statistiske regressionsfænomen.

Vi har faderens højde angivet ud ad X-aksen og sønnens ud ad Y-aksen. Et punkt angiver således et enkelt par af højde for henholdsvis far og søn. Hvis vi har at gøre med normalfordelte størrelser med en vis korrelation, så vil punktsværmene for alle datapar befinde sig inden for et ellipseformet område. I figuren er indtegnet det mest almindelige, nemlig 95 %-konfidensområde. Det vil sige det område, hvor vi har 95 % af alle punkter.

I diagrammet er regressionslinjen defineret som middelværdien af alle Y-værdier for en givet X-værdi. Her altså middelværdien af sønnens højde på den lodrette linje, der går igennem en bestemt værdi for faderens højde. Når vi tegner en lodret linje gennem faderværdien på 170 cm, så har vi en fordeling af sønnens højde i intervallet fra 160 til 190 cm, middelværdien bliver da 175 cm, altså 5 cm over faderværdien. Tilsvarende for linjen

gennem faderværdien 190 cm, hvor vi har intervallet 170 til 200 cm, med middelværdi 185, dvs. 5 cm under faderværdien.

Den stiplede regressionslinje ligger således for højt for faderværdier under gennemsnittet (her 180), og for lavt for faderværdier over gennemsnittet. Det betyder, at regressionslinjen får en mindre hældning end identitetslinjen, som angiver den korrekte relation mellem faderens og sønnens højde, nemlig at der ikke er nogen systematisk forskel. At den førnævnte middelværdi af sønnens højde for faderværdien 170 cm var på 175 cm, og at den tilsvarende middelværdi for faderværdien 190 cm var 185 cm, betyder en middelforskel på kun 10 cm i sønnens højde svarende til forskellen mellem 170 og 190 hos faderen, altså 20 cm. Det indebærer igen, at stigningstakten kun er 50 %, hvilket i regressionstermer angives ved, at regressionskoefficienten ligeledes er 0,50.

Regressionsfænomenet er erfaringsvis svært at forstå, men det har en ret enkelt forklaring: Når vi tager en plusafviger, altså et individ, der befinder sig i den øverste halvdel af sin population, vil det være en forholdsvis stor sandsynlighed for, at egenskaben er præget af tilfældigheder, der har bevirket den store værdi. Ved højde kan det således ud over genetiske faktorer, som måske ligger forholdsvis fast, være særlige vækstbetingelser under svangerskab eller opvækst, således højden fx er blevet 195 cm, hvoraf de 190 er genetisk bestemt, mens de sidste 5 altså tilskrives særlige betingelser under svangerskab eller under opvækst. Sønnen arver nu den genetiske disposition, men vil i almindelighed ikke have de samme gunstige betingelser som sin far under svangerskab og opvækst. Tilsvarende for en minusafviger, altså en far med en lille højde. Ud over de genetiske anlæg vil der være større sandsynlighed for forhold under svangerskab og opvækst som har været ugunstige for højdevækst, og disse vil ikke som de genetiske gå videre til sønnen.

Klinisk regression

En særlig form for regressionseffekt er den *kliniske regression*. Den forekommer i psykologiske forsøg, hvor forsøgspersonerne er udvalgt pga. af en forholdsvis ekstrem startværdi. Hvis vi fx undersøger psykologisk behandling af patienter med dermatitis (børnesår), hvor udvælgelsesgrundlaget er henvendelse til sundhedssystemet om denne lidelse, kan vi forvente, at lidelsen er særlig fremtrædende på henvendelsestidspunktet, og derfor kan vi også forvente, at den generelt vil være i bedring på et senere tidspunkt.

En observeret bedring kan altså lige så godt være en regressions- som en behandlingseffekt.

Her vil et design med forsøgs- og kontrolgruppe kunne give en kontrol af en evt. regressionseffekt, idet ækvivalente grupper vil have samme regressionseffekt. En tilsvarende kontrol kan vi imidlertid ikke uden videre påregne med naturlige grupper.

Udvælgelse

Udvælgelsesaspektet vedrører kravet om udvælgelsesækvivalens, denne er inden for sandsynlighedsregningens rammer sikret gennem randomisering, men altså ikke med de naturlige grupper i kvasieksperimentet. Og dette er en vigtig grund til, at Campbell og Stanley afviser at opfatte det som et eksperiment.

Bortfald

Som vi tidligere har været inde på, kan et *selektivt bortfald* i lige så høj grad som en dårlig udvælgelse skabe en skæv stikprøve. På grund af ækvivalensen i det ægte eksperiment vil man her i det mindste, gennem sammenligning af grupperne efter interventionstidspunktet, kunne bedømme interventionens virkning på den del af populationen, som ikke er udsat for bortfald. Dette er i sig selv en alvorlig begrænsning, som dog kan mildnes ved at undersøge, om der er overensstemmelse mellem observationerne før interventionen for de personer, som er forsvundet fra forsøget, og for dem, som stadig er med. Truslen mod den interne validitet bliver endnu større, hvis der er interaktion mellem intervention og bortfald, altså hvis interaktionen i forsøgsgruppen enten forstærker eller formindsker bortfaldet i forhold til kontrolgruppen. Disse bortfaldsproblemer findes allerede i det ægte gruppeeksperiment, men de må forventes at blive forstærket i kvasieksperimentet, hvor der ikke er sikret ækvivalens mellem grupperne, og hvor vi derfor må regne med en mulighed for systematisk forskel med hensyn til bortfald. Her er der altså et skærpet behov for kontrol ikke bare hvad angår empirisk ækvivalens mellem grupperne før interventionstidspunktet, men også hvad angår ækvivalens med hensyn til bortfald.

Den eksperimentelle metodes problemer anskuet symptomalt

Ud over de mere tekniske problemer, som er knyttet til de forskellige typer af eksperimentelle metoder, er der et metodologisk forhold, der skal fremhæves i dette afsnit. Når vi ser bort fra demonstrationeksperimentet, hvis ærinde er et andet end de kausale eksperimenter, så gælder det for de sidstnævnte, at det faktisk kun er det klassiske, individbaserede eksperiment, som slipper helskindet gen en metodologisk analyse, og det kun inden for dets anvendelsesområde, som er begrænset til den del af det almenpsykologiske område, der grænser op til det biologiske.

Når vi ser på mere højtorganiserede og mere komplekse fænomener, så viser der sig en række metodologiske problemer, der ikke har samme tekniske karakter, men som må opfattes som symptomer på en principiel modsigelse ved anvendelse af kausale eksperimenter til afprøvning af hypoteser inden for personligheds-, social- og udviklingspsykologi og til praktisk undersøgelse af intervention af pædagogisk, klinisk og arbejdspsykologisk art.

Problemerne med alle de fejlkilder, vi gennemgik i forrige kapitel: altså interviewereffekten, fordringseffekten, selvrepræsentationseffekten, de to Rosenthal-effekter og Hawthorne-effekten, peger alle på en barriere for den kausale reduktion indbygget ikke bare i den epistemiske, men også i den ontologiske grundantagelse (se s. xx i kapitel 1) i ethvert kausalt eksperiment, nemlig at der *ikke* foreligger samme forhold mellem forsøgsleder og forsøgsperson som mellem eksperimentator og forsøgsopstilling i naturvidenskaben.

Da Galilei (1957) foretog sit berømte snit mellem *primære og sekundære* kvaliteter udelukkede han imidlertid med føje subjektive forhold fra sin nye naturvidenskabelige metodologi. Han undersøgte objektive kvaliteter ved rene objekter. Entiteter der ikke bare besidder objekt karakter i relativ, men i 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 vores biologiske udstyr.

Når vi ønsker at undersøge fænomener fra psykologiens kerneområder, kan vi enten anskue selve relationen mellem forsøgsleder og forsøgsperson som en principielt uoverstigelig metodologisk barriere eller som et grundlæggende forhold ved selve de psykologiske fænomener, vi studerer. Når

vi ikke længere studerer naturvidenskabelige entiteter, hvis basale natur er os fremmed, så gælder digteren John Donnes (1957) ord om, at *intet menneske er en ø*.¹⁰ Vi kan ikke konsistent betragte en forsøgspersonen som et objekt isoleret fra forsøgslederen, for denne antagelse strider imod selve det vilkår for psykologiske undersøgelser, at den består i en relation mellem disse to.

Man kan måske sige, at jo mere man forsøger at betragte forsøgspersonen som et isoleret objekt og forsøgslederen som en uafhængig variabel uden anden forbindelse med forsøgspersonen end den at være observatør, jo mere vikler man sig ind i metodologiske problemer, når der ikke er tale om snævert psykofysiologiske fænomener.

Alle de nævnte metodeproblemer udgør dermed en del af selve det psykologiske eksperiments socialpsykologi. De er symptomer på den principielle modsigelse i at gennemføre en kausal reduktion af den menneskelige intentionalitet. Dette indebærer ikke, at de eksperimentelle metoder er uanvendelige, blot at de har den begrænsning, at de ikke kan sikre en kausal reduktion af de egentlige psykologiske fænomener.

Derfor kan de heller ikke stå alene. De er særdeles nyttige redskaber i psykologiens værktøjslager, men må suppleres med mere subjektive metoder, hvor den kompleksitet, der er indbygget i relationen mellem forsøgsleder og forsøgsperson, ikke blot er en fejl, men i lige så høj grad en informationskilde.

Litteratur

- Asch, S.E. (1951). Effects of group pressure upon the modification and distortion of judgements. I: H. Guetzkow (ed.), *Groups, Leadership, and Men*. Pittsburgh; Carnegie Press.
- Asch, S.E. (1952) Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs* 70. Hele nr. 416.
- Asch, S.E. (1952). *Social Psychology*. New York: Prentice Hall.
- Asch, S.E. (1955). Opinions and social pressure. *Scientific American*, 193:31-35.
- Baade, W. (1915). Aufgaben und Begriff einer „darstellenden“ Psychologie. *Zeitschrift für Psychologie*, 71:356-367.
- Baade, W. (1916) Über psychologische Darstellungsexperimente *Archiv für die gesamte Psychologie*, 36:1-23.
- Bruner, J.S. (1992). Another look at New Look. *American Psychologist*, 47:780-783.
- Bruner, J.S. & Goodman, C.C. (1947) Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology*, 42:33-44.

- Bundesen, C. & Larsen, A. (1975). Visual transformations of size. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 3:214-220.
- Bundesen, C., Larsen, A. & Farrel, J.E. (1981). Mental transformations of size and orientation. I: J. Long & A. Baddeley (eds.). *Attention and Performance IX*, s. 279-294. Hillsdale, N.J.: Lawrence Erlbaum.
- Campbell, D.T. & Stanley, J.C. (1966). *Experimental and Quasiexperimental Designs for Research*. Chicago: Rand McNally.
- Cannon, W.B. (1929). *Bodily Changes in Pain, Hunger, Fear, and Rage*. 2nd ed. New York & London: D. Appleton & Co.
- Donne, J. (1967 [1624]). *Essays in Divinity*. ed. by Evelyn M. Simpson. Oxford: Clarendon.
- Fechner, G.T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf & Härtel.
- Festinger, L. & Carlsmith, J.M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58:203-210.
- Galileo, G. (1957 [1623]). *The Assayer*. Genoptrykt i: Stillman Drake, *Discoveries and Opinions of Galileo*. New York: Doubleday.
- Galton, F. (1889). *Natural Inheritance*. London: Macmillan.
- Golding, W.G. (1999 [1959]). *Lord of the Flies*. London: Penguin.
- Herrnstein, R.J. & Boring, E.G. (eds.) (1965). *A Source Book in the History of Psychology*. Cambridge, MA: Harvard University Press.
- Grossman, Jason (2005). The randomized controlled trial: gold standard, or merely standard? *Perspectives in Biology and Medicine*, 48(4):516-534.
- Lange, C.G. & James, W. (1967 [1922]). *The Emotions (Volume I)*. Baltimore: Williams & Wilkins.
- Milgram, S. (1974). *Obedience to Authority*. New York: Harper and Row.
- Nielsen, T.I. (1978). *Handler: Subjektproblematiserende analyser og synteser*. København: Dansk Psykologisk Forlag.
- Schachter, S. (1964). The interaction of cognitive and physiological determinants of emotional state. I: L. Berkowitz (ed.), *Advances in Experimental Social Psychology*. New York: Academic Press.
- REF = Schachter, S. & Singer, J.E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 69:379-99.
- Shepard, R.N. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171:701-703.
- Weber, E.H. (1835). Über den Tatsinn. *Archiv für Anatomie, Physiologie und wissenschaftliches Medizin, Medical Müller's Archives*, 152-159.

Noter

Note 1, s. Ved "demonstration" (Darstellung) forstår man omsætningen af reale eller abstrakte ting, sagforhold, begivenheder eller endda følelser til konkrete tegn og modeller. (Kilde: Wikipedia).

Note 2, s. Ernst Heinrich Weber var en af de første, der som professor i anatomi

og fysiologi i Leipzig fra 1821 udviklede det psykofysiske eksperiment med sine tærskelundersøgelser af følesansen (Weber, 1835). Han opdagede her, at tærskelværdien stiger med stimulus (in casu vægt), idet forholdet mellem de to størrelser er konstant. På denne baggrund formulerede Gustav Theodor Fechner i *Elemente der Psychophysik* (1860) en lov om en logaritmisk sammenhæng mellem stimulus og respons. Dette er også beskrevet af Herrnstein og Boring (1965:66-75).

Note 3, s. I begrebet *Darstellung* ligger også betydningen fremstilling gennem skuespil.

Note 4, s. Navnet skyldes det kronologiske sammenfald, at 1947 ikke bare var året for en revolution i psykologien, men også i moden, idet Christian Dior samme år chokerede verden med indførelsen af en mere ekstravagant og luksuriøs designstil til afløsning af den strenge og asketiske stil fra krigsårene. Måske lå der også i dette lån en antydning af, at Bruner og hans medarbejdere på tilsvarende måde satte noget mere levende og udfordrende i stedet for den asketiske eksperimentpsykologi.

Note 5, s. Den aldrende Bruners egen beretning findes i artiklen "Another look af New Look" (Bruner 1992).

Note 6, s. "Post hoc" betyder *efter dette* og betegner hypoteser eller forklaringer, som først bliver fremsat efter den empiri, de omhandler, og som derfor er i umiddelbar modstrid med Poppers hypotetisk-deduktive metode, ifølge hvilken man først skal foretage en teoretisk udledning af en hypotese og derefter afprøve den eksperimentelt. En undtagelse fra den problematiske praksis at komme med en forklaring bagefter er dog det, som den amerikanske filosof og videnskabs-teoretiker Pierce kalder abduktion. Her bygger hypotesen ganske vist på en empirisk iagttagelse, men giver anledning til en teoriopstilling, som indebærer uafhængige konsekvenser, der kan afprøves empirisk. Den egentlige problematiske form for *post-hoc*-hypoteser er derfor dem, som er *ad hoc*. "Ad hoc" betyder *til dette*, og henviser til de forklaringer eller hypoteser, der for det første opstilles efter en uventet iagttagelse (måske ligefrem en iagttagelse, der går imod den teoretisk forudsagte), og som ganske vist giver en slags forklaring på denne iagttagelse, men som er snævert udformet med dette formål alene. Om der i dette tilfælde var tale om abduktion eller om en ren *ad hoc*-hypotese må afhænge af, om der faktisk blev foretaget en generalisering og efterfølgende afprøvning. Således må hypotesen indebære, at der var forskel på de drenge, der faktisk havde visse erfaringer med den mest værdifulde mønt og dem, der ikke havde det. Der kunne således være tale om drenge, som havde småjobs i forretninger, hvor de så halvdollar-mønter. En sådan afprøvning er mig bekendt

ikke foretaget, men da hypotesen har konsekvenser, der kan afprøves selvstændigt, er den nok abduktiv snarere end ad hoc.

Note 7, s. Asch var i 1920 som 13-årig indvandret fra Polen med sin jødiske familie.

Han var tæt knyttet til flere af de tyske gestaltpsykologer, som måtte flygte under krigen, således var Wertheimer hans læremester og Köhler i en periode hans kollega. Som det fremgår af hans socialpsykologi, var den turbulente periode med de totalitære styrers fremvækst, anden verdenskrig og den kolde krig i 1950'erne baggrunden for en meget reel bekymring over konformitetspres. Det er sigende, at Asch kalder sin monografi om eksperimentet for *Studies in Independence and Conformity: A Minority of One against a Unanimous Majority*.

Note 8, s. Den kronometriske metode var nogle år forinden indført af Shepard (1971), som forfatterparret havde besøgt under studieophold i USA.

Note 9, s. Af eng. *random* = tilfældig. Se nærmere forklaring i afsnittet om randomiseret stikprøve, s. xxx.

Note 10, s. Citatet lyder i al sin vælde: "Intet menneske er en ø, komplet i sig selv ... ethvert menneskes død formindsker mig, for jeg er en del af menneskeheden; det er derfor ikke mig forundt at vide, for hvem klokken ringer; den ringer for dig." Meditation XVII, fra John Donnes Essays, skrevet 1624 (Donne, 1957).

KAPITEL 3

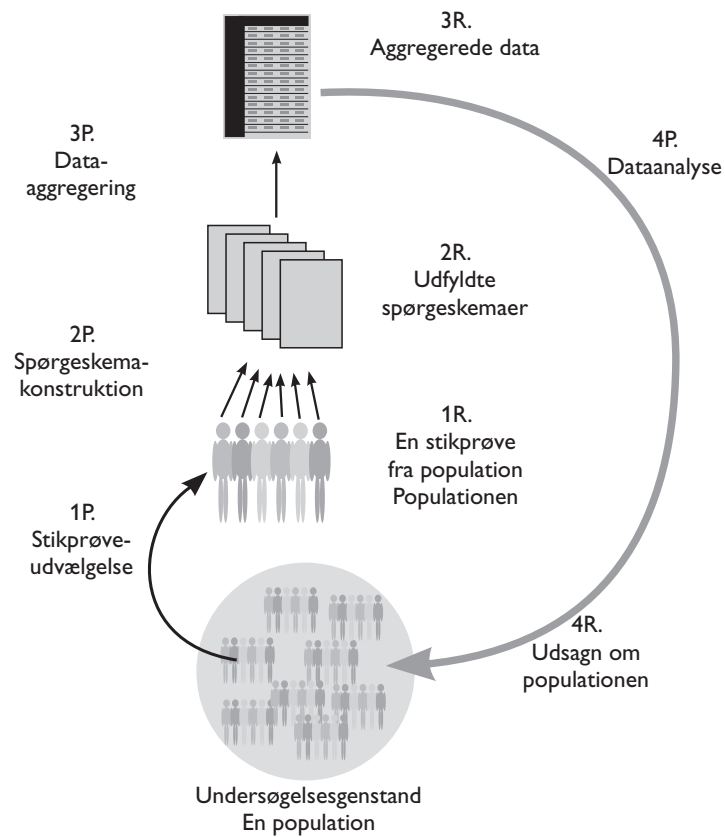
Surveys med kvantitative data

En karakteristik af surveyundersøgelsen som genre

Som vi så i kapitel 1 er survey- eller spørgeskemaundersøgelsen slet ikke en specifik psykologisk undersøgelsesmetode, men er i sit udgangspunkt en måde at studere et samfundsmæssigt fænomen på. *Sigtet* med en surveyundersøgelse, er således typisk at undersøge adfærd, forestillinger eller holdninger i en bestemt befolkning eller befolkningsgruppe, hvad enten temaet er videnskabeligt, administrativt eller kommercielt. Den *epistemiske undersøgelsesgenstand* er altså den befolkning eller befolkningsgruppe, hvis karakteristika vi vil udforske. Den *empiriske undersøgelsesgenstand* er derimod den *stikprøve* af befolkningen eller befolkningsgruppen, der faktisk bliver udvalgt som undersøgelsespersoner. Ved undersøgelsen går vi fra et *sociologisk aggregat* (kollektiv størrelse) over *stikprøven* bestående af *individdata*, og videre fra stikprøven til *aggregerede data*. De *aggregerede data* er resultatet af en *datareduktion*, hvor individuelle oplysninger om stikprøvens enkeltindivider bliver samlet til generelle oplysninger om hele stikprøven. Til sidst går vi fra de aggregerede data over til udsagn om den population, som jo er den epistemiske genstand for vores undersøgelse. Figur 3.1 ■ viser denne proces.

I diagrammet er gangen i en surveyundersøgelse opdividset i fire faser:

1. Stikprøveudvælgelse.
2. Datagenerering: spørgeskemakonstruktion og udfyldning.
3. Dataaggregering.
4. Dataanalyse.



Figur 3.1. Forløbet i en survey

For hver fase har vi en proces (P) og resultatet heraf (R).

1. Udvalgelse:
P = Valg af udvælgelsesmetode og selve udvælgelsen,
R = Stikprøve.
2. Datagenerering:
P = Spørgeskemakonstruktion,
R = Udfyldte skemaer.
3. Dataaggregering:
P = Dataaggregering,
R = Aggregerede data.
4. Dataanalyse:
P = Dataanalyse,
R = Udsagn om populationen.

I udvælgelsesfasen opstilles en udvælgelsesmetode, og med denne udvælgelses stikprøven fra populationen. I datagenereringsfasen konstrueres der et spørgeskema, som bliver udleveret til udfyldelse af de udvalgte undersøgelsespersoner. Dataaggregering indebærer, at de individuelle data fra skemaerne samles i statistiske opstillinger, aggregerede data, hvor kun informationen om den samlede stikprøve foreligger. Det er til syvende og sidst på grundlag af disse aggregerede data, at vi opnår information om populationen. Der sker en sociometrisk datareduktion, hvor indviddata er inddata og aggregerede data er uddata. En datareduktion er en bearbejdning af data, hvor omfanget og kompleksiteten formindskes ved en sammentælling eller sammenlægning af tallene. En sådan statistisk bearbejdning kaldes psykometrisk, hvis den tilhører det psykologiske forskningsområde, og sociometrisk, hvis den hører under det egentlig sociologiske, socialvidenskabelige område. Vi skal i næste kapitel gennemgå psykometriske metoder.

Et eksempel på en surveyundersøgelse

Surveymetoden er en af de mest brugte og samtidig en af de i psykologisk sammenhæng mest problematiske metoder. Den er for det første problematisk, fordi den ligger ret skævt i forhold til psykologiens genstandsområde og kun er rigtig adækvat ved samfundsvidenskabelige problemstillinger. For det andet er den oftest ret blind eller i det mindste nærsynet mht. til procesforløb og egentlig årsagssammenhæng. „Survey“ betyder om allerede nævnt *overblik* eller *oversigt*, og surveymetodens mål er netop at give et vist overblik over en social fænomenkreds.

Vi indleder med en eksemplarisk gennemgang af den foreliggende forskningsmetode. Vi vil i den forbindelse se på den store amerikanske familievoldsundersøgelse (Strauss et al., 1980), som da også er eksemplarisk i begge ordets betydninger. Den er et eksempel på en surveyundersøgelse, og den er et eksempel til efterfølgelse, når det gælder surveys.

Der er tale om en stor repræsentativ undersøgelse, hvor stikproven er udvalgt blandt de 46 mio. familier, der i 1976 levede i USA.

Undersøgelsens baggrund og sigte

Baggrunden for undersøgelsen var den debat, som kvindebevægelsen, i USA kaldet Women's Lib., rejste om vold mod hustruer. En debat, der bl.a. førte til oprettelsen af tilflugtssteder, på engelsk kaldet „shelters“. ¹ Da det var et hidtil neddysset fænomen, var der stor usikkerhed, hvad angik omfanget af denne vold, dens karakter og dens årsager. En række forskere, der gennem 1970'erne havde været pionerer mht. dette forskningsområde, fik derfor en meget stor bevilling til at lave en national undersøgelse i USA.

Undersøgelsens sigte er defineret gennem fire undersøgelsesmål:

1. Fastlæggelse af omfanget af vold i amerikanske familier. (Hvor meget vold der er, hvilke slags familier der er voldelige, og hvad der forårsager vold har indtil denne undersøgelse alene været undersøgt gennem små ikke-repræsentative undersøgelser).
2. Fastlæggelse af spektret af familievold. Hvilke voldsrelationer er der tale om, fx: mand mod kone, kone mod mand, forældre mod børn, og søskende indbyrdes. Også alvoren eller grovheden af volden undersøges.
3. Hvad betyder volden for de implicerede, hvilke familier er voldelige, og hvilke er ikke?
4. Hvad er årsagerne til volden?

Voldsbegrebet

Der skelnes i undersøgelsen mellem tre primære konfliktniveauer, ikke voldelig aggressivitet, normal vold, grov vold (grænsen mellem kategori 2 og 3 svarer til forskellen i politi- og domstolssprog mellem legemskrænkelse og legemsbeskadigelse).

Udvælgelsesmetoden

Når man skal lave en repræsentativ undersøgelse, kan man gøre det på to måder:

- Simpel tilfældig udvælgelse: Her trækker man simpelt hen lod i hele populationen (de 46 mio. familier).
- Stratificeret, tilfældig udvælgelse. Her opdeler man først populationen

i strata (lag), fx efter geografiske, etniske eller sociale kriterier og trækker så lod i hver stratum.

Det er den sidstnævnte metode Familievoldsundersøgelsen har anvendt. Konkret blev 103 primærområder, amter eller grupper af amter, udvalgt blandt de 3.142 amter i USA ud fra en stratifikation omfattende geografi, samfundsmæssige og befolkningsmæssige karakteristika. Inden for disse 103 primærområde blev der valgt 300 sekundærområder som interviewlokaliteter. Hver interviewlokalitet blev så opdelt i 10-25 boligkvarterer, hvorfra familierne så til sidst blev udvalgt.

Der var følgende inklusionskriterier:

- Kun intakte familier (mand og kone samlevende, ikke nødvendigvis gift).
- Ingen familier med børn under tre år (de mentes at være for stressede og derfor evt. mere voldelige end andre).

Der er åbenbart blevet satset på en stikprøve på ca. 3.300, hvilket svarer til en familie ud af hver 15.000. Da svarprocenten blev 65, er der faktisk inddraget 2.143 familier i undersøgelsen.

Der er to bemærkninger til svarprocenten.

For det første er den imponerende høj. Det skyldes tre forhold:

1. Det er en interviewundersøgelse, ikke en post-spørgeundersøgelse.
2. De udvalgte familier er besøgt fire gange, og har fået adskillige breve.
3. Der blev tilbudt økonomiske incitamenter (som det udtrykkes i bogen).

Den anden bemærkning er, at det lige er nok til at redde undersøgelsen. Sagen er den, at hvis der er flere, der svarer, end som undlader at svare, så er der meget ofte grund til antage, at nægterne er helt anderledes end dem, der svarer, og så kan resultatet være helt misvisende. Det kan også være tilfældet her, men antallet af nægtere er trods alt kun det halve af antallet af deltagere i undersøgelsen.

Udspørgemetoden

Der anvendtes et fuldt struktureret interviewskema (med et større hold af specielt trænede interviewere). Interviewpersonen var altid en far eller mor, i halvdelen af tilfældene var manden udvalgt, i den anden halvdel var det konen.

Kernen i interviewet var *Conflict Tactics Scale*. Dette skema (se tabel 3.1) blev benyttet om tre slags konflikter, i nævnte rækkefølge:

1. Konflikter mellem søskende.
2. Konflikter mellem forældre og børn.
3. Konflikter mellem ægtefæller (samlevende).

I tilfælde 1 gjaldt spørgsmålene et bestemt af børnene, i tilfælde 2 og 3 både interviewpersonen som gerningsmand (m/k) og den anden part. I tabel 3.1 er det den udgave, som omhandler punkt 3, der er bragt. Det fremgår, at aggressioner under voldtærsklen er angivet på hvid baggrund, mindre alvorlig (normal) vold er angivet på lysegrå baggrund og grov vold på mørkegrå baggrund.

Læg mærke til den gradvise eskalation mht. alvoren. Og det endda i to dimensioner.

For det første i voldsrelationen, der går fra børnene indbyrdes til konflikter mellem forældre og børn, hvor mange stadig anså korporlig afstraffelse for berettiget, og endelig til konflikter mellem mand og kone.

For det andet i selve konflikttrappen, fra den fredelige diskussion over hævet stemme og kasten med harmløse genstande, til knytnæver, slag med hårde genstande og kulminationen i form af brug af dødbringende våben.

CTS-skemaets validitet

CTS-skemaet er blevet valideret, dvs. afprøvet for, om det også peger på de reale fænomener, det er udviklet til at registrere.

Der er benyttet en snedig metode, hvor studenter fra nogle sociologikurser blev bedt om at angive, hvad deres far og mor havde gjort i det foregående år inden for skemaets kategorier. Samtidig blev et skema tilsendt faderen og moderen til besvarelse.

Der var en svarprocent på næsten 100 for de 110 studenter, der deltog, og en svarprocent på 72 hos forældrene.

Tabel 3.1. Skala for konflikthåndtering.

Tabel 3.2.

Skala for konflikthåndtering	Korrelation for ægtemænd	Korrelation for hustruer
Diskussion	0,19	-0,12
Verbal aggression	0,51	0,43
Vold	0,64	0,33

Resultatet af denne validering ses i ■ tabel 3.2 (Strauss 1980, s. 262).

Der er altså ingen enighed omkring opfattelsen af diskussioner mellem børn og forældre. Derimod er der en klar korrelation mht. verbal aggression. Hvad angår vold er det interessant, at sammenhængen ser ud til at være størst med faderens angivelse. Man kunne måske få den mistanke, at en del af konerne er tilbøjelige til at idyllisere.

Undersøgelsens databehandling

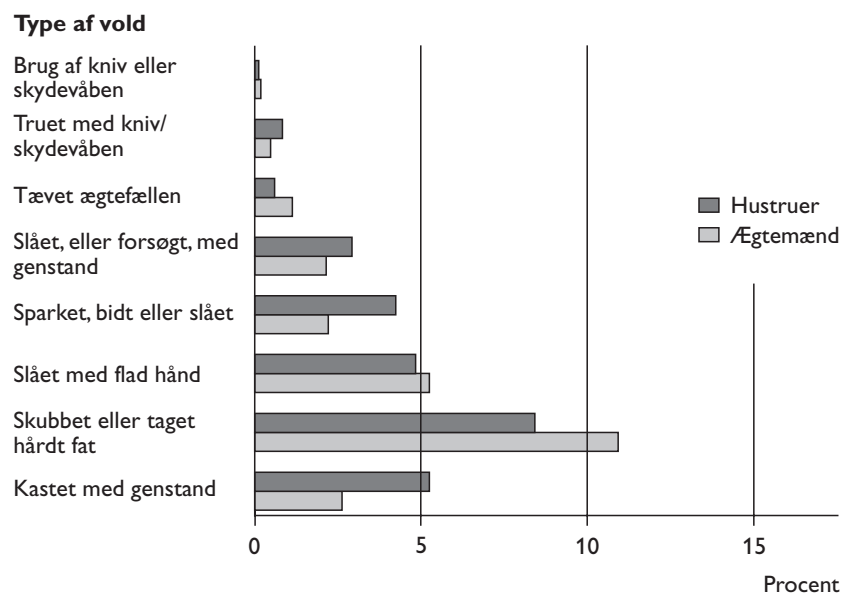
I databehandlingen er der først foretaget en ren datareduktion i form af en simpel optælling af frekvenser, altså forekomster af bestemt typer af konfliktadfærd og vold i familien. Derefter er der foretaget en egentlig analyse med et dybere teoretisk sigte.

Vi skal først kigge på undersøgelsens resultater og se på omfanget af CTS's aggressionsformer i relationen mellem mand og kone (figur 3.2).

Her er kategorierne anbragt, så de alvorligste ligger øverst, og vi ser, at hyppigheden stort set falder med grovheden. Der er 10 % af såvel mænd som koner, der har skubbet eller hevet i hinanden i det forløbne år, mens der er under en procent, der har brugt kniv eller skydevåben mod hinanden.

Med hensyn til kønsforskellen er der i det store og hele en forbavsende grad af ligestilling. Men hvert køn har sine favoritter. I den mere fredelige ende er mændene gladere end konerne for at skubbe og tæve, mens konerne har større hang til at kaste og slå med genstande og true end mændene.

Mest alvorligt er det nok, at 1 % i undersøgelsesåret har gennembanket deres kone. Det er altså ca. en halv million, det drejer sig om.



Figur 3.2. Aggressionsformer i relationen mellem mand og kone.

Jeg sagde i indledningen, at surveyens særlige form indebærer nogle metodiske svagheder. Det drejer sig især om følgende:

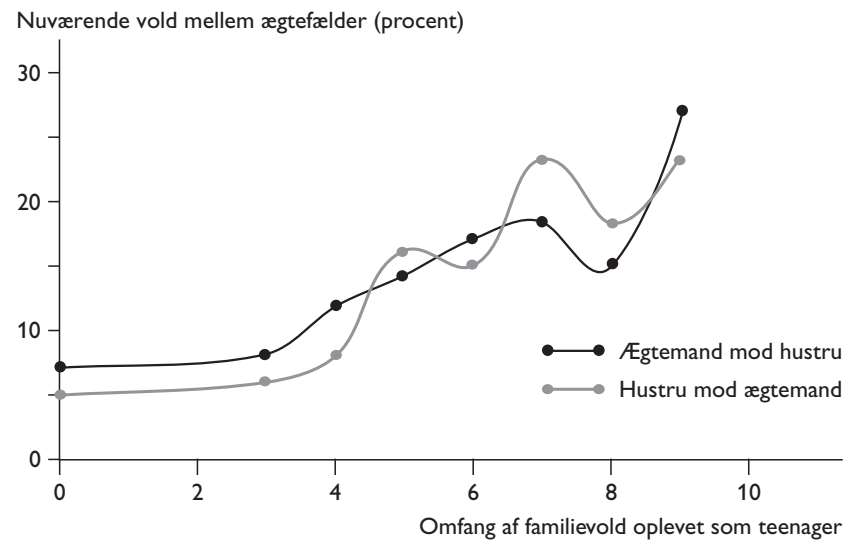
1. Individet forsvinder i mylderet af informanter.
2. Sammenhængene (den sociale kontekst bliver atomiseret i enkeltspørgsmål).
3. Tværsnitsdesignet fortæller ikke noget om udvikling eller om processer i tid i det hele taget.
4. Spørge- eller interviewskemaformen gør det (principielt) umuligt at gå fra statistiske sammenhænge til årsagssammenhænge.

