Kønsdiskriminerende reklame

– kommentarer til en empirisk undersøgelse

Niels J. Blunch og Kai Kristensen

1. Indledning

Rapporten *Kønsdiskriminerende Reklame* [1]), der er udarbejdet af en af Forbrugerombudsmanden nedsat arbejdsgruppe, har opnået en omtale, som kun sjældent bliver sådanne rapporter til del.

Såvel i begrundelsen for arbejdsgruppens konklusion (rapportens kapitel VIII) som i den almindelige debat, der har fulgt på rapportens offentliggørelse, indtager en af *Preben Sepstrup* gennemført *»undersøgelse af mands- og kvindebilledet i den danske magasin- og dagspresseannoncering« en central plads.* For en vurdering af såvel denne undersøgelses resultater som af de konklusioner, der kan drages – henholdsvis er draget – af disse, har naturligvis kvaliteten af selve undersøgelsens udformning og gennemførelse en afgørende betydning.

Da annonceundersøgelsens resultater og konklusioner *kan blive bestemmende for udformningen af eventuelle fremtidige lovindgreb på området,* finder vi det nødvendigt, såvel af hensyn til den fortsatte faglige debat omkring dette betydningsfulde emne som af hensyn til de mennesker, der skal have ansvaret for udarbejdelsen af retningslinierne for sådanne eventuelle indgreb, at foretage en vurdering af den af Sepstrup anvendte undersøgelsesmetode.

Vi vil således *ikke* i denne artikel beskæftige os med undersøgelsens resultater og konklusioner, men *udelukkende* med *kvaliteten* af de frembragte resultater og hermed også med *grænserne* for de *konklusioner,* der rent principielt kan drages af disse resultater. Vor vurdering bygger på den af Preben Sepstrup udarbejdede *Tekniske rapport*[2]), der indeholder en grundig beskrivelse af projektets metoder og tekniske gennemførelse.

1) Kønsdiskriminerende reklame. Rapport udarbejdet af én af Forbrugerombudsmanden nedsat arbejdsgruppe. København, januar 1979.

2) Preben Sepstrup: Reklamen som socialisationsfaktor – Teknisk rapport. Handelshøjskolen i Århus, 1979.

2. Om videnskabelig metode

Den videnskabelige metode er en erkendelsesmetode, en metode til at erkende indretningen af den verden, der omgiver os, og holdbarheden af de forestillinger, vi gør os om den. Den er dog naturligvis ikke den eneste metode, vi benytter os af for at opnå disse formål. Der er principielt fire baser, på hvilke man kan begrunde sandheden af en fremsat påstand:

1. Tradition. Påstanden accepteres, såfremt den er i overensstemmelse med traditionelle opfattelser.

2. Autoritet. Påstanden accepteres, såfremt den er fremsat af en person (eller persongruppe), hvis meninger inden for det pågældende område vurderes højt.

3. Intuition. Påstanden accepteres, såfremt den er i overensstemmelse med ens umiddelbare opfattelse og »Fingerspitzengefühl«.

4. Videnskabelig metode. Påstanden accepteres, såfremt objektive målinger af operationelt definerede variable verificerer den. Accepten er foreløbig og under forbehold af ændringer som følge af fremkomsten af yderligere relevant information.

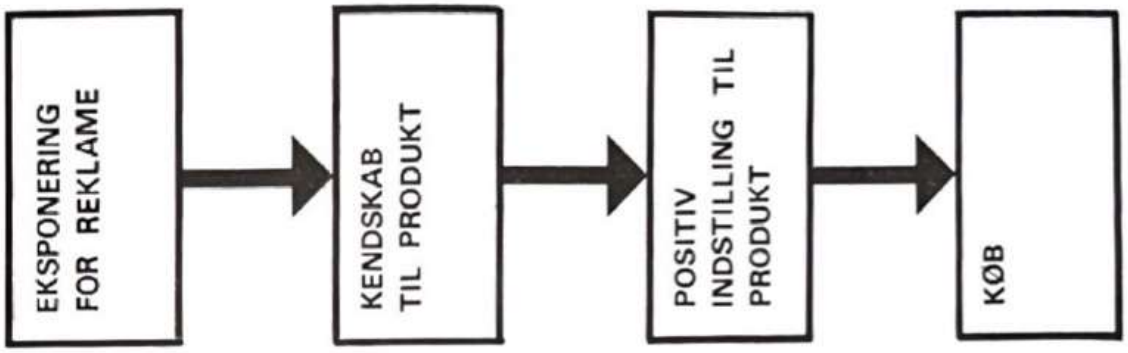
Der er under punkt 4 ovenfor benyttet en del nye begreber, som vil blive uddybet nedenfor, men det er uden videre klart, at den videnskabelige metodes acceptkrav er en konfrontering af den fremsatte påstand med empiriske data indsamlet og bearbejdet efter visse, nogenlunde alment accepterede retningslinier.

Dette kan synes så selvfølgeligt, at der næppe skulle være grund til at ofre mange ord på det, ligesom acceptkravene 1, 2 og 3 ved en sådan opremsning let kan få et latterligt skær. Det er derfor vigtigt at påpege, at kravet om en påstands konfrontation med empiriske data – dvs. med »virkeligheden« – som grundlag for påstandens accept ikke til alle tider og på alle steder har været betragtet med den selvfølgelighed, som tilfældet er i vor tids vestlige kultur. Velkendt er f. eks. den rolle, som de græske filosoffer – først og fremmest Aristoteles – spillede som traditionsbegrundelse gennem store dele af den europæiske middelalder.

Et eksempel:

Som grundlag for planlægning af og kontrol med reklameindsatsen må man gøre sig nogle forestillinger om den virkning, reklamen må formodes at øve på modtagerens adfærd. En opfattelse, der i snart mange år har haft ret stor udbredelse, er, at reklamemodtageren efter reklamepåvirkningen bevæger sig gennem en række psykologiske stadier, før det endelige køb finder sted. Antag, at vi ønsker at verificere denne opfattelse, således som den er afbildet i figur 1.

Figur 1.



Grundlaget for den videnskabelige verifikation er, at *hypotesen* om opfattelsens rigtighed skal kunne underbygges med *empiriske data.* Vi må derfor starte med at spørge: »Hvis figur 1 er rigtig, hvilke empiriske følger har da dette?«

Af den *teoretiske hypotese*

H_1: Den i figur 1 afbillede opfattelse er »sand«,

kan vi aflede en række *empiriske hypoteser,* f. eks.:

H_2: Kendskabet til produktet er større blandt personer, der har set annoncen, end blandt personer, der ikke har.