Vi skal imidlertid nu se, at denne survey præsterer det kunststykke at forholde sig til processer, der oven i købet forløber over hele tre generationer. Dermed angriber undersøgelsen selve spørgsmålet om familiernes såkaldt *sociale*, eller rettere *familiemæssige* arv.²

Det kan den gøre, fordi den har oplysninger, som omfatter relationen mellem forældre og børn, i og med at der spørges om vold i to generationer inden for samme familie. Desuden spørges den voksne (forælder)-inform-

ant også om volden i sin egen barndomsfamilie, og dermed går vi yderligere en generation tilbage.

Vi skal nu se på det sidste, altså relationen fra bedsteforældregenerationen til forældregenerationen (figur 3.3).■

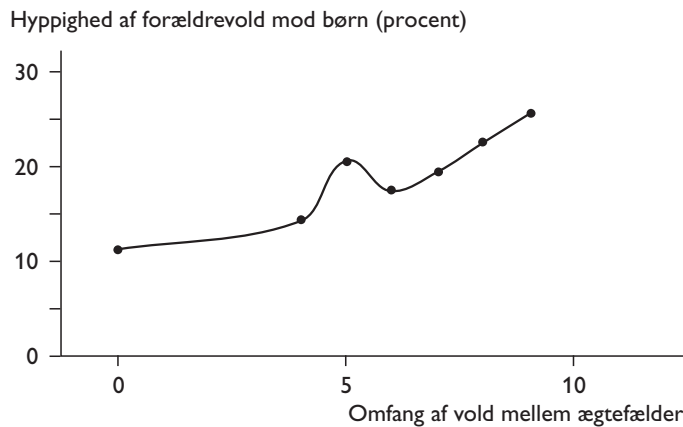


Figur 3.3. Intergenerationsammenhængen mellem vold i informaternes opvækstfamilie og ægtefælddevold i deres nuværende familie.

Vi ser her, at der med voksende *omfang af vold* i opvækstfamilien er en tendens til et voksende omfang af vold mellem de nuværende ægtefæller. Der ser således ud til at være en social arv fra den foregående til den nuværende generation (figur 3.4).■ Men går stafetten nu videre til børnene?

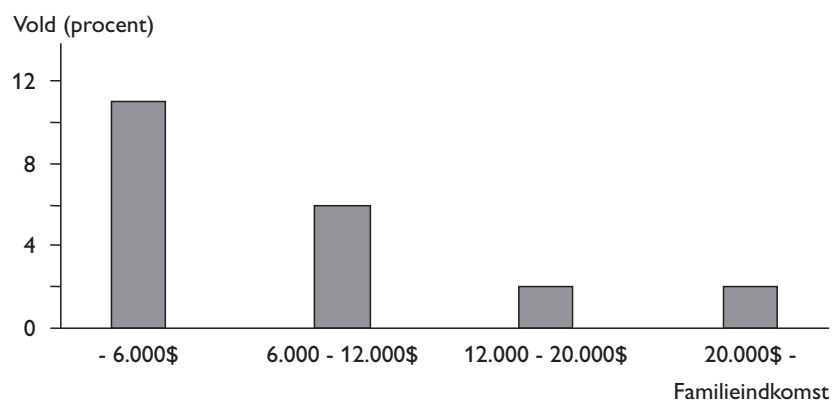
Det ser ud til, at vi har en ubrudt kæde over tre generationer, men er det nu også rigtigt? I en survey kan vi aldrig være sikre. Især skal vi passe på tilslørende faktorer (såkaldte *confounders*). Det vil sige skjulte eller ukendte variable, som styrer såvel den formodede uafhængige (forårsagende) som den afhængige (forårsagede) variabel. Hvis der nu var en variabel, der styrer volden i alle tre generationer, så er det den og ikke voldsmønsterets intergenerationelle transmission, der er på spil.

En sådan faktor kunne være socioøkonomisk status. Dette bekræftes ved en enkel sammenlignings med indkomst (figur 3.5):■ Når krybben er tom, bides hestene. Og de nøjes desværre ikke altid med at bide.



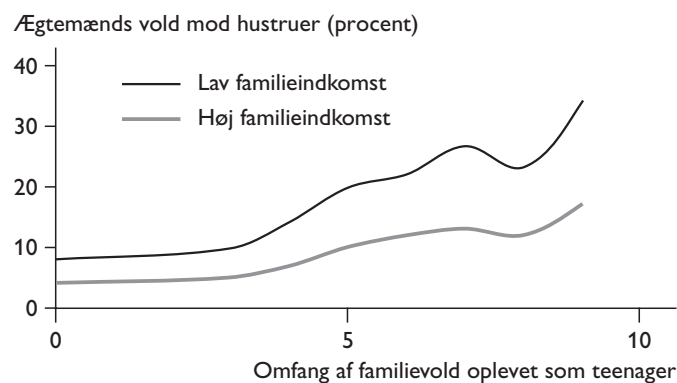
Figur 3.4. Sammenhængen omfanget af ægtefældrevold i nuværende familie og vold mod børnene.

Er indkomst så en tilslørende eller skjult faktor? Dette spørgsmål kan vi faktisk undersøge empirisk. Ved at bruge såvel den formodede tilslørende faktor (altså indkomst) som den tidligere udnævnte årsagsvariabel som uafhængige variable samtidigt, kan vi få syn for sagn.



Figur 3.5. Sammenhængen mellem familieindkomst og ægtemands vold mod hustruer.

Vi ser i figur 3.6 ■, at godt nok er den familiemæssige sociale vold ikke hele forklaringen, men det er indkomst på den anden side heller ikke. Hvis det forholder sig på denne måde (og her slår undersøgelsens i øvrigt

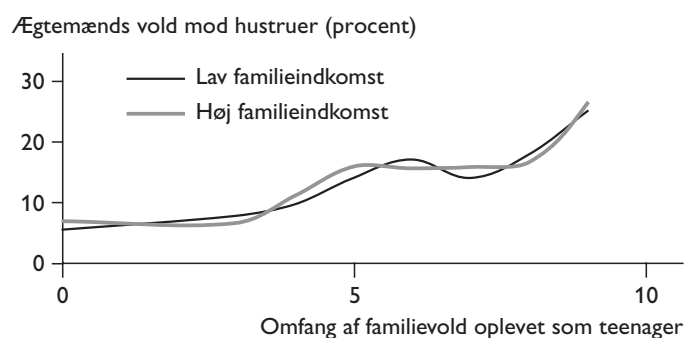


Figur 3.6. Trivariat analyse af indkomst og vold i opvækstfamilie som faktorer over for ægtefælle vold (konstruerede data med bivariat effekt).

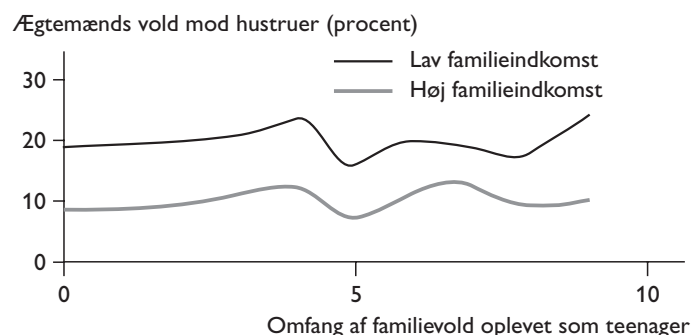
sofistikerede design desværre ikke til), så er der på den ene side den simple sociale afhængighed af indkomst (som i øvrigt også har sin egen sociale arv), og på den anden side den intergenerationelle transmission af familievoldsmønstre. Der er altså i hvert fald tale om en *to-faktoriel sammenhæng*.

Med disse to faktorer er der en række andre tænkelige muligheder. For eksempel at kun den ene, men ikke den anden var virksom. Hvordan havde det set ud, hvis det kun var et spørgsmål om intergenerationel transmission (familiemæssig arv) og ikke om socio-økonomisk status? (figur 3.7). ■

Det kunne imidlertid også være lige omvendt, at kun økonomien havde betydning (figur 3.8). ■



Figur 3.7. Trivariat analyse af indkomst og vold i opvækstfamilie som faktorer over for ægtefælle vold (konstruerede data med effekt kun af oplevet vold under opvækst).



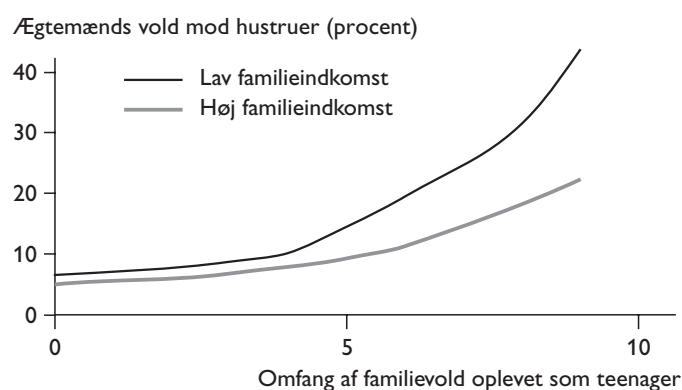
Figur 3.8. Trivariat analyse af indkomst og vold i opvækstfamilie som faktorer over for ægtefælle vold (konstruerede data hvor kun økonomien har effekt).

I dette diagram er der forskel på lav- og højindkomstfamilier, således at forekomsten af vold er størst hos de førstnævnte. Derimod er den anden faktor, omfanget af vold oplevet som teenager, uden betydning.

Den tredje mulighed var den, vi fik præsenteret i det første diagram (figur 3.6), hvor begge faktorer er virksomme samtidig og fungerer additivt, således at virkningen er noget i retning af summen af de to faktorer hver for sig.

Der er imidlertid også en fjerde mulighed, som slet ikke er utænkelig. Det kan være, at de to faktorer ud over deres specifikke effekt også forstærker hinanden (figur 3.9). ■

Her ser vi, at med en lav grad af vold under opvæksten er der ingen



Figur 3.9. Trivariat analyse af indkomst og vold i opvækstfamilie som faktorer over for ægtefælle vold (konstruerede data med to kausalfaktorer med interaktion).

forskel på lav- og højindkomstfamilier, mens der ved en meget høj grad af vold under opvæksten er dobbelt så meget vold (fra mandens side over for konen) i lavindkomstfamilierne som i højindkomstfamilierne (ca. 40 mod ca. 20 %). Et sådant fænomen kaldes i statistikken for *bifaktoriel interaktion*. Det vil sige, at de to faktorer har et kompliceret indbyrdes samspil. De kan fx forstærke hinanden. I et sådant tilfælde kunne man tale om en synergi-effekt. Hvis dette er tilfældet, gælder det ikke bare, at hestene bides, eller at hingsten i hvert fald bider hoppen, når krybben er tom. Den fattige hingst bliver så meget mere voldelig end den rige hingst, når den har oplevet vold som føl.

Vi ser i dette eksempel, at en velplanlagt og veludført survey kan komme ganske langt med at grave ned under de rene procentforekomster for at finde dynamiske sammenhænge. Der er imidlertid stadig den begrænsning, at vi ikke rigtig kan få at vide, hvorfor det forholder sig, som det nu gør. Mønstrer i det sidste eksempel med interaktion mellem familieøkonomi og vold under opvækst vil fx være meget interessant såvel praktisk som teoretisk. Men når vi begynder at spørge, *hvorfor* det nu forholder sig på denne måde, og *hvad* det er, der fremkalder synergien, bliver surveymetoden tavs. Man kan så gå videre med en udforskning af dette hvorfor på flere måder. For det første kan man søge at udvikle en teoretisk forklaring, som derefter kan afprøves med den hypotetisk-deduktive metode. For det andet kan man supplere med en kvalitativ metode, fx en semistruktureret interviewundersøgelse eller en feltundersøgelse, hvor man måske har en mulighed for at finde ud af, hvad det er der sker. (Se sidste afsnit i dette kapitel, s. xxx).

Surveydesign

Udvælgelsen af stikprøven

Et afgørende led i en survey er, som det fremgår af det indledende afsnit, udvælgelsen af den stikprøve, ud fra hvilken vi søger at få viden. Det er her væsentligt for undersøgelsens gyldighed, at stikprøven er *repræsentativ* for populationen, dvs. at den ikke er skævt udvalgt og dermed giver et fortegnet eller falsk billede af denne.

Der er her forskellige muligheder for udvælgelse:

1. Simpel randomiseret stikprøve.
2. Stratificeret randomiseret stikprøve.
3. Klyngeudvælgelse.

Simpel randomiseret stikprøve

Den enkleste og også mest udbredte metode er simpel randomisering, dvs. at man ud fra et register over alle individer i populationen foretager en tilfældig udtrækning. Tilfældig betyder her ikke principløs og arbitrær, men derimod at udvælgelsen sker gennem en form for lodtrækning, hvor alle individer i populationen har samme chance for at blive udtrukket. Har man således en population af andengenerationsindvandrerpiger fra 14 til 18 år i 8.-10. klasse, hvis uddannelsesplaner man ønsker at undersøge, så udgør denne population ca. 3.000 individer. Lad os nu sige, at man har ressourcer til en spørgeskemaundersøgelse omfattende 300 personer. Man kan da ud fra et register med 3.000 personer foretage en udtrækning, hvor hver person i registret har en chance for udvælgelse på 10 procent. Hvis man fx giver alle populationens individer løbenr. fra 1 til 3000, kan man med en billig lommeregner forsynet med den funktion, der hedder *random*, foretage en række udtrækninger af tilfældige tal mellem 0 og 1, som man hver gang multiplicerer med 3.000.

Når jeg nu på min 100 kroners lommeregner (figur 3.10) ■ aktiverer knappen RAND (for random) får jeg et tal mellem 0 og 1. Første gang er tallet (tilfældigvis) 0,5341. Det bliver ganget med populationsstørrelsen 3.000 til: 1602 (afrundet).

Jeg har nu den første i stikprøven og trækker det næste tilfældige tal, som denne gang bliver 0,6924, der giver nummeret 1498. Når jeg møder et

Figur 3.10. Randomknappen

allerede udtrukket tal, så trækker jeg bare et nyt, indtil der fremkommer et endnu ikke udtrukket. Og sådan fortsætter vi, til der er udtrukket de 300 personer, vi skal bruge.

Stratificeret randomiseret stikprøve

Lad os nu sige, at de 3.000 andengenerationsindvandrerpiger har en omtrentlig sammensætning som vist i (tabel 3.3).■

Tabel 3.3.

Eksjugoslavien	50
Tyrkiet	700
Libanon	30
Pakistan	100
Øvrige Asien	40
Afrika	30
Latinamerika	50
Europa (uden Eksjugoslavien og Tyrkiet) + Nordamerika	2000

Tabel 3.4.

Oprindelsesland	Antal i population	Antal i stikprøve
Eksjugoslavien	50	5
Tyrkiet	700	70
Libanon	30	3
Pakistan	100	10
Øvrige Asien	40	4
Afrika	30	3
Latinamerika	50	5
Europa (uden Eksjugoslavien og Tyrkiet) + Nordamerika	2000	200
I alt	3000	300

Ved en stratificering fordeler vi nu vores stikprøve på 300 forholdsmæssigt (Tabel 3.4).■

Ved den stratificerede udvælgelse trækker vi nu som før, men vi tager først 5 fra Ex-Jugoslavien, så 70 fra Tyrkiet osv.

Klyngeudvælgelse

Ved klyngeudvælgelsen er vores udvælgelsesenhed ikke individer, men større samfundsmæssige enheder, som fx familier, husstande eller skoler. Således er klyngen i familievoldsundersøgelsen en komplet familie med to forældre og hjemmeboende børn. I dette tilfælde er der tale om først en stratificeret klyngeudvælgelse af kommuner i USA, hvor 103 blev udvalgt. Dernæst en udvælgelse af 300 lokaliteter, som hver blev opdelt i boligområder af 10-25 husstande, hvor inklusionskriteriet var et samlevende par af en mand og en kvinde. Når en sådan husstand var udvalgt blev der trukket lod mellem manden og konen.

Klyngeudvælgelse kan både anvendes i tilfælde, hvor det er sagligt påkrævet som her ved familieundersøgelser, og i tilfælde hvor det gør udvælgelsen af individer lettere at foretage.

Udformning af spørgsmål og svarkategorier

En survey er en meget begrænset og formaliseret form for kommunikation, hvor den direkte og personlige kontakt er erstattet af en indirekte og upersonlig. Det indebærer store risici for, at kommunikationen går galt, således at resultatet af undersøgelsen bliver misvisende, biased. I værste fald sker der det, at respondenterne ikke forstår undersøgerens intention med spørgsmålet, og at undersøgeren til gengæld ikke forstår, hvad respondenterne mener med sit svar. En måde at forebygge dette på, er at gennemtænke og afprøve udformningen af spørgeskemaets spørgsmål og svarkategorier.

At stille spørgsmål

Et spørgsmål skal være så klart formuleret, at respondenterne er i stand til at besvare det på en måde, som giver et adækvat udtryk for hans eller hendes opfattelse, holdning eller adfærd. Denne formulering er jo ikke særlig veloperationaliseret, så lad os i stedet give en negativ anvisning ved at pege på de fælder, der erfaringsvis belaster eller ødelægger et spørgeskemas pålidelighed og gyldighed.

Tabel 3.5.

	Oprindelige spørgsmål	Reformulerede spørgsmål
A. Dårligt formulerede		
1. Vagthed	Har du ofte hovedpine?	Hvor mange dage havde du hovedpine sidste måned?
2. Indviklethed	I diskussionen om forholdet mellem EU's og nationalstatens suverænitet er der nogle, der ønsker at tillægge Europaparlamentet større kompetence, mens andre vil give det mindre kompetence. Går du ind for det første eller det sidste?	Synes du, at Europarlamentet skal have: a. Større kompetence? b. Uændret kompetence c. Mindre kompetence?
3. Dobbeltløbende spørgsmål	Synes du, at indvandrere har for stor indflydelse på dansk politik, og at de modsætter sig integration?	To spørgsmål: a. Har indvandrere for stor indflydelse på dansk politik? b. Modsætter indvandrere sig integration?
4. Dobbelt-negation	Er du uenig med de mennesker, der bekæmper genmodificeret landbrug?	Er du tilhænger eller modstander af genmodificeret landbrug?
B. Ledende		
1. Normative	Går du ind for militærtjeneste som en patriotisk pligt?	Skal militærtjeneste være obligatorisk eller frivillig?
2. Propaganderende	Bør der bruges flere midler på de mange sindslidende, som i dag forsømmes i dette samfund?	Er det beløb, der afsættes til sindslidende: a. for lille? b. passende? c. for stort?
3. Indbygget antagelse	Hvornår holdt du op med at afstraffe dine (nu voksne) børn?	
C. Umulige at besvare		
1. Fjern fortid	Hvor tit pjækkede du i skolen?	Har nogen sinde pjækket i skolen?
2. Fjern fremtid	Hvilken pensionsordning påtænker du (til 20-årige)?	Har du gjort dig overvejelser om din fremtidige pension?

- A. Spørgsmål, der er dårligt formulerede
- B. Spørgsmål, der er ledende
- C. Spørgsmål, der er umulige at besvare for respondenterne.

I kategori A og C vil reliabiliteten være lav, idet respondenterne vil svare mere eller mindre hen i vejret. I kategori B vil validiteten være lav, fordi svaret bliver fremtvunget af formuleringen.

I tabel 3.5 er der både anført et dårligt spørgsmål og et reformuleret, bedre spørgsmål.

Eksemplerne på ledende spørgsmål er jo påfaldende, men der er også mere subtile måder at inducere en bestemt holdning til et problemområde på. Schuman og Presser (1981) gennemgår nogle empiriske undersøgelser af svareffekter, nemlig enighedseffekt, tilladt/forbudt-effekt og konteksteffekt.

Enighedseffekt

Ved *enighedseffekt* forstås, at det ved et polært spørgsmål, altså hvor to forskellige holdninger modstilles, er markeret, hvilken holdning der er den socialt acceptable (tabel 3.6., efter Schuman & Presser, 1981).

Tabel 3.6.			
A		B	
Enkeltindivider er i højere grad end sociale forhold årsag til kriminalitet og lovløsheder her i landet		Sociale forhold er i højere grad end enkeltindivider årsag til kriminalitet og lovløshed her i landet	
Enig	59,6% (Enkeltindivider har skylden)	Enig	56,8% (Sociale forhold har skylden)
Uenig	40,4% (Sociale forhold har skylden)	Uenig	40,4% (Enkeltindivider har skylden)
(Undersøgelsen gennemført i USA i 1976. N = 409/432; „Ved ikke“ udeladt.)			

I begge formuleringer er det altså det alternativ, der er fremhævet som det socialt acceptable, der tiltrækker majoriteten af respondenter.

Tilladt/forbudt-effekt

Den næste type effekt ligner meget den forrige, men her består det ledende i, at de fleste mennesker umiddelbart har en mere positiv holdning til at

tillade noget end til at forbyde noget (tabel 3.7., efter Schuman & Presser, 1981).

Tabel 3.7.			
A		B	
Mener du, at USA bør forbyde offentlige taler til fordel for kommunismen?		Menner du, at USA bør tillade offentlige taler til fordel for kommunismen?	
Ja	39,3%	Nej	56,3%
Nej	60,1% (ikke forbyde)	Ja	43,8% (tillade)

Vi ser her, at majoriteten, som var imod at forbyde kommunistiske taler, svinger, når der spørges om at tillade noget sådant. En ting er at skride til forbud, noget andet er noget så overdrevent liberalt som at tillade samme ting. Hvad disse respondenter dybest set mener, er et spørgsmål, der blæser i vinden.

Et tilsvarende eksempel ses i en dansk undersøgelse (tabel 3.8., efter Olsen, 1998).

Tabel 3.8.			
A		B	
Mener De, at foreninger og partier, der udbreder racediskriminerende (racistiske) synspunkter bør være forbudt i Danmark?		Mener De, at foreninger og partier, der udbreder racediskriminerende (racistiske) synspunkter bør være tilladt i Danmark?	
Ja	56%	Nej	66%
Nej	31% (ikke forbudt)	Ja	20% (tilladt)

(Olsen, 1998. Undersøgelsen gennemført i Danmark i 1996. N = 968/920; „Ved ikke“ = 13/14 %).

Når man ændrer spørgsmål fra at omhandle forbud til at dreje sig om tilladelse, falder tilslutningen altså til en femtedel.

Konteksteffekt

Den allermest underfundige form for skjult holdningspåvirkning er nok konteksteffekten, hvor indflydelsen ikke er knyttet til selve udformningen

af det enkelte spørgsmål, men derimod til den kontekst, det indgår i (tabel 3.9., efter Schuman & Presser, 1981).

Tabel 3.9.

Kontekst 1. Spørgsmål A stilles alene

A. Mener du, at USA bør lade journalister fra kommunistiske aviser i andre lande komme her og sende nyheder, sådan som de opfatter dem, tilbage til deres aviser?

Ja 35,5%

Kontekst 2. Spørgsmål A stilles efter spørgsmål B

B. Mener du, at et kommunistisk land som Sovjetunionen bør lade amerikanske journalister komme ind og sende nyheder, sådan som de opfatter dem, tilbage til USA?

A. Mener du, at USA bør lade journalister fra kommunistiske aviser i andre lande komme her og sende nyheder, sådan som de opfatter dem, tilbage til deres aviser?

Ja 73,1%

Efter at være blevet gjort opmærksom på modstanderens kritisable holdning til pressefrihed bliver tolerancen over for kommunistiske journalister altså fordoblet.

At udforme svarmuligheder

Svarmuligheder som et kategorisystem

Da respondentens eneste måde at give sin opfattelse til kende på begrænser sig til nogle svarmuligheder, der på forhånd er opstillet af undersøgeren, er det vigtigt, at disse muligheder er klare og tilsammen udfylder det svarunivers, som svarer til forhåndsopfattelserne i den undersøgte stikprøve.

Der skal i første række foreligge et kategorisystem, dvs. at svarmulighederne må opfylde følgende formelle krav:

1. Gensidig udelukkelse: det må ikke være muligt at vælge flere svarmuligheder.

2. Fuldstændighed: det skal være muligt at svare dækkende ved at vælge en enkelt svarmulighed.

Derudover gælder de samme formuleringskrav som for spørgsmålene.

Ad 1. Hvis det er relevant at lade respondenter vælge flere muligheder, er der ikke tale om et svar på et enkeltspørgsmål, men om en såkaldt checkliste, hvor der kan sættes flere krydser (se tabel 3.10■):

Ad 2. Om svarmulighederne er udtømmende vil ofte være et empirisk snarere end et logisk spørgsmål, da undersøgerens viden om svaruniverset kan vise sig at være for begrænset. Det kræver derfor en pilotundersøgelse at sikre sig fuldstændighed.

Tabel 3.10.

Forkert:

Hvad mener du om EU?

Rigtigt:

Hvilke af følgende udsagn om EU er enig eller uenig i?

1. Det er for bureaukratisk	a. Enig	b. Uenig
2. Det koster for mange penge	a. Enig	b. Uenig
3. Det tiltager sig magt på Danmarks bekostning	a. Enig	b. Uenig
4. Det sikrer freden i Europa	a. Enig	b. Uenig
5. Det skaber mere velstand	a. Enig	b. Uenig
6. Det er en modvægt mod supermagten USA	a. Enig	b. Uenig

Svarmuligheder udformet som en ordinalskala

En meget udbredt måde at udforme svarmulighederne på er ordinalskalaen, hvor kategorisystemet er opbygget som et ordningssystem. Man følger her logikken i de såkaldt relative adjektiver, som muliggør en sammenligning mellem forskellige bedømmelsesobjekter på den måde, det kommer til udtryk i den komparative form (se tabel 3.11■).

Tabel 3.11.

Hvor enig er du i en politik, der føres af statsministeren?

Sæt kryds ved den svarmulighed, du tilslutter dig.

☐	☐	☐	☐	☐
fuldstændig enig	delvis enig	hverken enig eller uenig	delvis uenig	fuldstændig uenig

For at bortvejre en tvivl om ordningsrelationen kan man nummerere
mulighederne:

☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
fuldstændig enig	delvis enig	hverken enig eller uenig	delvis uenig	fuldstændig uenig

Peter er mere enig med statsministeren end Hans er

Det er denne ordningslogik, der er indbygget i de fleste ordinalskalaer, fx i den meget populære udformning, der efter ophavsmanden kaldes Likert-skalaen.

Registrering og indkodning af oplysninger og opstilling af database

Ved surveymetoden er der oftest tale om papir og blyant-skemaer, som udfyldes af informanten. Det er dog i stigende grad blevet almindeligt, at lade informanterne logge ind på selve databasen via internet eller pc, således at datagenerering bliver identisk med indkodning.

Uanset hvilken form for inddatering der anvendes, er det afgørende at forsyne hvert skema med et identifikationsnummer for hver person for at sikre en entydig identifikation. Dette gælder i særlig grad, hvis der finder en fletning af flere datakilder sted. Det vil sige, at der sker en sammensætning af flere skemaer, idet der for hver information foreligger forskellige spørgeskemaer. Således sker der fletning i følgende tilfælde:

- Ved længdesnitsdesign, fx effektundersøgelser.
- Ved særskilte oplysninger fra anden kilde, fx baggrundsoplysninger, journaldata.
- Ved oplysninger ud fra åbne spørgsmål kategoriseret til variable.

Et andet tilfælde, hvor identifikationsnummer er afgørende, er anonyme databaser, hvor dataloven kræver, at personidentifikation kun er mulig via en liste, som forskningslederen alene har adgang til.

Opstillingen af database sker nu i langt de fleste tilfælde i et statistikprogram, fx SPSS, hvor der opbygges en særlig datanøgle eller kodebog, som definerer formatet og betegnelserne for de variable og værdier, der skal indkodes efter (se fx Nielsen & Kreiner, 2003).

Statistiske sammenhænge

Når surveyundersøgelsen er gennemført og databasen etableret, kommer den afgørende fase, hvor datamaterialet skal analyseres. Denne bog er ikke en lærebog i statistik, idet der her skal henvises til (Kreiner, 2001 ??). Derimod skal vi gennemgå de særlige fælder, der er knyttet til analyse af multivariate sammenhænge.

Som nævnt i kapitel 1 indebærer den kvantitative metoder en vidtgående dekontekstualisering og atomisering, hvor helheder bliver isoleret og opdelt i enkelttegenskaber, der operationaliseres som variable. Denne atomisering gælder ikke kun individet på det specificerende plan, men også på det almene plan, hvor teoretiske forklaringer atomiseres og operationaliseres i form af relationer mellem variable. Når statistikken opstiller og afprøver hypoteser, der angiver sådanne relationer mellem variable, er problemet, at man altid dekontekstualiserer en sådan hypotese fra den større sammenhæng, den er placeret i. Specielt vil hypotesen kun omfatte et begrænset antal variable, således at man implicit eller eksplicit antager, at øvrige variable ikke influerer på det de variable, man analyserer. En afgørende side af hypotesen er således dens uafhængighed af andre variable, der er med i datamaterialet, men som ikke indgår i analysen, eller som slet ikke indgår i undersøgelsen.

Derfor er det i forbindelse med statistisk sammenhængsanalyse afgørende at forholde sig til denne dekontekstualisering. Hvis vi siger, at der i en sammenhængsanalyse altid på den ene side er nogle udvalgte variable og

relationer mellem disse og på den anden side de ikke inddragede variable og relationer, kan vi kalde disse to felter af variable og relationer for henholdsvis den *analytiske forgrund* og den *analytiske baggrund*.

Selve begreberne forgrund og baggrund henviser for så vidt et almen erkendelsesteoretisk eller metodisk forhold. Ved enhver videnskabelig undersøgelse vil der være tale om en udvælgelse af bestemte undersøgelsesaspekter, idet man udelukker en række andre aspekter, som man enten finder irrelevante for den aktuelle problemstilling, eller som det er praktisk umuligt eller uoverkommeligt at inddrage.

Dette gælder i første omgang selve den empiriske design- og indsamlingsproces, hvor problemstillingen opstilles og datamaterialet genereres. Det gælder i anden omgang desuden meget ofte den efterfølgende *analyse* af materialet. I hvert fald ved surveyundersøgelser vil der generelt indgå flere variable, end det er statistisk muligt at analysere på en gang. Dels bliver selve den beregningsmæssige side af analyserne hurtigt uoverkommelig, når der inddrages flere og flere variable, dels bliver det undersøgte antal informanter spredt stadig tyndere ud over mere og mere omfattende tabeller.

Der er nu to forskellige måder, man kan forholde sig på i analysen: en naiv og en reflekteret.

- *Den reflekterede sammenhængsanalyse*

Ved den reflekterede sammenhængsanalyse ekspliciterer man før den aktuelle analyse sin *hypoteses* og efter analysen sine *funds* foreløbige karakter. De har kun den angivne mening under den forudsætning, at den analytiske baggrund viser sig ikke at have nogen betydning. Derefter forsøger man, hvis det er muligt, at inddrage den analytiske baggrund, således at den nu bliver en del af den udvidede forgrund.

- *Den naive sammenhængsanalyse*

Ved den naive sammenhængsanalyse ignorerer man den analytiske baggrund og postulerer, at de fund, man har gjort med hensyn til forgrunden, har en almen og ubetinget gyldighed.

Selve opdelingen i forgrund og baggrund har her to betydninger:

- *Den analysemæssige betydning af opdelingen i forgrund og baggrund*

Her menes der med forgrund simpelthen det, der udvælges i en

foreliggende analyse, mens baggrund er alt det, der ikke udvælges på dette tidspunkt, men som måske inddrages senere.

- *Den teorimæssige betydning af opdelingen i forgrund og baggrund*

I denne betydning har man en teoretisk viden eller formodning om, at der er bestemte relationer mellem variablene, således at nogle er kausale i forhold til andre. De kausale kaldes baggrundsvariable i forhold til dem, som de indvirker på.

Denne kausale orden kan bero på teoretisk viden og årsagssammenhænge. I tilknytningsteorien er barnets tilknytning til omsorgspersonen en kausal variabel i forhold til det lille barns emotionelle og sociale udvikling.

Den kausale orden kan også være rent logisk baseret. Her er tidsrækkefølge en prækausal relation, idet en variabel, der går tidsmæssig forud for en anden, kan være en *mulig*, men ikke *nødvendig* årsag til efterfølgende, men ikke omvendt. Man kan således undertiden have nytte af at gøre tidsmæssigt senere foreliggende variable til analytisk baggrund for forudgående variable; denne opdeling er da kun relativ, men må inddrages i en bredere, absolut, sammenhæng for at blive meningsfuld.

I denne indledning til afsnittet om sammenhængsanalyse vil vi først i et afsnit om *forskellige betingningsformer* gennemgå de relationsformer, der kan være mellem to variable A og B, som udgør forgrunden, mens variablen C udgør baggrunden. Derefter vil et afsnit om *binære fejlslutninger* indeholde en gennemgang af de måder, hvorpå en umiddelbar og naiv analyse af en forgrund bestående af de to variable A og B, kan føre til misvisende konklusioner.