H_3: Positiv indstilling til produktet er større blandt personer, der har set annoncen, end blandt personer, der ikke har.

Disse er kun to eksempler (som desuden hver især kun har relation til dele af figuren), og læseren kan selv aflede yderligere empiriske hypoteser. Sådanne empiriske hypoteser har den egenskab, at deres sandhed – i hvert fald i princippet – lader sig konstatere empirisk, og konstatering af deres sandhed tjener herved til at underbygge den oprindelige teoretiske hypotese.

Der er imidlertid ét meget væsentligt aspekt, som man bør være opmærksom på: *Selv om en empirisk hypotese viser sig at være sand, behøver den teoretiske hypotese, den er afledt af, ikke at være det.* Det er nemlig muligt, *at en empirisk hypotese lader sig aflede af mere end én teoretisk hypotese.* Således kan såvel H_2 som H_3 afledes af følgende teoretiske hypotese:

H_x: En potientiel reklamemodtager, der kender en given vare og har en positiv indstilling til den, opsøger (bevidst eller ubevidst) annoncer for den pågældende vare.

Det gælder med andre ord om at aflede nogle empiriske hypoteser, der lader sig forene med så få teoretiske hypoteser som muligt. En mulig kandidat er

H_4: Der kan konstateres en tidsforskel mellem observation af annoncer og kendskab til varen, således at observation kommer *før* kendskab.

H_4 lader sig ikke aflede af H_x (tværtimod!), så hvis H_4 konstateres, er H_x falsificeret, og (en del af) H_1 må accepteres, *indtil der fremkommer nye teoretiske hypoteser, der mindst lige så godt kan være ophavsmand til de testede og accepterede empiriske hypoteser. Bemærk, at verifikation af en teoretisk hypotese foregår ved forkastelse af alternative hypoteser.* Det er bl. a. derfor, man kalder statistikken for »the science of disproof«.

Før en verifikation efter disse retningslinier kan finde sted, må de begreber, der er indeholdt i de empiriske hypoteser, imidlertid gives et empirisk indhold. Ser vi således på hypoteserne H_2 og H_3, indeholder disse begreberne *kendskab til produkt, set annoncen* og *positiv indstilling til produkt.* Disse begreber skal altså ikke kun defineres *teoretisk* derved, at videnskabsmanden – i fald der må antages at være udbredt uenighed om begrebernes almindelige indhold – må søges at fastlægge dette så éntydigt som muligt ved at referere til andre begreber, om hvis mening, der råder større enighed. Begreberne skal også gives et *empirisk* indhold gennem en fastlæggelse af målemetoder, hvorved begreberne defineres *operationelt.* Ved en *operationel* definition forstås en definition af begrebet ved hjælp af den operation, hvorved det måles. Et nærliggende eksempel er definition af temperatur som den egenskab, man måler med et nærmere specificeret termometer. I vort eksempel må de operationelle definitioner fastlægges ved specifikation af spørgeskema og kontaktform.

Før disse målinger (opfattet i den videste betydning af ordet) kan gennemføres, må vi tage stilling til hos hvem, vi er interesseret i at kunne foretage sådanne målinger; er det hos hele den danske befolkning eller hos enkelte geografisk, aldersmæssigt eller på anden måde afgrænsede grupper? Vi må definere *populationen af analyseenheder.*

Sædvanligvis kan ikke alle analyseenhederne i undersøgelsens population gøres til genstand for måling, men undersøgelsen må foregå på *stikprøvebasis.* Næste trin bliver derfor at udarbejde en *stikprøveplan.* Det må her tilstræbes, at videnskabsmanden eller andre, der er involveret i undersøgelsen, ikke bevidst eller ubevidst får mulighed for gennem påvirkning af valget af analyseenheder til stikprøven at øve indflydelse på undersøgelsens resultat. Den simpleste måde at imødegå denne risiko for farvning af resultaterne på er at udvælge stikprøven ved én eller anden form for *lodtrækning.* På lignende vis må populationen af produkter og populationen af annoncer defineres. Disse begrænsninger er medbestemmende for det omfang, i hvilket de fundne resultater kan generaliseres.

Herefter må *kontaktformen* fastlægges: Skal der benyttes postudsendte

spørgeskemaer, personligt interview eller telefoninterview? I visse tilfælde kan også simpel observation af respondenten komme på tale.

Efter at data er indsamlet, skal de *analyseres* og *fortolkes*, idet spørgsmålet er: I hvilket omfang verificerer de indsamlede data de afledede empiriske hypoteser? Det må her erindres, at data altid er behæftet med en vis usikkerhed (f. eks. stikprøveusikkerhed). *Valg af dataanalysemetoder* må derfor specificeres og begrundes, og fortolkningen af resultaterne foregå ved fremgangsmåder, der opfylder kravet om at være uafhængig af analytikernes (eller andre involverede personers) præferencer for specifikke konklusioner.

Den her beskrevne verifikationsproces kan afbildes som vist i figur 2 og betragtning af indholdet i de enkelte kasser – og forbindelsen mellem dem – vil vise, at den egentlige »filosofi« bag den videnskabelige metode er at fremskaffe viden, der så éntydigt som muligt *kan kommunikeres* til interesserede modtagere i en form, der sikrer, at modtageren – i fald han skulle være interesseret – *kan gentage (reproducere)* hele den proces, hvorved den pågældende viden blev frembragt, med størst mulig sikkerhed for at nå frem til samme resultat og konklusion. *Resultater, der ikke kan »genfremskaffes«, er uinteressante.* Dette er formuleret på den måde, at *den videnskabelige metode er et middel til at dele en erfaring.*

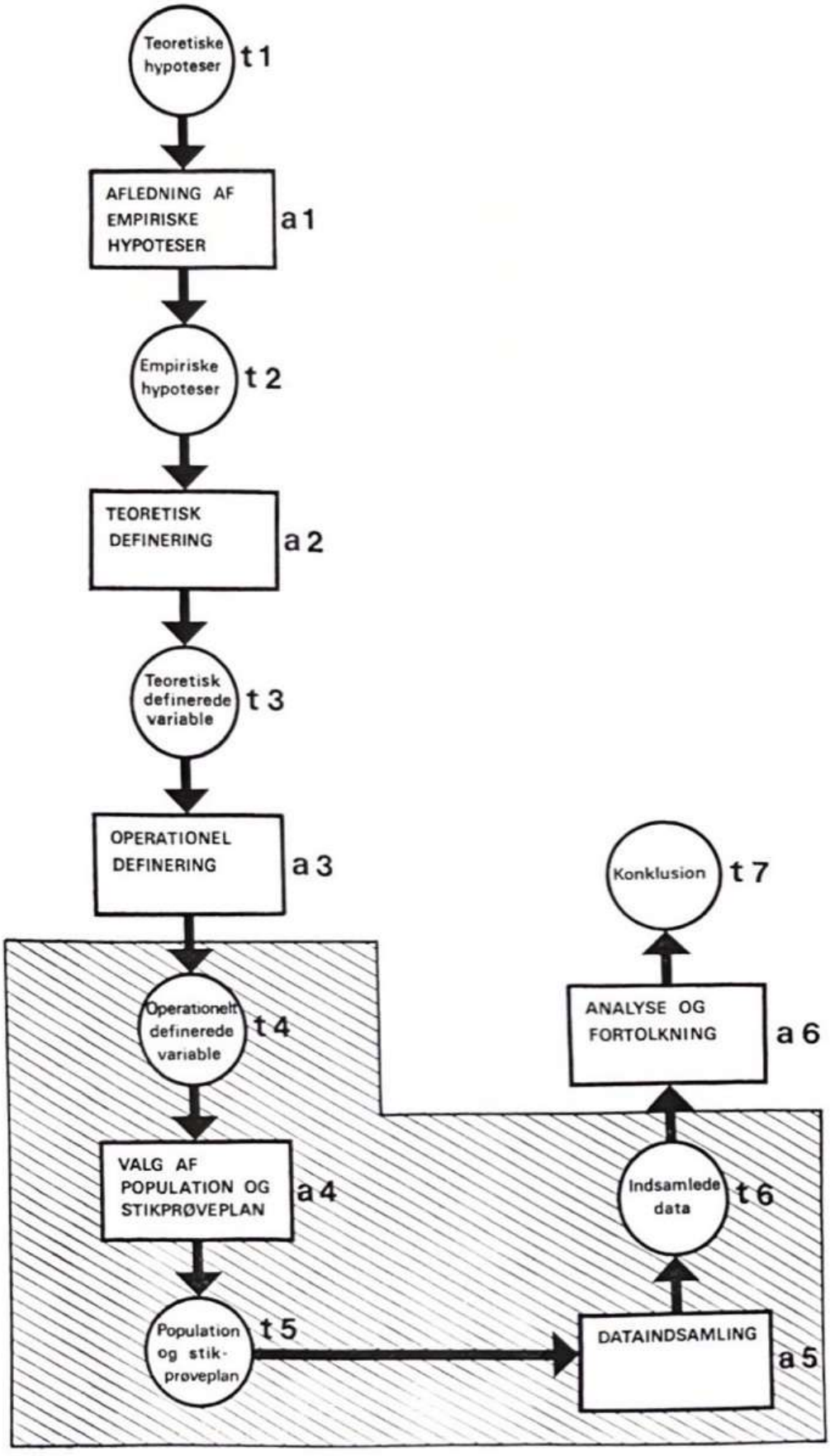
Har en undersøgelsesmetode stor evne til at frembringe (næsten) identiske data ved gentagelse, siges undersøgelsens *reliabilitet* at være høj.

3. Om kønsdiskrimineringsundersøgelsens reliabilitet

I hvilket omfang lader nu kønsdiskrimineringsundersøgelsen sig gentage med samme resultater til følge? For at besvare dette spørgsmål vil vi tage udgangspunkt i de operationelt definerede begreber (variable) – bolle t4 i figur 2 – og følge processen til fremkomsten af de data, hvorpå analysens konklusioner skal bygges, svarende til bolle t6 i figur 2. Som allerede nævnt vil vi ikke i denne artikel beskæftige os med de aktiviteter, der ligger efter bolle t6 i figuren, medens de tidligere aktiviteter – herunder især hypoteseopstillingen – vil blive behandlet i artiklens næste afsnit (afsnit 4).

Rent konkret er den foreliggende kønsdiskrimineringsundersøgelse foregået på den måde, at fem såkaldte *kodere* for hver af 4.238 annoncer i danske media har udfyldt et *spørgeskema* med 508 spørgsmål, der refererer til annoncernes indhold. De 508 spørgsmål betragtes af Sepstrup som lige så mange operationelt definerede egenskaber (variable) ved en stikprøve på 4.238 *analyseenheder* (annoncer).

Figur 2.


Teoretiske hypoteser
t 1
AFLEDNING AF EMPIRISKE HYPOTESER
a 1
Empiriske hypoteser
t 2
TEORETISK DEFINERING
a 2
Teoretisk definerede variable
t 3
OPERATIONEL DEFINERING
a 3
Operationelt definerede variable
t 4
VALG AF POPULATION OG STIKPRØVEPLAN
a 4
Population og stik-prøveplan
t 5
DATAINDSAMLING
a 5
Indsamlede data
t 6
ANALYSE OG FORTOLKNING
a 6
Konklusion
t 7

Hvad skulle hindre, at en gentagelse af denne del af processen fører til identiske data? To forhold springer her i øjnene:

For det første er der tale om en stikprøveundersøgelse, og man må naturligvis forvente, at udvælgelse af en ny stikprøve af annoncer vil føre til en anden svarfordeling. Denne *stikprøveusikkerhed* kan naturligvis formindskes ved at øge stikprøvestørrelsen (dvs. at undersøge flere annoncer), men den kan kun forsvinde helt, hvis der ikke foretages en stikprøveundersøgelse, men en totalundersøgelse, dvs. at alle annoncer i populationen undersøges. Heldigvis lader stikprøveusikkerhedens størrelse sig vurdere ved velkendte statistiske metoder baseret på sandsynlighedsregningen, hvorfor der kan tages hensyn til denne, når data skal fortolkes, og slutninger drages. Dette er Sepstrup naturligvis klar over, og lad os tilføje, at vi ikke mener, at stikprøveusikkerheden *i sig selv* er så stor, at den skulle blive noget væsentligt problem ved datas fortolkning.

For det andet kan forskellige kodere fortolke annoncernes indhold forskelligt. Koderreliabilitetens størrelse lader sig i realiteten kun vurdere ved at lade et antal forskellige kodere vurdere de samme annoncer (helst uden at de er klar over, at det er de samme annoncer!). En sådan realibilitetskontrol gennemføres også, og Sepstrup ofrer – i realistisk erkendelse af problemets afgørende betydning – store anstrengelser på en sådan kontrol, ligesom diskusionen af koderrealibiliteten fylder ca. en trediedel af tekstsiderne i den tekniske rapport.