Når der her er uoverensstemmelse mellem den umiddelbart fremtrædende forgrund og det, der faktisk viser sig at gælde, når man inddrager baggrunden, er der altså tale om et erkendelsesproblem. Dette løses gennem en analyseproces, hvor man supplerer den foreløbige undersøgelse af den umiddelbart fremtrædende forgrund med en analyse af såvel forgrund som baggrund. Herved opnår man et middelbart analyseresultat, som dels forklarer den første umiddelbare og misvisende fremtræden, dels må antages at være om ikke fuldt dækkende for det faktisk foreliggende, så i hvert fald mere dækkende end den oprindelige, umiddelbare fremtræden.

Forskellige betingningsformer

Vi vil her skelne mellem forskellige måder, hvorpå den analytiske baggrund kan influere på relationen mellem to forgrundsvariable A og B.

1. For det første kan der være tale om betingning med en baggrundsvariable. Her vil der generelt være forskel på den *ubetingede* og den *betingede* relation mellem de to forgrundsvariable.
2. For det andet kan en betinget relation være af forskellig karakter, nemlig enten *lokal* eller *global*, og i sidste tilfælde enten *homogen* eller *heterogen*.

Forskel mellem ubetingede og betingede relationer

Den mest grundlæggende forskel mellem en ubetinget og en betinget relation består i, at den ene viser afhængighed og den anden uafhængighed. Vi vil her anvende følgende notation:

A-B betyder, at der er en ubetinget afhængighed mellem A og B

A|B betyder, at der er en ubetinget uafhængighed mellem A og B

A-B|C betyder, at der er en betinget afhængighed mellem A og B givet C

A|B|C betyder, at der er en ubetinget uafhængighed mellem A og B givet C

Et tilfælde af denne art består i, at de to forgrundsvariable er ubetinget afhængige, men betinget afhængige, givet baggrundsvariablen C:

Vi har med andre ord to tilfælde heraf

Betinget uafhængighed – mediering

A-B betyder, at der er en ubetinget afhængighed mellem A og B

A|B|C betyder, at der er en ubetinget uafhængighed mellem A og B givet C

Den berømte danske statistiker og psykometriker Georg Rasch opdagede et eksempel på dette, da han i fyrerne udviklede BPP, den IK-test, der stadig anvendes af forsvaret. Her var man indstillet på ved udvælgelsen blandt rekrutter at lade inddelingen i menige, sergenter og officerer være bestemt af både *intelligens*, *uddannelse* og *urbanisering*, dvs. hvor henholdsvis landligt eller bymæssigt opvækstniveauet havde været.

Den kausale antagelse bag den analyse, vi her skal se, var, at intelligens var en funktion af bl.a. uddannelse og urbaniseringsgrad, idet disse blev

Tabel 3.12. En rekonstruktion af forholdet mellem urbaniseringsgrad og IK.

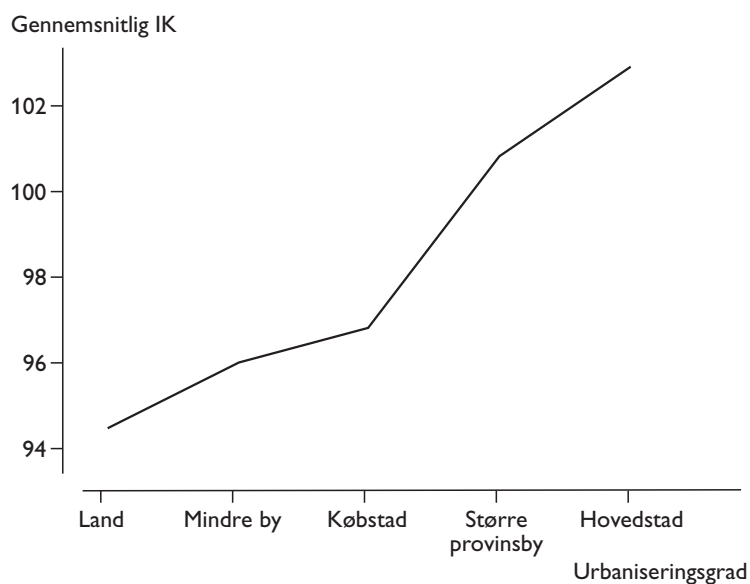
Urbanisering	Gennemsnitlig IK	N	Std. deviation
Land	94,3469	1031	15,97457
Mindre by	95,8856	991	16,99581
Købstad	96,7061	1038	17,61312
Større provinsby	100,6733	3051	19,35875
Hovedstad	102,7157	3889	19,33298
Total	99,9291	10000	18,86289

opfattet som prækausale i forhold til intelligens. Jeg har simuleret et materiale med 10.000 unge mænd undersøgt ved sessionen. Fordelingen efter urbanisering fremgår af tabel 3.12.

Figur 3.12 viser det samme i grafisk form.

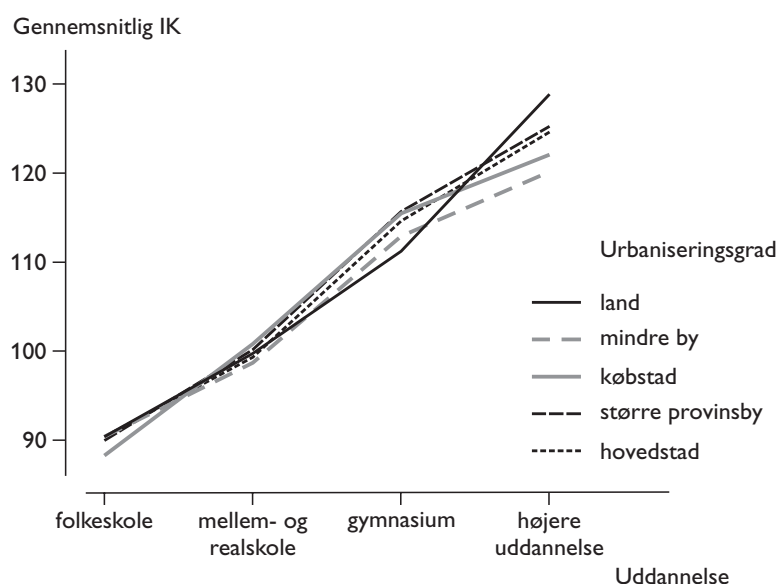
Korrelationskoefficienten er lig med 0,16, hvad der i betragtning af det store antal er meget stærkt signifikant.

Men dette var altså den umiddelbare, ubetingede sammenhæng. Hvis



Figur 3.12. Den tilsyneladende sammenhæng mellem urbaniseringsgrad og IK.

vi inddrager baggrundsvariablen uddannelse, får vi et helt andet billede, som det fremgår af figur 3.13.



Figur 3.13. Intelligens som funktion af uddannelse og urbanisering.

Vi ser her, at forskellene mellem urbaniseringsniveauerne er væk i de betingede fordelinger, således at IK-niveauet kun hænger sammen med uddannelse. Man siger, at den analytiske baggrundsvariabel, uddannelse, optræder som mediator mellem forgrundsvariablene.

En enkel statistisk teknik til at afprøve, om urbanisering således kan betinges væk, er udregning af den såkaldte partielle korrelationskoefficient, hvorved den del af korrelationen, som skyldes en baggrundsvariabel, elimineres.

Vi udregner først den ubetingede korrelation mellem urbanisering og IK, og her finder vi ligesom i figur 3.12 en klar sammenhæng (se tabel 3.13).

Tabel 3.13. Sammenhængen mellem urbanisering og IK.

Korrelation mellem urbanisering og IK	Korrelation	Signifikans (ensidig)	N
	0,157	0,000	10000

Tabel 3.14. Sammenhængen mellem urbanisering og IK, når der kontrolleres for uddannelsesniveau.

Partiel korrelationskoefficient for urbanisering og IK kontrolleret for uddannelsesniveau	Korrelation	Signifikans (ensidig)	f (frihedsgrader)
	0,302	0,005	9997

Når vi derimod går over til den betingede sammenhæng, får vi et helt andet billede (tabel 3.14).

Vi skal senere under afsnittet om statistiske fejlslutninger forklare, hvordan det kan gå til, at en tilsyneladende, men illusorisk sammenhæng som her mellem urbaniseng og IK, kan komme til at fremtræde som om, den er en realitet.

Betinget afhængighed – undertrykkelse (suppression)

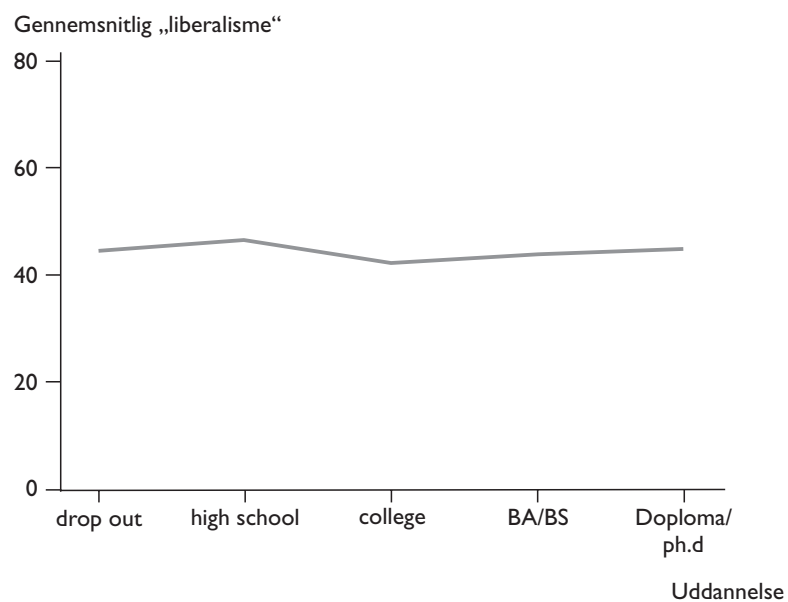
A|B betyder, at der er en ubetinget uafhængighed mellem A og B

A-B|C betyder, at der er en betinget afhængighed mellem A og B givet C

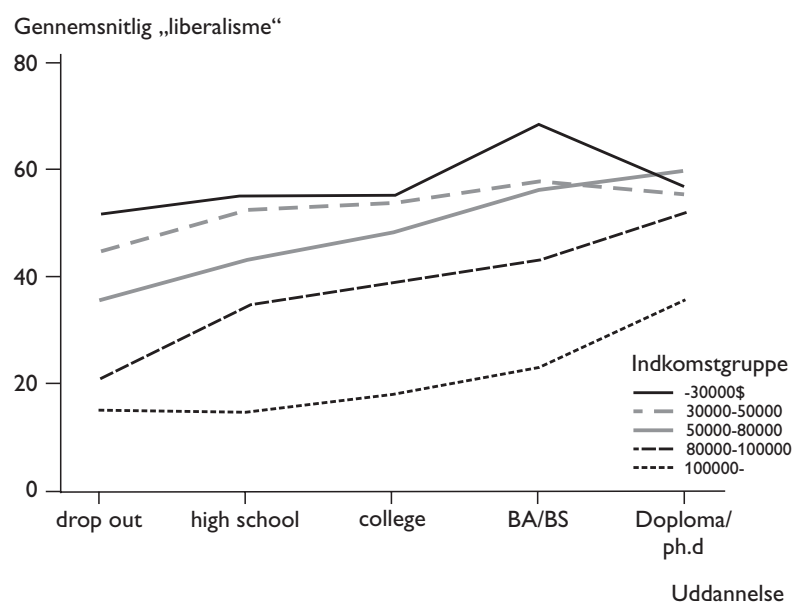
Det her anvendte eksempel er baseret på den officielle amerikanske indkomst- og uddannelsesstatistik,³ og på det i amerikansk samfundsvidenskab ofte iagttagne fænomen, at den politiske holdning, som kaldes *liberalism*, og som nærmest svarer til det, vi på dansk kalder venstreorientering eller måske snarere kulturradikalisme, er korreleret positivt med uddannelsesniveau, men negativt med indkomst. Sammenhængen mellem "liberalisme" i denne betydning af ordet og udannelse fremgår af figur 3.14.

Som det fremgår af denne figur, er der ingen sammenhæng at spore, men hvis vi inddrager baggrundsvariablen uddannelse, får vi et andet resultat (se figur 3.15).

Vi ser her, at de betingede fordelinger, i modsætning til den ubetingede, alle viser voksende sammenhæng mellem uddannelse og liberalisme. Samtidig ser vi et andet forhold, nemlig en direkte indvirkning af indkomstniveauet på graden af liberalisme, idet voksende indkomst hænger sammen med faldende liberalisme. Vi skal senere i afsnittet om statistiske



Figur 3.14. Sammenhængen mellem uddannelse og liberalisme.



Figur 3.15. Sammenhængen mellem uddannelse og „liberalisme“ kontrolleret for indkomst.

fejlslutninger vise, hvordan de komplicerede relationer mellem de tre variable bevirker, at sammenhængen mellem uddannelse og liberalisme bliver skjult eller undertrykt af den tredje variabel indkomst, der altså optræder som ”undertrykker” (eng. suppressor).

Lige som i det forrige afsnit kan vi nu bruge den partielle korrelationskoefficient som beregningsmæssig teknik til at vise forskellen mellem ubetinget og betinget sammenhæng. Når vi udregner den ubetingede korrelationskoefficient mellem uddannelse og liberalisme, finder vi – ligesom ved den grafiske analyse – ingen sammenhæng (se tabel 3.15).

Når vi imidlertid går over til at kontrollere for indkomst, giver den nu partielle korrelationskoefficient et ganske andet resultat (tabel 3.16).

Tabel 3.15. Sammenhængen mellem uddannelse og „liberalisme“.

Korrelation mellem uddannelse og „liberalisme“	Korrelation	Signifikans (p-værdi)	N
	0,012	0,652	1500

Tabel 3.16. Sammenhængen mellem uddannelse og „liberalisme“ – kontrolleret for indkomst.

Partiel korrelationskoefficient for uddannelse og „liberalisme“ kontrolleret for indkomst	Korrelation	Signifikans (p-værdi)	f (frihedsgrader)
	0,210	0,000	1497

Den indholdsmæssige forklaring på denne undertrykkelse (suppression) vil blive givet i næste hovedafsnit om statistiske fejlslutninger.

Forskellige typer af betingede sammenhænge

Global sammenhæng

Når sammenhængen er global, gælder den for alle betingelser af baggrundsvariablen:

$$A-B|C = [A-B|C = c_1, A-B|C = c_2, \text{ osv.}$$

Det eksempel (tabel 3.17), vi her skal vise, er hentet fra Svend Kreiners (2001, s. 360ff) lærebog i statistik.

Tabel 3.17. Udvikling i uddannelsesniveau for de to køn fra 1967 til 1984 (procent) (Kreiner, 2001).

År	Køn	< 9 år (8 år)	10-12 år (11 år)	13+ år (14 år)	I alt
1967	mand	35,5	55,1	9,4	100
	kvinde	70,6	26,1	3,3	100
	Total	53,4	40,3	6,3	100
1984	mand	33,8	52,2	14	100
	kvinde	53,3	35,2	11,5	100
	Total	43,6	43,7	12,7	100

Hvis vi her er interesseret i udviklingen med hensyn til uddannelsesniveau, vil det være naturligt at opfatte årstallet som en uafhængig variabel A, der påvirker fordelingen i uddannelsesniveau B. A og B er altså forgrundsvariable, mens C, køn, er en baggrundsvariabel. Derfor er $c_1 = \text{mand}$ og $c_2 = \text{kvinde}$.

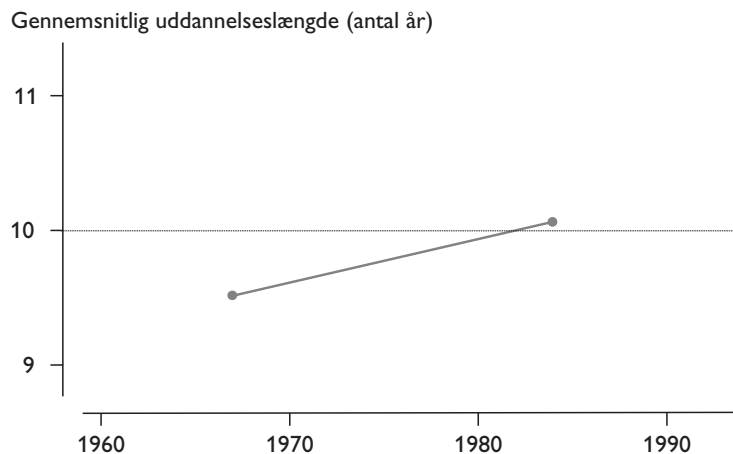
Vi kan nu begynde med at se på den ubetingede sammenhæng mellem de to forgrundsvariable (se tabel 3.18).

Tabel 3.18. Ændringen i uddannelsesniveau fra 1967 til 1984 (procent) (Kreiner, 2001).

År	< 9 år (8 år)	10-12 år (11 år)	13+ år (14 år)	I alt	Gennemsnit (antal år)
1967	53,4	40,3	6,3	100	9,5
1984	43,6	43,7	12,7	100	10,1

Hvis undersøger, om tiden påvirker uddannelsesniveauet, finder vi med korrelationsmålet (Kendalls) gammaværdien 0,21, som er meget stærk signifikant. Et simpelt, og stærkt forenklet, udtryk for denne ændring får

vi ved at se på et tentativt gennemsnitsmål for uddannelsen på de to tidspunkter (se figur 3.16).



Figur 3.16. Ændringen i uddannelsesniveau fra 1967 til 1984 (efter Kreiner, 2001).

Der sker en tilvækst med ca. et halvt uddannelsesår i perioden. Vi har altså en ubetinget sammenhæng mellem årstal og uddannelse, og vi kan nu spørge om denne sammenhæng også gælder betinget af baggrundsvariablen køn?

Lokal sammenhæng

Sammenhængen mellem de to variable A og B er betinget, dvs. kun gældende for visse betingelser:

$$A-B|C = c_1, A|B|C = c_2, \text{ osv.}$$

Sammenhængen mellem årstal og uddannelse er blot lokal, hvis den ikke gælder for begge køn, men kun for et af dem (se tabel 3.19).

Mens vi fandt en ubetinget sammenhæng mellem årstal og uddannelse, dvs. en generel stigning i uddannelsesniveaut, viser det sig, når vi nu betinger, at denne ubetingede sammenhæng dækker over en mere kompliceret betinget relation.

Lad os først se på mændene. Her er der ingen signifikant sammenhæng. Korrelationsmålet gamma er på kun 0,08 og slet ikke signifikant. For

Tabel 3.19. Kønsforskel i ændringer fra 1967 til 1984 (Efter Kreiner, 2001).

Køn	Korrelation (gammaværdi)	N	Signifikans
Mand	0,08	651	0,262
Kvinde	0,36	665	0,000

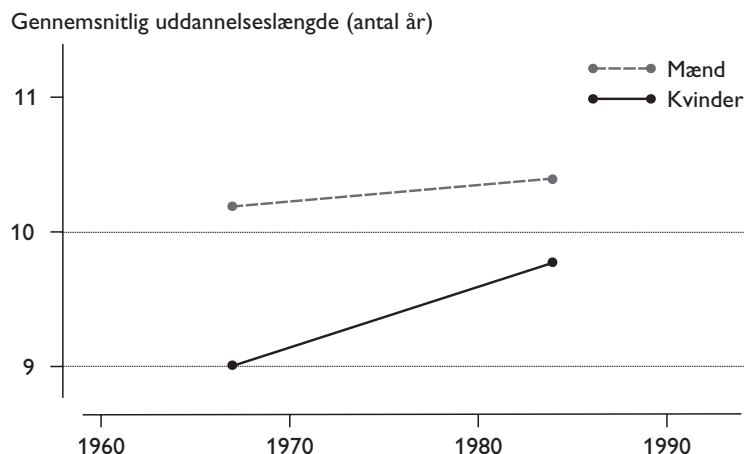
kvindernes vedkommende finder vi imidlertid en gammaværdi på 0,36, hvilket er overvældende signifikant.

Udregning af gennemsnitsværdier giver et indtryk af denne forskel, idet mændenes værdi kun stiger umærkeligt fra 10,2 til 10,4 år, mens der for kvinderne er tale om en stigning fra 9,0 til 9,8 år (tabel 3.20).

Tabel 3.20. Fordeling af uddannelsesniveau efter årstal og køn (procent) (Kreiner, 2001).

Køn	År	< 9 år (8 år)	10-12 år (11 år)	13+ år (14 år)	I alt	Gennemsnit (antal år)
Mand	1967	35,5	55,1	9,4	100	10,2
	1984	33,8	52,2	14	100	10,4
Kvinde	1967	70,6	26,1	3,3	100	9,0
	1984	53,3	35,2	11,5	100	9,8

Det samme udtrykkes grafisk i figur 3.17.



Figur 3.17. Udviklingen i uddannelsesniveau for mænd og kvinder fra 1967 til 1984 (efter Kreiner, 2001).

Homogen sammenhæng

Homogen sammenhæng foreligger, når der ikke er nogen styrkeforskelle i betinget sammenhæng mellem A og B givet C. Hvis $r(A,B)$ er udtryk for styrken i sammenhængen (korrelationen), har vi:

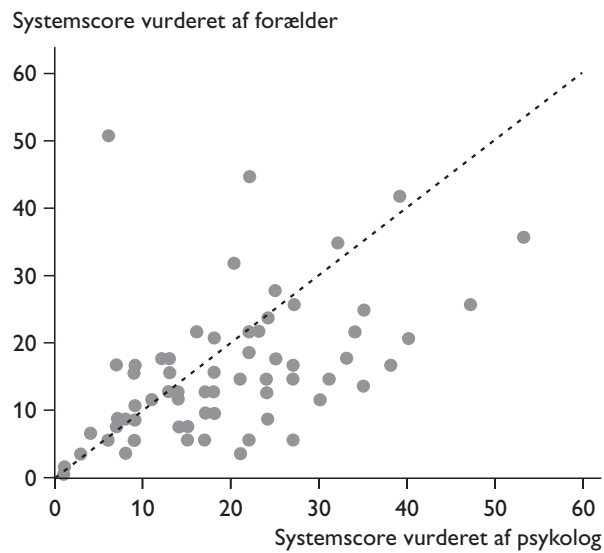
$$r(A,B|C = c_1) = r(A,B|C = c_2) = \text{osv.}$$

Som et eksempel på homogenitet vil jeg vise nogle resultater fra en evaluering, jeg har foretaget (Karpatschof & Helweg, 2003) af en psykologisk rådgivning for børn, der har mistet eller er ved at miste et nærtstående familiemedlem (dvs. en forælder eller en søskende).

I figur 3.18 ser vi samvariationen mellem den behandlende psykolog og den overlevende forælders vurdering af det sørgende barnets symptomer og problemer. Der er i alt vurderet 21 symptomer og problemer på en ordinalskala fra 0 til 5, således at scoren ligger mellem 0 og 65.

Vi ser en ganske nøje sammenhæng mellem de to slags vurderinger, og dette bekræftes af en korrelationsberegning (se tabel 3.21).

Man kan imidlertid forestille sig, at overensstemmelsen mellem psykolog og forælder afhænger af barnets alder. Vi vil derfor undersøge, om sammenhængen er homogen. Resultatet fremgår af tabel 3.22.



Figur 3.18. Sammenhængen mellem symptomscoren baseret på henholdsvis psykolog- og forældrevurdering.

Tabel 3.21. Sammenhæng mellem symptomscore baseret på psykolog- og forældrevurdering.

Korrelationskoefficienten for sammenhængen mellem symptomscoren baseret på psykolog- og forældrevurderinger	Korrelationskoefficient	Signifikans (p-værdi)	N
	0,482	0,000	64

Tabel 3.22. Sammenhængen mellem psykolog- og forældrevurdering opdelt efter aldersgruppe.

Aldersgruppe	Korrelationskoefficient	Signifikans (p-værdi)	N
-7	0,86	0,015	6
8-10	0,70	0,000	19
11-13	0,65	0,001	20
14-15	0,59	0,022	12
16-	-0,51	0,121	7

Denne tabel viser, at der er en solid og gennemgående korrelation uanset aldersgruppe, dog med undtagelse af gruppen af de ældste børn, hvor der slet ikke er nogen korrelation. Nu er det jo en velkendt sag, at unge mennesker og forældre har svært ved at forstå hinanden, og det kan jo være dette forhold, der viser sig her.

Tilbage er nu, at der faktisk også er en gennemgående tendens til, at korrelationen falder. Selv om antallene ikke gør det muligt at vurdere dette forhold statistisk, kan vi fortolke tabellen i retning af en systematisk inhomogenitet, således at der er meget stor enighed mellem de to bedømmere, hvad angår de små børn, mens overensstemmelsen falder for de lidt større børns vedkommende og helt er ophørt, når det drejer sig om de unge mennesker.

Heterogen sammenhæng

Heterogen sammenhæng foreligger, når der er forskelle i den betingede sammenhæng mellem A og B givet C. Hvis $r(A,B)$ er et udtryk for sammenhæng (korrelation), har vi for nogle værdier af C:

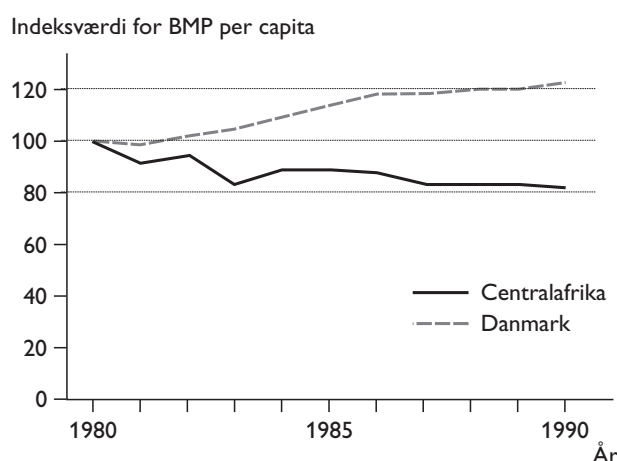
$$r(A,B|C = c_1) = r(A,B|C = c_2)$$

Vi siger ved heterogen sammenhæng, at C modificerer sammenhængen mellem A og B, eller at C er en modifikator. Der kan være tale om to typer:

Stærk heterogen sammenhæng eller absolut modifikation, hvor der er fortegnsskille i den betingede sammenhæng mellem A og B givet C:

$$r(A,B|C = c_1) > 0, \quad r(A,B|C = c_2) < 0, \quad \text{osv.}$$

Et sådant tilfælde foreligger ved økonomisk eller politisk polarisering, fx. når udviklingen i nogle befolkningsgrupper er positiv, mens den i andre er negativ. Således har gennemsnitsindkomsten været stigende i de industrialiserede lande, mens den i mange udviklingslande har været faldende, som det fremgår af figur 3.19.



Figur 3.19. Økonomisk udvikling i Centralafrika og Danmark.

Svag heterogen sammenhæng eller partiel modifikation. Selv om de betingede sammenhænge har samme fortegn, er der styrkeforskelle i den betinget sammenhæng mellem A og B givet C. Hvis $r(A,B)$ er udtryk for sammenhæng (korrelation), har vi for nogle værdier af C:

$$r(A,B|C = c_1) < r(A,B|C = c_2), \quad \text{osv.}$$

Lad os vende tilbage til opgørelsen af uddannelse for de to køn i 1967 og 1984 (jf. tabel 3.17). Da vi før så på årstal og uddannelsesniveau som forgrundsvariable med køn som baggrundsvariabel, fandt vi en betinget

sammenhæng for kvinder, men ikke for mænd. Sammenhængen mellem forgrundsvariablene var altså kun lokal.

Vi vil nu gøre køn til forgrundsvariabel, mens årstallet bliver baggrundsvariabel. Læg mærke til, at begge dele er indholdsmæssigt meningsfuldt, da både køn og årstal kan betragtes som determinerende. Det ville derimod være meningsløst, at gøre uddannelsesniveau til baggrundsvariabel, da uddannelsen hverken bestemmer observationstidspunktet eller kønnet. Tabel 3.23 viser den nye rollefordeling for vores variable.

Tabel 3.23. Uddannelsesfordeling ud fra årstal og køn.

År	Køn	Uddannelse						Total	
		< 9 år		10-12		13 +			
		N	%	N	%	N	%	N	%
1967	Mand	102	35,5	158	55,1	27	9,4	287	100
	Kvinde	211	70,6	78	26,1	10	3,3	299	100
	Total	313	54,4	236	40,3	37	6,3	586	100
1984	Mand	123	33,8	190	52,2	51	14,0	364	100
	Kvinde	195	53,3	129	35,2	42	11,5	366	100
	Total	318	43,6	319	43,7	94	12,7	730	100

Det er tydeligt, at mens forskellen mellem kønnene i 1967 var stor, således at der var en stærk sammenhæng mellem de to forgrundsvariable, er forskellen i 1984 mindsket, således at vi nu kun har en meget mindre grad af sammenhæng. Dette fremgår af tabel 3.24 over korrelationskoefficienterne angivet som gammaværdier.

Tabel 3.24. Forskellen i uddannelse mellem kønnene i 1967 og 1984 (angivet som gammaværdi).

År	Korrelation (gammaværdi)	N	Middelfejl
1967	0,59	651	0,054
1984	0,30	665	0,060

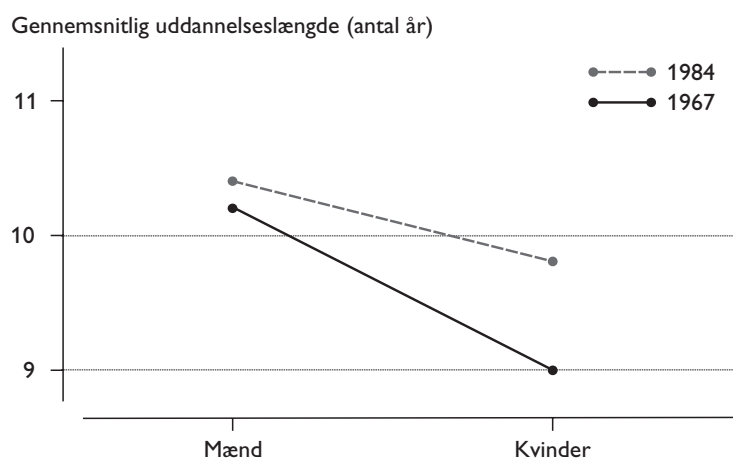
Vi ser her, at korrelationsmålet, gamma, falder numerisk fra 0,59 til 0,30, altså en halvering. Det fremgår af middelfejlen, som viser usikkerheden på disse tal, at forskellen må anses for at være mere end tilfældig. Der er altså tale om en relativ heterogenitet.

Vi kan også udregne den gennemsnitlige uddannelseslængde for de to køn i henholdsvis 1967 og 1984. Resultetet fremgår af tabel 3.25.

Tabel 3.25. Gennemsnitlig uddannelsestid for mænd og kvinder i henholdsvis 1967 og 1984.

År	Køn	Gennemsnitlig uddannelsestid (antal år)
1967	Mand	10,2
	Kvinde	9,0
1984	Mand	10,4
	Kvinde	9,8

Tabellen viser, at mens der i 1967 er en betragtelig forskel mellem kønnene, idet mændene har et højere uddannelsesniveau, er denne forskel stærkt reduceret i 1984. Dette forhold fremgår af de gennemsnitlige uddannel-



Figur 3.20. Formindskelsen af forskellen i uddannelsestid mellem kønnene.

sestider, hvor afstanden i 1967 var fra 9,0 hos kvinderne til 10,2 hos mændene, altså lidt over et års forskel. I 1984 var afstanden reduceret til godt et halvt år, nemlig fra 9,8 til 10,4.

Figur 3.20 er en grafisk fremstilling af dette.

Vi har altså her et tydeligt eksempel på en svag heterogen betinget sammenhæng, nemlig sammenhængen mellem variablene køn og uddannelse betinget af den relative modifikator årstal.

Binære fejlslutninger – tilsløring (confounding)

Mange statistiske fejlslutninger drejer sig om forskellige former for *tilsløring* (*confounding*); det vil sige, at en umiddelbart fremtrædende sammenhæng kun er *tilsyneladende*, idet den, når man inddrager baggrunden, viser sig at dække over en anden relation. Vi vil her skelne mellem to former for tilslørede sammenhæng og dermed mellem to former for fejlslutning.

1. Falsk (spuriøs) sammenhæng, som har to former:
 - a. illusorisk sammenhæng (mediation)
 - b. paradoksal sammenhæng (Simpson's paradoks)
2. Skjult sammenhæng – suppression

Falsk (spuriøs) sammenhæng

En tilsyneladende (prima facie) sammenhæng dækker over en anden relation, der viser sig, når man betinger med den *tilslørende faktor* (confounder).

Illusorisk sammenhæng (mediation)

En tilsyneladende sammenhæng mellem A og B forsvinder ved betingning med C. Der er her tale om en i første trin overset mediation:

1. **A-B: Der konstateres en ubetinget, marginal sammenhæng**

2. **A | B | C: Denne (ubetingede) sammenhæng forsvinder ved betingning**

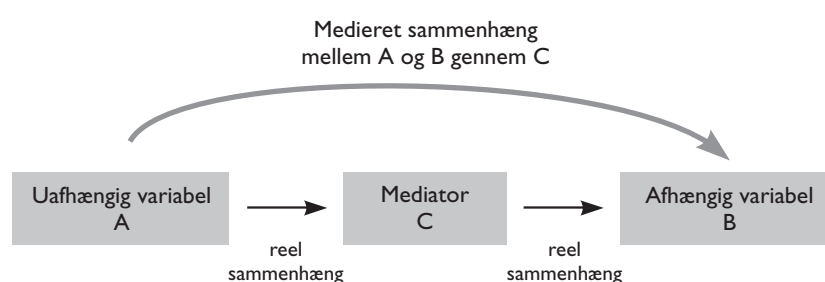
C er den mediator, som ved at forbinde begge forgrundsvariable forårsager

den illusoriske ubetingede sammenhæng. Der kan skelnes mellem to slags mediation: *intermediær* og *co-determineret*.

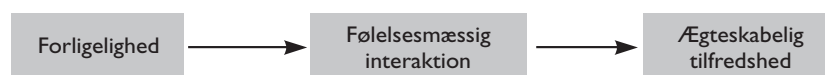
Intermediær mediation

I figur 3.21 er mediatoren C en mellemliggende størrelse i kausalkæden A,B,C.

A forårsager C, der atter forårsager B: $A \rightarrow C \rightarrow B$



Figur 3.21.

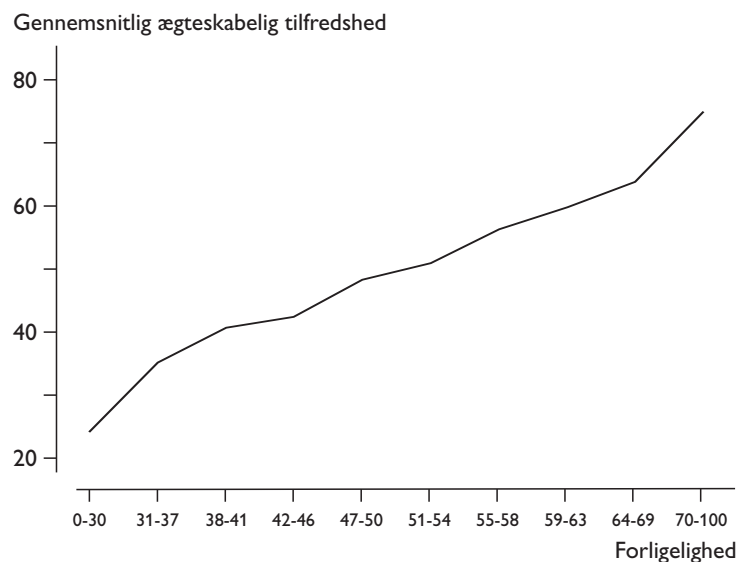


Figur 3.22. Et eksempel på ren mediation – med følelsesmæssig interaktion som mediator mellem forligelighed og ægteskabelig tilfredshed.

Ved den rene mediation er der mellem den uafhængige og afhængige variabel indskudt en mediator, også kaldet en intervenerende variabel, som formidler den kausale påvirkning fra den uafhængige variabel til den afhængige. I sin konsekvente udformning er der tale om en kædereaktion, hvor den uafhængige variabel først skaber eller indvirker på mediatoren, der så i næste omgang influerer på den afhængige variabel.

Det følgende eksempel (se figur 3.22 og 3.23) er et simuleret materiale på grundlag parrelationsforskeren Gottmanns undersøgelser.

En formentlig en udbredt forestilling, at forligeligheden er den afgørende faktor for ægteskabelig lykke. Lad os se på det simulerede materiale (figur 3.23).

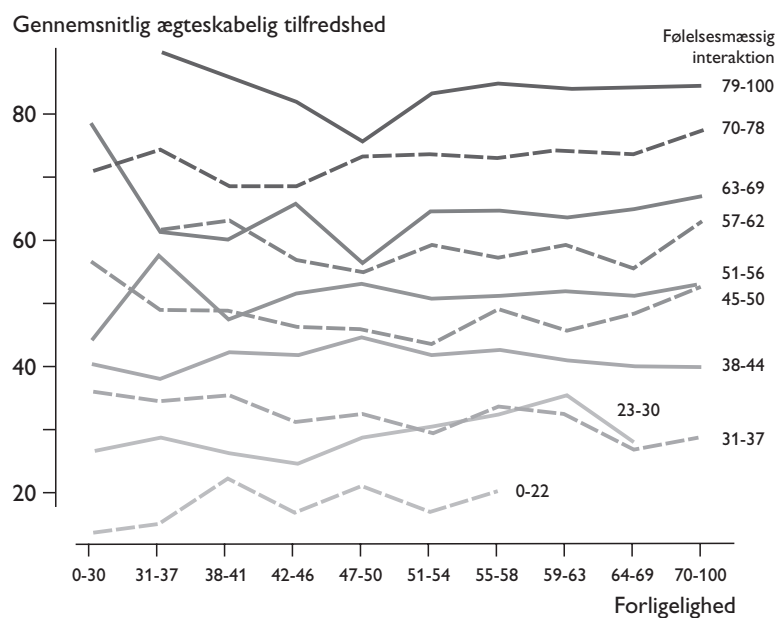


Figur 3.23. Sammenhængen mellem forligelighed og ægteskabelig tilfredshed.

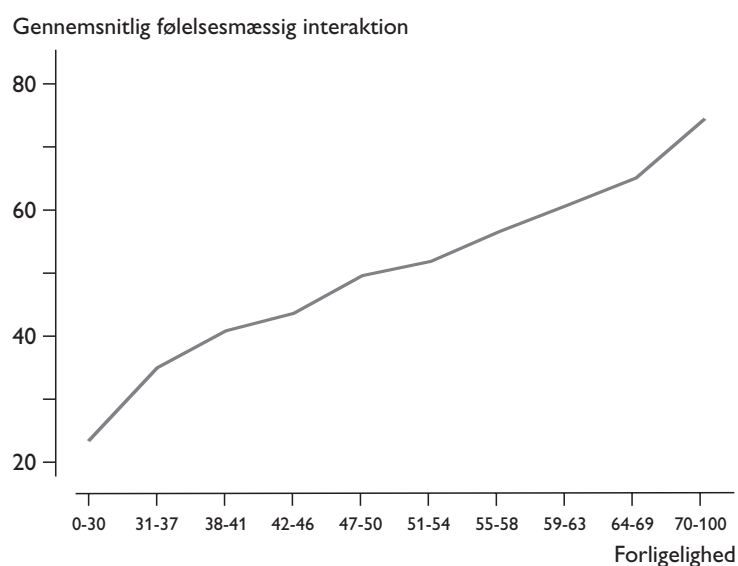
Denne sammenhæng ser jo ret overbevisende ud. Men den er ubetinget og kan derfor tænkes at være medieret af mellemliggende processer. Lad os antage, at forligeligheden har en determinerende virkning på den følelsesmæssige interaktion, der igen påvirker den ægteskabelige tilfredshed. I så fald kan vi undersøge, om der foreligger mediation, ved at betinge med den mulige mediator, som det vises i figur 3.24.

Vi ser her, at der slet ikke er nogen reel sammenhæng mellem forligelighed og ægteskabelig tilfredshed. For hvert niveau af følelsesmæssig interaktion har vi kurver som, bortset fra tilfældige variationer, er vandrette linjer. Derimod ser vi, at den følelsesmæssige interaktion nøje determinerer ægteskabelig tilfredshed, idet denne tilfredshed stiger fra en skalaværdi på ca. 20 for laveste niveau af interaktion til ca. 80 for højeste niveau af interaktion. Vi har nu vist, at den umiddelbare sammenhæng mellem forgrundsvariablene, forligelighed og ægteskabelig tilfredshed er en illusion.

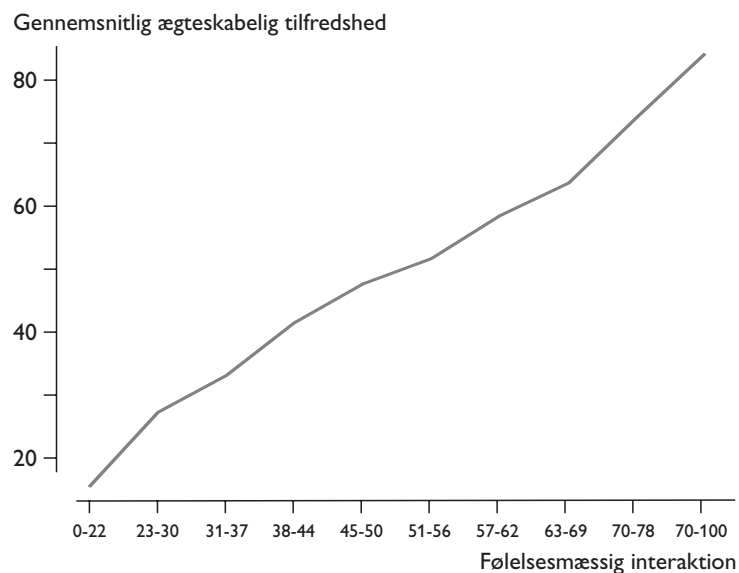
For at vise, at der faktisk er tale om mediering gennem baggrundsvariablen følelsesmæssig interaktion, mangler vi kun at demonstrere, at denne hænger sammen med begge de to andre. Det første led i denne mediering fremgår af figur 3.25, den anden af figur 3.26 (som altså bekræfter, hvad vi allerede så på figur 3.24).



Figur 3.24. Sammenhæng mellem forligelighed og ægteskabelig tilfredshed – betinget af følelsesmæssig interaktion.



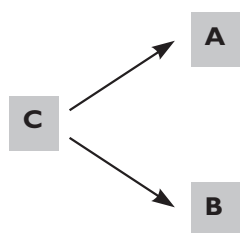
Figur 3.25. Sammenhængen mellem forligelighed og følelsesmæssig interaktion.



Figur 3.26. Sammenhængen mellem ægteskabelig tilfredshed og følelsesmæssig interaktion.

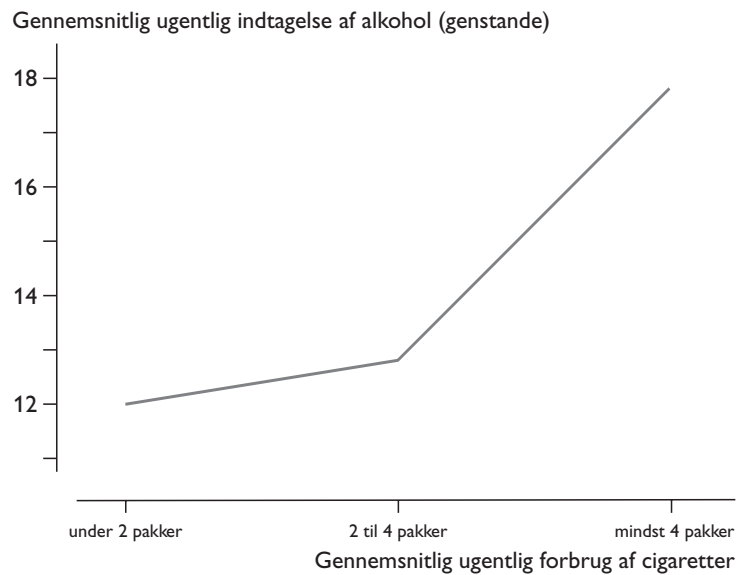
Kausal co-determination

Den anden form for mediering er den kausale co-determination. Her er mediatoren C en fælles kausal determinator for de to variabel A og B, som det vises i figur 3.27.

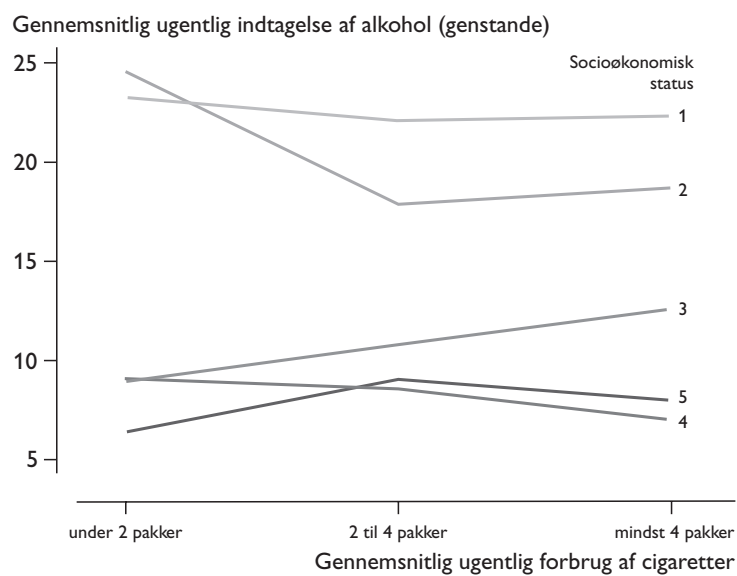


Figur 3.27.

Dette er tilfældet i de meget vigtige eksempler på tilsløring, hvor en sammenhæng mellem to størrelser fx er betinget af en social baggrundsva-riabel såsom socioøkonomisk status. Lad os forestille os, at der er foretaget en undersøgelse af såkaldt sundhedsmæssig risikoadfærd, og at man bl.a. har spurgt folk om forbrug af alkohol og tobak. Det kan nu være at man finder en sammenhæng mellem disse to størrelser – som vist i figur 3.28.



Figur 3.28. Den direkte sammenhæng mellem tobak- og alkoholindtagelse.



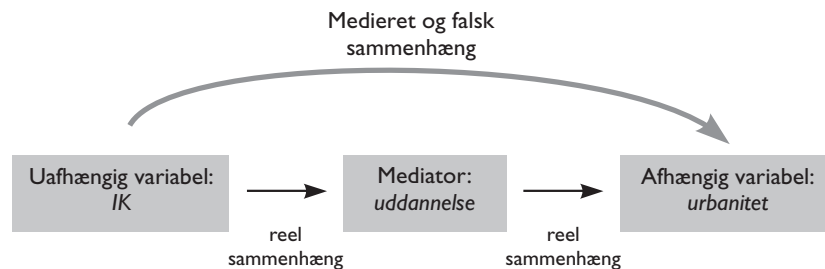
Figur 3.29. Den betingede sammenhæng mellem tobaks- og alkoholforbrug givet socio-økonomisk status.

Når vi derimod betinger med vores baggrundsvariabel, får vi et andet resultat (se figur 3.29).

Det viser sig, at sammenhængen mellem forbrug af de to nydelsesmidler var illusorisk, idet der foreligger en tilslørende baggrundsvariabel, nemlig den co-determinerende faktor socio-økonomisk status.

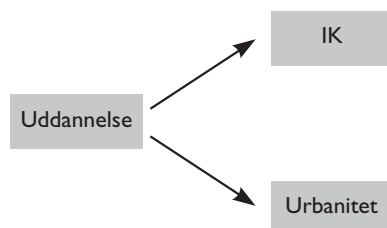
Ved gennemgangen af den betingede uafhængighed mellem urbanitet og IK – givet uddannelse – blev en indholdsmæssig forklaring stillet i udsigt. Da formålet med denne fremstilling ikke er uddannelsessociologisk, skal vi her kun vise, hvordan de to former for mediering evt. kan anvendes som hypotetiske forklaringsmodeller.

Vi kan for det første forestille os en kausal mediering, hvor IK er den uafhængige (og primære) kausalvariabel (se figur 3.30). Dette kunne være baseret på en teori om intelligensens medfødte karakter.



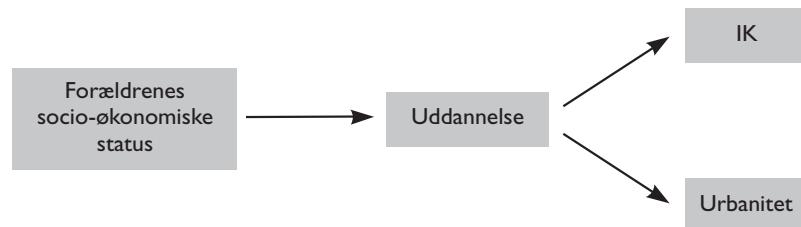
Figur 3.30. Kausalmodel med mediering mellem urbanisering og IK.

Som alternativ til dette kan vi i stedet opstille en sociologisk hypotese om uddannelsen som den fælles co-determinator (se figur 3.31). Her er antagelsen af uddannelse former intelligensudviklingen og bestemmer folks fremtidige bopæl.



Figur 3.31. Kausalmodel med urbaniseringsgraden som co-determinator for urbanitet og IK.

Vi kunne ligefrem stramme den sociologiske tendens ved at antage en determinator forud for co-determinatoren som vist i figur 3.32.



Figur 3.32. Kausalmodel med fælles baggrundsvariabel – som determinant for co-determinatoren for urbaniseringsgrad og IK.

Paradoksal sammenhæng (Simpsons paradoks)

Paradoksal sammenhæng vil sige, at en tilsyneladende sammenhæng skifter fortegn ved betingning med den *tilslørende faktor*, som da bliver til en *afslørende faktor*.

$$r(A, B) > 0$$

$$r(A, B|C) < 0 \text{ (eller omvendt, negative bliver til positive)}$$

Den alvorligste form for tilsløring er Simpsons paradoks, hvor det, der fremtræder som sort, faktisk viser sig at være hvidt. Simpsons paradoks er et statistisk fænomen, hvis eksistens umiddelbart forekommer ganske kontraintuitivt, idet hvad der ser ud som en bestemt, fx positiv sammenhæng mellem to variable, faktisk er udtryk for det stik modsatte, in casu altså en negativ sammenhæng. Her gælder ordsproget om ikke at skue hunden på hårene i den vidtgående form, at den tilsyneladende langhårede hund faktisk kan vise sig at være korthåret, eller omvendt. Vi skal først se et drastisk eksempel på dette.

I en amerikansk sydstat Missouri førte Amnesty en sag mod staten med

Tabel 3.26. Den oprindelige dødsdomsstatistik – baseret på morderens hudfarve (Lenza et al., 2003).

Gerningsmand	Dødsdomme	Sager	Procent
Hvid	284	3199	6,9 %
Sort	276	6644	4,2 %
Total	562	9843	100%

påstand om racediskrimination, idet Amnesty mente, at risikoen for dødsstraf var størst i mordsager, hvor gerningsmanden var sort. Imidlertid viste en opgørelse (se tabel 3.26) over 9843 mordsager (Lenza et al., 2003) det stik modsatte, idet sandsynligheden for dødsstraf faktisk var lavere for sorte mordanklagede end for hvide. Forskellen på godt 4 % for sorte og knap 7 % for hvide er stærk statistisk signifikant ($p < 0,001$).

For nu ikke at risikere et krav fra Ku Klux Klan om strengere domfældelse over for sorte, fik organisationen Amnesty nedsat en videnskabelig komité, der så lidt nærmere på tallene. Her er nemlig et godt eksempel på, at man ikke skal skue hunden på hårene, eller rettere sagt morderen på hans hudfarve. Der er nemlig et andet forhold, som må inddrages, *offerets* hudfarve. Når man i en tabel inddrager en ny variabel, taler man om at *nedbryde* eller *elaborere* tabellen. Da man nedbrød den ovenstående tabel efter denne nye baggrundsvariabel, fik man som vist i tabel 3.27 fire kombinationer af henholdsvis anklaget og offer:

Tabel 3.27. Den elaborerede tabel efter inddragelsen af offerets hudfarve (Lenza et al., 2003).

Gernings- mand	Offer								
	Hvid			Sort			Total		
	Døds- dom	Sager	%	Døds- dom	Sager	%	Døds- dom	Sager	%
Hvid	274	2945	8,4	12	254	4,7	284	3199	6,9
Sort	88	599	14,7	188	6045	3,1	276	6644	4,2
Total	362	3544	10,2	200	6299	3,1	562	9843	5,7

Her ser vi flere interessante ting:

- A. I mordsager er der en klar tendens til apartheid. Sorte mordere foretrækker sorte ofre, ligesom hvide mordere foretrækker hvide ofre. I de mordsager, hvor en sort var anklaget, var offeret i 91 % af tilfældene sort. Og tilsvarende: I de mordsager, hvor den anklagede var hvid, var 92 % af ofrene hvide. Det bliver altså i vid udstrækning, om ikke i familien så dog inden for racen.

- B. Prisen i mordsager, altså risikoen for dødsstraf, viser – hvor antallene er tilstrækkelige – følgende prisskala:

1. Dyrest er det, når sort myrder hvid: risiko = 14,7 %
2. Næstdyrest er det, når hvid myrder hvid: risiko = 8,4 %
3. Billigst er det, når offeret er sort: risiko = 3,1 %⁴

Dette mønster (med effekterne A og B) resulterer i en anden lige så ufuldkommen marginal tabel efter offerets hudfarve (tabel 3.28).

Tabel 3.28. En alternativ dødsdomsstatistik –baseret på offerets hudfarve (Lenza et al., 2003).

Offer								
Hvid			Sort			Total		
Dødsdom	Sager	%	Dødsdom	Sager	%	Dødsdom	Sager	%
362	3544	10,2	200	6299	3,1	562	9843	5,7

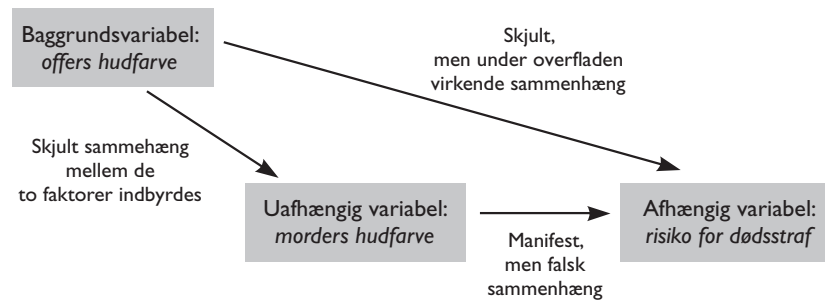
Der er altså tale om to modvirkende forhold, som tilsammen udligner hinanden, så der tilsyneladende ikke er nogen diskrimination. Det er diskriminerende består imidlertid i, at et hvidt liv er mere kostbart end et sort. Denne diskrimination giver imidlertid en svækket effekt, fordi hvide mordere tenderer at vælge hvide ofre, men sorte mordere foretrækker sorte ofre. Men den kontante forskelsbehandling ser vi i det faktum, at risikoen for at ende på dødsgangen efter at have ombragt en hvid person er op mod dobbelt så stor for en sort gerningsmand som for en hvid.

Vi har altså det paradoks, at umiddelbart når vi kun ser på én variabel, nemlig gerningsmandens hudfarve, får vi det resultat, at sorte slipper billigere fra et mord. Dette beror på en såkaldt marginal-tabel, som altså kan dække over både det ene og det andet, herunder det stik modsatte af det, som det ser ud til.

Når vi nedbryder denne marginaltabel ved at opdele efter en variabel, offerets hudfarve, så finder vi i den fremkomne krydstabel det resultat, at det forholder sig lige modsat. I de relevante sammenhænge gælder det, at der ikke er nogen forskelsbehandling når offeret er sort (for de tæller

alligevel ikke synderligt), men når offeret er hvidt, så straffes en sort gerningsmand hårdere end en hvid.

Dette paradoks kaldes for Simpsons paradoks (Simpson, 1951), hvorved forstås, at en overflade- eller prima facie-sammenhæng fremstår med en tendens, der er falsk, og som vender, når yderlige variable bliver inddraget. Figur 3.33 viser et diagram over dette.



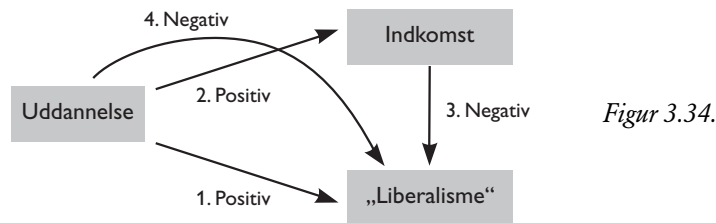
Figur 3.33. Eksempel på Simpsons paradoks.

Skjult sammenhæng – suppression

En tilsyneladende manglende sammenhæng mellem A og B viser sig ved betingning med den tilslørende faktor C at dække over en faktisk sammenhæng, der her optræder som *suppressor*.

1. $A \mid B$
2. $A-B \mid C$

Der er her givet et eksempel i afsnittet om „liberalisme“ (i amerikansk forstand) og uddannelse (se s. xx). Men hvordan kan man nu forestille sig en indholdsmæssig forklaring på dette fænomen? Der er faktisk belæg for at antage, at uddannelse påvirker studerende i retning af liberalisme. Når dette ikke bliver synligt, hænger det sammen med en modsatrettet medieret påvirkning, nemlig en voksende indkomst. Uddannelsen resulterer nemlig samtidig i en tendens til højere indkomst, og denne højere indkomst er tilbøjelig til at gøre folk mere konservative. Vi har altså her den hypotetiske kausalmodel, som vises i figur 3.34.



Figur 3.34.

Vi har her fire kausalpile:

1. Uddannelse → „Liberalisme“: direkte og positiv
2. Uddannelse → Indkomst: direkte og positiv
3. Indkomst → „Liberalisme“: direkte og negativ
4. Uddannelse → „Liberalisme“: medieret og negativ

Der er med andre ord to kausale påvirkninger fra uddannelse, der virker ind på liberalisme:

- a. Den direkte og positive pil (1)
- b. Den medierede og negative pil (4)

Når disse to påvirkninger indgår i en samlet effekt, ophæver de således hinanden. Resultatet er, at vi ikke ser nogen effekt af uddannelsen på liberalisme. Forbindelsen er nemlig skjult, ved at suppressoren indkomst medierer en negativ indflydelse, som ophæver den direkte positive effekt af uddannelsens liberaliserende tendens.

Sammenfatning om statistiske fejlslutninger

Om de tre gennemgåede fejltyper: den *falske*, den *paradoksale* og den *skjulte* sammenhæng kan vi sammenhængende sige:

1. Falsk sammenhæng svarer til det, man i metodelitteraturen kalder falsk positive fund.
2. Paradoksal sammenhæng omtales sjældent i metodelitteraturen, måske fordi der er tale om en så pinlig fejltype, at den bliver genstand for en kollektiv fortrængning.
3. Skjult sammenhæng svarer til det, man i metodelitteraturen kalder falsk negative fund.

Afsluttende bemærkninger om statistiske sammenhænge

Vi har talt om en enkelt betingende variabel C, men der kan være tale om en blok af variable, således at C kan tolkes som den betingende *kontekst*.

Vi har kun talt om to analytiske niveauer: forgrund og baggrund. Men der kan være (og er ofte) tale om mere end to. Således kan den afsløring, der sker, når den oprindelige ubetingede relation afsløres som misvisende gennem en betingning, resultere i en mere reflekteret analyse, som dog selv viser sig at være tilsløret, efter at der er blevet indført yderligere betingende variable ud over den initiale baggrundsvariabel. Imidlertid indebærer dette ikke en erkendelsesmæssig relativisme, for ved en sådan tretrins- (eller principielt mangetrins-) proces vil afsløringen ved trin 2 give et mere korrekt billede end ved trin 1, ligesom resultatet af trin 3 er endnu mere korrekt end ved trin 2 osv.

Survey-metodens indbyggede begrænsninger

Surveymetoden som kvantitativ metode

Vi skal først se på surveymetoden som en fundamentalt kvantitativ metode. I kapitel 1 (afsnittet om serialitet og kontekstualitet, s. xx) blev der træk, der karakteriserer henholdsvis kvalitative og kvantitative undersøgelser gennemgået. I overensstemmelse med definitionen af kvantitativ metode serialiserer surveyundersøgelsen sine genstande, så de bliver kontekstløse, uspecifikke og atomiseret til enkeltoplysninger. Metoder som arbejder konkret og ikke abstrakt, er fx casemetoden, interviewmetoden, og i en vis udstrækning feltundersøgelsesmetoden, som alle er kvalitative metoder.

Surveymetoden som tværsnitsmetode

Da en spørgeskemaundersøgelse normalt undersøger sine genstande på ét bestemt tidspunkt, evt. ved hjælp af en succession af forskellige stikprøver på hver sit tidspunkt, er den blind, hvad angår processer og udvikling. For at komme ud over dette kræves en længdesnitsundersøgelse, hvor hver genstand følges over tid.

Surveymetoden som en korrelationel metode

Surveymetoden er korrelationel, dvs. at den kun kan sammenholde sine enkeltoplysninger, der foreligger som perler på en snor. Hvis man skal afprøve om sammenhæng er kausal skal man strengt taget bruge en eksperimentel metode, hvis det da overhovedet har en mening at angribe fænomenet kausalt, og hvis det er rimeligt at afprøve det eksperimentelt.

Alternativer til surveymetoden

Det har altså flere gange været nævnt at survey-metoden har svære begrænsninger. Den er bedst til at belyse, snarere end at afdække et heterogent og ofte kompliceret område; når det drejer sig om at finde årsagsforklaringer og forstå den oplevelsesmæssige side af fænomenerne, er andre metoder mere velegnede. Vi skal her se på to alternative undersøgelser inden for området familievold.

“The marriage license as a hitting license”

I bogen *Behind Closed Doors* (Strauss, 1980) gives den delforklaring på familievold, at ægtepagten giver manden ret til at slå.

For denne tolkning taler retshistorien (i den vestlige verden), idet manden indtil for godt hundrede år siden havde revselsesret over for hustruen, en tradition, der indtil for ganske nylig blev videreført i rets- og politipraksis.

Men hvordan er den såkaldte retsopfattelse i befolkningen? Dette spørgsmål er undersøgt gennem et socialpsykologiskeksperiment foretaget i 1974 på Pennsylvania State University (Shotland & Straw, 1976). Forsøgspersonerne var 51 psykologistuderende, der en aften enkeltvis blev indkaldt under en dækinstruktion. Når forsøgspersonen var ankommet, startede der et simuleret skænderi lige uden for lokalet. Faktisk var der tale om to dramastuderende, der gennemspillede en diskussions- og voldsscene, som begyndte med, at manden beskyldte kvinden for at have taget en dollar, han havde tabt på gulvet. Efter en kort højroset diskussion begynder manden så et fysisk overgreb på kvinden, idet han rusker i hende, mens hun kæmper imod og skriger. Indimellem bønfalder hun manden om at give slip.

Den afgørende eksperimentelle variabel var nu, at der var to varianter af skuespillet. I den første udgave råbte kvinden, indimellem sine smer-

tenssskrig: "Jeg kender dig jo slet ikke." I den anden version, råbte hun: "Jeg fatter ikke, at jeg nogensinde har giftet mig med dig." Scenen, der kun varede et minut, blev enten afbrudt, når forsøgspersonen intervenerede, hvorefter den falske overfaldsmand flygtede, eller når der kom en person ud fra et nærliggende lokale og krævede, at overfaldsmand skulle holde op.

Derefter blev forsøgspersonen interviewet om alvoren af volden.

Der er således både en adfærdsmæssig og en oplevelsesmæssig afhængig variabel. Hvad angår den første, viste den eksperimentelle variabel sig at have stor betydning. Når parret fremtrådte som ægtepar, intervenerede kun 19 % af forsøgspersonerne, mens 65 % af dem greb ind, når parret blev opfattet som uden forudgående relation. Også graden af oplevet vold afhang af, hvordan relationen blev opfattet.

Den interpersonelle dynamik i hustruvold

Lenore Walker (1978) har foretaget en klinisk case-analyse på grundlag af mishandlede kvinder, hun havde i psykologisk behandling. Hun fremsætter i sin artikel hun den interessante hypotese, at der er et cyklisk forløb i parforhold med kronisk og alvorlig hustrumishandling. Hun opererer med tre faser.

- Spændingsopbygningen. Manden bliver mere og mere irriteret på konen. Hans tærskel for, hvad han opfatter som uacceptabelt, bliver stadig lavere. I begyndelsen er manden måske brændt på; senere er det nok, at det er en ret, han ikke kan lide.
- Den akutte voldsepisode.
- Genopbygningsfasen, hvor den forslåede kone udsættes for et romantisk stormløb af næsten samme intensitet som voldsepisoden.

Walker skriver, at når hun skulle besøge sine klienter på hospitalet efter overfaldet, behøvede hun ikke at spørge personalet, hvilken stue klienten lå på, for den lignede altid en blanding af en chokoladebutik og en blomsterforretning. Der er Walkers hypotese, som er blevet bekræftet af andre undersøgelser, at den mishandlede hustru ofte elsker den voldelige mand hinsides al fornuft. Og en vigtig faktor er her, den lidenskabelige måde, hvorpå han viser sin afhængighed af hende i den såkaldte genopbygningsfase.

Et slående eksempel på genopbygningsfasen findes i et interview med

en mishandlet hustru, der har søgt tilflugt i Danner-boligerne i 1982 (Karpatschof 1983, s. 331). Kvinden, der optræder under pseudonymet Alice, fortæller:

Når vi har vores gode perioder, ja så er han bare Da vi var forlovede, sagde jeg altid, at jeg havde vundet i lotteriet.

Hvorfor vælger du at fortsætte, hvorfor vælger du fx ikke at leve alene?

Jeg var overbevist om, at jeg ikke havde flere følelser for Viggo. Men da jeg så ham, da slog det mig: "Han ser jo fuldstændig ud, som jeg ser ud, når jeg er psykisk kørt ned." Han var langskægget og ligeglad med, hvordan han så ud. Han rystede på hænderne, og jeg sagde til ham: "Er det, fordi du har drukket, at du ryster på hænderne?" Nej, det var det ikke, han var så ulykkelig, og han havde ikke kunnet finde mig, og han kunne ikke forstå, at jeg kunne blive væk. Der var følelser tilbage. Og så tænkte jeg: "Halvanden år er jo ikke meget at give et ægteskab." Når to vildtfremmede mennesker, der ikke kender hinanden, flytter sammen, det er jo to verdener, der mødes. Når Viggo har det godt, og vi fx sidder ude på altanen, så kan han godt sidde og tage min hånd, der er jo altid en kærlig hånd, et klap. Og når han så tager min hånd og sidder og kigger på mig og siger: "Jeg elsker dig, Alice, jeg synes du er jordens dejligste kvinde."

Styrke og svaghed ved det kvalitative alternativ til surveymetoden

Hvad kan der komme ud af den supplerende viden, som fx Walkers kvalitative undersøgelse kan bidrage med? Det er klart, at den på den ene side kan fortælle os noget om den interpersonelle vold, der synes at gå mere i dybden, end surveyundersøgelsen gør. Den bringer os på sporet af en dynamik, som muligvis er en afgørende nøgle – både teoretisk og praktisk – til forståelse af den interpersonelle vold.