Hvad siger reliabilitetskoefficienten?

Som mål for reliabiliteten anvender Sepstrup *reliabilitetskoefficienten,* der beregnes som den *andel* af tilfældene, hvori koderne opnår overensstemmelse i kodningen. Gennemfører man således 18 parvise sammenligninger mellem kodernes besvarelse af spørgsmålet: »Indgår fysisk tilpashed/sundhed som værdi [i annoncen]?« med svarmulighederne »ja«, »nej« og »usikker« – idet koderne har fået samme annonce til vurdering – og deres svar stemmer overens i 15 af disse 18 tilfælde, siges *reliabilitetskoefficienten* at være 15/18=0,83.

Nu er det klart, at selv om koderne besvarer et spørgsmål (på annoncens vegne!) ved ren og skær lodtrækning mellem de tre svarmuligheder, må man forvente, at besvarelserne i et vist omfang stemmer overens, og dette forhold må *nødvendigves* tages i betragtning, når størrelsen af en aktuel reliabilitetskoefficient skal vurderes. Sepstrup er selv inde på dette, idet han udtrykkeligt omtaler en alternativ måde at beregne reliabilitetskoefficienten på, den såkaldte *Scott/Hellevik-formel,* der populært sagt kun er baseret på den del af kodernes overensstemmelse, der ligger ud over, hvad man kunne forvente som følge af rene tilfældigheder (lodtrækning!). Scott/Hellevik-koefficienten er altid *mindre* end den ordinære reliabilitetskoefficient. F. eks. er den i ovennævnte eksempel 0,56 mod den ordinære 0,83. Sepstrups begrundelse for at benytte

den ordinære koefficient og ikke Scott/Helleviks-koefficienten fortjener at blive bragt *in extenso* (pp. 25-26).

»1. Selv om det lyder umiddelbart rigtigt at tage hensyn til en »kunstig« eller tilfældig opnået del af reliabiliteten, så gælder det dog for en vurdering af den enkelte undersøgelses pålidelighed, at den konstaterede reliabilitet rent faktisk *er* opnået, hvilket er det væsentlige, når formålet med reliabilitetstestet er at give et bidrag til muligheden for at vurdere resultaternes pålidelighed. Det afgørende er, i hvilket omfang måleinstrumentet er anvendt ensartet – hvilket også, når der anvendes flere kodere, er en kraftig indikator på, at instruktionen er fulgt – og det er det i det omfang, som den i forhold til indholdsanalyser traditionelt beregnede reliabilitetskoefficient angiver. Hertil kommer at et islæt af tilfældig ensartethed ifølge sagens natur må antages at gentage sig, således at den traditionelle reliabilitetskoefficient udmærket kan fungere prognostisk.

2. Hvis Scott/Hellevik-formlen anvendes, må tolkningen af reliabilitetskoefficienten blive en anden end for den traditionelt beregnede koefficient. Det kommer hverken Scott eller Hellevik ind på. Det traditionelle mål siger noget om, hvor ofte ensartede målinger har givet samme resultat = hvor stor en andel af målingerne, der er ens. Den reliabilitetskoefficient, der tager hensyn til, hvor stor ensartethed der kan opnås ved tilfældigheder, må sige noget om, hvor stor en del af afstanden mellem det tilfældige og det perfekte, der er overvundet i målingen, en størrelse som det forståelsesmæssigt er betydeligt vanskeligere at håndtere, når det skal vurderes, hvad der er rimeligt. Hellevik og Scott kommer da heller ikke ind på dette problem.

3. Reliabilitetskoefficienten, der tager hensyn til ensartethed, der kan opnås ved tilfældigheder, er følsom over for antallet af kategorier (svarmuligheder) i måleinstrumentet, hvilket hænger sammen med beregningen af den tilfældige overensstemmelse, således at det er sværere at opnå en numerisk høj koefficient for måleinstrumenter med få kategorier.

4. Reliabilitetskoefficienten, der tager hensyn til den ensartethed, der kan opnås ved tilfældigheder, er meget følsom over for, at resultaterne har en spredning over alle svarmuligheder i måleinstrumentet. Også dette hænger sammen med beregningen af den tilfældige overensstemmelse. Hvis der er en ringe spredning, er det sværere at opnå en høj koefficient.«

Vor vurdering af disse begrundelser er:

ad 1. Begrundelsen bygger på manglende kendskab til den videnskabelige metodes formål og reliabilitetsbegrebet. Den videnskabelige metodes formål er (bl. a.) at sikre, at modtageren af rapporten (eller afhandlingen) kan gentage

undersøgelsen med rimelig sikkerhed for at opnå (nogenlunde) identiske resultater. Derfor er reliabilitet ikke et spørgsmål om graden af overensstemmelse mellem den foretagne undersøgelses kodere; men derimod om den grad af overensstemmelse, der kan forventes mellem den foreliggende undersøgelses resultater (med de her anvendte kodere) og resultaterne af en gentagen undersøgelse (med andre kodere). Iøvrigt vil vi overlade det til læseren at finde en mening i begrundelse 1's sidste punktum – vi har opgivet!

ad 2. Begrundelsen bygger på Sepstrups manglende evne til at fortolke Scott/Hellevik-koefficienten. Dette er naturligvis ikke ensbetydende med, at den ikke lader sig fortolke. Scott/Hellevik-koefficienten er principielt ikke vanskeligere at fortolke end ethvert andet statistisk associationsmål. At spørge: »Er denne reliabilitetskoefficient større, end rene tilfældigheder kan begrunde?«, afviger i princippet ikke fra spørgsmålet: »Er forekomsten af glamourisering i de undersøgte annoncer så meget større for kvinders end for mænds vedkommende, at det ikke kan skyldes tilfældigheder (stikprøveusikkerhed)?«. Spørgsmål af denne type besvares i dusinvis af Sepstrup i hans redegørelse for undersøgelsens resultater og i hans fortolkning af disse.

Den mest nærliggende forklaring på, at Scott og Hellevik ikke beskæftiger sig med »fortolkningsproblemet«, er nok, at det ikke er noget problem *for dem.*

ad 3 og 4. Begrundelserne bygger på kosmetiske hensyn: Scott/Hellevik-koefficienten giver ikke så »imponerende« numeriske værdier som den ordinære koefficient. Tillad os et citat til yderligere belysning af Sepstrups argumentation (p. 26):

> »Anvendelsen af Scott/Hellevik-formlen kan således antages at ville producere numerisk små koefficienter i sammenhængen her uden mulighed for at afgøre, hvad disse udtrykker om pålideligheden, men med en (psykologisk betinget) mulighed for uheldige associationer for den tankegang, der er vant til at beskæftige sig med normerne omkring den i indholdsanalysen traditionelle beregning af reliabilitetskoefficienten.«

Bedømmelsen af et sådant argument i en »videnskabelig« rapport vil vi overlade til læseren!