Det er imidlertid også klart, at den har nogle afgørende metodologiske svagheder: Den har ingen udvælgelsesplan, og resultatet kan derfor ikke generaliseres til en bestemt population; der er heller ikke opstillet apriori-hypoteser, så dens udsagn bliver ikke afprøvet.

Walkers undersøgelse er derfor i virkeligheden en pilotundersøgelse, der både fortjener og behøver efterfølgende undersøgelser, som vil være egnede til at efterprøve dens resultater.

Litteratur

- Bandura, A. (1994). Self-efficacy. I: V.S. Ramachaudran (ed.), *Encyclopedia of Human Behavior*, vol. 4, s. 71-81. New York: Academic Press.
- Baron, R. & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51: 1173-1182.
- Grau, R. et al. (2001). Moderator effects of self-efficacy on occupational stress. *Psychology in Spain*, 5(1): 63-74.
- Karpatschof, L. (1980). *Kvindens svaghed – en undersøgelse af hustrumishandling*. København: Antropos.
- Kreiner, S. (2001). ??????????
- Lenza, M., Keys, D. & Guess, T. (2003). *The prevailing injustices in the application of the death penalty in Missouri (1978-1996)*. University of Missouri, St. Louis. <http://www.umsl.edu/~phillips/dp/Lenza1.html>
- Nielsen, T. & Kreiner, S. 2003, *SPSS – Introduktion til databehandling & statistisk analyse*, 2. udg. København: Jurist- og Økonomforbundets Forlag.
- Olsen, H. (1998). *Tallenes talende tavshed. Måleproblemer i surveyundersøgelser*. København: Akademisk Forlag.
- Schuman, H. & Presser, S. (1981). *Questions and Answers in Attitude Surveys*. New York: Academic Press.
- Shotland, R.L. & Straw, M.K. (1976). Bystanders' response to an assault: When a man attacks a woman. *Journal of Personality & Social Psychology*, 34(5): 990-999.
- Simpson, E.H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society, Ser. B*, 13: 238-241.
- Strauss, M.A. et al. (1980). *Behind Closed Doors*. New York: Doubleday.
- Van de Ven, A.H. & Ferry, D.L. (1980). *Measuring and Assessing Organizations*. New York: Wiley.
- Walker, L.E. (1978). Battered women and learned helplessness. *Victimology*, 2: 3-4.
- Walker, L.E. (1979). *The Battered Woman*. New York: Harper & Row.

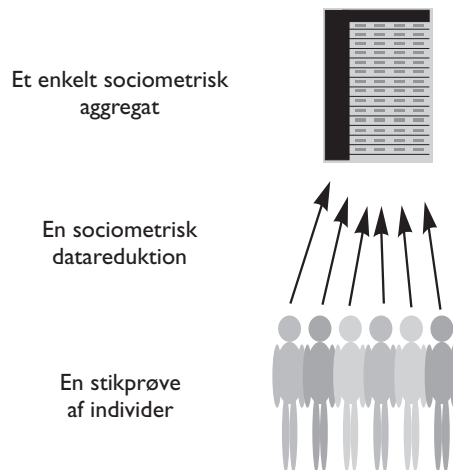
Noter

- Note 1, s. 142 Herhjemme førte denne bevægelse til oprettelse af de såkaldte krisecentre, hvoraf den største og mest kendte er krise- og kvindecetret Grevinde Danner, oprettet 1979.
- Note 2, s. 147 Forskellen mellem begreberne *social* og *familiemæssig* arv er, at den første term henviser til barnets hele opvækstmiljø, dvs. alle dets sociale og kulturelle omgivelser, mens den sidste term går på, hvad der specifikt kendetegner den enkelte familie.
- Note 3, s. 168 <http://www.census.gov/population/www/pop-profile/educattn.html>. <http://www.census.gov/hhes/income/income00/inctab7.html>.
- Note 4, s. 189 Forskellen mellem sort myrder sort og hvid myrder sort er ikke signifikant.

KAPITEL 4

Psykometriske metoder

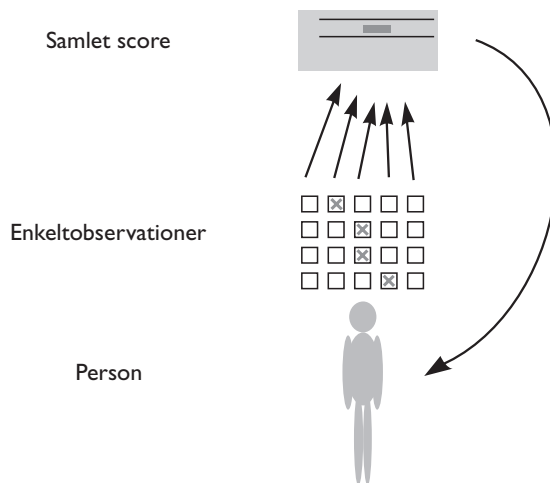
Psykometri som genre



Figur 4.1. Den sociometriske datareduktion.

I kapitel 3 så vi på den sociometriske datareduktion, der forekommer, når vi har surveyoplysninger fra en række enkeltpersoner, som statistisk sammentælles eller sammenregnes, så individerne forsvinder, mens et sociometrisk resultat foreligger, fx:

- Så mange procent i den og den sociale gruppe gør eller mener sådan og sådan.
- Hvis man gør eller mener A i den pågældende gruppe, er der stor sandsynlighed for, at man også gør eller mener B.



Figur 4.2. Den psykometriske datareduktion.

Ved den psykometriske datareduktion (figur 4.2) er sigtet derimod at udtale sig om enkeltindividet, fx:

- Hvilke egenskaber karakteriserer personen generelt?
- Hvilken tilstand er personen i på et bestemt tidspunkt?
- Hvilken forandring er der sket med personen fra tidspunkt t1 til tidspunkt t2?

Samtidig er man i psykometrien som regel ude efter at sammenligne eller adskille en række enkeltindivider mht. bestemte egenskaber. Sigtet er med andre ord *differentialpsykologisk*. Vi karakteriserer således den enkelte person ved at angive, hvad der adskiller de enkelte personer fra hinanden.

En psykometrisk undersøgelse beskæftiger sig normalt med et af to hovedområder:

- A. Mentale evner (intelligens)
- B. Personlighedstræk (herunder holdninger)

Uanset området er der typisk tale om følgende størrelser:

1. En stikprøve fra en såkaldt *population* (udsnit af en befolkning)
2. Et testbatteri bestående af forskellige delprøver

3. Hver delprøve består igen af en række enkeltopgaver, items:
Ved område A, undersøges mentale evner ved, at respondenter søger at løse en række opgaver
Ved område B, undersøges personlighedstræk ved, at respondenter besvarer en række spørgsmål om sine karakteristika eller præferencer, eller ved at en observatør angiver bestemte karakteristika ved personens adfærd eller koder hans/hendes adfærd
4. Hver enkeltopgave eller spørgsmål bliver scoret efter på forhånd angivne principper:
Ved område A er der typisk tale om rigtigt (scores som 1) eller forkert (scores som 0).
Ved område B kan der fx anvendes en såkaldt Likert-skalering, hvor man skal vurdere, hvor enig man er i et bestemt udsagn, idet man angiver graden af enighed:
 1. Helt uenig
 2. Ret uenig
 3. Lidt uenig
 4. Hverken uenig eller enig
 5. Lidt enig
 6. Ret enig
 7. Helt enig

Selve scoringsproceduren består af følgende tre skridt:

1. Man scorer, som vist i figur 4.3, hvert enkelt spørgsmål ved at fastlægge nogle scoringsværdier for de forskellige svarmuligheder (her altså 7). Dette kan gøres på en af nedenstående måder:
 - a. Enten med de eksisterende numre, *skala-a* (figur 4.3). Disse tal har ikke nogen indlysende mening.
 - b. Eller med tallene fra 0 til 6, *skala-b*, som er en *monopolær* skala, der strækker sig fra 0 til en stor værdi.
 - c. Eller med tallene fra -3 til 3, *skala-c*, som er en *bipolær* skala, der strækker sig fra en stor negativ til en stor positiv værdi.
 - d. Eller man finder ved hjælp af en psykometrisk model den underliggende empiriske struktur i resultaterne, *skala-d*, hvor der i dette eksempel således ikke er ækvidistance, men større afstand mellem den yderste og den næstyderste kategori i begge ender.

Svarkategorier	Helt uenig	Ret uenig	Lidt uenig	Hverken uenig eller enig	Lidt enig	Ret enig	Helt enig
Skala-a: Naturlige tal	1	2	3	4	5	6	7
Skala-b: Monopolær	0	1	2	3	4	5	6
Skala-c: Bipolær	- 3	- 2	- 1	0	1	2	3
Skala-d: Empirisk skala	- 3,2	- 1,8	- 0,9	0	0,9	1,8	3,2

Figur 4.3. Fire forskellige måder at score en Likert-skala.

2. Efter fastlæggelsen af scoringsværdierne summerer man alle enkeltresultaterne fra de enkelte items til en *råscore* for delprøven.
3. Man bruger endelig en psykometrisk metode til at finde en overordnet struktur i hele testbatteriet.

Psykometriens begyndelse og tidlige udvikling fra Galton til Binet

Der er som sagt to hovedområder af almen interesse, som er undersøgt i psykometrien: Mentale evner (intelligens) og personlighedstræk. Vi skal først se på intelligensforskningen og derpå den psykometriske udforskning af personlighedstræk. Metodologien er for så vidt den samme, men sigtet og resultaterne har vist sig at være temmelig forskellige.

Sir Francis Galton – Intelligens hos hunde og hos mennesker

Intelligensforskningen begyndte med englænderen Galton, der måske ikke helt retfærdigt er gået over i videnskabshistorien som Darwins sandsynligvis mere intelligente,¹ men mindre vedholdende fætter (eller rettere halvfætter). Det tilføjes også ofte, at han spredte sig over så mange videnskaber, at han ikke rigtig slog igennem nogen steder. Han gjorde sig således gældende som opdagelsesrejsende, meteorolog, statistiker, biometriker og

Classification Of Men According To Their Natural Gifts.

Grades of natural ability, separated by equal intervals.		Numbers of men comprised in the several grades of natural ability, whether in respect to their general powers, or to special aptitudes.							
Below Average	Above average	Proportionate, viz. one in	In each million of the same age.	In total male population of the United Kingdom, viz. 15 millions, of the undermentioned ages :—					
				20—30	30—40	40—50	50—60	60—70	70—80
a	A	4	256791	651000	495000	391000	268000	171000	77000
b	B	6	162279	409000	312000	246000	168000	107000	48000
c	C	16	63 563	161000	123000	97000	66000	42000	19000
d	D	64	15696	39800	30300	23900	16400	10400	4700
e	E	413	2423	6100	4700	3700	2520	1600	729
f	F	4300	233	590	450	355	243	155	70
g	G	79000	14	35	27	21	15	9	4
x	X								
all grades below e	all grades above G	1000000	1	3	2	2	2	—	—
On other side of average			500000	1268000	964000	761000	521000	332000	149000
Total, both sides			1000000	2536000	1928000	1532000	1042000	664000	298000

The proportions of men living at different ages are calculated from the proportions that are true for England and Wales. (Census 1861, Appendix, p. 107.)

Example.—The class F contains 1 in every 4,300 men. In other words, there are 233 of that class in each million of men. The same is true of class f. In the whole United Kingdom there are 590 men of class F (and the same number of f) between the ages of 20 and 30; 450 between the ages of 30 and 40; and so on.

Tabel 4.1. Galtons intelligensklasser.

genetiker. Blandt andet på baggrund af sin interesse for spørgsmålet om arveligheden af mentale egenskaber, blev han grundlægger af psykometrien som særlig psykologisk disciplin.

Galton lavede ganske vist nogle ret elendige intelligensprøver, fordi han ledte efter intelligensen gennem tærskelbestemmelser og reaktionstider, men han fik afgørende betydning på psykometrien gennem sine statistiske ideer, især begreberne korrelation og regression. Han var den første, der overførte brugen af normalfordelingen til mentale egenskaber. Astronomen Queteles havde med held beskrevet belgiske rekrutters højdefordeling med den fra naturvidenskaben kendte normalfordeling, og Galton var overbevist om, at intelligens måtte være normalfordelt, hvad der dog var en rent spekulativ forestilling, i og med at hverken han eller andre i samtiden kunne måle intelligens.

Man kan sige, at hvis der er et punkt, hvor den konstruktivistiske retning inden for psykologiens videnskabsteori har ret i sin kritik af realismen, så er det med hensyn til intelligensbegrebet. Ligesom det hos Pirandello er syv personer, der søger en forfatter, er det hos Galton et begreb, der søger sin genstand. Han fandt den aldrig, men han efterlod et *skema* for konstruktion af intelligens.

I sit værk *Hereditary Genius* (Galton, 1869) søgte han ud fra optegnelser om Englands store mænd at påvise intelligensens arvelighed, og til det

formål opstillede han på baggrund af normalfordelingen nogle spekulative beregninger over intelligensens fordeling (se tabel 4.1).■

Galton inddelte den formodede normalfordelte intelligensfordeling i lige store intervaller, øjensynlig med en bredde på 0,7 gange standardafvigelsen. Dem over gennemsnittet gav han store bogstaver, dem under små bogstaver. A er således gruppen lige over gennemsnittet, og a er gruppen lige under gennemsnittet.

Disse, på hver sin side, gennemsnitsmennesker, vil der være ca. en fjerdedel af, mens der er ca. en sekstendedel af de to næste grupper, som altså ligger hhv. lige over og lige under gennemsnitsgrupperne. Disse grupper hedder så B og b.

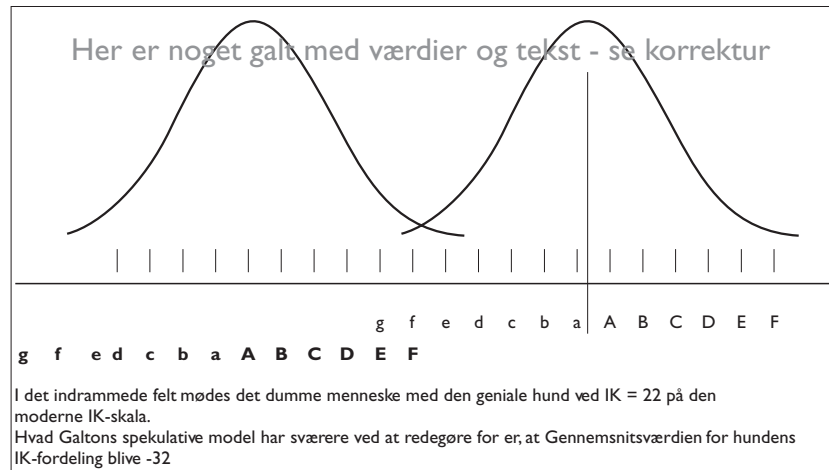
Galton konkluderer nu med en karakteristisk sammenblanding af en rent matematisk modelbetragtning og et empirisk udsagn om intelligensen forekomst:

Vi når således frem til den ubestridelige, men uforklarede konklusion, at eminent begavede mennesker befinder sig lige så højt hævet over middelmådigheden, som idioter befinder sig under den; en kendsgerning, der i betydelig grad udvider vore forestillinger om de kolossale forskelle mellem mennesker, hvad angår intellektuel begavelse (ibid. s. 36).

Han fortsætter sine spekulationer på følgende måde:

Jeg vil gå ud fra, at klassen F blandt hunde og andre af de mere begavede dyr nærmest er sammenlignelig med klassen f blandt mennesker i henseende til hukommelse og tankekraft. Klasse G blandt sådanne dyr er utvivlsomt langt overlegne i forhold til gruppe g blandt mennesker (ibid.).

Jeg har tidligere moret mig over dette stykke komparative kognitionspsykologi og set det som et udtryk for Viktoriatiden, hvor en gentleman havde mere respekt for sin jagthund end for sin kammertjener. Men hvis vi omsætter Galtons beregninger til den IK-fordeling, der gælder for Stanford-Binet-testen,² som vi senere skal beskæftige os med, så er Galtons menneske- og hundeklasser fordelt, som det fremgår af figur 4.4.■



Figur 4.4. Galtons intelligensklasser hos mennesker og hunde,

Der, hvor de meget uintelligente mennesker møder de geniale hunde, er i henholdsvis klasse f for mennesker og klasse F for hunde. Det er omkring $IK = 34$. Og der, hvor Galton er sikker på, at hunde rangerer over mennesker, er i klasse G for hunde og g for mennesker, hvilket svarer til henholdsvis $IK = 46$ for den geniale hund, og $IK = 22$ for det vitterligt meget dårligt begavede menneske, som Galton kalder en idiot, hvad der også var gængs psykiatrisk jargon indtil 1960'erne. Hvis vi fortsætter oversættelse af Galtons model til den moderne IK -skala, finder vi, at hundens gennemsnitlige IK er -32, hvad der jo er mindre indlysende.

Uanset Galtons åbenbare undervurdering af dårligt begavede mennesker og hans overvurdering af godt begavede hunde er der en afgørende fejl i hans ræsonnement, som fører til den absurditet, at hundene får en intelligensmæssig middelværdi på -30. Et resultat, der ville have været helt uacceptabelt for hundevennen Galton. Fejlen i ræsonnementet er, at Galton implicit antager, at hundenes intelligensfordeling har samme spredning som hos mennesker. Men det er øjensynligt ikke tilfældet. Og det er en lige så afgørende forskel som hundenes lavere niveau. Hvis vi foregriber IK -begrebet, der er baseret på begrebet intelligensalder, så var antagelsen, at mennesker udvikler sig, til de omkring 15 år, således at en person med gennemsnitsintelligens, og dermed $IK = 100$, ender med en intelligensalder på 15. Hunden udvikler sig derimod ikke meget efter sin etårs fødselsdag.

Ud fra min rekonstruktion af Galtons tabel, svarer klassen f af mennesker til en IK på 40 og dermed en intelligensalder på 6. Dette er for så vidt ikke særlig urealistisk. Derimod er det særdeles urealistisk, at selv en helt usædvanlig begavet hund, nemlig en fra Galtons klasse F, skulle kunne udvikle en sådan intelligensalder.

Det er ud fra denne socialdarwinistisk inspirerede despekt for de dårligt begavede ikke uforståeligt, at Galton gik skridtet videre med hensyn til overførslen af begrebet *kunstig udvælgelse* (eng. artificial selection), som var velkendt netop i forbindelse med forædlingen af racehunde. Halvfættteren Darwin overførte ideen til evolutionsteorien, hvor den affødte begrebet *naturlig udvælgelse*. Galton overførte den rent praktiske idé til menneskeavlen, idet han blev racehygiejnens ophavsmand med grundlæggelsen af *The Eugenic Society*.

Binet og den første intelligensprøve

Med Binet tog intelligensforskningen en helt anden retning. Binet var ligesom Galton genetisk orienteret, men i den stik modsatte betydning af ordet. Hvor Galtons genetiske perspektiv var arvebiologisk, var Binets udviklingspsykologisk. Galton så intelligens som en essentielt biologisk nedarvet egenskab, mens Binet opfattede intelligens som den samlede disposition til at forholde sig til et givet problem, en disposition, der udviklede sig gennem personens liv og særligt markant gennem barndommen.

Binets udviklingspsykologiske forståelse af intelligens var allerede funderet på videnskabelige studier. Den kom nu til fuldt udtryk, da den franske undervisningsminister kort efter 1900 nedsatte en undersøgelseskommision, der skulle tage stilling til, hvad der kunne gøres for de svagt begavede elever, der med den nyindførte almindelige skolepligt var blevet et pædagogisk problem.

Denne kommission bad Binet og hans videnskabelige partner Simon om at udvikle en undersøgelsesmetode til at udskille de elever, hvis intellektuelle niveau lå så meget under deres jævnaldrende kammerater, at de måtte have en særlig tilrettelagt undervisning.

Binet og Simon (1977a, 1977b) udviklede en mængde prøver omfattende stort set alle perceptuelle, hukommelsesmæssige og tænkemæssige funktioner og afprøvede dem på såvel normale som intellektuelt retarderede børn i alderen fra tre til 12 år. Herudfra opbyggede de deres berømte test,

som består af fem prøver på hvert alderstrin. Ved en enkel, men samtidig meget effektiv scoringsmetode findes nu barnets *mentale alder* – det begreb, der siden er blevet omdøbt til *intelligensalder*. Man indleder med at teste barnet på et niveau, man er sikker på, at barnet kan klare, og bliver ved, til det ikke kan løse nogen opgaver. Barnet tildeles nu en mentalalder svarende til det aldersniveau, hvor det sikkert eller næsten sikkert kan klare fire ud af fem opgaver. Hertil lægges et år for hver yderligere fem prøver, det samlet har besvaret korrekt.

Det psykometrisk interessante er her, at den kvantitet, man måler, ikke er nogen okkult essens som "intelligens i sig selv", men en simpel og i princippet velkendt egenskab, nemlig alder.

Det raffinerede er dels, at der altså er tale om en slags aldersnorm, dels, at det på en tilsyneladende enkel, men i virkeligheden særdeles sofistikeret måde er lykkedes at konstruere en sådan aldersnorm. Det psykologisk interessante er det ofte kritiserede forhold, at intelligens med føje er defineret, ikke som den enkle, rendyrkede egenskab, Galton søgte, men netop som noget sammensat, noget komplekst, noget grundlæggende livspraktisk og dermed også almenpsykologisk grundlæggende, nemlig hvordan man er i stand til at løse de problemer, man møder her i livet.

Intelligenskvotientens ankomst

Binet fandt altså frem til at måle den mentale alder og kunne dermed også rent praktisk beregne det retarderede barns handicap som det antal år, det lå efter sin aldersnorm. Det var imidlertid den tyske psykolog Stern (Robinson, 1977), der i 1912, året efter Binets død, foreslog, at man i stedet for at opgøre overensstemmelsen eller uoverensstemmelsen mellem mental alder (intelligensalder) og kronologisk alder som en *differens*, skulle beregne en *kvotient*, den berømte *intelligenskvotient*:

$$IK = IA/KA$$

Her er intelligenskvotienten (IK) defineret, så tallet 1 står for det alderssvarende niveau, mens en IK mindre end 1 angiver en intelligens lavere og en IK større end 1 en intelligens højere end normen for en given aldersgruppe.

Da amerikanerne overtog Binets robuste omnibustest til udskillelse af dårligt begavede skoleelever, skete der både en tilbagevenden til Galtons

arvebiologiske intelligensopfattelse og til hans psykometriske idé om, at intelligens følger en normalfordeling. Derimod fulgte amerikanerne Binets brede og ikke Galtons snævre psykofysiske forsøg på operationalisering af intelligens.

Det var Terman (Terman & Merrill, 1937), der i 1916 gjorde Binet-testen amerikansk i form af Stanford-Binet-testen. Han inkorporerede begejstret Sterns idé om en intelligenskvote, der nu blev multipliceret med 100, så man (efter afrunding) fik et fint heltal.

I al sin matematiske enkelthed har IK dog den indbyggede begrænsning, at Binets test jo var designet til at finde de retarderede børn, hvilket fint lader sig gøre med prøver op til fx 15-årsalderen.

Når man som de mere socialdarwinistiske amerikanere lægger større vægt på de velbegavede, får man imidlertid nogle alvorlige problemer. Man kan ikke bestemme en mentalalder præcist, når den overstiger fx 14 år. To fejlbesvarede opgaver på 14-årstrinnet kan ikke kompenseres ved en efterfølgende rigtigt besvaret opgave. Desuden bliver det et akut problem at definere mentalalder for de meget velbegavede. Hvis vi forestiller os en intelligens på 150, så skulle en 14-årig have en mentalalder på 21, men hvad vil det sige?

Da amerikanerne udviklede voksenprøver, gjorde de derfor det, at de forlod Binets begreb om mental- eller intelligensalder, men beholdt begrebet IK som en nærmest indholdstømt rest.

Da Wechsler (1939) moderniserede Stanford-Binet i trediverne, forlod han den nøje aldersmæssigt opdelte, men funktionsmæssigt omnibusprægede opbygning af prøverne til fordel for kun tre aldersopdelte intelligensbatterier, hver sammensat af to funktionsbestemte delprøver, en verbal og en manuel/(performance)-mæssig:

- Wechsler Pre-school and Primary Scale of Intelligence (WPPSI) – 3-7 år
- Wechsler Intelligence Scale for Children (WISC) – 7-16 years
- Wechsler Adult Intelligence Scale (WAIS) 16 år og opefter.

Hos Wechsler er IK-begrebet en ren fetich – en magisk størrelse, som reelt er ensbetydende med en normrelateret intelligensopfattelse. Proceduren er nemlig den, at man i hver delprøve finder testpersonens score, dvs. antallet af rigtigt løste opgaver, og så oversætter man denne score til en såkaldt percentil, dvs. personens placering i scorefordelingen for vedkommendes aldersgruppe. Har en person således i en delprøve med 29 opgaver

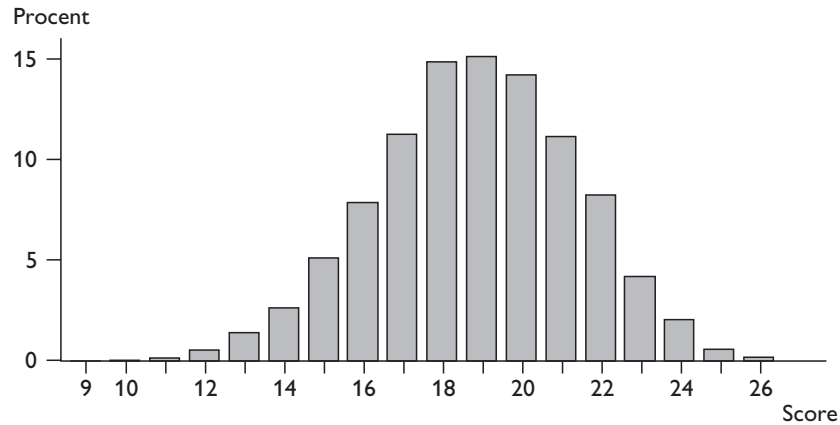
Tabel 4.2. Et konstrueret eksempel på en scorefordeling i en intelligensprøve.

Score	Antal	Procent	Kommuleret procent
8	0	0	0
9	1	,0	,0
10	6	,1	,1
11	17	,2	,2
12	55	,6	,8
13	144	1,4	2,2
14	266	2,7	4,9
15	515	5,2	10,0
16	789	7,9	17,9
17	1129	11,3	29,2
18	1489	14,9	44,1
19	1515	15,2	59,3
20	1421	14,2	73,5
21	1118	11,2	84,7
22	825	8,3	92,9
23	422	4,2	97,1
24	207	2,1	99,2
25	59	,6	99,8
26	22	,2	100,0
27	0	0	100,0
Total	10.000	100,0	100,0

løst de 21, og dette svarer til, hvad kun de 16 % bedste kan klare, dvs. at 84 % klarer sig dårligere, så har vedkommende en percentil på 84. Percentilen måles altså som den procentandel, der har dårligere resultat end den pågældende person. Tabel 4.2 ■ viser hvordan.

Denne tabel er et eksempel på resultatet af standardiseringen af testen, hvor 10.000 repræsentative amerikanere i den pågældende aldersgruppe er testet.

Den første kolonne angiver scoren. Den anden viser antallet af personer med en bestemt score. I tredje kolonne er disse antal omsat til procent. Og i sidste kolonne er procenterne adderet begyndende med de laveste værdier. Det er den operation, der hedder *kumulering*. Når vi går tilbage fra



Figur 4.5. Fordelingsdiagram for scorefordeling i dominanstest

den sidste kolonne (kumuleret procent) til den første, får vi percentilen, dvs. den observation (score), der svarer til en bestemt kumuleret procent.

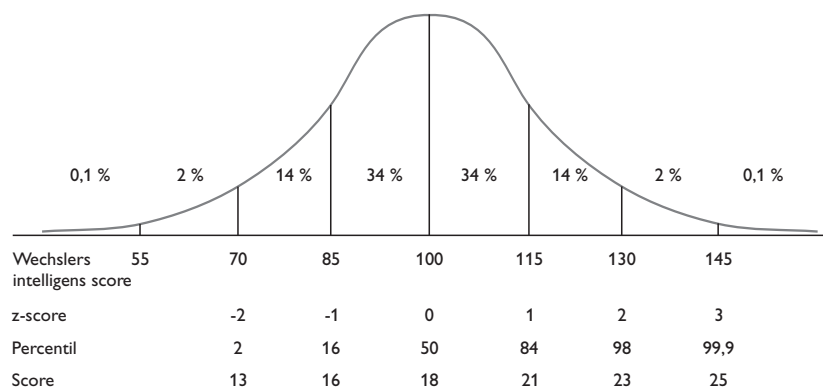
Scoren fordeler sig her i intervallet fra 9 til 26, med en midte, dvs. median eller 50-percentil omkring 18. Vi kommer nemlig til en scoreværdi omkring 18, når vi går til venstre fra de 50 % i sidste kolonne. Hvis vi interpolerer, hvad der er en lidt tvivlsom operation, fordi decimaler ikke kan forekomme ved score-værdier, får vi medianen til 18.4.

Vores testperson med scoren 21 ligger altså på 84,7-percentilen som markeret i tabel 4.2. Det betyder, at der er 84,7 procent, som har scoret til og med de 21 point. Mens der er 15,3 procent med en score bedre end dette.

Figur 4.5 viser, hvordan fordelingen ser ud i grafisk form■.

Læg mærke til, at dette diagram viser, at fordelingen er lidt skæv, nemlig venstreskæv. Den går længere til venstre fra sin midte, som er 18.4, end til højre. Til venstre ender fordelingen ved 9 og til højre ved 26. Den går således 9.4 til venstre og kun 7.6 til højre. Det er ikke meget, men med 10.000 personer i stikprøven er det absolut signifikant, dvs. at det ikke er nogen tilfældighed.

Efter denne sammenligning mellem vores testperson og normfordelingen af de 10.000 personer i standardiseringsmaterialet kan vi nu finde den eftertragtede intelligens, her kaldet IK.



Figur 4.6. Wechsler's Intelligensfordeling. Percentil og score jf. tabel 4.2.

- Testpersonens percentil, 84,7-percentilen bliver oversat til IK, ved at man på fordelingsdiagrammet figur 4.6 finder den IK-værdi, percentilen svarer til.

Den normalfordeling, som Galton var så glad for og ville bruge på såvel mennesker som hunde, er her optegnet sådan, som Wechsler fandt det passende for den første kategori (hundene holdt han sig klogeligt fra). Wechsler fastholdt Sterns IK, men som et betydningsforladt levn, der alene er et udtryk for selve scorefordelingen. Ligesom i Sterns oprindelige, men nu forladte definition, skal IK have en middelværdi på 100. I en normalfordeling, som der her er tale om, er middelværdien det samme som 50-percentilen. Og ud fra tabellen (4.2), vi så på lige før, ligger denne midte ved en score på ca. 18. Vi kan ikke angive det helt præcist, da fordelingen jo springer med et point ad gangen, da den i modsætning til normalfordelingen ikke er en kontinuert, men en diskret fordeling.

Her skal man i øvrigt ikke lade sig dupere af den flotte klokkeformede normalfordeling. Sådan har Wechsler nemlig selv lavet den. Som vi så i det konstruerede eksempel, er den faktiske scorefordeling jo slet ikke symmetrisk og derfor heller ikke normalfordelt.

Wechsler fastholdt ud over den galtonske forudsætning, at IK-fordelingen skulle være en normalfordeling, også forudsætningen fra Sterns oprindelige IK-definition, at middelværdien og også medianen skulle være 100. Desuden bestemte han, at spredningen eller standardafvigelsen skulle være

15. Standardafvigelsen i en normalfordeling angiver fordelings to vendepunkter, der svarer til hhv. 16- og 84-percentilen. Det betyder, at hovedparten af testpersonerne, nemlig ca. 2/3 ligger mellem disse vendepunkter, som dækker IK-intervallet fra 85 til 115. Med to gange spredningen, altså 30 til begge sider for midten, får vi et IK-interval fra 70 til 130, som indeholder 95 % af personerne. Det svarer til scoreintervallet fra 13 til 23. Går vi endelig videre til tre gange spredningen, altså 45, så får vi et IK-interval fra 55 til 145, og her har vi så 99,9 % af alle personer. Kun en forsvindende minoritet på en promille ligger udenfor, og er altså ud fra almindelige intelligensgrænser enten klart intelligens-handicappede eller usædvanligt højt begavede.

Da selve designet af IK-skalaen er helt vilkårligt, er den mest relevante skala for så vidt den såkaldte z-score, der bygger på den enkleste form for normalfordeling, det vil sige den standardiserede normalfordeling, som har middelværdien 0 og spredningen 1. Z-scoren er angivet umiddelbart under Wechslers IK-kurve i figur 4.6.

Vi kommer nu til årsagen til, at spredningen i Wechslers IK-fordeling definitorisk er fastsat til 15. Sagen er nemlig, at Wechsler måtte sikre sig mod negative IK-værdier. Og nu er det sådan, at hvis vi efter at have set på intervaller omkring midten på en, to og tre gange spredningen ser på fem gange spredningen, så vil der kun være under 1 milliontedel af personerne, der befinder sig uden for. Det vil sige, at intervallet fra $100 - 75 = 25$ op til $100 + 75 = 175$ ikke bare indeholder hovedparten af personerne, men det vil også være de færreste psykologer, der i deres karriere kommer ud for nogen med en IK under 25 eller over 175. Og tager vi et skridt til og siger seks gange spredningen, så får vi IK-intervallet fra 10 til 190, og så er chancen for at finde en person, som ligger uden for, 1 til 500 millioner. Testpsykologen kan altså være forholdsvis sikker på at undgå et sådant chok ved fx at finde en person, der ligger så meget lavere end scoren svarende til en IK på 10, at vedkommendes IK kunne gå hen og blive negativ.