Vore kommentarer ovenfor skal ikke tages som udtryk for, at vi betragter Scott/Hellevik-formlen som et genialt reliabilitetsmål, men som en påpegning af, at Sepstrups begrundelse for at foretrække den ordinære koefficient er utilfredsstillende, idet den bygger på *manglende viden* (begrundelse 1), *manglende evner* (begrundelse 2) og *kosmetiske hensyn* (begrundelserne 3 og 4).

En *rationel* anvendelse af et hvilket som helst reliabilitetsmål må *nødvendigvis* tage hensyn til, at reliabiliteten i et vist omfang kan være opnået ved til-

fældigheder; Scott/Hellevik-koefficienten er et forsøg i denne retning – omend måske ikke det mest vellykkede.

Måske tror Sepstrup at have taget hensyn til den »tilfældige« reliabilitet ved kun at medtage spørgsmål med en ordinær reliabilitetskoefficient på mindst 0,80 i analysen? Grænsen 0,80 er imidlertid fastlagt ganske arbitrært, *og de statistiske konsekvenser af dette valg er ikke undersøgt*. Ganske vist er hans valg af 95% sikkerhedsgrænser for stikprøveusikkerheden lige så arbitrært, men det har dog en klar (omend måske ikke altid lige relevant!) statistisk fortolkning. At 18% ± 1,5% af personerne i de undersøgte annoncer er afbildet i »arbejdssituationer« betyder, at der er 95% sandsynlighed for, at intervallet 16,5–19,5% indeholder det procenttal, som man ville finde ved en totalundersøgelse (i modsætning til en stikprøveundersøgelse). Hvad er den tilsvarende statistiske konsekvens af at benytte en maksimal reliabilitetskoefficient på 0,80?

Tillad os endnu et citat (p. 26-27):

> »Pålideligheden af de resultater, der kommer ud af spørgsmål 35, kan endelig også udtrykkes på den måde, at de 4 kodere alle besvarede spørgsmålet med »nej« for de to annoncers vedkommende. For den tredie annonce svarede de 3 kodere »ja« og den 4. koder »nej«, og det lyder jo egentlig som et ganske pålideligt måleinstrument, der kun – under forudsætning af, at flertallet har ret – laver én fejl ved 12 målinger (8,3%).

Vi vil overlade det til læseren selv at fundere over den mellem tankestregerne anførte sætning (den kunne nok fortjene det!) og blot undre os over den dybere mening med tallet 8,3%. Er dette et uenigheds- eller fejlmål? Tænker vi os, at to kodere besvarer samme spørgsmål i relation til samme annonce med henholdsvis »ja« og »nej«, bliver det tilsvarende tal 50%. Lad det være koderne en trøst, at de aldrig kan blive helt uenige, men *mindst* halvvejs enige.

Behandling af spørgsmål med uacceptabel reliabilitet er et helt kapitel for sig. Som nævnt medtages kun spørgsmål, hvor reliabiliteten er mindst 0,80, idet Sepstrup *definerer begrebet spørgsmål så bredt som til at omfatte punkter i en variabels udfaldsrum*. Eksempelvis dækker et spørgsmål om illustrationens størrelse over fire delspørgsmål som følger: »Under en fjerdedel at annoncen«, »Mellem en fjerdedel og halvdelen«, »Mellem halvdelen og tre fjerdedele« og »Over tre fjerdedele«, hver med svarmulighederne »ja« eller »nej« og »usikker«. For hvert af disse bedømmes reliabiliteten med følgende resultat i den ovenfor nævnte rækkefølge: 0,92; 0,46; 1,00 og 1,00. Den variabel, som angiver ilustrationens størrelse, har altså ikke én men fire reliabiliteter, hvilket i sig selv kan være slemt nok. Det bliver dog værre, thi ét af udfaldene må se sig henvist til klassen af undermålere og må derfor udelades med den absurde konsekvens, at størrelsesvariablen videreføres med kun tre punkter i sit udfaldsrum. Principielt er der tale om en variabel med fire gensidigt udelukkende kategorier, hvoraf én naturligvis altid er residualt bestemt. På den ene

side har det derfor ingen effekt at udelukke en kategori, på den anden side er det meningsløst at gøre det, thi svarmulighedernes indbyrdes afhængighed udelukker, at et af svarene kan behandles uafhængigt af de øvrige. En indførelse af svarmuligheden »usikker« på de fire delspørgsmål berører ikke dette ræsonnement.

Denne og andre underfundigheder – f. eks. at pulterkammeralternativet »andet« ofte udviser højere reliabilitet end de konkrete svarmuligheder – gør rapporten til morsom læsning, *men styrker ikke troen på materialets anvendelighed til en videnskabelig belysning af reklamens mange facetter.*

Som ovenfor nævnt indgår kun spørgsmål med en ordinær reliabilitetskoefficient på mindst 0,80 i analysen. Denne regel er dog ikke uden undtagelser, idet det hedder (p. 33):

> »Hvor der alene er tale om *sammenligninger* mellem for eksempel typer af personer eller mellem mænd og kvinder medtages også spørgsmål med en reliabilitetskoefficient på under 0,80. Resultaterne fra sådanne spørgsmål er acceptable i sammenligningsøjemed, da der ikke foreligger antagelser om systematik i produktionen af de fejl, der har medført den for lave reliabilitet«.

Dette ræsonnement er så fejlagtigt, som det overhovedet kan være. Såfremt man sammenligner to usikre tal – f. eks. ved at se på differencen mellem dem – bliver usikkerheden på sammenligningen *større* end på de to tal hver for sig, *med mindre* usikkerheden er af »systematisk« art.

Den statistisk kyndige læser vil fra sin barnelærdom huske formlen: Var (X–Y) = Var (X) + Var (Y)–2 Cov (X,Y), og et tilsvarende asymptotisk resultat kan udledes, hvis vi i stedet for at se på differencen ser på forholdet X/Y. Hvis Sepstrup derfor har ret i, at der ikke kan antages at være »systematik i produktionen af de fejl, der har medført den for lave reliabilitet«, så vil troværdigheden af en sådan sammenligning være mindre end troværdigheden af den af de to målinger, der er mindst troværdig! Konsekvenserne heraf for konklusionerne i en rapport, der fortrinsvis har til formål at belyse kønsdiskriminering, dvs. *forskelle* i behandlingen af mænd og kvinder, vil vi overlade til læsernes fantasi!