Hvis normalfordelingens antagelser er korrekte, vil man med den her opgivne spredning på 15 altså være temmelig sikker på at undgå overraskelser som IK under 10 eller over 190. Sådanne resultater vil være så usandsynlige, at de svarer til 1 til 1.000.000.000.

Vi har her fulgt udviklingen af test efter den psykometriske linje, der går fra Binet til Wechsler. Det er en linje, der er orienteret mod anvendt

psykologi, og som ikke er synderlig optaget af matematiske principper. Vi indledte imidlertid med en helt anden linje i intelligensforskningen, nemlig den, der begyndte med Galton. Det er denne anden linje, vi nu skal følge, for det er den, der fører videre til faktoranalysen.

Faktoranalyse

Spearman og G-faktoren

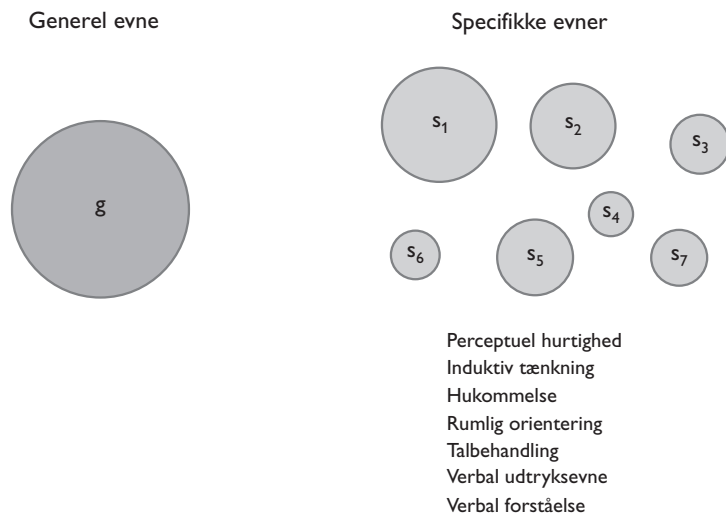
Spearman fulgte på mange måder i Galtons fodspor og blev den første betydelige psykometriker. Da han begyndte at indsamle resultater fra de intelligensprøver, som efter Binets eksempel blevet udviklet i begyndelsen af 1900-tallet, fandt han ud af, at de gennemgående havde en ret høj indbyrdes korrelation. Når det var intelligensprøver foretaget på skolebørn, viste de en god sammenhæng med børnenes karakterer i skolen. Herudfra udviklede Spearman (1927) sin berømte tofaktorteori for intelligens. Faktisk er der kun tale om en enfaktorteori, for der er kun én generel og afgørende faktor i denne teori, og den kaldte han netop for G-faktoren, idet G står for *generel*.

En faktor skal her forstås som en hypotetisk skala, hvorpå individerne fordeles. I dette tilfælde er der tale om intelligensskalaer, eller dimensioner. Ifølge Spearmans teori er der i alle intelligensprøver to typer af faktorer. Den første type udgøres af G-faktoren, dvs. den universelle evne til at løse problemer, som ikke er knyttet til problemets særlige art. Den anden type er specifikke faktorer, der er bundet til problemets særlige substans eller udformning.

I illustrationen (figur 4.7) opereres der med S-faktorer for følgende områder: Verbal forståelse, Verbal udtryksevne, Talbehandling, Rumlig orientring, Hukommelse, Induktiv tænkning, Perceptuel hurtighed.

Spearman opfandt i sit arbejde med tofaktorteorien den faktoranalytiske metode, der har følgende trin:

1. Man udregner korrelationerne mellem alle prøverne indbyrdes og opstiller dem i en kvadratisk tabel (matrix).
2. Man leder efter et mindre antal hypotetiske faktorer, der kan forklare disse korrelationer (en faktor er en sammenvejning af resultaterne for de enkelte prøver).



Figur 4.7. Spearmans tofaktorteori.

3. Man udregner vægtene (ladningerne) for faktorerne i hver prøve.
4. Man udregner for hver faktor den såkaldte *sande score* for alle individerne, dvs. individernes placering på faktordimensionerne.

Modellen i G-faktorteorien er nu, at enhver intelligenstag, T, har to bestanddele, dels prøven, som har med den *generelle* intelligens at gøre – G-faktoren – dels alt det i prøven, som i og for sig er uinteressant, fordi det alene karakteriserer den *specifikke* udformning af prøven. Denne prøvespecifikke faktor kaldes S-faktoren. Spearman antager nu, at resultatet af personen p's testresultat i testen T er:

$$\text{Score} = (\text{personens placering på G-skalaen}) \times (\text{testens ladning af G}) + (\text{personens placering på S-skalaen}) \times (\text{testens ladning af S}) =$$

$$F_G \times g_T + F_S \times s_T$$

Det skal her tilføjes, at i den faktoranalyse, Spearman grundlagde, er det af matematiske bekvemmelighedsgrunde skik at erstatte råscoren, som jo er et pointtal, med en transformeret størrelse, hvor middelværdien er 0 og spredningen 1.

I denne formel er F_G altså personens placering på G-skalaen. F_G udgør en normalfordelt tilfældighedsvariabel, som her kaldes en faktor. Denne faktor er altså netop G-faktoren.

Tilsvarende er F_S personens placering på S-skalaen, som er en anden normalfordelt tilfældighedsvariabel. Denne faktor, S-faktoren, angiver personens standpunkt i den pågældende prøve.

G-skalaen er altså den egentlige intelligensskala, mens S-skalaen er et udtryk for den særlige karakter af den enkelte prøve. Hvis fx prøven går ud på at manipulere med geometriske figurer, vil S her være en skala, som vedrører denne færdighed, rensat for generel intelligens. Vi kan således have en autist som ligger meget højt på en sådan skala, men lavt på G-skalaen.

Spearman antager som sagt, at personers evner tenderer mod at være normalfordelt på såvel G- som S-skalaerne, der ligesom den transformerede G-score, skal være standardiserede, dvs. have middelværdien 0 og spredningen 1, som vi så det med z-score-skalaen i figur 4.6. Han antager også, at der ikke er nogen som helst sammenhæng mellem G- og S-skalaerne. Det ligger for så vidt i påstanden om, at S-skalaerne er helt specifikke for hver prøve. Og endelig antager han, at de forskellige S-skalaer er uden indbyrdes sammenhæng. Der skal med andre ord være stokastisk (eller statistisk) uafhængighed mellem alle disse skalaer. Disse antagelser er afgørende for den berømte formel, Spearmans tetradiske formel, som vi nu skal se.

Hvis vi i et intelligensbatteri har fire forskellige test T_1 , T_2 , T_3 og T_4 , som har G-ladningerne g_1 , g_2 , g_3 , g_4 , så kan man vise, at vi får følgende korrelationskoefficienter, der traditionelt betegnes med det græske bogstav, ρ .

Tabel 4.3. Spearman tetradiske formel

$$\rho(T_1, T_2) = g_1 \times g_2$$

$$\rho(T_1, T_3) = g_1 \times g_3$$

$$\rho(T_4, T_2) = g_4 \times g_2$$

$$\rho(T_4, T_3) = g_4 \times g_3$$

Disse ligninger udgør Spearman tetradiske³ formel, som med et begrænset kendskab til matematisk statistik kan bevises ud fra hans forudsætninger.⁴

Ud fra Tabel 4.3 ■ kan vi vise følgende to ligninger:

$$\rho(T1,T2) / \rho(T1,T3) = g2 / g3 \text{ (nr. 2 divideret med nr. 1)}$$

$$\rho(T4,T2) / \rho(T4,T3) = g2 / g3 \text{ (nr. 3 divideret med nr. 4)}$$

Når vi nu kombinerer disse to ligninger gennem division får vi ligningen:

$$\rho(T1,T2) / \rho(T1,T3) = \rho(T4,T2) / \rho(T4,T3)$$

Spearman's G-faktor-model er således empirisk afprøvelig, idet vi kan undersøge, om denne ligning holder empirisk. Oven i købet på en særdeles enkel måde. Dette er en betydningsfuld og sjælden kvalitet ved G-faktor-teorien. En både smuk og psykologisk interessant teori. Den har imidlertid – sine smukke egenskabe, at den ikke passer på faktiske intelligensprøver.

Faktorerne begynder at yngle

Man var derfor efter Spearman nødt til at operere med nogle faktorer mellem G-faktoren og de helt specifikke faktorer. Disse mellemliggende størrelser kaldte man *gruppefaktorer*. På et tidspunkt viste dette redningsforsøg viste sig uhåndterligt, og de fleste afskaffede G-faktoren, som blev erstattet af gruppefaktorerne alene. Det kan fx være verbal og matematisk-logisk intelligens.

Hermed er vi fremme ved den faktoranalytiske model, som er en modifikation af G-faktorligningen ovenfor. Lad os antage, at der netop er to faktorer F1 og F2, som jo kunne være de netop nævnte former for intelligens, verbal og matematisk-logisk (P er personen; T er en delprøve):

Score for P i T =

$$(P's \text{ placering på F1-skalaen}) \times (T's \text{ ladning af F1}) + \\ (P's \text{ placering på F2-skalaen}) \times (T's \text{ ladning af F2})$$

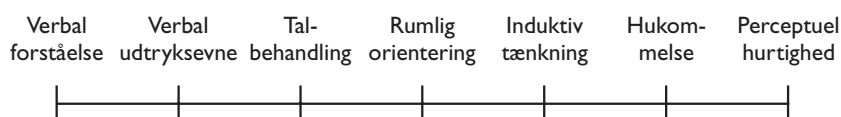
Ladningerne er talmæssige udtryk for, hvor stor del af den givne test der er bestemt af de til ladningerne svarende faktorer. Og hvad er så en faktor? Teoretisk er det en personskala knyttet til en bestemt egenskab, her altså en særlig slags intelligens. Empirisk er det en operationalisering af denne teoretiske størrelse, som er et vægtet gennemsnit af de test, som indgår i en faktoranalyse.

En illustration af Spearman's faktoranalytiske metode kan findes i

følgende empiriske materiale fra 220 irske skoleelever (Lawley & Maxwell, 1971) med karakterer i seks fag. Tre sprogligt-humanistisk orienterede fag og tre matematiske fag.

Da jeg ikke er besiddelse af originalmaterialet, har jeg rekonstrueret fordelingerne og omsat dem til den danske 13-skala (tabel 4.4)■.

Blandt de første 25 elever har jeg med lysegrå markeret to, der er bedst til sprogligt orienterede fag, mens to, der er bedst til de matematiske fag, er markeret med mørkegrå.



Figur 4.8. Multi-faktorteori. (Der er ingen henvisning til denne figur i teksten■)

Den hypotese, som her er eksemplificeret, går ud på, at der er to forskellige dygtighedsfaktorer, den ene omfattende sprogligt orienterede fag, den anden matematisk orienterede fag. Vi kan afprøve denne hypotese ved at udregne korrelationsmatricen, dvs. den tabel, hvor fagenes indbyrdes sammenhæng er opstillet (tabel 4.5).

Læg mærke til, at fagene åbenbart falder i hovedgrupper: de sprogligt-humanistiske fag gælisk, engelsk, historie, og de matematiske fag aritmetik, algebra og geometri. Her er der rimelige korrelationer indbyrdes, det mørkegrå felt inden for det sproglige område, og det mellemgrå felt inden for det matematiske område, mens korrelationerne generelt er lavere på tværs af gruppegrænserne, det lysegrå felt.

Den mest almindelige faktoranalyse, *principal component*-metoden, giver nu det resultat, som fremgår af tabel 4.6.

Her ser den første faktor ud til at kunne kandidere som G-faktoren, mens faktor 2 opdeler i ikke-matematiske og matematiske fag. Faktor 2 er en såkaldt bipolar faktor, som i øvrigt er orienteret mod *ikke-matematik*.

I sig selv er faktorerne jo beregnet helt i blinde, så det er helt tilfældigt, hvad der bliver positivt, og hvad der bliver negativt på en bipolar skala. Der er ingen indholdsmæssig tolkning i faktoranalysen; en indholdsbestemmelse er overladt til os, der betragter resultaterne.

Men hvordan bliver faktorerne overhovedet konstrueret?

Tabel 4.4. Karakterfordeling i tre sprogligt-humanistiske og tre matematisk prægede fag.

Individ nr.	Gælisk	Engelsk	Historie	Aritmetik	Algebra	Geometri
1	7	8	6	10	7	6
2	5	8	5	7	8	8
3	9	13	7	10	11	10
4	10	11	3	11	10	10
5	10	10	9	10	10	9
6	7	10	11	8	10	10
7	7	7	5	8	7	6
8	7	10	8	7	8	8
9	9	5	7	6	8	8
10	8	8	8	10	11	11
11	7	10	5	7	11	13
12	8	3	6	10	8	7
13	10	11	9	11	11	11
14	10	9	8	10	9	9
15	8	10	10	10	8	8
16	8	8	8	7	5	5
17	8	7	6	9	9	9
18	8	9	5	13	10	9
19	11	8	10	10	13	13
20	7	9	9	9	10	9
21	11	10	9	10	10	10
22	10	10	8	9	9	10
23	5	6	3	7	9	10
24	7	6	8	8	5	6
25	9	9	13	9	8	8

Computeren bruger en bestemt optimaliseringsmetode til at finde den vægtede kombination af alle variablene, her de seks skolefag, som giver den bedste redegørelse for resultaterne.

Da vi har seks variable eller skalaer kan vi højst få seks faktorer, og de udregnes af faktoranalyseprogrammet efter faldende forklaringsværdi, som det ses af tabel 4.7.

Vi ser her, hvor stor en andel af variationen, variansen, de enkelte

Tabel 4.5. Korrelationer mellem fagene.

	Gælisk	Engelsk	Historie	Aritmetik	Algebra	Geometri
Gælisk	*	0,439	0,410	0,228	0,329	0,248
Engelsk		*	0,351	0,354	0,320	0,329
Historie			*	0,164	0,190	0,181
Aritmetik				*	0,595	0,470
Algebra					*	0,464
Geometri						*

Tabel 4.6. Fagenes faktorladninger.

	Faktorlading for F1	Faktorlading for F2
Gælisk	0,553	0,429
Engelsk	0,568	0,288
Historie	0,392	0,450
Aritmetik	0,740	-0,273
Algebra	0,724	-0,211
Geometri	0,595	-0,132

faktorer forklarer, og de to første forklarer hen ved to tredjedele af hele variationen. Den første faktor eller komponent, som den hedder i denne metode, forklarer 43 % af variansen og den næste 21 %. Da der er seks variable, og dermed seks mulige faktorer, må de hver i gennemsnit forklare $1/6 = 16,7$ % af variansen. De to første forklarer altså mere end gennemsnittet, og de næste fire forklarer mindre. Et mål for en faktors forklaringsværdi er derfor at udregne, hvor meget den ligger over eller under gennemsnittet. Denne størrelse hedder faktorens *eigenvalue* – et begreb, der, som navnet antyder, stammer fra en tid, da videnskaben, herunder matematikken, var tysk domineret. Når den første faktor har en eigenvalue på omkring 2,5, betyder det, at dens forklaringsværdi på godt 42 % er to en halv gange så stor som gennemsnittet på 16,7 %.

Disse to førende faktorer, som vi jo endnu ikke kender, kalder vi

Tabel 4.7. Faktorernes forklaringsværdi.

Total Variance Explained			
Component	Indledende „Eigenvalues“		
	Total	% af varians	Kummuleret %
1	2,56	43	43
2	1,28	21	64
3	0,79	13	77
4	0,58	10	87
5	0,45	7	94
6	0,34	6	100

Extraction Method: Principal Component Analysis.

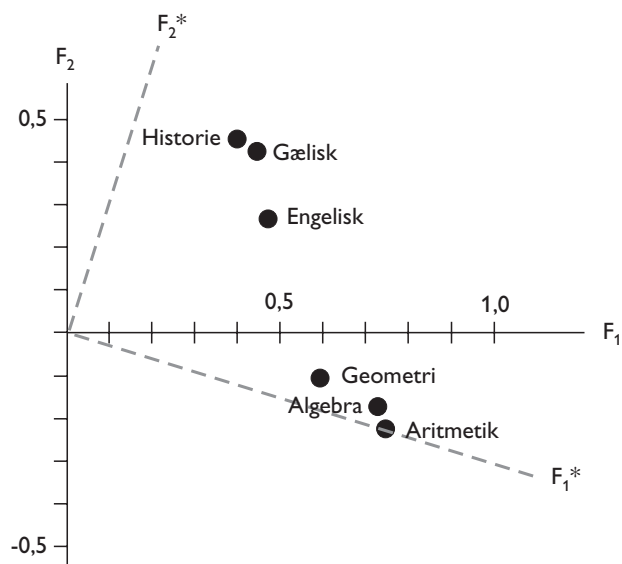
simpelthen F_1 og F_2 , og hvis vi opstiller et koordinatsystem med disse to faktorer, ser billedet ud som på figur 4.9. ■

Når vi her foretager en rotation af det ret tilfældigt orienterede koordinatsystem, vi har lagt ud med, kan vi få vores seks fag beskrevet på en mere forståelig måde.

Det vil sige, at vi udfører en drejning af akserne, så de tre sprogligt orienterede fag ligger langs den ene akse og de tre matematisk orienterede langs den anden akse. Faktoranalysen er principielt den samme, uanset hvordan man orienterer akserne, men undertiden giver det indholdsmæssigt mere mening, hvis man roterer. Den her angivne rotation er valgt ud fra det hensyn, at der skal være rene prøver, altså prøver som overvejende er ladet med kun én faktor.

Efter en sådan rotation bliver faktorladningerne for de ny faktorer som angivet i tabel 4.8.

Her er det nærliggende at opfatte F_1^* som faktor for matematisk begavelse og F_2^* som udtryk for verbal intelligens, Spearmans teori var som tidligere nævnt både psykologisk tiltalende og empirisk afprøvelig efter rent popperske principper.⁵ Desværre viste den sig altså at være forkert. Problemet var, at der optrådte en tredje slags faktorer, der hverken var generelle eller specifikke, men nærmest midt imellem. De såkaldte gruppefaktorer, som måske godt kunne svare til de syv nævnte områder, men



Figur 4.9. Fagene tegnet i et faktordiagram.

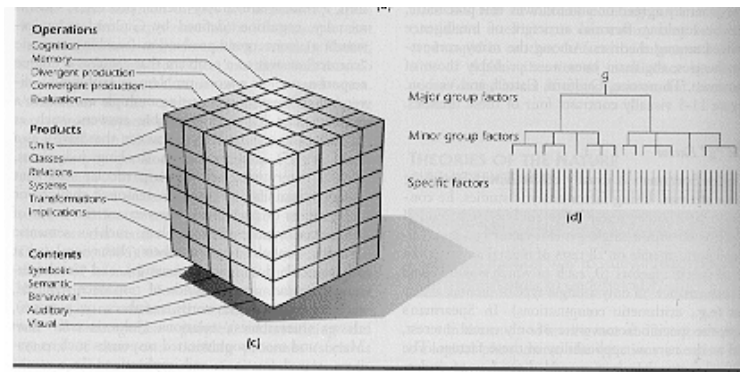
som altså ikke var specifikke for den enkelte prøve, men fælles for alle prøver inden for et givet område.

Efterhånden ynglede faktorerne, og der opstod et utal af faktoranalytiske intelligensmodeller. Dette billede, hvor snart sagt hver enkelt intelligensforsker havde sin egen faktoranalyseteori, skabte i lang tid en betydelig skepsis i den psykologiske verden over for såvel den faktoranalytiske metode som de faktoranalytiske teorier, der kom ud af den.

Denne skepsis har været ved, indtil der for nogle år siden opstod en

Tabel 4.8. Fagenes roterede ladninger.

	Faktorlading for F_1^*	Faktorlading for F_2^*
Gælisk	0,369	0,594
Engelsk	0,433	0,467
Historie	0,211	0,558
Aritmetik	0,789	0,001
Algebra	0,752	0,054
Geometri	0,604	0,083



Figur 4.10. Thurstones model – faktoranalysens Keopsyramide. (ulæselig)

hypotetisk-deduktiv form for faktoranalyse, som muliggør en stringent afprøvning af en på forhånd opstillet teoretisk model. Det er den såkaldte konfirmatoriske faktoranalyse, som er omtalt i det sidste afsnit af dette kapitel. Her vender vi for så vidt tilbage til Spearmans form for modelbrug, men altså med mere end en enkelt faktor.

Det er ikke i denne bog opgaven at gennemgå intelligensforskningen i sin noget oppustede helhed, og slet ikke dens mange vildveje. Her skal blot fremhæves et par markante træk.

En vej har været at dække intelligensområdet ved at have rigeligt mange faktorer, og her betegner Thurstone (1938) et næsten skræmmende højdepunkt med modellen, som vises i figur 4.10. I denne model indgår der over hundrede faktorer, som Thurstone da også har været nødt til at organisere i en hierarkisk model med tre niveauer:

- A. Dimension: indhold, operation, produkt
- B. Kategori inden for dimension:
 - 1. Indhold (symbolsk, semantisk, adfærdsmæssig, auditiv, visuel)
 - 2. Operation (kognitiv, hukommelsesmæssig, divergent og konvergent problemsløsning, evaluering)
 - 3. Produkt (enhed, klasse, relation, transformation, implikation).
- C. Specifik faktor, fx symbolsk mht. indhold, kognitiv mht. operation, enhed mht. produkt

Dette giver mindst 150 kombinationer, som Thurstone da også forestiller sig inddelt efter hoved, undergruppe og specifik faktor.

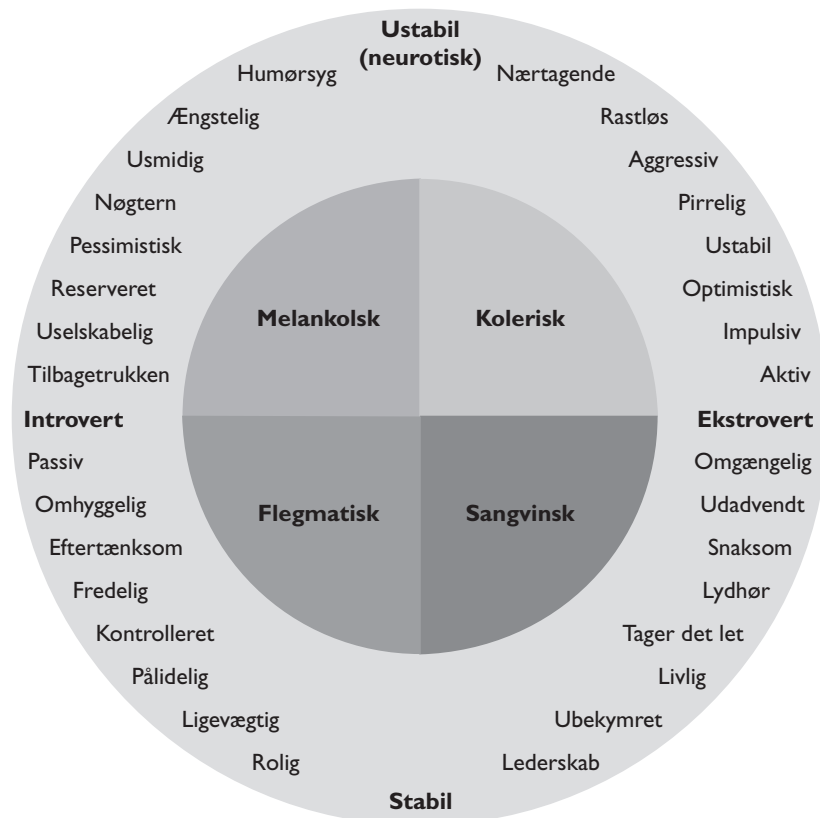
En af Spearmans elever Raymond Cattell (1987) modificerede lærerme-sterens teori på en måde, der på en gang i empirisk henseende løsnede den endimensionale intelligensteoris urealistiske matematiske greb og samtidig modificerede selve intelligensopfattelsen på en teoretisk frugtbar måde. Cattell spalter G-faktoren i to bestanddele, som han kalder *krystalliseret* (eng. crystallized) og *flydende* (eng. fluid) intelligens, forkortet **Gc** og **Gf**. Herved giver han et konstruktivt bidrag til striden om intelligensens henholdsvis medfødte og erhvervede karakter. Den flydende intelligens er nemlig den spontane evne til at løse problemer, uafhængigt af allerede erhvervet viden, mens den krystalliserede intelligens er resultatet af ens hidtidige intellektuelle udvikling.

Den personlighedspsykologiske faktoranalyse

Samtidig har Cattell også været en pioner inden for den faktoranalytiske personlighedsforskning. Han fandt lige efter anden verdenskrig frem til 16 basale personlighedstræk (Cattell, 1950), hvilket resulterede i en stadig meget anvendt personlighedstest, nemlig Cattells 16 PF-test (1970), hvor PF står for *personlighedsfaktorer*. Nu er 16 forskellige faktorer en del mere, end hvad vi mennesker kan have i hovedet på en gang, så tendensen har været at forenkle denne personlighedsbeskrivelse til at omfatte et mindre antal faktorer.

Eysenck foretog noget af en genistreg ved sammenfatte de to mest kendte personlighedsteorier, nemlig Freuds og Jungs, med et prægnant træk fra psykiatrien. Freud inspirerede ham til at finde en *neuroticisme*-skala, Jung til en *extroversions-introversions*-skala, og psykiatrien til en *psykoticisme*-skala. Denne trefaktorteori afprøvede han yderligere empirisk (Eysenck, 1947, 1985). Og som glasuren på kagen fik han de to psykologiske dimensioner – neuroticisme og extroversion – til at passe med antikkens lære om de fire temperamentter (se figur 4.11). Og så er der her tale om et meget populært antal dimensioner, nemlig tre, det er lige til at huske og heller ikke helt plat.

Netop i de senere år er der sket en interessant nyudvikling inden for den faktoranalytiske personlighedsforskning, idet to forskere McCrea og Costa (1990) har udbygget Eysencks teori til en slags de luxe-model, der stadigvæk har et begrænset antal personlighedsfaktorer, nemlig fem, som ifølge dem dækker alle personlighedstest af betydning. De har bevaret Eysencks to psykologiske dimensioner Neuroticisme (N) og Extroversion (E) og sup-



Figur 4.11.

pleret med følgende tre: Åbenhed for nye erfaringer (O for "openness"), Venlighed (A for "agreeableness") og Samvittighedsfuldhed (C for "conscientiousness"). Denne model har fået kælenavnet, *The Big Five* eller *Femfaktormodellen*, og ud fra de tre første dimensioner kaldes testen NEO-PI, NEO-Personality Inventory og i seneste udgave NEOPI-R, NEO-Personality Inventory, Revised.

Vi skal imidlertid nu forlade faktoranalysen og først se på en anden psykometrisk metode, itemanalysen.

Itemanalyse og itemresponsmodeller

Rasch og itemanalysen

I den danske psykologis mere end hundredårige historie har der kun været to internationale notabiliteter. Den ene er Edgar Rubin, kronologisk nummer to i rækken af professorer i universitetsfaget psykologi; han huskes i dag bedst for opdagelsen af *figur-grund-forholdet* ved visuel perception (Rubin, 1967). Den anden er slet ikke psykolog, men derimod psykometriker, og på dette område en af de allermest betydelige. Det drejer sig om statistikeren Georg Rasch. Han koncentrerede sig ligesom Spearman om måling af intelligens og færdigheder.

Rasch blev omkring 1950 statistisk makker for testpsykologen Børge Prien i forbindelse med nogle psykometriske projekter inden for militær- og skolepsykologi. Det drejede sig om hhv. udviklingen af en intelligensprøve for rekrutter og om standspunktsprøver for elementærfagene i skolen. Det første projekt førte til udviklingen af forsvarets intelligensprøve, BPP, som nu gennem mere end et halvt århundrede har været brugt til at sortere ungersvende på session. Det andet projekt førte til en lang række pædagogiske prøver udviklet af Danmarks Pædagogiske Institut (DPI), som blev oprettet i 1954. (For en gennemgang af Rasch's bidrag se Olsen, 2003).

Når vi ser på et videnskabeligt begreb, fx intelligens eller regnefærdighed, er der to måder, vi kan forsøge at bestemme begrebet på. Vi kan foretage en teoretisk bestemmelse af begrebet, og hvis det teoretiske grundlag er godt nok, kan vi gå i gang med en empirisk udforskning af det fænomen, som begrebet refererer til.

Vi kan også gribe problemet an på en måde, der giver afkald på en teoretisk forståelse og blot fortæller, hvordan vi skal forholde os rent empirisk. Altså hvilke empiriske operationer vi skal anvende for at beskrive et foreliggende fænomen. Denne sidste fremgangsmåde kaldes derfor *operationel*, og den videnskabsteoretiske og metodologiske retning, der bekender sig til denne fremgangsmåde som et universelt princip, kaldes *operationalismen*. Et typisk eksempel på operationalisme er den påstand, at intelligens blot er det, vi måler med en intelligensprøve, som psykologen E.G. Boring udtrykte det i en populær artikel (Boring, 1923:35).

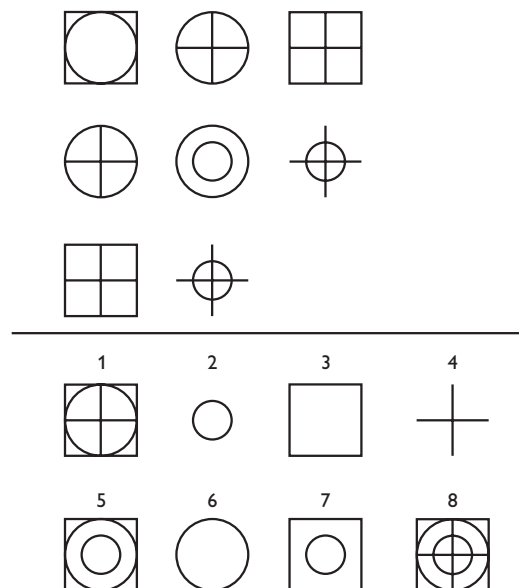
Det enestående ved Rasch er, at han på en måde vælger en tredje vej – den, man måske kunne kalde den metodologiske analyse. Han hævder

ikke, at vi skal gå ud fra et begreb om, hvad det er, vi observerer, og heller ikke, at vi skal nøjes med at opstille en bestemt operationel definition.

Det spørgsmål, han stiller, er:

- Under hvilke betingelser kan et teoretisk begreb på konsistent vis modsvarer en bestemt operation?

Lad os gå tilbage til intelligensprøverne og se på Raven-prøven, der anses for at være noget af det nærmeste, man kan komme en måling af G-faktoren (Figur 4.12).■



Figur 4.12. Et item fra Ravens progressive matrices

Raven-prøven (se Raven, 1998) er opbygget med geometriske figurer, der er opstillet i en tre gange tre matrix, hvor der mangler en figur i nederste højre hjørne. Under stregen er der nu en række figurer, man kan vælge til den tomme plads. En opgave fra Raven-testen er vist i figur 4.12. Når prøven kaldes progressiv, er det fordi den ligesom Binet-prøven er opbygget med stigende sværhedsgrad.

Det udsagn, at Raven-prøven måler G-faktoren, indeholder faktisk forskellige påstande:

1. En teoretisk påstand: Denne prøve måler en bestemt endimensional psykologisk egenskab.
2. En operationel påstand: Graden af denne egenskab findes ved at tælle point.

Rasch's psykometriske præstation er bl.a. at have vist, at betingelserne for påstand 1 er identiske med betingelserne for påstand 2.

Dygtigheds- og lethedsgad

Enhver samvittighedsfuld matematiker – og det var Rasch i høj grad – søger ved ethvert problem efter de nødvendige og tilstrækkelige betingelser for løsningen. Rasch fandt først en tilstrækkelig betingelse for, at vi ved en simpel scoring i en test med to svarkategorier, fx rigtigt og forkert, finder en veldefineret endimensional egenskab.

Det er en tilstrækkelig betingelse, at vi for hver testperson har et enkelt tal, som udtrykker personens placering på dimensionen, og et enkelt tal, der udtrykker den generelle tendens til, at et givet item, altså en enkelt opgave eller prøve, bliver besvaret positivt. Forudsætninger er, at disse to tal samvirker på den matematisk simpleste måde, og det gør de, hvis tendensen til, at den givne testperson løser den givne opgave, netop er produktet af disse to størrelser, som i statistikken kaldes parametre. I en færdigheds- eller intelligensprøve vil det sige, at vi skal multiplicere personens dygtighedsgrad med itemets lethedsgad.

Rent empirisk har vi ved intelligens- og færdighedsprøver at gøre med tre skalaer:

1. Personskalaen – her dygtighed
2. Itemskalaen – her lethedsgad⁶
3. Svarskalaen – her rigtighed som binært kategorisystem, ja-nej

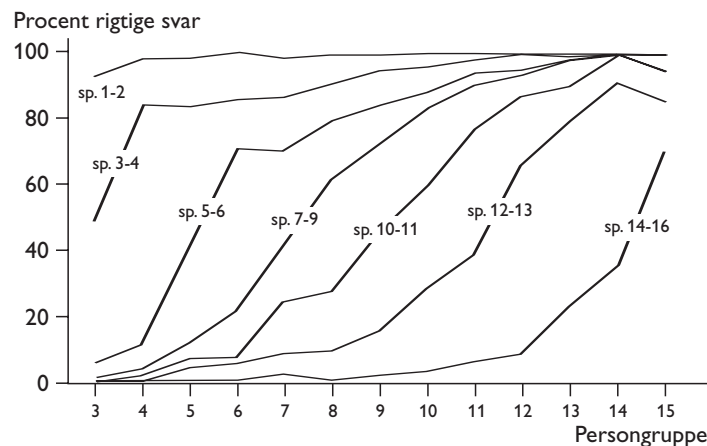
Den metode Rasch indledte med i sit arbejde med den militærpsykologiske intelligensprøve, BPP, var at opdele testpersonerne i scoregrupper, som repræsenterer den første skala, person- (her dygtighedsskalaen), og derefter udregne og optegne de såkaldte itemresponskurver, dvs. de kurver, vi finder i et koordinatsystem, hvor x-aksen angiver scoregruppen, mens y-aksen

viser løsningsprocenterne i hver af disse scoregrupper, og hvor disse forskellige løsningsprocenter er forbundet med en streg fra et punkt i en bestemt scoregruppe til punktet i næste scoregruppe. En sådan responskurve tegnes nu for hvert item (opgave),

Vi skal her se Rasch's første itemresponskurver fra 1953. Det drejer sig om den ene af de fire intelligensprøver i BPP, nemlig T-prøven, dvs. en talprøve, hvor en række tal er vist, og hvor man så skal finde frem til det næste tal i rækken. Altså en række af typen:

6, 5, 4, 3, ?

Der er 17 sådanne opgaver med stigende sværhedsgrad.



Figur 4.13. Rasch's første itemresponskurver fra BPP's T-prøve (efter Kousgaard, 2003:22).

Vi ser her (Figur 4.13)■, at der er tale om velordnede kurver,⁷ hvor løsningsprocenten vokser dels med persongruppen, altså *dygtighedsgrad*, udtrykt ved score, dels med *lethedsgaden*, idet de letteste items er til venstre og de sværeste til højre. Rasch har her, på grund af det begrænsede antal testpersoner, slået items sammen, to eller tre ad gangen for på denne måde at få mere sikre procenter. (Item 17 er udgået af analysen, fordi opgaven er så svær, at der ikke var nok, der havde kunnet løse den).

Det næste eksempel er hentet fra den personlighedspsykologiske type af test, som bl.a. omfatter erhvervspsykologiske test til fx lederudvælgelse. I

disse er *dominans* et træk, der på godt og ondt spiller en væsentlig rolle. Lad os kikke på en til dette formål skitseret dominanstest.

Test for dominans bestående af 20 spørgsmål med svarmulighederne:
Ja eller Nej.

Instruktion:

Passer nedenstående karakteristikker på din person?

De første seks items:

- Item1: Jeg kan godt lide at lede en gruppe.
- Item2: Jeg tager gerne ordet i en forsamling.
- Item3: Hvis nogen fornærmer mig, giver jeg igen med samme mønt.
- Item4: Jeg hader at være nummer to.
- Item5: Jeg har ikke let ved at tage imod ordrer.
- Item6: Jeg har let ved at komme i kontakt med andre.

Hvis vi ser på itemanalysens tre skalaer har vi her:

1. Personskalaen – personparameteren er her dominans
2. Itemskalaen – spørgsmålsparameteren er her attraktivitet⁸
3. Svarskalaen – responsdimensionen udgøres her af graden af enighed

Lad os antage, at vi har lavet en grundig konstruktion og afprøvning med 1200 repræsentative testpersoner (det er fint, og ikke usædvanligt for de seriøse konsulentfirmaer).

I et statistikprogram som SPSS får vi den simpleste repræsentation, når vi koder 0 for Nej og 1 for Ja.

Fra et simuleret datamateriale vises i tabel 4.9 et udsnit for svarene fra de første 25 personer på disse seks spørgsmål.

Det springende punkt er, om vi kan lægge de 20 nuller og ettaller for en given person sammen, og derved opnå et gyldigt mål for dominans. Ja, faktisk har vi to spørgsmål, et psykometrisk og et (personligheds)psykologisk.

Det psykometriske spørgsmål er:

- Kan vi lægge de 20 tal sammen (score)?

Tabel 4.9. 25 personers svar på dominanstest.

	Item1	Item2	Item3	Item4	Item5	Item6
Person	Jeg kan godt lide at lede en gruppe	Jeg tager gerne ordet i en forsamling	Når jeg fornærmes giver jeg igen med samme mønt	Jeg hader at være nr. 2	Jeg har ikke let ved at tage mod ordrer	Jeg har let ved at komme i kontakt med andre
1	1	1	1	0	1	0
2	1	0	0	0	1	1
3	0	0	1	0	1	1
4	1	0	0	1	1	1
5	0	1	1	0	1	0
6	0	1	0	0	1	1
7	0	0	0	1	1	0
8	0	1	0	0	1	0
9	1	1	1	1	0	0
10	0	0	1	0	0	0
11	1	1	0	1	1	1
12	1	1	0	0	1	0
13	0	1	0	0	1	0
14	1	1	0	0	1	0
15	1	1	1	1	1	1
16	1	1	1	1	1	1
17	0	1	0	0	1	1
18	1	0	0	0	1	0
19	0	0	1	1	1	1
20	0	0	0	0	0	1
21	1	1	1	0	1	1
22	0	0	1	0	1	0
23	0	1	0	0	1	1
24	1	0	0	1	1	1
25	1	1	1	0	1	1

Det (personligheds)psykologiske spørgsmål er:

- Hvis det er rimeligt at score, måler denne score så dominans?

Tabel 4.10. Scorefordeling for dominanstest.

Scoregruppe	Antal	Procent	Kumuleret %
0	1	0,1	0,1
2	11	1,1	1,2
3	11	1,1	2,3
4	32	3,2	5,5
5	45	4,5	10,0
6	58	5,8	15,8
7	87	8,7	24,5
8	88	8,8	33,3
9	98	9,8	43,1
10	108	10,8	53,9
11	108	10,8	64,7
12	102	10,2	74,9
13	85	8,5	83,4
14	76	7,6	91,0
15	37	3,7	94,7
16	32	3,2	97,9
17	14	1,4	99,3
18	5	0,5	99,8
19	2	0,2	100,0
I alt	1000	100	

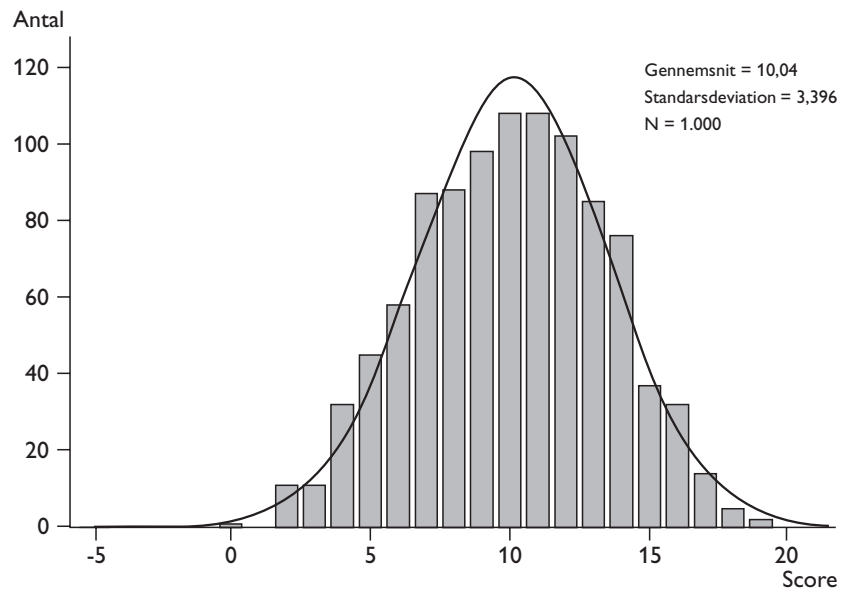
Det første spørgsmål går på, om vi måler noget som helst. Det vedrører altså det, vi kalder *reliabilitet*.

Det andet spørgsmål drejer sig om, hvorvidt det i givet fald netop er dominans, vi måler. Dette spørgsmål vedrører det, vi kalder *validitet*.

Lad os lige se på scoringsprocessen.

Da vi har sat nej til 0 og ja til 1, er det i et database- eller statistiksystem let at finde scoren som summen af de seks svar på spørgsmålene. Gør vi det, får vi følgende tabel (tabel 4.10). ■

Ligesom i tabel 4.2 er 2. kolonne antal personer med en given score, 3. kolonne procenterne, og sidste kolonne de kumulerede procenter. Læg mærke til, at der ikke er nogen personer, som har scoret 1 point. Sådan arbejder mange statistiksystemer, så man må selv se efter, om der er tomme værdier i tabellen.



Figur 4.14. Scorefordeling for dominanstest.

Tabel 4.11.

Scoregruppe	Antal	Procent	Kumuleret %
0-4	55	5,5	5,5
5-6	103	10,3	15,8
7	87	8,7	24,5
8	88	8,8	33,3
9	98	9,8	43,1
10	108	10,8	53,9
11	108	10,8	64,7
12	102	10,2	74,9
13	85	8,5	83,4
14	76	7,6	91,0
15-	90	9,0	100,0
Total	1000	100,0	

Grafisk præsenteret ses scorefordelingen i figur 4.14. ■ Vi har en klokkeformet, symmetrisk fordeling, der ikke på nogen væsentlig måde afviger fra normalfordelingen.

I praksis gør vi nu det (se tabel 4.11 ■), at vi slår de små yderkategorier sammen. Her er alle værdier fra 0 til 4 slået sammen, derefter er værdierne 5 til 6 slået sammen, værdier fra 7 til 14 har tilstrækkelige antal hver for sig, og til sidst er alle værdier fra 15 og opefter samlet.

Der er på denne måde dannet grupper på omkring 50 eller derover, og dermed er grupperne store nok til at være statistisk bæredygtige.

Itemresponskurverne

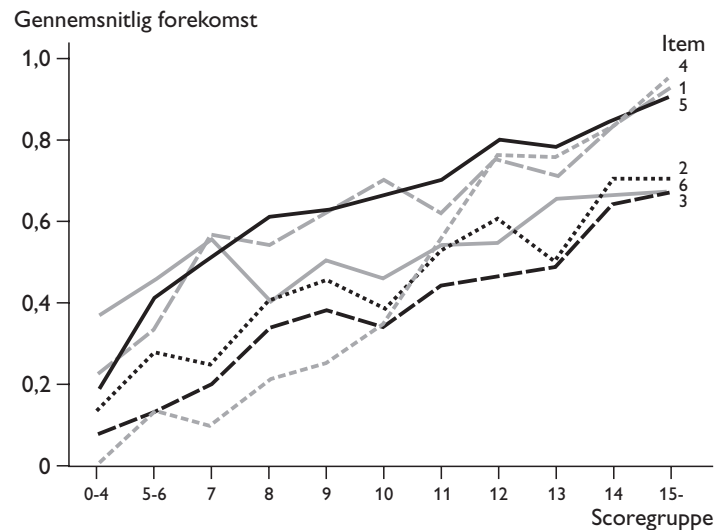
Vi skal nu gå i gang med den såkaldte itemanalyse. Spørgsmålet om den *interne reliabilitet* kan formuleres således: Hænger de 20 spørgsmål overhovedet sammen på en sådan måde, at det er meningsfuldt at score?

Alle itemanalyser går her ud fra det, som kaldes *itemresponskurverne*. Det vil sige, at vi begynder med at udregne den gennemsnitlige forekomst af positive svar.

Tabel 4.12 ■ viser, hvad vi kan kalde itemreponsprocenterne, som fremkommer ved, at man dividerer antallet af positive besvarelser i en

Tabel 4.12. Gennemsnitsrespons på de seks dominansspørgsmål i de forskellige scoregrupper.

Score-gruppe	Item 1 ledet en gruppe	Item 2 tage ordet	Item 3 give igen	Item 4 ikke nr. 2	Item 5 ikke ordrer	Item 6 let i kontakt
0-4	0,22	0,13	0,07	0,00	0,18	0,36
5-6	0,33	0,27	0,13	0,13	0,41	0,45
7	0,56	0,24	0,20	0,09	0,51	0,55
8	0,53	0,40	0,33	0,20	0,60	0,40
9	0,61	0,45	0,38	0,24	0,62	0,50
10	0,69	0,38	0,33	0,34	0,66	0,45
11	0,61	0,52	0,44	0,56	0,69	0,54
12	0,75	0,60	0,46	0,75	0,79	0,54
13	0,71	0,49	0,48	0,75	0,78	0,65
14	0,83	0,70	0,63	0,83	0,84	0,66
15-	0,92	0,70	0,67	0,94	0,90	0,67



Figur 4.15. Itemresponskurver for de seks første items.

scoregruppe med antallet af personer i denne scoregruppe. Der er her tale om decimalangivelser, fx viser øverste venstre hjørne tallet 0,22, hvilket er det samme som 22 %.

Vi overfører nu denne tabel til et diagram, som viser itemresponskurverne. Vi sætter scoregruppens scoreværdi som X-akse og bruger gruppens responsprocent som Y-værdi, sådan som det vises i figur 4.15. I dette diagram har vi alle itemanalysens tre dimensioner repræsenteret:

1. Personskalaen – her dominans – ved scoregrupperne, som er vandret akse
2. Itemskalaen – her spørgsmålet om attraktivitet – ved de forskellige kurver
3. Svarskalaen – her procentdelen af de personer i scoregruppen, der siger ja til et bestemt udsagn

Kurverne for de første seks items er her tegnet op (de øvrige er ikke medtaget, da de ville gøre diagrammet for uoverskueligt). Vi ser, at de seks kurver stort set ligger ordnet i lag. De mest attraktive items (som giver de stejleste kurver), ligger hele tiden over de mindre attraktive. Item nummer 5 ligger i toppen, og item nummer 3 ligger i bunden.

Tabel 4.13.

Score-gruppe	Item1 ledede en gruppe	Item2 tage ordet	Item3 give igen	Item4 ikke nr. 2	Item5 ikke ordrer	Item6 let i kontakt
0-4	0,22	0,13	0,07	0,00	0,18	0,36
5-6	0,33	0,27	0,13	0,13	0,41	0,45
7	0,56	0,24	0,20	0,09	0,51	0,55
8	0,53	0,40	0,33	0,20	0,60	0,40
9	0,61	0,45	0,38	0,24	0,62	0,50
10	0,69	0,38	0,33	0,34	0,66	0,45
11	0,61	0,52	0,44	0,56	0,69	0,54
12	0,75	0,60	0,46	0,75	0,79	0,54
13	0,71	0,49	0,48	0,75	0,78	0,65
14	0,83	0,70	0,63	0,83	0,84	0,66
15-	0,92	0,70	0,67	0,94	0,90	0,67
Total	0,62	0,45	0,38	0,45	0,65	0,53
Kommentar			Mindst attraktive: total = 0,38	Mest stejle: stign. = 0,94	Mest attraktive: total = 0,65	Mest flade: stign. = 0,31

Der er imidlertid to items, der afviger fra dette mønster, nummer 4 og nummer 6.

Nummer 4, der passende nok hedder ”Jeg hader at være nummer 2”, er en langsom starter, men så sætter den spurten ind, og det vinder faktisk på de sidste meter. For nummer 6: ”Jeg har let ved at komme i kontakt med andre”, er det lige omvendt. Dette item starter allerhøjest, men ender nede i bunden. Endnu mere forskelligt må vi vurdere disse to spørgsmål, hvis vi ser på stigningstakten. Nummer 4 har en meget høj stigningstakt, mens nummer 6 har en meget lav stigningstakt.

Denne stigningstakt dækker over et begreb, der i psykometrien kaldes itemets *diskriminationsgrad*. Det vil sige, i hvilken grad et item skiller færene fra bukkene, altså adskiller dem med lav dominans fra dem med høj dominans.

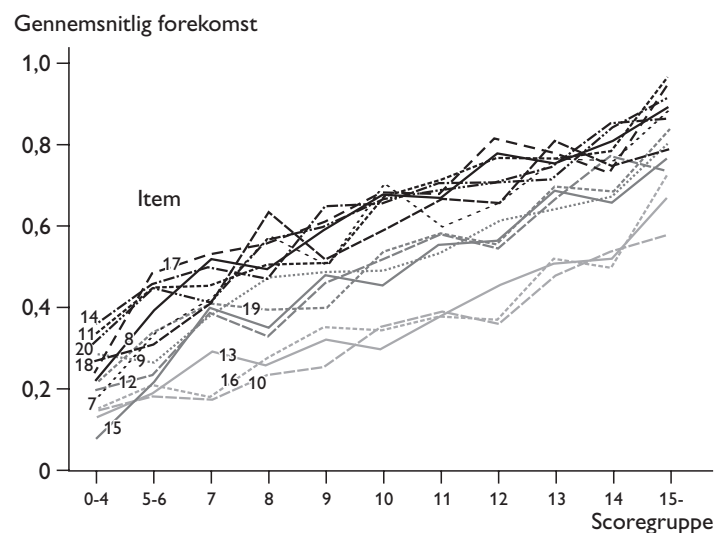
Vil der i dette eksempel være nogen psykologisk forklaring på denne forskel mellem de to items? Ja, item nummer 4 er efter alt at dømme dårligt

formuleret. Det handler nemlig ikke særligt meget om dominans, men snarere om kommunikation og kontakt. En faktoranalytiker ville sige, at det er en helt anden faktor, der er på spil.

For nummer 4's vedkommende er der derimod nok tale om et meget centralt spørgsmål. En faktoranalytiker ville sige, at det har en meget høj ladning af dominansfaktoren. Der er jo tale om selv at bestemme.

Selv om den grafiske fremstilling med itemresponskurverne giver et godt generelt overblik, lønner det sig at se på selve itemresponsprocenterne, som giver et mere præcist billede af de forskellige items egenskaber i relation til scoregrupperne.

I tabel 4.13 ■ får vi et numerisk udtryk for det, vi så i itemresponsdiagrammet. Fire items skilte sig ud. For det første de to, der lå i top og bund med hensyn til attraktivitet, henholdsvis item nummer 5 og 3. Hvis vi ser på totalforekomsten (over alle scoregrupper), ligger nummer 3 kun på 38 procent positive svar, mens nummer 5 har 65 procent positive svar. For det andet var der de to, som afveg men hensyn til stigningstakt eller stejlnhed, altså det, der kaldes diskrimination. Det var nummer 4, hvis stigningstakt kan illustreres ved at se stigningen fra laveste til højeste scoregruppe, hvilket giver 94 procent, mens nummer 6 kun har en stigning på 31 procent.



Figur 4.16. Item-responskurver for item nr. 7 til 20.

Vi har nu studeret egenskaberne for de første seks items og skal også lige se på de sidste 14.

Responskurverne (Figur 4.16) ■ ligger rigtigt pænt, alle ordnet i lag.

Endimensionalitet, konsistens og reliabilitet

Som nævnt i begyndelsen af denne gennemgang af itemanalysen, er ideen med denne analyseform simpelthen at undersøge om en række items, udsagn eller spørgsmål, er en del af en endimensional skala, dvs. at de operationaliserer en bestemt egenskab, som de testede personer kan ordnes efter. Når reliabilitetsundersøgelsen i itemanalysen falder positivt ud, har vi påvist testens konsistens, indre sammenhæng, hvilket atter indebærer, at den er reliabel/pålidelig og altså måler en bestemt veldefineret egenskab.

Lokal afhængighed

Det skal her indskydes, at der paradoksalt nok er en form for konsistens, som er for meget af det gode. Itemanalysemodellen forudsætter nemlig, at der ikke må være en for *direkte* sammenhæng mellem besvarelsen af flere forskellige items. De skal hver for sig kun afhænge af forholdet mellem person- og itemparameter. Sammenhængen mellem to items må alene bestå i, at de interagerer med de samme personer, hvor den enkelte person har en konstant personparameter. Disse respondenter må altså ikke lade sig påvirke af svaret på et tidligere spørgsmål. Hvis vi altså ved siden af udsagnet:

- Jeg kan ikke lide at være nummer 2

tilføjede spørgsmålet:

- Jeg kan godt lide at være nummer 1

ville der utvivlsomt være det, der i statistikken hedder *stokastisk afhængighed*. Svarene på de to spørgsmål ville hænge alt for godt sammen. Der ville ikke være den i itemanalysen forudsatte responsive uafhængighed mellem svarene på de enkelte items. Denne form for afhængighed hedder i itemanalysen *lokal afhængighed*. Imidlertid vil en sådan afvigelse fra den ægte itemanalysemodel ikke forhindre, at man stadig kan tale om, at spørgsmålene vedrører samme egenskab, og man vil stadig kunne foretage en scoring. Modellen er blot matematisk noget vanskeligere at håndtere,

og den opfylder ikke de idealkrav, som stilles til den ægte itemanalyse. Denne type model vil vi kalde for *generaliseret itemanalysemodel med lokal afhængighed*.

Konsistens i itemanalysen – specifik objektivitet

Rasch's krav til konsistens var meget vidtgående, således at de på den ene side er næsten umulige at opfylde i praksis, men på den anden side belyser begrebet endimensional skala på en forbilledlig klar måde. Han forestillede sig, at der til en bestemt skala ikke blot skulle være en tilfældig samling af items med tilhørende svarkategorier. Der skulle også være en veldefineret gruppe af items, som skulle være helt udskiftelige med hensyn til måling af personparameter, fx dygtighed, ligesom der skulle være en veldefineret gruppe af personer, der ligeledes var helt udskiftelige med hensyn til måling af itemparameter, fx lethed. Dette er i virkeligheden et krav, der går ud over psykometri og langt ind i psykologien. For afgrænsningen af såvel item- som af personpulje er jo en teoretisk opgave, der er relateret til selve begrebet for den givne skala. Det forhold, at en model imødekommer disse krav, kaldte Rasch for *specifik objektivitet*.

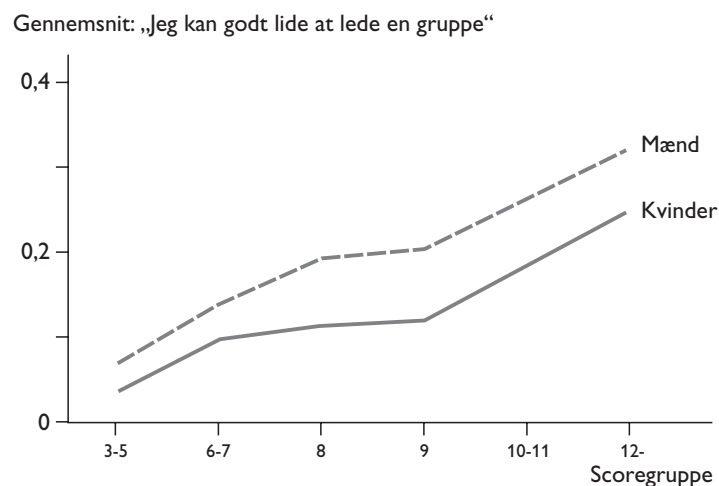
En eventuel inkonsistens eller inhomogenitet med hensyn til items vil fx vise sig i form af overkrydsende itemresponskurver som vist i figur 4.15.

Tilsvarende kan vi undersøge inhomogenitet inden for den undersøgte persongruppe ved at afprøve, om de registrerede baggrundsvARIABLE for personerne giver anledning til inkonsistens.

Itembias – differentiell itemfunktion

Vi har allerede set på den første type konsistensafprøvning i form af itemresponskurver, hvor vi især var opmærksomme på afvigelser fra konsistens i form af afvigende diskrimination.

Den anden type konsistensanalyse går på, at det enkelte item skal virke på samme måde over for alle personer i den forstand, at det skal have en bestemt lethed- eller attraktivitetsgrad, som sammen med personparametrene bestemmer svarsandsynlighederne. Det indebærer, at alle personer med samme score, når man ser bort fra tilfældigheder, skal opføre sig på *samme måde i forhold til samme item*, og det vil specielt sige, at baggrundsvARIABLE ikke må betyde noget. Der kan således godt være scoreforskelle mellem køns- eller aldersgrupper, men for samme scoregruppe (altså fx for mænd og kvinder med nogenlunde samme samlede score på testen) må disse baggrundsvARIABLE ikke have nogen indflydelse på et bestemt item.



Figur 4.17.

Dette krav til gruppemæssig homogenitet kontrollerer man ved at tegne itemresponskurver ganske som før, men nu med en opdeling af materialet efter grupper defineret ved baggrundsvARIABLE, fx køn. Lad os her tage det første item i dominanstesten "Jeg kan godt lide at lede en gruppe" (figur 4.17). ■

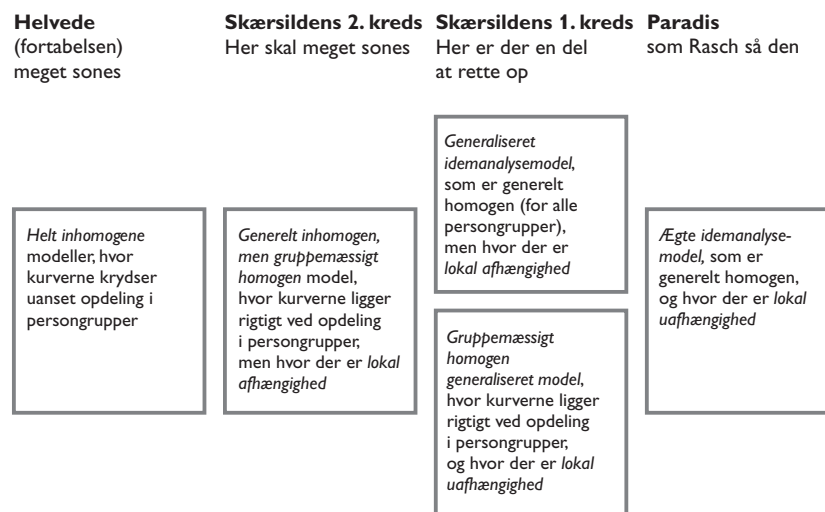
Når det drejer sig om at lede en gruppe, ser dette ud til at være mere populært hos mændene end hos kvinderne. Læg mærke til, at dette ikke har noget at gøre med, om mændenes dominansscore lå på et højere gruppeniveau end kvindernes. Forskellen mellem kønnene er nemlig på dette spørgsmål uafhængig af scoren. Den forskel, vi ser, er mere subtil. Den indebærer, at det pågældende item er relativt mere populært hos mændene end hos kvinderne – alt andet lige. For en bestemt dominansscore vil mændene således tendere mod at ligge højere på dette item.

Det givne item er altså ikke i overensstemmelse med itemanalysens strenge konsistenskrav. Kvinder og mænd opfører sig forskelligt. Lige som man ved intelligensprøver må være på vagt over for kulturel bias, har vi her noget tilsvarende. Dette item om ledelse af en gruppe ville faktisk stille kvinder dårligere end mænd, fordi kvinder på dette item scorer lavere end mænd med samme dominansgrad.

Når et sådant spørgsmål viser itembias eller differentiell itemfunktion, som man kalder dette fænomen, er itemanalysemodellen faktisk brudt sammen. Der er da to muligheder:

- Man opgiver at etablere den oprindelige skala ud fra de 20 spørgsmål ved at eliminere det spørgsmål, der giver itembias. Da vi i forvejen har spørgsmål med diskriminationsafvigelse, skal vi i øvrigt alligevel foretage en revision.
- Man opgiver at etablere en generel skala for de forskellige inhomogene grupper, her altså de to kønsgrupper, og forsøger i stedet at danne forskellige skalaer for de forskellige grupper. I dette tilfælde en mande- og en kvindeskala. Den sidste strategi vil i øvrigt stemme overens med den form for ligestillingspolitik, der bekæmper de forhold ved en test, som er præget af kønsbias.⁹

Vi kan karakterisere alle typer af modeller for items ud fra denne karakteristik, baseret på hhv. begrebet differentiell itemfunktion og lokal afhængighed, jf. figur 4.18 ■



Figur 4.18. Oversigt over itemanalysen og dens mere eller mindre vanartede slægtninge.

Helvede Længst til venstre har vi placeret itemanalysens pendant til Dantes helvede, hvor de helt uopretteligt defekte modeller er henvist. I hvert fald skæve items er uhjælpeligt fortabte (medmindre de bliver omformuleret).

Skærsildens 2. kreds Her har vi modeller, hvis defekter i princippet kan

tænkes repareret, selv om de er så alvorlige, at der foreligger to forbrydelser mod Rasch's krav til itemanalysen. For det første er de inhomogene, de har altså en differentiell itemfunktion, som kræver en gruppemæssig opdeling. For det andet har de lokal afhængighed.

Skærsildens 1. kreds Modellerne her har kun en af de to nævnte defekter og er således tæt på at være lydefri itemanalysemodeller.

Rasch's Paradis Her ender vi omsider med den rigtige og lydefri itemanalysemodel.

Ækvidistanceproblemet

Vi har nu set et eksempel fra en dominanstest, hvor svarskalaen var dikotom, men i personligheds- og holdningstest anvender man ofte en Likert-skala med fx tre eller fem kategorier, hvor man angiver, hvor enig eller uenig man er i et givet prædikat.

En Likert-skala med tre kategorier kan således være:

Enig
Hverken enig eller uenig
Uenig

Ved spørgsmål orienteret således, at enighedssvar udtrykker dominans, kan der scores:

Enig	2
Hverken enig eller uenig	1
Delvis uenig	0

Ved "omvendte" spørgsmål, altså dem, som er orienteret mod uenighed, vil man med den første kodning score:

Enig	0
Hverken enig eller uenig	1
Delvis uenig	2

En anden kodning er den bipolære, der giver visse kognitive fordele:

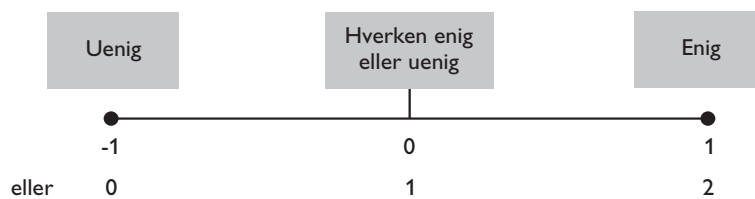
Her er først svarkategorier, hvor kodningen af spørgsmålene er rettet mod *enighed*

Enig	1
Hverken enig eller uenig	0
Delvis uenig	-1

Og her er kodningen af spørgsmålene rettet mod *uighed*:

Enig	-1
Hverken enig eller uenig	0
Delvis uenig	1

Men kan vi uden videre bruge sådanne kodninger i vores scoring. Kan vi med andre ord gå ud fra, at svarskalaen opfattes og anvendes ækvivalent? At skalaen med andre ord grafisk kan beskrives på følgende måde (figur 4.19).■



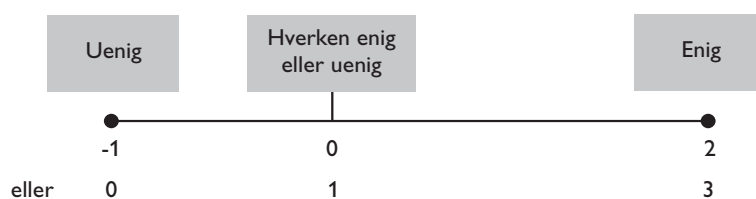
Figur 4.19. Scoring i en Likert-skala med tre kategorier.

Da Rasch i 1950'erne og 1960'erne arbejdede med sin itemanalyse, var computeren endnu kun i babyalderen, og det var endnu kun på store anlæg, såkaldte *mainframes*, at der kunne foretages mere komplicerede beregninger. Rasch holdt sig derfor mest til de dikotome materialer og modeller. I dag er der standardprogrammer til en almindelig pc, der kan operere med mange svarkategorier, sådan som i Likert-skalaen. Med disse programmer kan man indlede med at undersøge, om der overhovedet er tale om en homogen model, dvs. en ægte eller generaliseret itemanalyse, eller om det er muligt at opnå den ved at kassere visse items eller ved at opdele i undergrupper.

Hvis dette er tilfældet, har vi imidlertid i første omgang en flerdimensional model, der kræver en kompliceret scoringsfunktion, hvor der skal scores for alle svarkategorier, bortset fra en af dem. I tilfældet her med tre kategorier er der en todimensional scoringsfunktion, hvor vi fx må score

både antal *enige* og *uenige*. Her vil der altså slet ikke være én enkelt persondimension, eller hvad faktoranalytikere ville kalde en enkelt faktor, men to forskellige. Det vil indebære, at vi ikke har en simpel bipolar dimension, hvor uenighed er det modsatte af enighed. Man kan fx tænke på folks holdning til forskellige politiske partier, hvor der måske er nogle, man overvejende er meget enige med, andre man overvejende er meget uenige med, nogle man ikke har særlige markante holdninger til, men også nogle hvor man på nogle punkter er meget uenig og på andre meget enig. Her er der tale om et holdningsunivers, som ikke kan repræsenteres endimensionalt.

Det vi er en stor fordel, når vi kan reducere til en enkelt dimension sådan som i det dikotome tilfælde. I så fald får vi en type af *endimensionale* modeller. Disse vil i almindelighed have en empirisk scoringsfunktion, som vi ikke kan gå ud fra er ækvidistant.



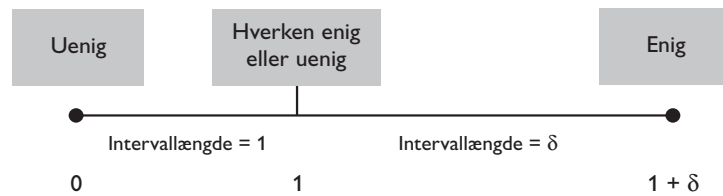
Figur 4.20. Eksempel på scoring i en ikke-ækvidistant Likert-skala med tre kategorier.

Scoringsfunktion kan fx se ud som i figur 4.20.■ Her er scoringsfunktionen udpræget ikke-ækvidistant. Afstanden mellem de to sidste punkter er dobbelt så stor som mellem de to første.

Men scoringsfunktion kan på den anden side jo også være ækvidistant. Indtil for nylig måtte spørgsmål om ækvidistance/ikke-ækvidistance opfattes som ret metafysiske. For hvordan skulle man kunne afgøre, om der er tale ækvidistance eller ej? I dag, hvor scoringsfunktion er en empirisk beregnelig størrelse, i hvert fald når det gælder et større antal respondenter, har spørgsmålet imidlertid antaget en empirisk karakter.