Selvom vi allerede nu har beskæftiget os mere med det Sepstrupske reliabilitetsbegreb, end begrebet kan bære, da vi senere må konkludere, at det i en videnskabelig sammenhæng er aldeles uinteressant, så fortjener dog endnu et punkt at blive påpeget. Det er iøjenfaldende, at medens der er en temmelig høj koderreliabilitet for så vidt angår en lang række spørgsmål, hvis besvarelse må kræve ret subjektive, holdningsafhængige vurderinger som f. eks. »Kan personen totalt set karakteriseres som romantisk« (r = 1) og »Er personen glamouriseret/stereotypiseret i forhold til traditionelle normer for udseende« (r = 0,91) så er det spørgsmål, der har den laveste reliabilitet, følgende: »Er

illustrationen større end en fjerdedel og mindre end halvdelen af annoncen«, hvor realibilitetskoefficienten er så lav som 0,46! Dette er ét blandt flere eksempler på, at koderne udviser større overensstemmelse i deres holdninger til undersøgelsesobjektet end i kodning af holdningsuafhængige spørgsmål, hvilket er et dårligt tegn, idet det rejser spørgsmålet: *I hvilket omfang er der tale om en undersøgelse af 4.238 annoncers indhold, og i hvilket omfang er der tale om en undersøgelse af de 5 koderes kvindepolitiske holdninger?* Under alle omstændigheder bør det konstaterede forhold få den kritiske læser til at interessere sig for, om de fem kodere kan formodes at udvise holdningsmæssige ligheder – *uanset hvilken holdning, der rent faktisk er tale om.*

De fem kodere, der alle er nævnt ved navn i rapporten, er alle kvinder, og det er ingen hemmelighed, i hvert fald ikke her på stedet, at de alle i deres virkelighedsopfattelse er nært beslægtede og næppe i overensstemmelse med den af Sepstrup i øvrigt så højt besungne virkelighed. Dette er der naturligvis intet odiøst i *per se,* men det fører uundgåeligt til, at den kritiske læser af rapporten må betvivle dens videnskabelighed. *Koderusikkerheden elimineres ikke ved at kræve høj reabilitet i Sepstrups forstand, eftersom denne på ingen måde beskytter mod, at koderne som gruppe betragtet er biased i forhold til de personer, det hele drejer sig om, nemlig dem, der udsættes for reklamen* (jfr. formålsbeskrivelsen, hvori det hedder, at formålet med undersøgelsen er at se, hvad det er, folk bliver udsat for).

Lidt skarpt kan man konkludere, at Sepstrup har udviklet en metode til indstillingsundersøgelser, der indeholder visse lovende aspekter. En bedre vurdering af metodens fortræffeligheder ville dog kunne opnås ved også at prøve den på »forsøgspersoner« med *forskellige* holdninger, så der forestår endnu et vist udviklingsarbejde.

Som indstillingsundersøgelse lider metoden dog under den ulempe, at den er ret bekostelig: Antager vi, at de fem »forsøgspersoner« hver har kodet godt 850 annoncers »svar« på 508 spørgsmål, har hver »forsøgsperson« været udsat for mere end 430.000 stimulti. Det er vist verdensrekord for en indstillingsundersøgelse!

4. Undersøgelsens referenceramme og hypotesehierarki

En høj reliabilitet er kun en nødvendig, men ikke en tilstrækkelig betingelse for, at en videnskabelig undersøgelse kan tillægges værdi. Undersøgelsen må også have høj *validitet,* dvs. man må være nogenlunde sikker på, at man måler det, man ønsker at måle – det er ikke tilstrækkeligt, at man får identiske måleresultater ved gentagne målinger! Validiteten afhænger i vidt omfang af overensstemmelsen mellem de teoretiske og operationelle definitioner på de egenskaber, der er genstand for undersøgelse. De teoretiske definitioner fortæller, *hvad* der skal måles, og de operationelle, *hvordan* der skal måles. Det kan være fristende at

undlade de teoretiske definitioner og nøjes med de operationelle, men så kan undersøgelsesresultaterne principielt ikke generaliseres.

Skal en undersøgelses validitet og mulighederne for at generalisere dens resultater vurderes, må hele hypotesehierarkiet, både de teoretiske og de heraf afledede empiriske hypoteser samt disses indbyrdes relationer, underkastes en nærmere prøve.

Som det fremgår af figur 2, starter den videnskabelige verifikationsproces med de opstillede teoretiske hypoteser. Udvælgelsen af disse falder uden for den videnskabelige metodes område, og de valg, man her står overfor, kan ikke afgøres ud fra de videnskabelige spilleregler, f. eks. ud fra reliabilitets- og validitetskriterier. Vi må dog nødvendigvis også her beskæftige os med selve opstillingen af de teoretiske hypoteser.

Udgangspunktet for enhver videnskabelig undersøgelse er undersøgerens *referenceramme*. Herved forstås de forestillinger, som denne har om de fænomener, der er genstand for undersøgelse. Det er her hensigtsmæssigt at skelne mellem to typer af forestillinger:

1. De forestillinger, som ønskes testet ved den forestående undersøgelse. Der er her tale om foreløbige *antagelser,* der eventuelt kan have form af hypoteser.

2. De forestillinger, som ikke ønskes testet ved den forestående undersøgelse, fordi deres »rimelighed« synes selvfølgelig. Disse forestillinger kaldes undersøgelsernes *forudsætninger.*

Den praktiske konsekvens af forholdet mellem antagelser og forudsætninger er naturligvis, at de konklusioner, der efter undersøgelsens gennemførelse kan drages med hensyn til antagelsernes holdbarhed, kun gælder under de valgte forudsætninger. Det er derfor vigtigt, at forudsætningerne er realistiske, dvs. er i overensstemmelse med den nyeste (videnskabeligt verificerede) faglige viden på området, og ikke blot hviler på forestillinger, der er populære i det miljø, undersøgeren tilhører.

Under de begrænsninger, som udgøres af undersøgelsens forudsætninger, opstilles altså en række teoretiske hypoteser; af disse afledes en række empiriske hypoteser, hvis begreber defineres teoretisk og operationelt. Da undersøgelsesresultaterne nødvendigvis er behæftet med en vis usikkerhed (f. eks. stikprøveusikkerhed), må mæste skridt være at afgøre, hvilke undersøgelsesresultater der skal føre til forkastelse, og hvilke til accept af de empiriske hypoteser. Man må altså definere de målte variables *udfaldsrum* og afgrænse den del af udfaldsrummet, som vil føre til forkastelse af de empiriske hypoteser, det såkaldte *kritiske område.*

Det er naturligvis særdeles vigtigt at skelne skarpt mellem begreber som forudsætning, teoretisk og empirisk hypotese, variabel, udfaldsrum og kritisk

område, ligesom det er af afgørende betydning, at begreberne er logisk forbundne, således at datafortolkningen kan føres tilbage til de teoretiske hypoteser, men Sepstrups bestræbelser i disse henseender er ikke præget af ubetinget succes.

Der tages i rapporten udgangspunkt i en grundlæggende antagelse om, at reklamen har en socialiserende virkning, der kan være i modstrid med såvel efterstræbelsesværdige samfundsmæssige som personlige målsætninger. Denne antagelse konkretiseres i projektets hoved- og delproblemstillinger, der stort set er indeholdt i et ønske om at undersøge, om reklamen afspejler virkeligheden. Vi skal ikke nærmere kommentere dette formål ud over at påpege, at en undersøgelse af, om reklamen gengiver en begrænset del af virkeligheden synes ret overflødig – kunne man forestille sig, at reklamen gengav hele virkeligheden – hvordan dette tågede begreb så skulle defineres?

Efter omformulering af hoved- og delproblemstillingerne fremkommer et sæt af såkaldte »grundlæggende hypoteser«, der senere danner basis for afledning af tre »hovedhypoteser«, 31 »specifikke hypoteser« og et for os ukendt antal »endelige, helt specifikke hypoteser«. Størrelsen af den sidste mængde af hypoteser har Sepstrup valgt ikke at belaste læseren med. Han skriver (p. 15):

> »Denne detaljerede hypoteseudformning er .... enormt omfattende hvortil kommer, at den fremgår af kodeskemaet ..., hvis spørgsmål (og krydstabuleringsmuligheder) *alle* [vor fremhævelse] er udtryk for detaljerede hypoteser omkring de forskellige elementer og variabler. Da kodeskemaets indhold på den anden side må anses for tilstrækkeligt motiveret igennem den indtil nu foretagne hypoteseopstilling, er det valgt ikke at belaste denne rapport med den helt detaljerede hypoteseopstilling«.

Hvis Sepstrup mener, hvad han skriver, men det gør han nu nok ikke, må vi indrømme, at der er tale om et enormt arbejde. Taget ordret efter citatet er der tale om 508 spørgsmål og 128.778 krydstabuleringsmuligheder. En reklameverden, der kan danne basis for 129.286 hypoteser er i sandhed rigt facetteret! Denne beregning er naturligvis absurd, men den illustrerer ganske godt, hvor vigtigt det er at adskille begreber som hypotese, variabel og udfaldsrum.

Vi kan i vor redegørelse passende tage udgangspunkt i de såkaldte »grundlæggende hypoteser« (pp. 8-9):

> »A. Til ethvert produkt er knyttet en brugsværdi, der udgør køberens motiv til at købe. Producentens motiv til at sælge er produktets byttеværdi. Til fremme af producentens interesse formidles der gennem reklamen et skin af brugsværdi (æstetik) ofte i form af bestemte værdier.
>
> B. I nær sammenhæng med A antages det, at reklamen ikke afspejler virkeligheden, men en gengivelse (en fordrejet delmængde), der er i

overensstemmelse med dens formål, herunder dens interesse i nogle værdier og normer frem for andre.

C. I sammenhæng med A og B antages det, at reklamen formulerer sig forskelligt om og til mænd og kvinder, henholdsvis tilskriver mænd og kvinder forskellige værdier og roller.«

Om formålet med disse hypoteser skriver Sepstrup (p. 9):

»Formålet med at opstille disse 3 hypoteser er ikke at teste dem i traditionel forstand i netop den formulering, hvori de fremtræder her. Det er derimod meningen, at disse hypoteser skal virke som styringsinstrument i forhold til et antal mere specificerede hypoteser, der kan belyse den eller de problemstillinger, som de grundlæggende hypoteser er udtryk for«.

Meningen er ikke ganske klar. Er der tale om forudsætninger (som ikke skal testes, men kun styre), eller er der tale om hypoteser, som skal testes partielt i form af test af del-hypoteser afledt af de »grundlæggende hypoteser«?

Hypotese A fremhæver, at Sepstrups virkelighedsfortolkning bygger på begreberne brugsværdi, bytteværdi og vareæstetik. Der er ikke tale om en hypotese i den sædvanlige brug af ordet, men om en valgt forudsætning, der ikke senere underkastes prøvning – og som bekendt gælder en videnskabelig undersøgelses resultater kun under de valgte forudsætninger. Hypoteserne B og C må opfattes som (teoretiske) hypoteser i sædvanlig forstand, da de direkte kan relateres til undersøgelsesformålet. Betragtes de som forudsætninger, kan analysens formål ikke realiseres. Bemærk imidlertid, at begge hypoteser er af beskrivende art. Selvom det skulle lykkes at verificere dem i traditionel forstand, kan der ikke heraf drages kausale slutninger, dvs. man kan ikke på dette grundlag udtale sig om reklamens socialiserende virkninger.

Af de »grundlæggende hypoteser« afledes som nævnt tre såkaldte »hovedhypoteser« (pp. 9-10):

»1. Reklamens verden er en ung verden med smukke, glade mennesker, en ferie-, fritids- og fornøjelsesverden, hvor arbejde, fattigdom, alderdom, dårlige sociale forhold, konflikter, sorg og ulykke ikke forekommer, en idyllisk og harmonisk verden, hvis problemer ikke er af struktuel eller økonomisk karakter.

2. Producenterne bedrager gennem reklamen til at lære folk bestemte holdninger og en bestemt adfærd af direkte relevans for deres forbrug og indkøb. Det drejer sig om folks forhold til deres egen krop og seksualitet og til socialt samvær, om tingsliggørelse, om undertrykkelse af modsætninger mellem køber og sælger, om købsargumenter, om ressourcebevidsthed m. m.