Jeg har opstillet en simuleringsmodel med de nævnte tre svar, hvor placeringen af den sidste i forhold til de to andre systematisk er blevet varieret (figur 4.21).■

Afstanden mellem de to første værdier i scoringsfunktionen, dvs. afstanden mellem kategorien *uenig* og kategorien *hverken enig eller uenig* kan defineres helt frit og er her sat til 1. Afstanden mellem de to næste



Figur 4.21. Generelt billede af scoring i en ikke-ækvivalent Likert-skala med tre kategorier.

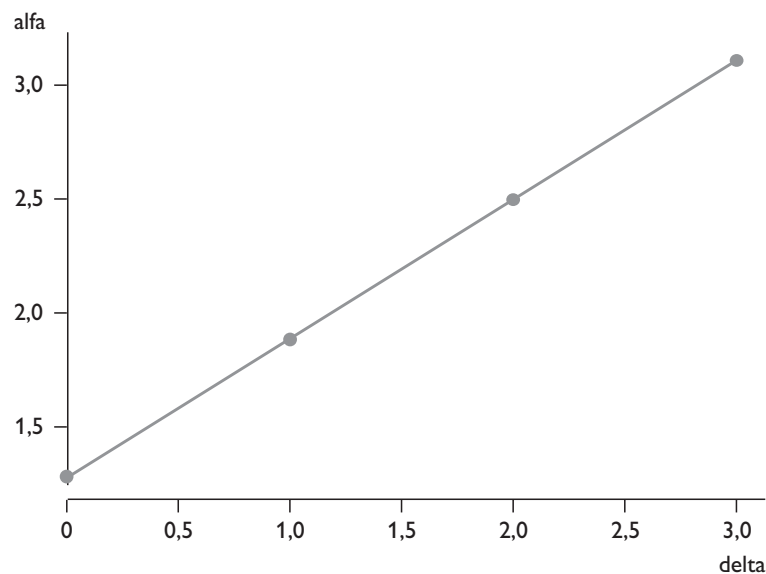
kategorier er angivet med en parameter, betegnet med det græske bogstav δ .

Det viser sig nu, at denne scoringsparameter δ kommer til udtryk, når vi bruger en ækvivalent scoringsfunktion til at beregne kategori 3-forekomsten for itemresponsene givet scoren, og dernæst foretager en bestemt

Figur 4.22. Eksempler på hældninger, der viser graden af ikke-ækvivalens for Likert-skala med tre kategorier.

transformation, nemlig den såkaldt logistiske,¹⁰ for såvel middelfresponserne som for scoren (figur 4.22).■

Vi ser altså et systematisk fald af hældningen for de logistiske itemresponsekurverdiagrammerne, når ikke-ækvilidistance-parameteren delta aftager. Dette fremgår af figur 4.23.■



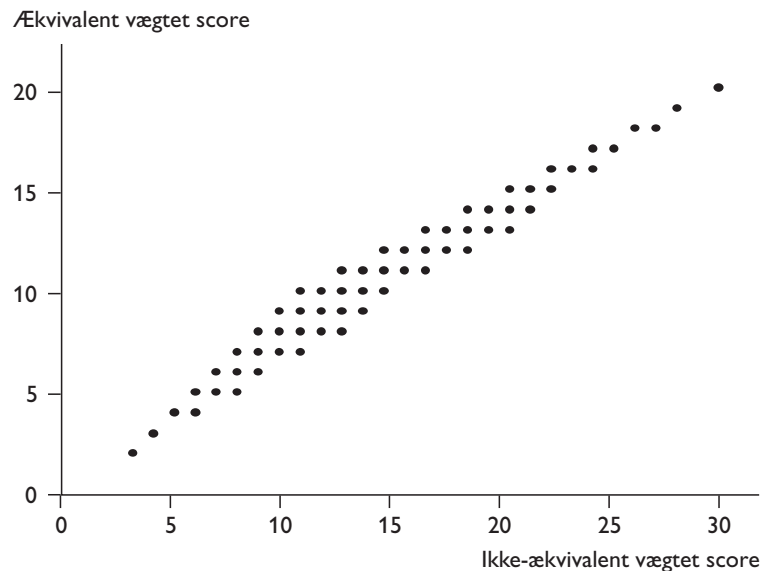
Figur 4.23. Sammenhængen mellem den teoretiske ikke-ækvilidistance og den empiriske hældning i diagrammerne ovenfor.

Denne sammenhæng mellem hældning og delta kan sammenfattes i følgende funktion:

$$\text{alfa} = 1,3 + 0,6 \cdot \text{delta}$$

Når vi for et empirisk materiale har tegnet de logistiske itemresponsekurver og fundet deres hældning, kan vi således kontrollere, om der er tale om ækvilidistance, eller om der foreligger en anden scoringsfunktion. Men i øvrigt spiller det slet ikke nogen praktisk rolle, om der er ækvilidistance eller ej.

Vi kan således i tilfældet med $\text{delta} = 2$, hvor afstanden mellem kategori 2 og 3 er dobbelt så stor som mellem kategori 1 og 2, udregne to forskellige



Figur 4.24. Diagram over sammenhængen mellem scores med ikke-ækvivalent og ækvivalent scoringsfunktion.

Tabel 4.14. Tabel over scoringsværdier ved ikke-ækvivalent og ækvivalent scoringsfunktion

	Uenig	Hverken enig eller uenig	Enig
Korrekt scoring vægtet ikke-ækvivalent	0	1	3
Ukorrekt scoring vægtet ækvivalent	0	1	2

scorer, nemlig dels den korrekte, ikke-ækvivalente, dels den forenklede ækvivalente.

Det vil sige, at vi har de to scoringsfunktioner, som fremgår af tabel 4.14.

Vi finder da den relation mellem den vægtede og altså korrekte score med scoringsfunktionen (Figur 4.24).■

Der er her tale om en simulering, hvor jeg er i den heldige situation at kende de sande personparametre, og det fremgår af tabel 4.15■, at både

den korrekte og den ukorrekte score har praktisk taget samme korrelation med disse sande personparametre. Dermed har de to scoringsmetoder også næsten samme reliabilitet, nemlig mellem 0,60 og 0,70. Når vi udregner den indbyrdes korrelationskoefficient mellem de to scores, viser denne sig da også at være så tæt på 1, som den næsten kan komme. På trods af den betragtelige afvigelse fra ækvidistance her, hvor afstanden mellem kategori 2 og 3 er den dobbelte af afstanden mellem kategori 1 og 2, er forskellen mellem de to scoringer altså langt mindre end den statistiske usikkerhed, middelfejlen.

Tabel 4.15. Korrelationstabel for personparameter, uvægtet og vægtet score

Correlations			pers	score	sc_vægt
pers	Pearson Correlation		1	,621**	,700**
	Sig. (1-tailed)		.	,000	,000
	N		10000	10000	10000
score	Pearson Correlation		,621**	1	,982**
	Sig. (1-tailed)		,000	.	,000
	N		10000	10000	10000
sc_vægt	Pearson Correlation		,700**	,982**	1
	Sig. (1-tailed)		,000	,000	.
	N		10000	10000	10000

** . Correlation is significant at the 0.01 level (1-tailed)

I tabel 4.15 har vi personparameteren i øverste linje og første kolonne; derefter kommer den uvægtede score, som jo er baseret på en ukorrekt antagelse om ækvidistance, og til sidst den vægtede score baseret på den korrekte antagelse om en bestemt ikke-ækvidistance. Vi ser, at den vægtede og således korrekte score er en brøkdel bedre end den simple, men ukorrekte, uvægtede score. Denne højere reliabilitet for den vægtede score er uden praktisk betydning.

Derimod kan det være metodologisk og teoretisk interessant at få konstateret, at en Likert-skala *ikke* er ækvidistant. Det kan meget vel betyde, at udformningen af svarkategorierne er defekt.

Helvede
(fortabelsen)
meget sones

Skærsildens
2. kreds
Her skal meget sones

Skærsildens
1. kreds
Her er der en del
at rette op

Paradis' 3. kreds
almindelige engle

Paradis' 2. kreds
ærkeengle

Paradis' 1. kreds
Vorherre selv

Helt inhomogene
modeller, hvor
kurverne krydser
uanset opdeling i
persongrupper;
flerdimensional
scoringsfunktion

Generelt inhomogen,
men gruppenæssig
homogen model, hvor
kurverne ligger rigtigt
ved opdeling i
persongrupper;
men hvor der lokal
afhængighed,
flerdimensional
scoringsfunktion

Generaliseret
identifikationsmodel,
homogen (for alle
persongrupper),
men med lokal
afhængighed,
flerdimensional
scoringsfunktion,
flerdimensional
scoringsfunktion

Gruppenæssigt
inhomogen,
generaliseret model,
homogen for den
enkelte persongruppe,
med lokal
afhængighed,
flerdimensional
scoringsfunktion

Ægte
identifikationsmodel,
som er generelt
homogen,
og hvor der lokal
afhængighed,
flerdimensional
scoringsfunktion

Ægte
identifikationsmodel,
som er generelt
homogen,
og hvor der lokal
afhængighed,
endimensional
scoringsfunktion,
men forskellig fra
item til item

Ægte
identifikationsmodel,
som er generelt
homogen,
og hvor der lokal
afhængighed,
endimensional
scoringsfunktion,
som er den samme
for alle item

Figur 4.25.

En hierarkisk oversigt over itemanalysemodeller

Når vi sammenfatter de modeller, der netop er gennemgået, har vi modelhierarkiet i figur 4.25, som er en udbygning af figur 4.18.

Til alle praktiske formål er der endimensionale modeller, der kan anvendes som målingsinstrumenter. Det afgørende er her, om der er den samme scoringsfunktion for alle items eller ej. Hvis dette er tilfældet, spiller det kun en teoretisk rolle, om denne er ækvidistant, medmindre der foreligger et helt usandsynligt tilfælde, hvor nogle af afstandene mellem nabokategorier er grotesk meget større end andre.

Validitet og validering

Det spørgsmål, der står tilbage, er imidlertid, hvilken egenskab det er, vi har målt?

Selv om vi har vist, at vi har målt noget, kan vi jo ikke være sikre på, at vi nu har målt *det, som vi tror, vi har målt*. I det her beskrevne tilfælde var der tale om en test for dominans. Men er det nu også lige dominans, vi har målt? Det kunne jo også være ambition eller måske selvtillid? Det spørgsmål, som melder sig, efter at vi har godtgjort, at testen måler et eller andet, er *hvad det så er, vi har målt*. Dette forhold, at testen altså faktisk måler, det vi har ønsket, kalder man for *validitet eller gyldighed*. Det er et betydeligt mere indviklet problem end reliabiliteten, som blot vedrører den rent talmæssige målingspålidelighed. Faktisk kan spørgsmålet om validitet kun sjældent besvares direkte, men der er dog forskellige måder at angribe det på.

Kriterievaliditet

Den mest overbevisende måde er bestemmelse af *kriterievaliditet*. Det går ud på at sammenholde testresultatet med en oplysning om de testede personer, som dels stammer fra en anden kilde end testen, dels på mere overbevisende måde end testen selv peger på den egenskab, vi ønsker at måle med vores test. Ofte søger man at relatere testen til et karakteristikum knyttet til testpersonernes status i den praktiske virkelighed, som testen til syvende og sidst skal kunne sige noget om.

Nu kan man jo indvende, at hvis man allerede havde fx en angivelse af

Tabel 4.16. Valideringskriteriet organisatorisk niveau

1	Topleder
2	Divisionsdirektør/funktionschef
3	Afdelings-/gruppechef
4	Supervisor
5	Formand
6	Specialist-/teknisk funktion

Tabel 4.17. Dominans og organisatorisk niveau – Kriterie-validitet

Organisatorisk niveau	Mean	N	Std. Deviation	Std. Error of Mean
Topleder	14,91	16	2,185	,546
Divisionsdirektør/funktionschef	14,52	23	2,728	,569
Afdelings-/gruppechef	13,27	39	2,892	,463
Supervisor/formand	9,57	15	3,550	,917
Specialist-/teknisk funktion	10,53	124	3,781	,340

dominans i folks praktiske liv, så var der slet ikke nogen grund til at lave en test. Og det er selvfølgelig netop problemet, at man endnu ikke har en på forhånd given bestemmelse af den eftersøgte egenskab. Men det kan være, at man trods alt har et karakteristikum, der måske ikke direkte kan verificere testens validitet, men i stedet, i overensstemmelse med Poppers videnskabsteori (Poppers 1963), kan afprøve validiteten indirekte gennem et forsøg på falsificere den.

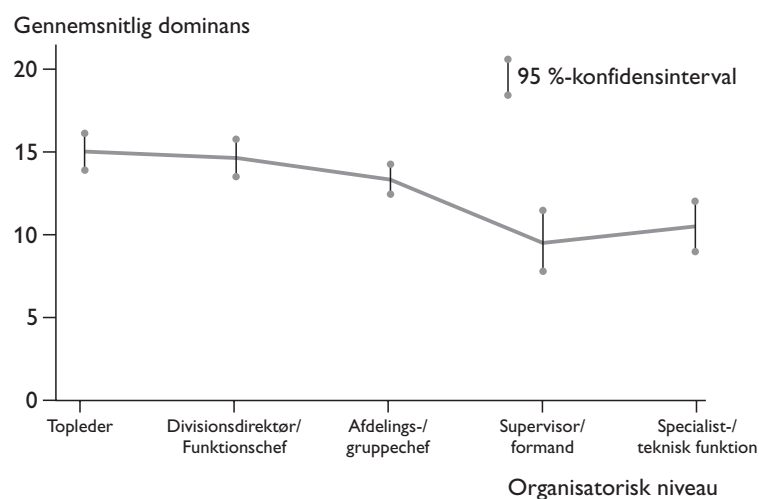
Jeg har her et eksempel på en validitetsafprøvning af en test for dominans, hvor der er oplysning om de testedes organisatoriske status i de firmaer, hvor de er ansat. Som det fremgår af tabel 4.16, er der en opdeling i seks statusniveauer.

Topledere er de allerøverste chefer, divisionschefer leder en større del af firmaet, som fx produktion, økonomi eller salg. En afdelingschef er souschef lige under en divisionschef. Så kommer supervisoren, som leder en lille gruppe underordnede, og til sidst det administrative og tekniske personale, som ikke har nogen ledelsesfunktion.

Det er indlysende, at en overordnet, som er sat til at lede en underordnet,

må fremvise en eller anden grad af dominans. Ud fra selve det hierarkiske princip i organisationen er det også plausibelt, at dominansen er mest udtalt jo højere oppe i det organisatoriske niveau, man befinder sig. Så det er den validitetsafprøvning, vi vil anvende.

Vi ser i tabel 4.17, hvordan 124 testpersoner fra disse forskellige organisatoriske niveauer fordeler sig med hensyn til dominans. Vi bemærker, at der er en ganske god sammenhæng, når vi ser på gennemsnitsresultatet på de fem niveauer. Hvis vi også inddrager standardafvigelsen og tegner intervaller på to gange (helt præcis 1,96 gange) denne til begge sider, får vi figur 4.26.



Figur 4.26. Diagram over dominans på forskellige organisatoriske niveauer.

Tabel 4.18. Korrelationen mellem dominans og organisatorisk niveau

Kendall's tau_b		Dominans
Organisatorisk niveau	Correlation Coefficient	-,334(**)
	Sig. (1-tailed)	,000001
	N	217

**. Correlation is significant at the 0.01 level (1-tailed).

Diagrammet viser, at sammenhængen i tabel 4.17 ikke kan betragtes som tilfældig. Der er en klar (signifikant) sammenhæng, i hvert fald når vi sammenligner de tre øverste niveauer med de to nederste,. At dette er tilfældet, bekræftes endelig af en sammenhængs- eller korrelationstest (se tabel 4.18).■

En særlig test for sammenhængen mellem organisatorisk niveau og dominans viser, at en klar korrelation som den fundne kun ville optræde ved en tilfældighed i ét ud af en million tilfælde, hvor vi foretog et sådant tilfældighedseksperiment. Validiteten er således bekræftet, og det er oven i købet en økologisk validitet, dvs. at den refererer til det pulveriserende liv.

Dette var et eksempel på kriterievaliditet.

Begrebsvaliditet som konvergent og divergent validitet

En anden form for validitet hedder konvergent og divergent validitet.

Her sammenligner man testen med andre test, hvoraf nogle forventes at ligne meget og derfor korrelere med det givne mål, mens andre forventes ikke at ligne eller i alt fald ikke at ligne særlig meget – og således ikke have nogen eller kun ringe korrelation.

*Tabel 4.19. Korrelationen mellem dominans og andre personlighedstræk
Begrebsvalidering gennem analyse af konvergens og divergens*

Correlations		Synlighed	Dristighed	Selvtillid	Psykologisk indsigt	Empati
Dominans	Pearson Correlation	,701**	,576**	,682**	,189**	,083
	Sig. (1-tailed)	,000	,000	,000	,002	,101
	N	239	239	239	239	239

**. Correlation is significant at the 0.01 level (1-tailed).

Vi vil således forvente, at dominans ligner egenskaber som synlighed, dristighed og selvtillid, mens de har mindre til fælles med psykologisk indsigt og empati. Med andre ord, at de første tre egenskaber skal vise konvergens, de sidste to divergens. Tabel 4.19■ viser, at dette også er

tilfældet. Vi har altså en bekræftelse af denne specifikke afprøvning af *begrebsvaliditet*, som den engelske term *construct validity* blev oversat af Kaj Spelling, tidligere professor i intelligensforskning.

Differentiel validitet

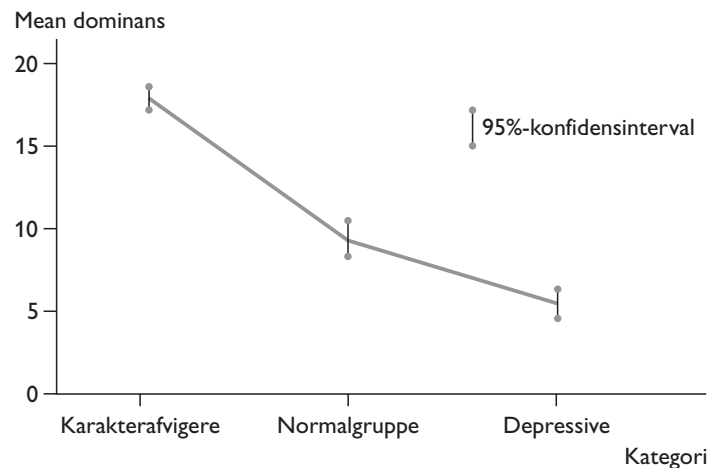
Den sidste form for validering, som også er en måde at afprøve begrebsvaliditet på, kan man kalde differentiel valididering. Det vil sige, at man har nogle underbyggede formodninger om, at der er visse grupper, som vil score højt, og andre, som vil score lavt. Lad os forestille os, at vi har to kliniske grupper, som er testet. Den ene er kriminelle karakterafvigere, der er mentalundersøgt i forbindelse med retlig undersøgelse (for at se om de er egnede til almindelig straffesonding), den anden er patienter indlagt på et psykiatrisk hospital for en alvorlig depression. Her vil en differentiel validering vise, at den første gruppe ligger højt og den anden gruppe lavt på dominanstesten. Vi kan endda udvide afprøvningen ved at medtage en normalgruppe, fx en gruppe patienter behandlet på en kirurgisk afdeling for forholdsvis harmløse sygdomme som fx blindtarmsbetændelse.

Lad os tænke os, at tabel 4.20 kom ud af denne undersøgelse.

Tabel 4.20. Begrebsvalidering gennem analyse af konvergens og divergens

Reportdominans				
Kategori	Mean	N	Std. Deviation	Std. Error of Mean
Karakterafvigere	17,77	30	1,768	,323
Normalgruppe	9,19	35	4,003	,677
Depressive	5,35	35	3,370	,570
Total	10,42	100	6,025	,603

Tabellen her viser de tre gennemsnit, som bekræfter den differentielle hypotese. Vi har altså en differentiel validering, og hermed yderligere et bidrag til vores begrebsvaliditet. Figur 4.27 er en grafisk fremstilling af dette med konfidens- eller sikkerhedsintervaller. Selv om grupperne er af beskeden størrelse, har vi således meget sikre gruppeforskelle.



Figur 4.27. Begrebsvalidering gennem analyse af differentiell diskrimination (diagram).

Begrebsvaliditet i itemanalysen

Psykometrikeren Svend Kreiner har givet en meget elegant og effektiv definition af begrebsvaliditet ved at inkorporere dette testpsykologiske begreb i itemanalyseteorien (Avlund et. al., 1993). Nærmere betegnet knyttet til begrebet differentiell itemfunktion. Ifølge Kreiners definition betyder begrebsvaliditet kort og godt, at man efter at have indført alle relevante kriterier som baggrundsvariable, får bekræftet, at der slet ikke er nogen form for differentiell itemfunktion knyttet til disse variable.

Det betyder nemlig, at den information, der ligger i disse baggrundsvariable, fx organisatorisk niveau, som det var tilfældet i vores eksempel på kriterievaliditet, kun kommer til udtryk i fordelingen af personparametre og dermed scorefordelingen i de forskellige grupper inden for den givne baggrundsvariabel. Hvis der nu var en differentiell itemfunktion, er der en information i baggrundsvariablen, som ikke kommer til udtryk i itemanalysens scoring; derfor er scoringen informationsmæssig ufuldstændig og ikke helt valid.

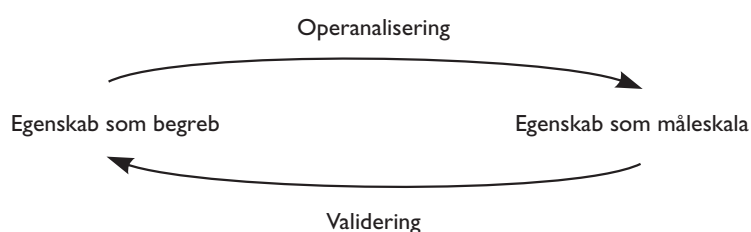
Operationalisering og validering

Skalavalidering kan ses i sammenhæng med selve den proces, hvor skalaen

kom til verden. Begrebsvalideringen af en bestemt skala er den omvendte proces i forhold til konstruktionen af skalaen, som er en operationalisering af et bestemt begreb:

- Ved *operationalisering* går vi fra *begrebet* for en egenskab til *konstruktionen* af en konkret måde at måle det på.
- Ved *valideringen* af en skala går vi fra den *konkrete måde at måle på* til en *rekonstruktion af skalaens begrebslige indhold*.

Denne dobbelte relation er grafisk illustreret i figur 4.28).■



Figur 4.28. Dialektikken mellem operationalisering og validering

Først med valideringen har vi således vist, at vores begrebsliggørelse er lykkedes. Det er ikke nok at kunne opbygge et begreb; det skal også kunne opfylde sin funktion.

Itemresponsteori

Efter at Rasch's itemanalyse er blevet anerkendt og udbredt over hele verden, er der gjort forsøg på at udvide denne ret snævre klasse af modeller, som Rasch selv kanoniserede, især i den enkle form, hvori han præsenterede sin teori i den grundlæggende bog fra 1960 (Rasch, 1980). Den teori, som har opbygget et sådant bredere sortiment af modeller, hedder itemresponsteori (Item Response Theory), og opererer efter min opfattelse på en noget ad hoc præget måde, med muligheder for flere typer af modelparametre. Vi har jo fx set, at der er et absolut tabu mod krydsende itemresponskurver. Den slags tilfælde med såkaldt variabel diskrimination ved de forskellige items blev, som vi har set, forvist til Rasch's udgave af

Dantes helvede. I itemresponsteorien er man ikke så religiøst anlagt som Rasch, hvis far var missionær (Olsen, 2003).

Her parametriserer man blot denne efter Rasch's opfattelse rene dødsynd, således at der ud over itemparameteren, henholdsvis lethedsværhedsparameter i intelligenstag og attraktionsparameter i personlighedstest, også er en diskriminationsparameter. Disse itemresponsmodeller kan sagtens bringes til at fungere på moderne computere. Spørgsmålet er blot, om man metodologisk finder dem acceptable. For en introduktion til IRT (Item Response Theory), se Baker (2001).

Den konfirmatoriske faktoranalyse

I det tidligere afsnit om faktoranalyse beskæftigede vi os kun med den såkaldte eksplorative type, som i udgangspunktet er helt ateoretisk, men evt. søger at opbygge en teori ad induktiv vej.

Denne metodologi har, som vi har set, ført til udvikling af talrige konkurrerende faktoranalytiske teorier om fx intelligens og personlighed, uden at der rigtigt har været nogen tegn på afklaring. Dette i forening med det frontalangreb, Rasch rettede mod den eksplorative faktoranalyse i 1950'erne, og de åbenlyse kvalitetsmæssige forbedringer, hans itemanalyse indebar, førte i 1970'erne til, at svenskeren Jöreskog (1971) udviklede en hypotetisk-deduktiv metodologi for faktoranalyse, hvor man ud fra en fagligt funderet teori udformer en eksplicit og apriorisk opstillet model, som man derefter tester empirisk. Denne metode kaldes for konfirmatorisk faktoranalyse, og selve modellen hedder LISREL, Linear Structural Equation Modelling.

Hermed bliver metodologien på mange måder en parallel til itemanalysen, og rent matematisk statistisk er hovedforskellen, at mens itemanalysen er en diskret model baseret på det kategoriale ved svarmulighederne, er LISREL ligesom den eksplorative faktoranalyse baseret på en antagelse om kontinuerte, nemlig normalfordelte variable.

Litteratur

Avlund, K., Kreiner, S. & Schultz-Larsen, K. (1993). Construct validation and the Rasch model: functional ability of healthy elderly people. *Scandinavian Journal of Social Medicine*, 4:233-46.

- Baker, F. (2001). *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- Binet, A. & Simon, T. (1977a). New methods for the diagnosis of the intellectual level of subnormals, (Methodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique*, 12:191-244). I: Robinson D.N. (ed.) (1977). *Significant Contributions to the History of Psychology 1750-1920, Series B, Psychometrics and Educational Psychology, Vol IV, A. Binet, T. Simon, L.W. Stern, F. Galton*, Washington: University Publications of America.
- Binet, A. & Simon, T. (1977b). The development of intelligence in the child, (Le developpement de l'intelligence chez les enfants. *L'Année Psychologique*, 14:1-94). I: Robinson D.N. (ed.) (1977). *Significant Contributions to the History of Psychology 1750-1920, Series B, Psychometrics and educational psychology, Vol IV, A. Binet, T. Simon, L.W. Stern, F. Galton*, Washington: University Publications of America.
- Boring, E.G. (1923). Intelligence as the tests test it. *New Republic*, 35:35-37.
- Cattell, R.B. (1950). *Personality: A Systematic, Theoretical and Factual Study*. New York: McGraw Hill.
- Cattell, R.B. (1987). *Intelligence: Its Structure, Growth and Action*. New York: Elsevier.
- Christensen, K.B., Bjørner, J.B., Kreiner, S. & Petersen, J.H. (2002). Testing unidimensionality in polytomous Rasch models. *Psychometrika*, 67(4):563-574.
- Eysenck, H.J. (1947). *Dimensions of Personality*. New York: Methuen.
- Eysenck, H.J. & Eysenck, M.W. (1985). *Personality and Individual Differences – A Natural Science Approach*. New York: Plenum Press.
- Galton, F. (1869). *Hereditary Genius*, London: Macmillan.
- Jöreskog, K.G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36:409-426.
- Kousgaard, E. (2003). *50 år med intelligensprøven BPP*. København: Forsvarsakademiet, Institut for Militærpsykologi.
- Kreiner, S. & Christensen, K.B. (2004). Analysis of local dependence and multidimensionality in graphical loglinear Rasch models. *Communications in Statistics. Theory and Methods*, 33(6):1239-1276.
- Lawley, D.N. & Maxwell, A.E. (1971). *Factor Analysis as a Statistical Method*. London: Butterworth.
- McCrea, R.R. & Costa, P.T. (1990). *Personality in Adulthood*. New York: Guildford.
- Olsen L.W. (2003). Essays on Georg Rasch and his contributions to statistics. Ph.d. afhandling???, Københavns Universitet.
- Pearson, K. (1914). *The Life, Letters and Labours of Francis Galton*. London: Cambridge University Press.
- Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*. London: Routledge.
- Rasch G. (1980). *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: Chicago University Press, 1960???
- Raven, J.C., Styles, I., & Raven, M.A. (1998). *Raven's Progressive Matrices: SPM Plus Test Booklet*. Oxford, England: Oxford Psychologists Press; San Antonio, TX: The Psychological Corporation.
- Robinson D.N. (ed.) (1977). *Significant Contributions to the History of Psychology 1750-1920, Series B, Psychometrics and educational psychology, Vol IV, A. Binet, T. Simon, L.W. Stern, F. Galton*, Washington: University Publications of America.

- Rubin, E. (1967). *Synsoplevede Figurer: Studier i psykologisk Analyse: Del 1*. København: Akademisk Forlag, 1915 -???
- Spearman, C.E. (1927). *The Abilities of Man, their Nature and Measurement*. New York: Macmillan.
- Stern, L.W. (1977). The psychological methods for testing intelligence, I: Robinson D.N. (ed.) (1977). *Significant Contributions to the History of Psychology 1750-1920, Series B, Psychometrics and educational psychology, Vol IV, A. Binet, T. Simon, L.W. Stern, F. Galton*, Washington: University Publications of America.
- Terman, L.M. (1917). The intelligence quotient of Francis Galton in childhood. *American Journal of Psychology*, 28:209-215.
- Terman, L.M. & Merrill, M.A. (1937). *Measuring Intelligence*. Cambridge, Mass.: The River Press.
- Thurstone, L.L. (1938). *Primary Mental Abilities*. Chicago: University of Chicago Press.
- Wechsler, D. (1939). *The Measurement of Adult Intelligence*. Baltimore: Williams and Wilkins.

Noter

Note 1, s. 200 Galtons efterfølger, Lewis Terman (1917), anslog på grundlag af Karl Pearsons (1914) oplysninger om Galtons præstationer i barndommen, Galtons IQ til omkring 200.

Note 2, s. 202 Her er middelværdien 100 og spredningen 16.

Note 3, s. 213 Tillægsordet tetradisk er dannet ud fra det græske ord for 4.

Note 4, s. 213 Beviset er:

$$\rho(T1, T2) = \text{Middelværdi af } (score1 \times score2) =$$

$$\text{Middelværdi } (FG \times g1 + FS1 \times s1) \times (FG \times g2 + FS2 \times s2) =$$

Her er:

$$\text{Middelværdi } \{(FG \times FG \times g1 \times g2) +$$

$$(FG \times FS1 \times g1 \times sT + FS1 \times FG \times sT \times g2 + FS1 \times FS2 \times s1 \times s2)\} = g1 \times g2$$

I overstående ligning er nemlig

$$\text{Middelværdi } \{FG \times FG\} = (FG)^2 = \text{Variansen på } FG = 1,$$

pga. Standardiseringsantagelsen,

mens

$$\text{Middelværdi } \{FG \times FS1\} = \text{Middelværdi } \{FG \times FS2\} =$$

$$\text{Middelværdi } \{FS1 \times FS2\} = 0,$$

på grund af uafhængighedsantagelserne.

Note 5, s. 218 Poppers videnskabsteori (Popper, 1963) kræver af en videnskabelig teori, at den selv kan udlede empiriske betingelser for at afprøve dens holdbarhed. Dette krav lever Spearmans teori op til.

Note 6, s. 225 Undertiden fremstilles itemanalysemodellen i stedet med en positiv personparameter, in casu dygtighed eller intelligens, og en negativt fungerende

itemparameter, in casu sværhedsgrad, således at vi i stedet skal dividere den første parameter med den anden. Vi har da:

Sandsynlighed for pos. svar for person P på item I =
 $(\text{dygtighed for P}) / (\text{sværhedsgrad af I}) / [1 + (\text{dygtighed for P}) / (\text{sværhedsgrad af I})]$

Denne parametrisering giver de to skalaer, person- og itemskala, en simpel begrebsmæssig form, hvor vi måler dygtighed og sværhedsgrad på samme skala fra 0 og til uendelig. At en persons dygtighed er identisk med sværhedsgraden af et bestemt item betyder, at personens sandsynlighed for at besvare spørgsmålet korrekt er præcis 50 procent. Hvis vi derimod tager et item J, der er dobbelt så svært som det førstnævnte, og altså med en sværhedsgrad der er det dobbelte af personens dygtighedsgrad, så får vi:

Sandsynlighed for pos. svar for person P på item J = $1 / (1 + 1) = 1/3$.

Nu er sandsynligheden altså faldet til ca. 33 procent.

Note 7, s. 226 (Kousgaard 2003:22).

Note 8, s. 227 For eksempel vil et spørgsmål om, hvorvidt man har Saddam Hussein som ledelsesmæssigt forbillede, nok være mindre attraktivt, end hvis der var tale om Nelson Mandela.

Note 9, s. 238 Eksemplet med at ændre en test på grund af differentiell itemfunktion er ikke et akademisk påfund. I USA blev selve brugen af test, især intelligenstest, i en længere periode blokeret af retssager ført af borgerrettighedsgrupper, som procederede for, at testene var racemæssigt eller kønsmæssigt diskriminerende. I sagen Griggs v. Duke Power Co. førte den afroamerikanske Griggs en sag mod firmaet Duke Power Co. med påstand om, at brugen af intelligensprøver i sig selv var racediskriminerende. Griggs fik her medhold ved højesteret.

Se http://supct.law.cornell.edu/supct/html/historics/USSC_CR_0401_0424_ZO.html

Note 10, s. 243 Den logistiske transformation overfører en hyppighed, h, på en intervalskala (fra minus til plus uendelig) gennem transformationen:

$\text{logit}(h) = \log(h/(1-h))$, hvor log er logaritmefunktionen.