3. En række adfærdstræk, mål og interesser tilskrives helt eller overvejende hvert af de to køn. Specielt for kvinden gælder, at hun er ung og smuk, at hun lever socialt problemløst og selvoptaget, at hun forbindes med det statiske i tilværelsen, at hun er isoleret til det private rum og primært defineres via sin krop, og at hun er et genstandsagtigt objekt til iagttagelse og erotisk pirring. Kvindens interesser er helbred, udseende, husholdning og barnepleje, medens de overvejende dynamiske, aktive, prestigesøgende mænd bevæger sig i offentlighed og er koncentreret om hobby, arbejde og karriere«.

Her er alle »hypoteser« virkelig hypoteser. Bemærk dog, at medens hovedhypotese 1 og 3 er beskrivende, så er hovedhypotese 2 kausal. Den lægger herved op til belysning af reklamens socialiserende virkning. Det er imidlertid ret overraskende at møde en kausal hypotese på dette sted. For det første er hovedhypoteserne 1, 2 og 3 – ifølge Sepstrup – at opfatte som »specifikationer« af de tilsvarende grundlæggende hypoteser A, B og C, og det er vist første gang, vi har set en beskrivende hypotese »specificeret« i en kausal hypotese – det modsatte er nok mere almindeligt. For det andet er det principielt ikke muligt at foretage kausalanalyser på grundlag af den foreliggende undersøgelse. Denne må karakteriseres som en beskrivende engangsundersøgelse, medens en kausalanalyse må bygge på data fremskaffet eksperimentelt, eller på en måde, der fra en teoretisk synsvinkel kan sidestilles hermed. Det må kraftigt understreges, at den foreliggende undersøgelse ifølge hele sin karakter ikke kan give grundlag for at udtale sig om årsager og virkninger, men kun kan beskrive! Hovedhypotese 2 må derfor nok gives en mere ydmyg, beskrivende form.

Ifølge Sepstrup skal ej heller hovedhypoteserne testes, de skal også »styre« nemlig de såkaldte »specifikke hypoteser«. Disse er nærmest, hvad vi med almindelig sprogbrug ville kalde teoretiske hypoteser. Nogle kan dog i realiteten betragtes som empiriske hypoteser, medens andre nærmest er af hybrid art. Et blandt mange eksempler herpå er (p. 13):

> »H.2.8: Reklamen socialiserer til et forvredent, kunstigt forhold til naturlige kropsfunktioner (sved, menstruation) og -tilstande (skaldethed, rynker) ved at fremhæve disse som noget negativt, der skal skjules eller ændres«.

Skal formuleringen have mening (hvilket vi forudsætter), er der tale om en teoretisk hypotese og en heraf afledt empirisk hypotese (sidstnævnte begyndende efter den sidste parentes). Bemærk desuden, at hypotesen – som de fleste »specifikke hypoteser« afledt af hovedhypotese 2 – er af kausal form: der postuleres en virkning, »reklamen socialiserer«. Som tidligere nævnt er det principielt ikke muligt at drage kausale slutninger af denne undersøgelse. Det skal des-

uden bemærkes, at denne hypotese – som de fleste andre – indeholder en række begreber (her »forvredent« og »kunstigt«), som ikke er teoretisk defineret.

Endnu et citat (p. 10):

> »Såvel de grundlæggende hypoteser som hovedhypoteserne er sammenhængende og delvis også overlappende. Der er derfor et vist arbitrært element i, hvilke specifikke hypoteser der i det efterfølgende henføres til hver af hovedhypoteserne. Forholdet spiller ikke nogen væsentlig rolle, da de specificerede hypoteser ikke skal bruges til direkte testning af hovedhypoteserne, idet konstruktionen af et »hypotesehierarki« primært sigter mod en så gennemarbejdet baggrund som muligt for beskrivelsen af reklameindholdet (registreringsskemaet)«.

Bemærk, at Sepstrup her brænder broerne bag sig. Selv om samtlige specifikke hypoteser skulle kunne verificeres, er det umuligt på grundlag af dette at sige noget om hovedhypotesernes »sandhed«. Sepstrup opfatter tilsyneladende de specifikke hypoteser som teoretiske, hvorefter den spændte læser venter en afledning af et sæt konkrete, empiriske hypoteser. Som tidligere nævnt skuffes man dog på dette punkt, idet man må lade sig nøje med kodeskemaet og de muligheder, fantasien åbner.

Konklusion: Det Sepstrupske hypotesehierarki kan karakteriseres som følger:

1. Uoverensstemmelse mellem undersøgelsens formål og undersøgelsesplanen.
2. Uheldig sammenblanding af begreber.
3. Flere »brud« og »manglende led« i overgangen fra teoretisk hypotese til specifik måling.
4. Manglende teoretiske definitioner.

Hertil kommer, at såvel de 31 »specifikke hypoteser« som spørgeskemaets 508 »helt specifikke hypoteser« myldrer af kåde og barokke indfald, som nok ville have moret læseren af disse linier, men som vi af pladshensyn må udelade.

Sammenfattende kan de afgøres, at hypotesehierarkiets konstruktion forhindrer læseren af rapporten i at bedømme, hvilke hypoteser der skal testes, og på hvilket grundlag dette test skal finde sted. *De konklusioner, undersøgelsen måtte blive årsag til, vil derfor uundgåeligt komme til at stå i et uvidenskabeligt skær. Desuden er der intet – hverken teoretisk eller empirisk – belæg for at drage slutninger af kausal natur.*

5. Afsluttende bemærkninger

En gennemgang som denne kommer let til at virke ensidigt kritisk, hvilket i det foreliggende tilfælde er uheldigt, eftersom rapporten indeholder visse kvaliteter,

ligesom det ej heller må underkendes, at den repræsenterer et endog meget stort arbejde. *Det står dog fast, at den måde, hvorpå undersøgelsen er gennemført, forhindrer, at dens resultater og konklusioner generaliseres, og disse kan derfor ikke danne grundlag for udformningen af eventuelle lovindgreb på reklameområdet.*

Samtidig hermed ønsker vi at påpege, at Statens Samfundsvidenskabelige Forskningsråd efter vor opfattelse så rigtigt, da man bevilgede støtte til projektet. Reklamen udgør en betydelig facet af dagliglivet i dagens Danmark, og det siger sig selv, at det er vigtigt, at denne facet belyses både grundigt og omhyggeligt. Det er derfor beklageligt, at Sepstrup har forvaltet de betroede talenter på en så lidet overbevisende måde, som tilfældet er.

Endelig bør det nævnes, at ingen af os nærer specielt varme følelser for reklamebranchen, men branchen har som enhver anden i et demokratisk samfund ret til en »fair trial«, og det føler vi ikke, der er lagt op til med denne rapport.